

GERARD, F.-M., HUGONNIER, B. & VARIN, S. (2018). *Mesure de la qualité des systèmes éducatifs des pays de l'OCDE*, in ADMEE-Europe, *L'évaluation en éducation et en formation face aux transformations des sociétés contemporaines, Actes du colloque, Esch-sur-Alzette : Université de Luxembourg*, pp. 131-143.

Mesure de la qualité des systèmes éducatifs des pays de l'OCDE (7499)

François-Marie Gerard*, Bernard Hugonnier** & Sacha Varin***

*Belgique

**France

***Suisse

Mots-clés : système éducatif, évaluation, comparaisons internationales

En 1990, la *Déclaration mondiale sur l'Éducation pour tous* adoptée à Jomtien (Thaïlande), sous l'égide de l'UNESCO, soulignait la nécessité de fournir à tous les enfants, à tous les jeunes et à tous les adultes une éducation qui réponde à leurs besoins et qui soit pertinente pour leur vie. Cette déclaration ouvrait la voie au concept de qualité conforme à des critères fondés sur des besoins. Elle permettra de redéfinir l'objectif de nombreux systèmes éducatifs en faisant en sorte que les compétences, les connaissances, les valeurs et les attitudes, que l'enseignement et l'apprentissage encouragent, reflètent et prennent en compte les besoins et les attentes des personnes, des pays, de la population mondiale et du monde du travail aujourd'hui.

Il s'en suit que la qualité des systèmes éducatifs est une exigence fondamentale. Il reste que ce concept n'est pas encore clairement défini et que la question de l'évaluation de cette qualité reste ouverte. Plusieurs auteurs se sont efforcés de délimiter ce que recouvrent la qualité de l'éducation et son évaluation²³.

De ces travaux, il ressort que la qualité d'un système éducatif est un concept polysémique. Son évaluation doit donc prendre en compte plusieurs facettes. C'est d'ailleurs le cas d'autres phénomènes mesurés dans les décennies récentes [développement humain (ONU) ; compétitivité économique (World Economic Forum, WEF) ; « vivre-mieux » (OCDE)]. À cette polysémie du concept s'ajoute – dans le cas de la qualité – une difficulté supplémentaire : comment quantifier une mesure de la qualité qui, par nature, est qualitative ? Tout cela explique qu'à ce jour, il n'existe aucune mesure validée de la qualité des systèmes éducatifs.

Jusqu'à présent, cette qualité a principalement été mesurée par le niveau de performance des élèves. Or, dans les tests, tant nationaux qu'internationaux, comme PIRLS, TIMSS ou PISA, la variable latente n'est ni les performances des élèves, ni la qualité des systèmes éducatifs des pays (principalement pour ce qui nous concerne ici des pays de l'OCDE), mais le niveau de connaissances et de compétences des élèves. L'écart méthodologique, entre l'objectif visant à mesurer la qualité des systèmes éducatifs et

²³ Voir notamment Behrens (2007) ; Bouchard et Plante (2000, 2003) ; Chatel (2006) ; De Ketele et Gerard (2007) ; Hanushek & Wößmann (2012) ; Sall et De Ketele (1997).

l'usage qui est fait des tests passés par les élèves, est donc considérable et les erreurs d'interprétation qui s'en suivent dans les classements peuvent être d'un degré très élevé.

Notre problématique – initiée au sein du Séminaire international « École et République » qui s'est tenu au Collège des Bernardins (Paris), de 2014 à 2016 (Hugonnier & Serrano, 2017) – était ainsi de développer un indicateur dont l'objet était de mesurer réellement la qualité des systèmes éducatifs des pays de l'OCDE.

1. Les critères d'évaluation de la qualité des systèmes éducatifs

Mesurer la qualité d'un système éducatif nécessite de disposer d'informations sur un nombre très élevé de facteurs : la qualité des moyens (économiques, financiers, culturels, humains) qui font fonctionner le système ; mais aussi la qualité de la formation des enseignants et la qualité de leur engagement (par exemple, leur conscience professionnelle) ; l'origine sociale des élèves et des parents, la motivation des élèves, le soutien des élèves par leurs parents pour les devoirs à la maison ; la qualité de l'enseignement (qualité des programmes et de leur équilibre, des rythmes scolaires, des didactiques, des pédagogies, des relations élèves-enseignants, du climat scolaire) ; la qualité des résultats en termes de pourcentage de réussite, de taux d'échec et d'abandon, de niveau de performances ; les niveaux d'efficacité et d'équité ; enfin la qualité du contexte scolaire (situation sanitaire, l'environnement socio-économique, les conditions culturelles, l'autonomie des établissements)²⁴.

Cependant, réunir des informations statistiques fiables et comparables portant sur tous ces facteurs et pour l'ensemble des 35 pays de l'OCDE est strictement impossible à l'heure actuelle. En conséquence, pour structurer nos travaux, nous nous sommes dirigés vers le modèle en « E » proposé par Gerard (2001, 2015), largement inspiré par les travaux antérieurs, dont Sall et De Ketele (1997). « E » comme enseignement, évaluation, éducation, élève, enseignant, mais aussi envie, expérience, éthique, échec, énergie, empathie, éveil, entrave, élitisme, économie, encourager, exigence, être, examen, erreur, émotion, enthousiasme, école, esprit, excellence, émerveillement, emploi, émancipation, entrepreneuriat, évolution... Tous des éléments qui, d'une manière ou d'une autre, contribuent à la qualité d'un système éducatif.

²⁴ Scheerens, Luyten & Van Ravens, 2011.

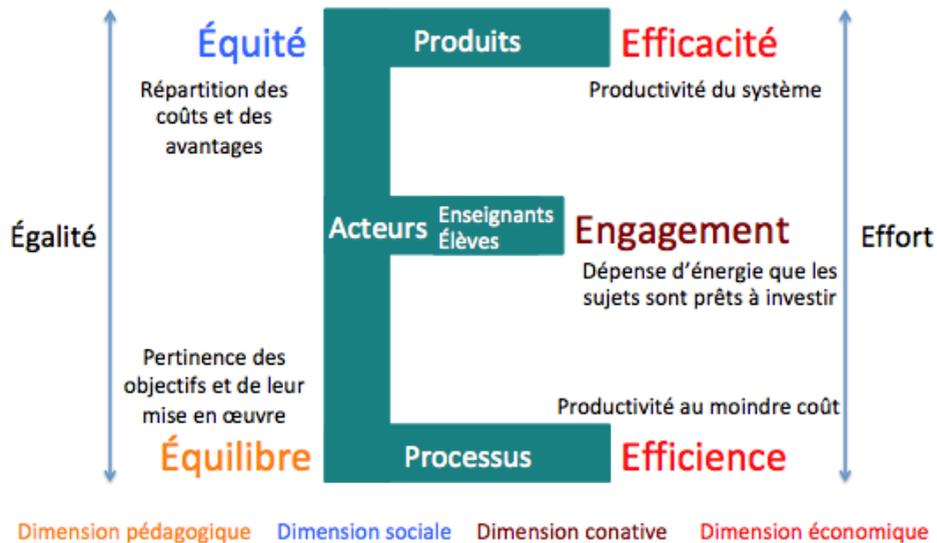


Figure 1 : Modèle en E pour l'évaluation de la qualité des systèmes éducatifs

Ce modèle en E, présenté en figure 1, permet d'organiser les critères constitutifs de la qualité d'un système éducatif autour de différents axes :

- les axes horizontaux concernent les *produits* du système, mais aussi les *processus* et les moyens mis en œuvre pour les atteindre, ainsi que – au centre du modèle – les *acteurs* qui les concrétisent, principalement les enseignants et les élèves ;
- les axes verticaux abordent, du côté droit, l'*effort* à fournir par le système pour atteindre les produits et mettre en œuvre les processus. Du côté gauche, à la charnière des axes, la notion d'*égalité* introduit des éléments essentiels pour la qualité du système.

À partir de ces axes, cinq critères d'évaluation de la qualité peuvent être identifiés, en lien avec quatre dimensions : l'efficacité (dimension économique), l'efficience (dimension économique), l'équité (dimension sociale), l'équilibre (dimension pédagogique) et l'engagement des acteurs (dimension conative).

Les critères retenus pour mesurer la qualité des systèmes éducatifs à travers l'**indicateur synthétique de qualité** (ISQ) rencontrent ce modèle, mais leur choix est cependant restreint par les statistiques permettant de procéder à l'évaluation puisque ces statistiques doivent être non seulement disponibles pour les 35 pays de l'OCDE, mais aussi fiables et comparables, y compris de manière diachronique. Ainsi, l'équilibre – critère essentiel pour évaluer la qualité d'un système éducatif dans la mesure où il est le seul qui porte spécifiquement sur la dimension pédagogique, au cœur des systèmes éducatifs – n'a pas été retenu dans ce cadre, car il n'existe actuellement pas de statistiques qui permettraient de le mesurer. Nous ne désespérons pas de pouvoir le faire, comme c'est d'ailleurs le cas actuellement dans une application de la démarche de l'indicateur synthétique de qualité aux systèmes éducatifs des cantons suisses (Varin, 2018).

Les cinq critères retenus dans la présente recherche sont dès lors les suivants (Gerard, Hugonnier & Varin, 2017) :

- l'*engagement des élèves* évalue combien ces derniers sont bien intégrés dans l'école et s'y trouvent heureux et mesure également leur motivation instrumentale et leur persévérance ;
- l'*engagement des enseignants* évalue la participation des enseignants à la vie scolaire, leur niveau de conscience professionnelle et leur attention vis-à-vis des élèves ;
- l'*efficacité des systèmes éducatifs* évalue la capacité de ces systèmes à conduire les élèves à de bons résultats, à limiter le pourcentage d'élèves en difficulté et à maximiser la proportion d'élèves très performants ;
- l'*efficacité des systèmes éducatifs* évalue le rendement des dépenses publiques d'éducation par rapport aux bénéfices apportés par les systèmes mesurés par leur efficacité ;
- l'*équité des systèmes éducatifs* évalue notamment la capacité de ces systèmes à limiter l'impact du milieu familial sur les performances des élèves.

2. Les statistiques utilisées pour évaluer les critères

Pour évaluer chacun de ces critères et les mesurer par un score, nous avons identifié les statistiques qui paraissaient les plus pertinentes. La première application de la démarche ISQ a porté sur les données issues de l'enquête PISA 2012, enrichies d'informations contenues dans *Regards sur l'éducation* (2013), également publiés par l'OCDE. Six statistiques ont été identifiées pour chacun des cinq critères, soit au total trente. Lors de la deuxième application, portant sur les données PISA 2015, nous avons constaté l'impossibilité de continuer à travailler avec ces trente statistiques, certaines d'entre elles n'étant plus disponibles. Le choix s'est donc limité à un total de 15 statistiques, à raison de trois statistiques pour chacun des cinq critères, non sans nous être assurés que ce passage de 30 à 15 statistiques ne changeait pas fondamentalement les résultats.

Le calcul de l'ISQ nécessite la standardisation des données, car toutes ne se présentent pas sous un format identique. Même si les valeurs aberrantes sont peu nombreuses, les distributions des données auxquelles nous avons à faire sont asymétriques et présentent souvent de longues queues. Dans de tels cas, il convient d'utiliser des estimateurs robustes et efficaces²⁵. Pour standardiser les données, nous avons dès lors opté, comme indice de tendance centrale, pour le *Hodges-Lehmann (HL) estimateur* qui est efficace à 95% et assez robuste, car il tolère 29% de valeurs aberrantes avant d'être inefficace. Comme indice de dispersion des données, nous avons eu recours au *ScaleTau2 estimateur (Robust Tau-estimate of scale)* qui est très efficace à 95% et robuste à 50%²⁶.

Le score de chaque critère est obtenu en faisant la moyenne des 3 statistiques qui le composent. L'indicateur synthétique de qualité (ISQ) correspond à la moyenne des cinq scores finaux des critères.

Chaque statistique a donc le même poids dans l'évaluation des critères. Ceux-ci sont également équipondérés. C'est un choix, relevant d'une certaine subjectivité inhérente à toute évaluation (Gerard, 2017). Techniquement parlant, il serait évidemment très facile de décider d'un autre choix, en donnant plus de poids à l'un ou l'autre critère, voire à l'une ou l'autre statistique. La question essentielle serait

²⁵ Varin et Panagiotakos, 2018.

²⁶ La standardisation effectuée transforme donc les données en soustrayant à chaque valeur le *HL estimateur* et en la divisant par le *ScaleTau2 estimateur*. Les scores standardisés sont ensuite retransformés de telle sorte que l'indice de tendance centrale (*HL estimateur*) soit égal à 500 et celui de dispersion (*ScaleTau2 estimateur*) égal à 100.

alors la pertinence de cette pondération différenciée. Si l'OCDE décidait d'utiliser l'ISQ, la décision quant à cette pondération serait évidemment du ressort de l'ensemble des pays membres de l'OCDE.

3. Résultats et utilisation de l'ISQ

L'indicateur synthétique de qualité débouche sur une évaluation chiffrée de la qualité des systèmes éducatifs des 35 pays de l'OCDE. Il permet d'établir un classement de ces 35 pays. Nous pensons que celui-ci n'est pas le plus important. De nombreuses critiques ont été adressées quant à la pertinence, la validité et la fiabilité des classements présentés par les enquêtes internationales (Champollion & Barthes, 2012 ; De Ketele, 2006 ; Roegiers, 2012). Les classements des pays n'apportent pas réellement d'information pertinente pour l'action, et ils risquent d'entraîner une course à la première place aussi vaine qu'inutile. En effet, tous les tests internationaux (PISA, PIRLS, TIMMS...) classent les pays suivant les scores des élèves. Les pays les moins bien placés s'intéressent dès lors aux politiques des pays les mieux classés pour éventuellement tenter de les dupliquer alors même que les contextes sont fondamentalement différents.

Au lieu de se limiter à faire de tels classements, l'indicateur synthétique de qualité développé ici propose trois nouvelles approches permettant une analyse politique de la qualité des systèmes éducatifs des pays de l'OCDE nettement plus constructive.

3.1. Première approche : une catégorisation de la qualité des systèmes éducatifs

Afin de permettre aux pays d'apprécier la qualité de leur système éducatif²⁷ et sans mettre en avant le classement de cette qualité que permet la valeur de l'ISQ, on pourrait se baser sur lui pour créer des catégories hiérarchisées correspondant à des niveaux de qualité. Une telle catégorisation pose cependant des difficultés pour déterminer les frontières entre chaque catégorie, surtout si on prend en compte les marges d'erreur²⁸.

Cependant, l'ISQ est composé de 5 critères dont l'indice de tendance centrale (*HL estimateur*) est à chaque fois de 500 points. La qualité des systèmes éducatifs pourrait dès lors être évaluée suivant le nombre de critères égaux ou supérieurs à ce score de 500.

On pourrait ainsi considérer que les pays qui ont 5 critères égaux ou supérieurs à 500 ont une haute qualité de leur système éducatif, alors que ceux qui en ont 4 seraient de très bonne qualité, 3 de bonne qualité, 2 de qualité moyenne et 1 ou 0 de qualité insuffisante.

Dans le tableau 1, les critères n'atteignant pas 500 sont indiqués en jaune.

²⁷ Les analyses réalisées globalisent chaque fois ici les données des pays, alors même que ceux-ci ont parfois des systèmes éducatifs régionalisés, par exemple en Allemagne, en Belgique, en Suisse...

²⁸ Une limite inhérente à toute étude basée sur l'établissement d'une mesure à partir de données échantillonnées est liée à la marge d'erreur. Dans le cas du PISA, cette marge d'erreur est relativement minime étant donné les procédures très rigoureuses d'échantillonnage et la taille importante des échantillons nationaux. Dans PISA 2015, pour les 35 pays de l'OCDE, la taille moyenne des échantillons est de 7104 élèves par pays (OCDE, 2016, p. 306). La marge d'erreur à 95% de confiance est dès lors de 2 ou 3 points par résultat sur l'échelle PISA. L'ISQ étant lui-même une moyenne de moyennes, ces marges d'erreur peuvent se cumuler, tout en restant relativement réduites. Il n'empêche que ces marges d'erreur existent et invitent toujours à être prudent dans l'analyse des résultats et dans les conclusions qui peuvent en être tirées.

Tableau 1 : Scores des 35 pays de l'OCDE pour chaque critère et l'ISQ

| Pays | Critères | Engagement des élèves | Engagement des enseignants | Efficacité | Efficiency | Equité | ISQ |
|---------------------|----------|-----------------------|----------------------------|------------|------------|--------|-----|
| Allemagne | | 544 | 403 | 561 | 609 | 496 | 523 |
| Australie | | 508 | 531 | 533 | 495 | 488 | 511 |
| Autriche | | 413 | 491 | 485 | 522 | 430 | 468 |
| Belgique | | 426 | 353 | 535 | 484 | 420 | 444 |
| Canada | | 500 | 522 | 636 | 562 | 582 | 560 |
| Chili | | 569 | 341 | 248 | 382 | 469 | 402 |
| Corée | | 527 | 529 | 614 | 516 | 531 | 543 |
| Danemark | | 484 | 431 | 543 | 459 | 539 | 491 |
| Espagne | | 595 | 520 | 482 | 566 | 594 | 551 |
| Estonie | | 512 | 565 | 639 | 692 | 634 | 608 |
| États-Unis | | 530 | 499 | 463 | 459 | 556 | 501 |
| Finlande | | 529 | 530 | 631 | 588 | 562 | 568 |
| France | | 321 | 441 | 502 | 524 | 326 | 423 |
| Grèce | | 586 | 498 | 324 | 389 | 449 | 449 |
| Hongrie | | 480 | 552 | 400 | 584 | 357 | 475 |
| Irlande | | 567 | 512 | 568 | 539 | 521 | 542 |
| Islande | | 561 | 513 | 432 | 404 | 553 | 493 |
| Israël | | 488 | 403 | 388 | 425 | 416 | 424 |
| Italie | | 450 | 445 | 452 | 548 | 514 | 482 |
| Japon | | 549 | 424 | 660 | 671 | 571 | 575 |
| Lettonie | | 561 | 486 | 458 | 515 | 596 | 523 |
| Luxembourg | | 434 | 519 | 442 | 452 | 372 | 444 |
| Mexique | | 568 | 492 | 116 | 282 | 538 | 399 |
| Norvège | | 516 | 480 | 544 | 467 | 554 | 512 |
| Nouvelle-Zélande | | 521 | 579 | 549 | 470 | 454 | 515 |
| Pays-Bas | | 537 | 398 | 561 | 536 | 469 | 500 |
| Pologne | | 407 | 510 | 540 | 587 | 520 | 513 |
| Portugal | | 531 | 535 | 508 | 512 | 551 | 527 |
| République slovaque | | 373 | 504 | 344 | 480 | 383 | 417 |
| République tchèque | | 371 | 493 | 477 | 603 | 369 | 463 |
| Royaume-Uni | | 562 | 483 | 520 | 442 | 551 | 512 |
| Slovénie | | 405 | 465 | 567 | 579 | 532 | 510 |
| Suède | | 460 | 517 | 502 | 493 | 466 | 488 |
| Suisse | | 420 | 528 | 553 | 482 | 470 | 490 |
| Turquie | | 453 | 543 | 158 | 334 | 588 | 415 |

Cette méthode présente une limite évidente qui est celle de considérer comme équivalent un pays dont les 5 critères auraient des scores proches de 600 et un autre dont les scores seraient égaux à 500 de même que de considérer comme étant dans des catégories différentes des pays ayant des scores juste supérieurs à 500 et d'autres avec des scores juste inférieurs à 500.

Après plusieurs essais d'autres catégorisations non convaincantes, nous avons combiné les deux entrées : l'ISQ et le nombre de critères égaux ou supérieurs à 500. Le graphique (figure 2) présente le croisement de ces deux entrées et permet de dégager six catégories :

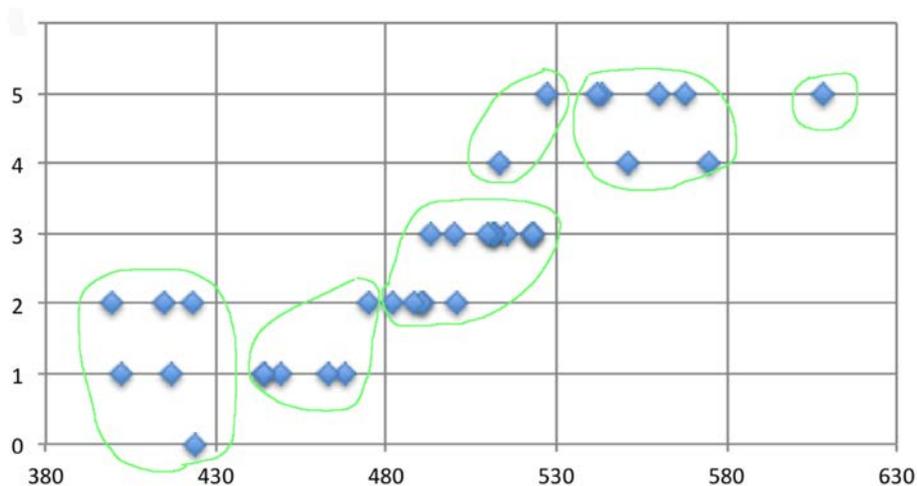


Figure 2 : Croisement de l'ISQ et du nombre de critères ≥ 500

Ces 6 catégories correspondent à différents niveaux de qualité des systèmes éducatifs. Ceux-ci peuvent dès lors être positionnés sur le continuum suivant (figure 3) :

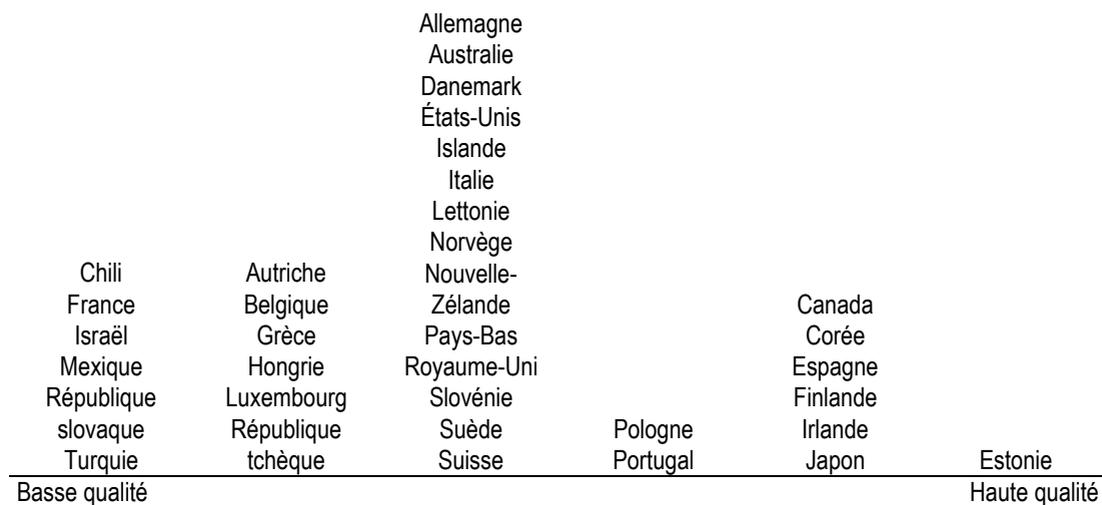


Figure 3 : Niveaux de qualité des systèmes éducatifs

Cette catégorisation permet d'indiquer aux pays qu'au regard des cinq critères utilisés, on peut qualifier la qualité de leur système éducatif dans la mesure où les pays qui ont les plus forts engagements de leurs élèves et de leurs enseignants et présentent des systèmes éducatifs qui ont l'efficacité, l'efficience

et l'équité les plus élevées, sont à notre sens, par construction, ceux qui disposent des meilleurs systèmes éducatifs.

Même si ce type de catégorisation est imparfait et ne permet sans doute pas de rendre compte de la qualité réelle des systèmes éducatifs, il invite cependant les pays concernés à se poser des questions et à prendre d'éventuelles mesures non pas pour améliorer leur classement, mais pour améliorer la qualité de leur(s) système(s) éducatif(s). C'est précisément l'objet de la seconde approche.

3.2. Deuxième approche : l'appréciation de chaque critère

L'indicateur synthétique de qualité (ISQ) apporte à chaque pays de l'OCDE des informations précieuses : le niveau d'engagement des élèves et celui des enseignants est-il plus ou moins haut ou bas, d'une part, et qu'en est-il de celui de l'efficacité, de l'efficience et de l'équité de leur système éducatif, de l'autre ? En d'autres termes, l'ISQ permet à chaque pays de déterminer ses forces et ses faiblesses propres.

À partir de leur analyse, chaque pays peut donc :

- capitaliser sur ses forces, en prenant les mesures adéquates pour renforcer ces points forts ;
- agir sur ses points faibles, en prenant toutes les mesures qui s'imposent pour limiter ses insuffisances.

Il en résultera vraisemblablement des politiques bien plus efficaces, car s'adressant aux problèmes spécifiques du pays, que celle consistant à s'inspirer des mesures prises par les pays qui se trouvent en haut du classement.

Chaque pays pourrait ainsi être analysé par un graphique « radar », comme le montre la figure 4 reprenant les scores par critère de l'Islande et du Danemark qui ont un ISQ équivalent²⁹.

²⁹ Si nous présentons ici deux pays dans le même graphique, c'est avant tout à des fins heuristiques. Dans la pratique et en cohérence avec la démarche, nous pensons qu'il est préférable que chaque pays ait son propre graphique afin d'analyser sa propre situation.

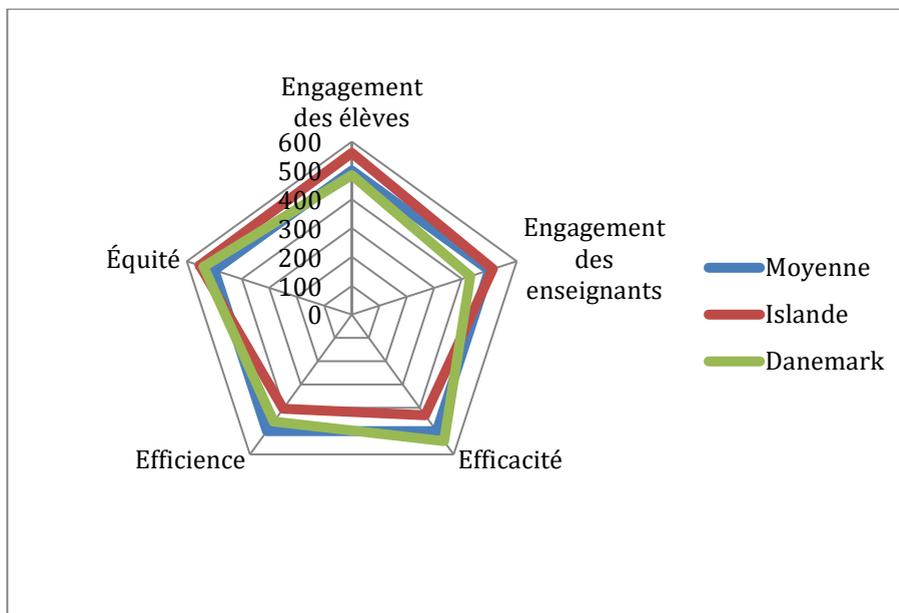


Figure 4 : Forces et faiblesses de l'Islande et du Danemark

On voit que l'Islande (en rouge) devrait améliorer l'efficacité et l'efficience de son système éducatif (critères qui sont tout deux inférieurs à la moyenne), et capitaliser sur ses forces qui sont un engagement des élèves très supérieur à la moyenne, sur une bonne équité et sur un engagement des enseignants qui sont dans la moyenne.

De son côté, le système éducatif du Danemark (en vert) présente des forces avec une efficacité et une équité supérieures à la moyenne, mais aussi un engagement des élèves et une efficience dans la moyenne, et une faiblesse avec l'engagement de ses enseignants.

Les améliorations que pourrait apporter chacun de ces pays pour améliorer la qualité de leur système éducatif et les points positifs sur lesquels ils pourraient s'appuyer sont donc très différents, bien au-delà d'un simple score basé sur les performances de leurs élèves.

L'analyse différenciée peut encore aller plus loin. Chaque critère est composé, dans cette deuxième application de l'ISQ, de trois statistiques. Ce sont donc trois informations supplémentaires dont disposent les pays pour distinguer, au sein de chaque critère, quelles sont, à chaque fois, les forces et les faiblesses sur lesquelles ils peuvent, à nouveau, agir.

Par exemple, si le score d'engagement des élèves du Danemark est juste en dessous de la moyenne (484), les résultats pour chacune des statistiques qui le constituent sont très différenciés :

- Pourcentage d'élèves manquant les cours : 577
- Pourcentage d'élèves se sentant bien intégrés à l'école : 447
- Indice synthétique de motivation instrumentale des élèves³⁰ : 427

³⁰ Selon PISA, si la motivation intrinsèque correspond à des énoncés du style « j'aime faire des sciences, car j'aime les sciences », la motivation instrumentale se caractérise par des énoncés tels que « j'aime faire des sciences, car j'ai compris qu'en étant bon en sciences, j'aurai, par exemple, plus de chances de passer dans la classe supérieure et de faire de meilleures études et d'avoir un meilleur emploi plus tard ».

Si le Danemark souhaite améliorer la qualité de son système éducatif, il pourrait donc s'appuyer sur une assiduité importante de ses étudiants (en s'interrogeant sur ce qui l'explique), mais devrait se poser des questions en ce qui concerne des éléments plus « affectifs » liés à la dimension conative de ses étudiants.

Par contre, l'Islande a un score d'engagement de ses élèves équivalent (488), mais réparti d'une manière très différente :

- Pourcentage d'élèves manquant les cours : 429
- Pourcentage d'élèves se sentant bien intégrés à l'école : 473
- Indice synthétique de motivation instrumentale des élèves : 561

Les actions à mener par le Danemark ou l'Islande pour améliorer la qualité de leur système éducatif au plan de l'engagement des élèves devraient donc être très différentes.

Ce type de constat est bien éloigné des conclusions qui pourraient être tirées d'un simple classement des pays en fonction des performances des élèves dans les trois disciplines étudiées par PISA.

3.3. Troisième approche : les corrélations entre critères

Les cinq critères peuvent être considérés comme des instruments politiques puisque – au-delà de leur mesure – chacun peut être une perspective d'action pour contribuer à l'augmentation de la qualité du système éducatif. Même si les statistiques sur lesquelles ils se basent correspondent à des mesures indépendantes l'une de l'autre, ces critères sont – dans la réalité complexe qu'est un système éducatif – naturellement interdépendants, de sorte qu'agir sur l'un peut entraîner un impact sur d'autres. Le tableau 2 fournit cette information (plus le chiffre est proche de 1, plus cet impact est élevé).

Tableau 2 : Matrice des corrélations croisées (Kendal-tau³¹ et MIC³²)³³

| | Efficacité | Efficience | Equité | Engagement des élèves | Engagement des enseignants |
|----------------------------|------------|------------|--------|-----------------------|----------------------------|
| Efficacité | 1 | 0.48 | 0.24 | 0.38 | 0.06 |
| Efficience | 0.48 | 1 | 0.38 | -0.13 | 0.07 |
| Equité | 0.24 | 0.38 | 1 | 0.31 | 0.16 |
| Engagement des élèves | 0.38 | -0.13 | 0.31 | 1 | -0.001 |
| Engagement des enseignants | 0.06 | 0.07 | 0.16 | -0.001 | 1 |

³¹ Kendal-tau plutôt que Spearman, car relations non linéaires mais monotones et petits échantillons (N=35 pays). Kendal-tau donne de meilleurs résultats que Spearman en cas de petits échantillons.

³² MIC pour Maximal information coefficient, en cas de relations non linéaires et non monotones.

³³ À titre indicatif, relevons que les intervalles de confiance calculés grâce à la méthode du *BCa bootstrap* (Varin, 2016) révèlent que la corrélation (-0.001) entre l'engagement des enseignants et l'engagement des élèves ou vice versa, la corrélation (0.06) entre l'efficacité et l'engagement des enseignants ou vice versa et celle entre l'efficacité et l'engagement des enseignants (0.07) ou vice versa ne sont pas statistiquement différentes de zéro à 95% de confiance.

Une lecture des corrélations deux à deux telles que présentées dans le tableau 2 fournit les informations suivantes³⁴ :

- l'efficacité a des effets marquants sur trois leviers (efficacité, engagement des élèves, et équité) ou vice versa. C'est un instrument intéressant ;
- l'efficacité n'a des effets marquants que sur deux leviers (efficacité et équité) ou vice versa, mais l'effet sur l'engagement des élèves est négatif. C'est un instrument à utiliser avec grande prudence ;
- l'équité a des effets positifs sur quatre leviers (efficacité, efficacité, engagement des élèves et engagement des enseignants) ou vice versa. C'est donc un instrument intéressant.

La conclusion qui se dégage est que deux instruments politiques semblent à privilégier : l'équité d'abord et l'efficacité ensuite. L'équité est en effet le seul des cinq instruments politiques qui conduit à des accroissements des 4 autres instruments ou vice versa (on remarque que la somme de leurs coefficients d'interdépendance est respectivement de 1,09 et de 1,16 ce qui est bien supérieur à la situation des trois autres outils). L'équité conduit directement à une augmentation marquante de l'efficacité et de l'efficacité ou vice versa, mais aussi de l'engagement des élèves et de celui des enseignants ou vice versa. C'est donc par essence le premier outil à utiliser. L'efficacité peut être également considérée, avec un effet positif sur l'efficacité, sur l'engagement des élèves et sur l'équité ou vice versa.

Ces constats confirment l'intuition défendue par Lessard et Meirieu (2004) que la norme ultime du pilotage réussi d'un système éducatif restera toujours l'efficacité en termes de performances des élèves ainsi que l'équité pédagogique.

Ajoutons que les analyses réalisées permettent de montrer

- qu'aucun des 35 pays de l'OCDE ne dispose d'un système éducatif qui soit en tête dans les 5 critères ;
- qu'il n'y a pas de caractéristiques similaires entre des pays proches, qu'ils soient anglo-saxons, nordiques, germaniques ou latins ;
- que chaque pays suit une stratégie propre pour atteindre l'excellence et qu'il n'y a pas une voie unique pour arriver au sommet.

4. Conclusion

La mesure de la qualité des systèmes éducatifs est et restera un défi permanent tant le concept est, comme on l'a vu, polysémique et par nature difficilement quantifiable. Les résultats de la recherche donnée ici doivent donc être poursuivis et ce d'autant plus que l'ISQ présente certaines limites : par manque de données, certains facteurs que nous jugeons essentiels (comme par exemple, la qualité de la formation des enseignants, l'origine sociale des élèves, le soutien des parents, l'équilibre des programmes) ne sont pas analysés. La qualité du processus et du contexte scolaires n'est pas non plus mesurée. La qualité est mesurée en fin de parcours et certains pourraient faire valoir qu'il conviendrait de la mesurer à chaque palier (la maternelle, le primaire, le collège, le lycée). Enfin, question d'importance, l'ISQ pourrait-il valoir pour tous les pays dans le monde ?

³⁴ Pour rappel, il s'agit de liens corrélationnels et nullement de relations de causalité. En effet, s'agissant d'une étude observationnelle et non pas expérimentale (dans laquelle nous aurions pu « manipuler » les niveaux de l'une des deux variables), nous ne pouvons pas apporter de conclusion de causalité aux relations observées.

Force est cependant de reconnaître que l'indicateur synthétique de qualité (ISQ), qui a été présenté dans ces lignes, apporte une contribution importante à la recherche, même si la démarche doit encore être améliorée.

D'un point de vue méthodologique, ce travail met, en effet, en avant :

- une démarche holistique en portant un regard qui englobe
 - la finalité fondamentale de tout système éducatif qui est que les élèves apprennent (efficacité),
 - une perspective sociale afin que tous les élèves réussissent (équité),
 - des exigences économiques en fonction de ressources limitées (efficacité),
 - le vécu des acteurs – élèves et enseignants – à la fois moteur du processus et résultat de ce qui existe dans le système (engagement) ;
- une démarche innovante, en ne se limitant pas à un regard fondé sur une seule dimension - par exemple le niveau de compétences des élèves - et en cherchant à donner du sens à partir des différentes dimensions prises en compte ensemble ;
- une démarche scientifique, qui se base sur des statistiques reconnues et fiables, avec un traitement et une analyse variée des données ;
- une démarche diagnostique et diachronique, permettant un diagnostic différencié, au-delà des classements, à différents moments, pour affiner le pilotage ;
- une démarche évolutive, ouverte à la discussion – les critères et les statistiques qui les composent ne sont pas figés dans un moule immuable – et ayant des évolutions futures possibles vers la prise en compte d'autres critères (efficacité externe, équilibre, durabilité...).

D'un point de vue des politiques d'éducation, la recherche fournit pour chaque pays des informations innovantes à trois niveaux :

- où se situe le niveau de qualité de son système éducatif comparé à celui des autres pays ;
- où se trouvent ses forces et ses faiblesses parmi les cinq critères ;
- dans ces derniers, qui sont chacun composés de trois statistiques, à nouveau quelles sont ses forces et ses faiblesses ? En conséquence, où agir pour capitaliser sur les unes et limiter les autres ? Cela établit une autre approche que celle souvent menée dans certains pays, consistant, plutôt que de développer sa propre stratégie, à copier celle des pays les mieux placés dans les classements internationaux.

Bibliographie

Behrens, M. (Éd.). (2007). *La Qualité en éducation. Pour réfléchir à la formation de demain*. Québec : Presses de l'Université du Québec, collection Éducation-Recherche.

Bouchard, C. & J. Plante (2000). « La qualité : sa définition et sa mesure », *Service social*, 47(1/2), p. 27-62.

Bouchard, C. & J. Plante (2003). « La qualité : mieux la définir pour mieux la mesurer », *Les Cahiers du Service de Pédagogie Expérimentale*, (11/12), Service de pédagogie expérimentale de l'université de Liège, p. 219-236.

Chatel, É. (2006). « Qu'est-ce qu'une éducation de 'qualité' ? Réflexion à partir de l'évaluation comparée des deux diplômes technologiques et professionnels tertiaires », *Éducation et sociétés*, 2006/2 (n° 18), p. 125-140.

Champollion, P. & Barthes, A. (2012). « De l'usage et du mésusage du classement par rang en matière de médiatisation de l'évaluation internationale PISA », *Questions vives – À propos de PISA : pourquoi, pour qui, évaluer et comparer les compétences des élèves*, Vol.6 n°16, p. 15-27.

- Conférence mondiale sur l'éducation pour tous (1990). *Déclaration mondiale sur l'Éducation pour tous et Cadre d'action pour répondre aux besoins éducatifs fondamentaux*, Jomtien-Paris : UNESCO.
- De Ketele, J.-M. (2006). « Mais qu'évaluent donc les enquêtes internationales ? », *Revue internationale d'éducation de Sèvres*, CIEP, n°43, 141-157.
- De Ketele, J.-M. & Gerard, F.-M. (2007). « La qualité et le pilotage du système éducatif », in M. Behrens (Éd.). *La Qualité en éducation. Pour réfléchir à la formation de demain*. Québec : Presses de l'Université du Québec, collection Éducation-Recherche, Chap.1, p. 19-38.
- Gerard, F.-M. (2001). « L'évaluation de la qualité des systèmes de formation », *Mesure et Évaluation en Éducation*, Volume 24, n°2-3, 53-77.
- Gerard, F.-M. (2015). « Le néolibéralisme se niche-t-il au sein même des dispositifs pédagogiques scolaires ? », in J.-L. Chancerel & S. Varin, *Néolibéralisme et éducation - Éclairages de diverses disciplines*, Louvain-la-Neuve : Academia-L'Harmattan, p. 135-154.
- Gerard, F.-M. (2017). « Objectiver la subjectivité », in D. Leduc & S. Béland (Dir.), *Regards sur l'évaluation des apprentissages en arts à l'enseignement supérieur, Tome 1*. Québec : Presses de l'Université du Québec, p. 29-47.
- Gerard, F.-M., Hugonnier, B. & Varin, S. (2017). « La qualité des systèmes éducatifs des pays de l'OCDE enfin mesurée », in B. Hugonnier & G. Serrano (dir.). *Réconcilier la République et son école*. Paris : Éditions du Cerf, p. 61-73.
- Hanushek, E. A. & Wößmann, L. (2012), *Education Quality and Economic Growth*, Washington, DC : The World Bank.
- Hugonnier, B. & Serrano, G. (dir.) (2017). *Réconcilier la République et son école*. Paris : Éditions du Cerf.
- Lessard, C. & Meirieu, P. (dir.) (2004). *L'obligation de résultats en éducation*. Bruxelles : De Boeck.
- OCDE (2013). *Regards sur l'éducation*, Paris : Éditions OCDE, <http://dx.doi.org/10.1787/eag-2013-fr>
- OCDE (2014). *Résultats du PISA 2012 : Savoirs et savoir-faire des élèves : Performance des élèves en mathématiques, en compréhension de l'écrit et en sciences (Volume I)*, PISA, Paris : Éditions OCDE. <http://dx.doi.org/10.1787/9789264208827-fr>
- OCDE (2016). *Résultats du PISA 2015 (Volume I) : L'excellence et l'équité dans l'éducation*, PISA, Paris : Éditions OCDE, <http://dx.doi.org/10.1787/9789264267534-fr>
- Roegiers, X. (2012). « L'influence des épreuves standardisées internationales sur l'évolution des dispositifs et des représentations en matière d'évaluation des acquis des élèves », *Questions vives – À propos de PISA : pourquoi, pour qui, évaluer et comparer les compétences des élèves*, Vol.6 n°16, p. 29-41.
- Sall, H.N. & De Ketele, J.-M. (1997). « Évaluation du rendement des systèmes éducatifs : apports des concepts d'efficacité, d'efficience et d'équité ». *Mesure et Evaluation en Education*. Vol. 19, n°3.
- Scheerens, J., Luyten, H. & Van Ravens, J. (Eds.) (2011). *Perspectives on Educational Quality : Illustrative Outcomes on Primary and Secondary Schooling in the Netherlands*. Dordrecht, Heidelberg, London, New York : SpringerBriefs in Education.
- Varin, S. (2016). *Étude quantitative des données basée sur les statistiques et le logiciel R. Approches descriptive et inférentielle*. Hallennes-lez-Haubourdin : The bookedition.com.
- Varin, S. (2018, à paraître). « La qualité du collège enfin mesurée : 11 cantons suisses sous la loupe », *L'éducateur*.
- Varin, S. & Panagiotakos, DB. (2018, to be published). « The use of robust regression, in biomedical and social sciences research », *Journal of Data Science*.