

# Indicateur synthétique de la qualité des systèmes éducatifs des pays de l'OCDE : comparaison des résultats 2015 et 2018

**François-Marie Gerard**

BIEFOR : Louvain-la-Neuve

**Bernard Hugonnier**

University catholique de Paris

**Sacha Varin**

Collège Villamont

## Résumé

L'indicateur synthétique de qualité (ISQ) des systèmes éducatifs a été développé pour la première fois sur la base de données recueillies en 2012 (Gerard, Hugonnier et Varin, 2017). Il se fondait sur cinq critères : l'efficacité, l'efficacité et l'équité des systèmes, l'engagement des élèves et celui des enseignants. Une deuxième évaluation de l'ISQ a été réalisée à partir de données relatives à 2015 (Gerard, Hugonnier et Varin, 2018).

Grâce à une troisième évaluation de l'ISQ portant sur des données établies en 2018, la validité de trois importantes hypothèses peut être vérifiée :

- 1) les systèmes éducatifs n'évoluent pas tous au même rythme ni dans le même sens ;
- 2) l'ajout de l'engagement des parents en tant que sixième critère a un impact sur la qualité des systèmes éducatifs ;
- 3) l'ISQ est un indicateur valide et fiable.

L'objet de cet article est de tester la validité de ces trois hypothèses.

**Mots-clés** : qualité, ISQ, indicateur synthétique, systèmes éducatifs, PISA

# Synthetic indicator of the quality of educational systems in OECD countries: comparison of 2015 and 2018 results

## Abstract

The synthetic quality indicator (SQI) of education systems was first developed on the basis of data collected in 2012 (Gerard, Hugonnier and Varin, 2017). It was based on five criteria: system effectiveness, efficiency, equity, student engagement and teacher engagement. A second SQI evaluation was conducted using data from 2015 (Gerard, Hugonnier and Varin, 2018).

With a third SQI evaluation using 2018 data, three important hypotheses can be tested for validity:

- 1) not all education systems evolve at the same pace nor in the same direction;
- 2) the addition of parent engagement as a sixth criterion has an impact on the quality of education systems;
- 3) the SQI is a valid and reliable indicator.

The purpose of this paper is to test the validity of these three hypotheses.

**Keywords :** quality, SQI, synthetic indicator, education systems, PISA

## 1. Introduction

La mesure de la qualité des systèmes éducatifs de l'OCDE est encore récente ; une telle information n'est en effet disponible – grâce au développement d'un indice synthétique de qualité (ISQ) de ces pays – que sur la base des données PISA 2012 (Gerard, Hugonnier et Varin, 2017) et 2015 (Gerard, Hugonnier et Varin, 2018). La parution des résultats du PISA 2018 permet à nouveau de rendre compte de la qualité de tels systèmes et de mettre en évidence de possibles progrès ou inversement des régressions dans le niveau de qualité de certains.

La disponibilité des données et l'interprétation des résultats permet également de vérifier ou d'invalider trois hypothèses essentielles :

- 1) les systèmes éducatifs n'évoluent pas tous au même rythme ni dans le même sens, tout au contraire, certains pouvant – entre 2015 et 2018 – progresser de 5% ou plus tandis que d'autres régressent du même pourcentage. La vérification de cette hypothèse permet de plus d'identifier les critères ayant le plus influé sur ces évolutions ;
- 2) l'ajout de l'engagement des parents en tant que sixième critère dans les données de 2018 a un impact sur la qualité des systèmes éducatifs. En effet, si un tel ajout modifie le score et le classement des pays de l'OCDE, cela établirait que l'engagement des parents est un critère significatif dans la mesure de la qualité des systèmes éducatifs et donc doit être pris en compte ;
- 3) l'indicateur synthétique de qualité est un indicateur fiable au sens où il reflète fidèlement les performances des systèmes éducatifs telles que mesurées par l'ISQ, ce qui valide sa composition.

L'objet de cet article est de vérifier la validité de ces trois hypothèses.

## 2. Cadre théorique

### 2.1 Critères retenus pour la recherche

Plusieurs auteurs se sont efforcés de délimiter ce que recouvrent la qualité d'un système éducatif et son évaluation (Bautier et Marchand, 2020 ; Behrens, 2007 ; Bouchard et Plante, 2000, 2003 ; Chatel, 2006 ; Conférence des ministres de l'éducation des pays ayant le français en partage [CONFEMEN], 2011 ; Corriveau, Boyer et Fernandez, 2009 ; De Ketele et Gerard, 2007 ; Dominguez-Gil, Segovia-Gonzalez & Contreras (2022) ; Gerard, 2001 ; Hanushek et Wößmann, 2012 ; Murafa, Ikonomi, Todorakov, Mickovksa-Raleva, Brankovic, Georgiana Radu, Ivan, Stankovic, Radisic et Klemencic, 2012 ; Sall et De Ketele, 1997 ; Varin, 2021a, 2021b).

De ces travaux, il ressort que la qualité d'un système éducatif est un concept polysémique : il recouvre plusieurs facettes, depuis la prise en compte de la performance du système jusque la nécessité d'inclure des dimensions telles que l'équité, la participation des acteurs, l'approche pédagogique, la durabilité des effets...

Sur la base de la littérature, les critères pour évaluer la qualité d'un système éducatif sont nombreux. Ainsi, des propositions construites sur la base du modèle d'évaluation de Stufflebeam et ses collaborateurs (1980) retiennent jusqu'à une quinzaine de critères (De Ketele et Gerard, 2007 ; Roegiers, 1997) alors que Dominguez-Gil, Segovia-Gonzalez & Contreras (2022) n'en utilisent que trois. Pour l'ISQ, la sélection s'est faite sur la base de la

pertinence des critères et de leur disponibilité tout en veillant à ne pas multiplier les critères et à disposer de données valides et fiables.

## 2.2 Définition d'un système éducatif de qualité

Comme le relèvent Foin et Gauthier (2020), la qualité d'un système éducatif n'est pas en soi une donnée objective. Sa mesure dépend de la définition qu'on lui donne ; celle retenue par des experts internationaux réunis de 2014 à 2016 au Collège des Bernardins (Hugonnier et Serrano, 2017) est la suivante : un système éducatif de qualité est celui qui est à la fois efficace, efficient, équitable et où l'engagement des élèves, des enseignant·e·s et des parents est élevé (voir 2.4).

## 2.3 Critères permettant d'évaluer la qualité d'un système éducatif

Initialement, les experts des Bernardins n'avaient retenu que les cinq critères suivants à partir de données statistiques de l'OCDE et de l'UNESCO, de 2012 et de 2015 :

- l'*efficacité* des systèmes éducatifs en tant que capacité de ces systèmes à répondre avec satisfaction aux objectifs qui leur ont été fixés et notamment de permettre aux élèves d'avoir de bonnes performances scolaires ;
- l'*efficience* des systèmes éducatifs définie comme le rapport entre l'efficacité et les moyens utilisés pour les atteindre, que ces moyens soient d'ordre financier, matériel, humain, temporel, etc. ;
- l'*équité* des systèmes éducatifs en tant que capacité d'un système éducatif à limiter l'impact de l'origine sociale des élèves sur leurs performances scolaires et de garantir l'égalité des chances ;
- l'*engagement des élèves* en tant que volonté de ceux-ci de faire les investissements et les efforts nécessaires pour améliorer leurs performances scolaires ;
- l'*engagement des enseignant·e·s* en tant qu'implication de leur part pour permettre à tous les élèves de progresser dans leurs études.

## 2.4 L'engagement des parents comme sixième critère

La disponibilité récente de statistiques sur l'engagement des parents (Organisation de coopération et de développement économiques [OCDE], 2019) a permis d'ajouter dans le calcul de l'ISQ 2018 un sixième critère, l'engagement des parents. En effet, dans le domaine de l'éducation, ce facteur peut jouer un rôle important. Par exemple, à la suite de constats issus de l'étude PISA de 2009 (Programme international pour le suivi des acquis des élèves, 2011), Borgonovi et Montt (2012) ont pu montrer, dans une étude portant sur 14 pays, que l'engagement parental peut améliorer les résultats à la fois cognitifs et non cognitifs des élèves, et contribuer à la réduction des écarts de performance entre les différents groupes socioéconomiques. Même si, en théorie, un système éducatif n'a pas pour objectif d'avoir des parents engagés, on peut constater que les écoles, en France par exemple, développent désormais une politique d'incitation des parents pour qu'ils s'intéressent davantage à l'école et à son fonctionnement, mais également pour participer plus souvent aux réunions de parents d'élèves et passer davantage de temps avec leurs enfants, non seulement pour les aider à faire leurs devoirs à la maison, mais également pour avoir des relations suivies avec les enseignants et contribuer aux activités culturelles de l'école ou à l'organisation de voyages à l'étranger. Des politiques proactives à l'égard des parents peuvent permettre d'améliorer les performances des élèves, mais aussi l'équité de l'éducation. Il s'ensuit que l'engagement des parents, devrait être pris en compte en tant qu'indicateur important de la qualité d'un système éducatif.

Cependant, cette information n'étant pas disponible en 2015, deux ISQ ont été calculés pour 2018, l'un sans et l'autre avec l'engagement des parents ; la comparaison entre les résultats permet de mesurer l'influence de ce facteur sur la qualité des systèmes éducatifs. Ce critère est évalué sur la base d'informations portant sur la participation des parents dans la vie scolaire de leurs enfants en général, tant à l'école qu'à la maison.

## 2.5 La validité et la fiabilité de l'ISQ

Pour tout outil de mesure, la validité – degré d'adéquation entre ce que l'on déclare faire (évaluer telle ou telle dimension) et ce que l'on fait réellement ; entre ce que l'outil mesure et ce qu'il prétend mesurer (Laveault et Grégoire, 1997, 2002) – et la fiabilité – degré de confiance que l'on peut accorder aux résultats observés, indépendamment du moment ou de l'observateur – sont chacune des caractéristiques importantes qu'il s'agit de tester tout au long du processus d'élaboration et d'utilisation. La disponibilité de données diachroniques permet d'apporter des éléments de réponse à ces questions.

Depuis Campbell (1957), les travaux sur la validité (De Landsheere, 1982 ; De Ketele et Roegiers, 1996 ; Fortin, 1994 ; Institut national d'excellence en santé et en services sociaux [INESSS], 2015 ; Van Driel, 2004) ont montré que celle-ci recouvre différentes dimensions, dont notamment la validité théorique ou de construit, la validité critérielle, la validité interne, la validité externe...

La dimension testée ici est la *validité interne empirique*. Il s'agit d'évaluer la mesure dans laquelle les corrélations entre les statistiques formant un des critères de l'ISQ sont plus élevées que celles entre les statistiques de deux critères (De Ketele et Gerard, 2005). En effet, si les statistiques relatives à un critère convergent, avec une corrélation importante, c'est qu'elles mesurent la même dimension. En revanche, les statistiques relatives à deux critères différents devraient être peu convergentes et peu corrélées, car elles sont censées mesurer des dimensions différentes. Il s'agit donc de vérifier que les corrélations propres à un critère sont maximisées (les statistiques évaluent bien le même critère) alors que les corrélations relatives à deux critères sont minimalisées (les statistiques évaluent deux critères différents).

Enfin, la comparaison des ISQ 2015 et 2018 devrait nous apporter des informations en termes de fiabilité, dans la mesure où l'on peut supposer qu'une période de 3 ans n'introduit pas de variation fondamentale dans la réalité des systèmes éducatifs. Si les données diachroniques convergent, cela refléterait la fiabilité de la mesure.

## 3. Méthodologie de validation des trois hypothèses

Dans cette partie, sont présentées l'instrumentation et les méthodes d'analyse des données utilisées dans le but de vérifier la validité des trois hypothèses. Tous les calculs statistiques sont établis en utilisant le logiciel R (R Core Team, 2020 ; Asselin de Beauville & Varin, 2014 ; Varin, 2016).

### 3.1 La population et l'échantillon

La population est constituée de 37 pays au total. Les données de la Lituanie et la Colombie étaient incomplètes en 2018. Notre échantillon non probabiliste de convenance est dès lors constitué de 35 pays, raison pour laquelle seules les statistiques descriptives seront présentées et discutées.

### 3.2 Les données utilisées

### ***3.2.1 Provenance des données et présentation des statistiques***

Pour évaluer chacun des critères par un score, les statistiques les plus pertinentes ont été identifiées parmi celles disponibles dans les enquêtes PISA et dans les données de l'UNESCO. Compte tenu de leurs origines, nous considérons comme acquis que ces données sont valides et fiables.

Chacun des 6 critères est constitué de trois statistiques, soit un total de 18 présentées ci-dessous. Les résultats de l'ISQ auraient-ils été différents si d'autres statistiques avaient été utilisées ? Cela a été testé lorsque l'ISQ a été utilisé pour plusieurs cantons suisses et sur plusieurs pays européens en utilisant d'une part dix-huit statistiques, puis quarante-trois (Varin, 2018 ; Varin, 2021). Il s'avère que les résultats sont sensiblement les mêmes. On peut en déduire que l'utilisation de dix-huit statistiques est suffisante. De leur côté, Dominguez-Gil, Segovia-Gonzalez & Contreras (2022) n'utilisent que 3 critères et 7 statistiques pour élaborer leur propre indicateur composite de la qualité des systèmes éducatifs des pays de l'OCDE.

Le tableau 1 présente les statistiques utilisées pour calculer l'ISQ.

**Tableau 1 - Six critères, constitués chacun de trois statistiques, soit au total dix-huit statistiques**

Critères	Statistique 1	Statistique 2	Statistique 3
<b>Efficacité</b>	Moyenne des élèves en littératie dans les enquêtes PISA 2015 et 2018	Pourcentage d'élèves présentant un score en littératie inférieur au niveau 2 dans PISA 2015 et 2018	Pourcentage d'élèves présentant un score en littératie supérieur au niveau 5 dans PISA 2015 et 2018
<b>Efficience</b>	Rapport entre le score d'efficacité et les dépenses d'éducation primaire et secondaire au PIB en 2014 et en 2017	Rapport entre le score d'efficacité et les dépenses de l'enseignement primaire et secondaire au budget de l'État en 2014 et en 2017	Rapport entre le score d'efficacité et les dépenses d'éducation primaire et secondaire au PIB par habitant·e en 2014 et en 2017
<b>Équité</b>	Score en littératie associé à une augmentation d'une unité de l'indice ESCS (indice mesurant le statut économique, social et culturel des élèves) dans PISA 2015 et 2018	Différence de performance en littératie entre les élèves du quart supérieur et les élèves du bas de l'ESCS dans PISA 2015 et 2018	Pourcentage d'élèves résilient·e·s parmi les élèves défavorisé·e·s dans PISA 2015 et 2018
<b>Engagement des élèves</b>	Pourcentage d'élèves qui manquent des cours dans PISA 2015 et 2018	Pourcentage d'élèves déclarant : « Je me sens appartenir à l'école » dans PISA 2015 et 2018	Index de motivation instrumentale en sciences dans PISA 2015 et 2018
<b>Engagement des enseignant·e·s</b>	Pourcentage d'enseignant·e·s ne répondant pas aux besoins individuels des élèves dans PISA 2015 et 2018	Pourcentage d'enseignant·e·s absent·e·s dans PISA 2015 et 2018	Pourcentage d'enseignant·e·s n'étant pas bien préparés pour leurs cours dans PISA 2015 et 2018
<b>Engagement des parents</b>	Pourcentage de parents qui, de leur propre initiative, discutent des progrès de leur enfant avec un·e enseignant·e dans PISA 2015 et 2018	Pourcentage de parents qui participent aux décisions scolaires locales dans PISA 2015 et 2018	Pourcentage de parents bénévoles dans des activités physiques ou extrascolaires dans PISA 2015 et 2018

### 3.2.2 Standardisation et pondération des statistiques

Le calcul de l'ISQ nécessite la standardisation des données, car toutes n'ont pas les mêmes unités. Même si les valeurs aberrantes sont peu nombreuses, les distributions des données sont asymétriques et présentent souvent de longues queues. Dans de tels cas, il convient d'utiliser face aux valeurs aberrantes des mesures robustes et efficaces (variance de l'estimateur la plus petite) (Panagiotakos & Varin, 2020).

Pour standardiser les données, dès 2015, l'indice de tendance centrale retenu est le *Hodges-Lehmann (HL) estimateur* (qui remplace la moyenne), estimateur efficace à 95%. Il est aussi assez robuste, car il tolère 29% de valeurs aberrantes avant d'être inefficace. Comme indice de dispersion des données, il a été fait recours au *ScaleTau2 estimateur (Robust Tau-estimate of scale)* (qui remplace la variance) qui est très efficace à 95% et robuste à 50%. Les scores standardisés sont ensuite retransformés en échelle « PISA » de sorte que l'indice de tendance centrale (*HL estimateur*) est égal à 500 et celui de dispersion (*ScaleTau2 estimateur*) égal à 100.

Le score de chaque critère est la moyenne arithmétique des trois statistiques qui le composent. L'ISQ est alors le résultat de la moyenne arithmétique des scores finaux des cinq/six critères.

Chaque statistique a ainsi le même poids dans l'évaluation des critères qui sont eux-mêmes équipondérés. C'est un choix effectué lors des travaux du Séminaire qui a initié l'ISQ. Il est inhérent à toute évaluation qui relève toujours d'une certaine subjectivité (Gerard, 2017).

### **3.2.3 Prise en compte inversée de la valeur des statistiques**

Six statistiques sur les dix-huit retenues pour notre étude ont subi une attribution de points inversée (plus la valeur des dites statistiques est élevée, moins le score de l'ISQ l'est). Il s'agit des statistiques suivantes : pourcentage d'élèves peu performant·e·s en littéracie au PISA 2015 et 2018 (c'est-à-dire en dessous du niveau 2 dans l'échelle PISA qui en comporte 5) ; pourcentage d'élèves qui manquent des cours ; pourcentage d'enseignant·e·s absent·e·s ; pourcentage d'enseignant·e·s n'étant pas bien préparé·e·s pour leurs cours ; le score en littéracie associé à une augmentation d'une unité de l'indice ESCS (indice mesurant le statut économique, social et culturel des élèves) et la différence de performance en littéracie entre les élèves du quart supérieur et les élèves du bas de l'ESCS.

## **3.3 Méthodes d'analyse des données**

### **3.3.1 Outils pour mesurer l'évolution des scores des pays entre l'ISQ 2015 et l'ISQ 2018**

Les paragraphes suivants présentent les outils statistiques permettant de vérifier la validité de la première hypothèse : « les systèmes éducatifs n'évoluent pas tous au même rythme ni dans le même sens, tout au contraire, certains pouvant progresser de 5% tandis que d'autres peuvent au contraire régresser du même pourcentage ». L'intérêt de cette vérification est qu'elle permet d'identifier les critères ayant le plus influencé ces évolutions.

Afin de déterminer si la moyenne des scores de 2018 a augmenté par rapport à celle de 2015, sont calculées la moyenne arithmétique ainsi que les variations positives ou négatives en pourcentage, ce qui est souvent plus facile à interpréter (Tableau 2, présenté en 4.1.1). La même méthode est utilisée pour les moyennes des cinq critères (Tableau 3, présenté en 4.1.1). Afin de discerner si les scores de l'ISQ 2015 sont corrélés à ceux de l'ISQ 2018, un graphique (nuage de points et droite de régression non paramétrique – qui minimise l'erreur quadratique moyenne – entre les deux séries de scores) s'avère être un outil performant pour l'analyse (Varin, 2020). Puisque la relation est non linéaire et monotone (Figure 1, présentée en 4.1.1), pour déterminer si les scores ISQ 2015 et ISQ 2018 sont corrélés, le coefficient tau de corrélation de Kendall est le plus efficace (Croux et Dehon, 2010 ; Varin, 2016) pour mesurer l'association entre les deux séries de scores (ISQ 2015 et ISQ 2018). Il est compris entre -1 et +1 et s'interprète comme un coefficient de Pearson.

### **3.3.2 Outils pour mesurer l'influence du critère « engagement des parents »**

Pour vérifier la validité de la deuxième hypothèse « l'introduction de l'engagement des parents en tant que sixième facteur explicatif de la qualité des systèmes éducatifs des pays de l'OCDE modifie le classement et le score des pays », les instruments statistiques suivants sont utilisés :

le coefficient W de Kendall analyse la concordance entre les rangs, en permettant, grâce à une variante, d'intégrer les scores ex-aequo. Le coefficient W de Kendall varie entre 0 (discordance parfaite) et 1 (concordance parfaite). La relation entre les scores ISQ 2018 incluant le critère d'engagement des parents et les scores ISQ 2018 l'excluant, est non linéaire et monotone comme en témoigne la

Figure 2 (nuage de points et droite de régression non paramétrique) (voir 4.2.3).

### 3.3.3 Outils pour vérifier si l'ISQ est un outil valide et fiable

Afin de valider la fiabilité l'ISQ, sont calculés les coefficients de corrélation partielle de Kendall, car les relations entre les différentes statistiques ne sont pas linéaires mais monotones. Les coefficients de corrélation partielle mesurent la liaison entre deux variables lorsque l'influence d'une troisième (voire quatrième, cinquième, etc.) est gardée constante sur les deux premières qui sont comparées. Il n'est en effet pas rare qu'une (ou plusieurs) autre(s) variable(s) faussent la corrélation entre deux variables, laissant penser à tort à l'existence (ou l'absence) d'une liaison entre les deux variables. La « moyenne » calculée des corrélations a été également élaborée (Bushman et Wang, 1995 ; Corey, Dunlap et Burke, 1998).

## 4. Résultats

### 4.1 Évolution des scores moyens de l'ISQ 2015 et 2018 concernant les 35 pays

#### 4.1.1 Résultats et analyses

Le Tableau 2 ci-après montre que la moyenne des scores de l'ISQ 2018 n'a que légèrement augmenté, de 4 points (+0,8%) par rapport à la moyenne de l'ISQ 2015. Le tau de Kendall s'élève à 0,72, preuve d'une liaison positive assez forte entre les deux séries. Cependant, des différences importantes peuvent être relevées dans les scores de certains pays comme le montre le Tableau 2.

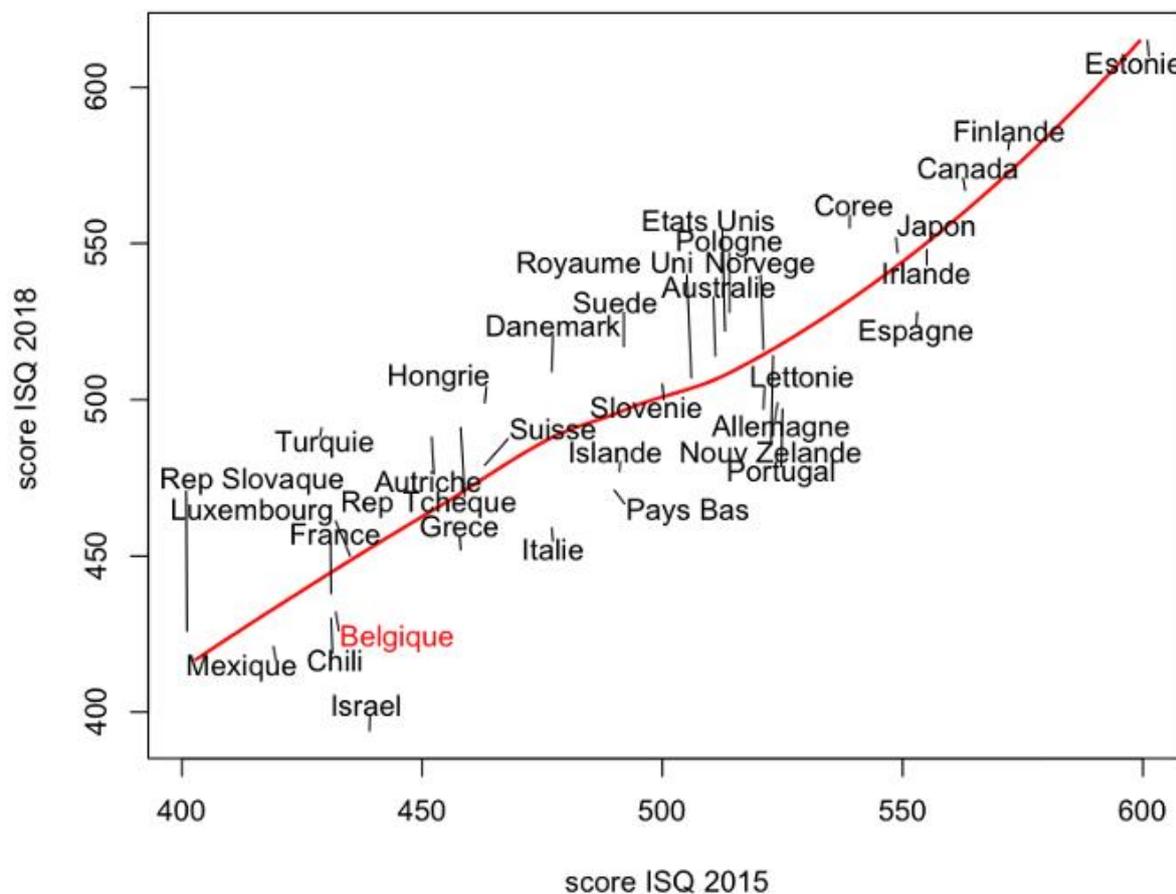
Tableau 2 - Scores ISQ 2015, score ISQ 2018 et différence en pourcentage entre le score 2015 et 2018

Pays	ISQ 2015	ISQ 2018	Différence en % entre ISQ 2015/2018
Allemagne	524	499	-4,8%
Australie	511	514	0,6%
<b>Autriche</b>	452	488	<b>8%</b>
Belgique	432	432	0,0%
Canada	563	567	0,7%
Chili	431	430	-0,2%
Corée	539	555	3%
<b>Danemark</b>	477	509	<b>6,7%</b>
Espagne	553	528	-4,5%
Estonie	601	615	2,3%
Etats-Unis	513	522	1,7%
Finlande	572	580	1,4%
France	431	438	1,6%
Grèce	458	452	-1,3%
<b>Hongrie</b>	463	499	<b>7,8%</b>
Irlande	555	548	-1,3%
Islande	491	477	-2,8%
<b>Israël</b>	439	394	<b>-10,2%</b>
Italie	477	459	-3,8%

Japon	549	547	-0,4%
Lettonie	521	497	-4,6%
Luxembourg	435	450	3,4%
Mexique	419	421	0,5%
Norvège	521	516	-0,9%
Nouvelle-Zélande	523	514	-1,7%
Pays-Bas	490	471	-3,9%
Pologne	514	528	2,7%
<b>Portugal</b>	525	497	<b>-5,3%</b>
<b>République slovaque</b>	401	426	<b>6,2%</b>
<b>République tchèque</b>	458	491	<b>7,2%</b>
Royaume Uni	506	507	0,2%
Slovénie	500	505	1%
<b>Suède</b>	492	517	<b>5,1%</b>
Suisse	463	479	3,5%
<b>Turquie</b>	429	491	<b>14,5%</b>
<i>Moyennes</i>	492	496	0,8%

Au total, 7 pays ont enregistré une hausse notable de plus de 5% de leur score et 2 seulement une baisse supérieure à 5%. Comme le montre la Figure 1, les pays se situant au-dessus de la courbe rouge sont ceux qui ont augmenté leur score et inversement pour ceux se trouvant sous cette courbe. La Belgique, qui se trouve sur la courbe, est le seul pays dont le score ISQ n'a pas évolué de 2015 à 2018.

**Figure 1 - Courbe de régression non paramétrique entre ISQ 2015 et ISQ 2018**



Les sept pays présentant des pourcentages d'évolution positive supérieurs à plus de 5% sont l'Autriche, le Danemark, la Hongrie, la République tchèque, la République slovaque, la Suède et la Turquie (14,7%) ; et les deux pays présentant une baisse supérieure à 5% sont Israël et le Portugal. Il est à noter que les résultats de la Turquie doivent être considérés avec une extrême prudence tant certaines évolutions spécifiques prêtent à questionnement, comme le montre le Tableau 3.

Afin d'identifier les critères dont l'évolution entre 2015 et 2018 ont entraîné ces hausses et ces baisses notaires, le Tableau 3 détaille les scores 2015 et 2018 pour chacun des cinq critères et les évolutions en pourcentage. Pour rappel, l'engagement des parents ne figure pas dans ce tableau, car pour l'année 2015, faute de données disponibles, nous n'avions pas pu retenir ce critère.

**Tableau 3 - Scores des critères en 2015 et en 2018 et écarts en pourcentage entre les scores de 2015 et 2018**

Pays	Efficacité			Efficience			Equité			Engagement des élèves			Engagement des enseignant-e-s		
	2015	2018	Δ%	2015	2018	Δ%	2015	2018	Δ%	2015	2018	Δ%	2015	2018	Δ%
Allemagne	568	533	-6,2	608	578	-4,9	498	450	-9,6	544	529	-2,8	403	405	0,5
Australie	541	562	+3,9	488	486	-0,4	487	507	4,1	508	505	-0,6	531	508	-4,3
Autriche	451	459	+1,8	473	465	-1,7	430	470	9,3	413	530	28,3	491	516	5,1
Belgique	510	505	-1	449	428	-4,7	419	416	-0,7	426	419	-1,6	353	390	10,5
Canada	649	640	-1,4	564	570	1,1	583	622	6,7	500	486	-2,8	522	520	-0,4
Chili	329	319	-3	447	377	-15,7	469	470	0,2	568	608	7,0	341	375	10,0
Corée	605	606	+0,2	502	496	-1,2	530	570	7,5	526	598	13,7	529	506	-4,3
Danemark	511	539	+5,5	420	474	12,9	540	538	-0,4	484	519	7,2	431	478	10,9

Espagne	488	499	+2,3	567	547	-3,5	594	498	<b>-16,2</b>	595	583	-2,0	520	513	-1,3
Estonie	609	651	<b>+6,9</b>	685	744	<b>8,6</b>	637	661	3,8	512	486	<b>-5,1</b>	565	534	<b>-5,5</b>
Etats-Unis	513	571	<b>+11,3</b>	468	485	3,6	556	476	<b>-14,4</b>	530	542	2,3	499	535	<b>7,2</b>
Finlande	643	635	-1,2	597	614	2,8	563	564	0,2	529	532	0,6	530	553	4,3
France	530	505	-4,7	534	506	<b>-5,2</b>	328	405	<b>23,5</b>	322	289	<b>-10,2</b>	441	485	<b>10,0</b>
Grèce	366	344	<b>-6</b>	389	478	<b>22,9</b>	448	487	<b>8,7</b>	586	474	<b>-19,1</b>	498	478	-4,0
Hongrie	371	422	<b>+13,7</b>	557	600	7,7	356	363	2,0	480	571	<b>19,0</b>	552	542	-1,8
Irlande	611	623	+2	564	597	<b>5,9</b>	521	570	<b>9,4</b>	567	458	<b>-19,2</b>	513	491	-4,3
Islande	442	424	-4,1	385	338	<b>-12,2</b>	552	552	0,0	561	546	-2,7	513	523	1,9
Israël	437	420	-3,9	449	364	<b>-18,9</b>	417	348	<b>-16,5</b>	488	476	-2,5	403	363	<b>-9,9</b>
Italie	445	429	-3,6	530	491	<b>-7,4</b>	513	540	<b>5,3</b>	450	376	<b>-16,4</b>	445	456	2,5
Japon	589	556	<b>-5,6</b>	615	625	1,6	570	583	2,3	548	474	<b>-13,5</b>	424	498	<b>17,5</b>
Lettonie	456	435	-4,6	510	518	1,6	594	583	-1,9	561	419	<b>-25,3</b>	486	529	<b>8,8</b>
Luxembourg	437	406	<b>-7,1</b>	412	405	-1,7	372	389	4,6	435	479	<b>10,1</b>	519	574	<b>10,6</b>
Mexique	179	178	-0,6	321	325	1,2	537	508	<b>-5,4</b>	567	615	<b>8,5</b>	492	478	-2,8
Norvège	587	542	<b>-7,7</b>	469	427	<b>-9,0</b>	554	567	2,3	516	532	3,1	479	512	<b>6,9</b>
Nouvelle-Zélande	580	571	-1,6	480	479	-0,2	454	475	4,6	521	473	<b>-9,2</b>	579	571	-1,4
Pays-Bas	540	473	<b>-12,4</b>	509	445	<b>-12,6</b>	469	519	<b>10,7</b>	537	507	<b>-5,6</b>	397	411	3,5
Pologne	540	597	<b>+10,6</b>	594	661	<b>11,3</b>	518	505	-2,5	407	392	-3,7	511	487	-4,7
Portugal	505	491	-2,8	506	454	<b>-10,3</b>	550	524	-4,7	531	496	<b>-6,6</b>	535	518	-3,2
République slovaque	310	348	<b>+12,3</b>	431	458	<b>6,3</b>	384	384	0,0	374	426	<b>13,9</b>	504	515	2,2
République tchèque	464	492	<b>+6</b>	593	590	-0,5	369	407	<b>10,3</b>	372	468	<b>25,8</b>	493	499	1,2
Royaume Uni	517	563	<b>+8,9</b>	418	467	<b>11,7</b>	552	552	0,0	562	444	<b>-21,0</b>	483	507	<b>5,0</b>
Slovénie	542	513	<b>-5,4</b>	555	540	-2,7	533	523	-1,9	406	480	<b>18,2</b>	465	467	0,4
Suède	525	576	<b>+9,7</b>	491	493	0,4	467	489	4,7	460	490	<b>6,5</b>	517	536	3,7
Suisse	483	465	-3,7	414	432	4,3	471	443	<b>-5,9</b>	420	499	<b>18,8</b>	528	555	<b>5,1</b>
Turquie	200	380	<b>+90</b>	359	526	<b>46,5</b>	590	531	<b>-10,0</b>	453	513	<b>13,2</b>	542	508	<b>-6,3</b>
<b>Moyenne</b>	<b>487,8</b>	<b>493,5</b>	<b>+1,17</b>	<b>495,8</b>	<b>499,5</b>	<b>0,75</b>	<b>497,9</b>	<b>499,7</b>	<b>0,36</b>	<b>493,1</b>	<b>492,4</b>	<b>-0,14</b>	<b>486,7</b>	<b>495,3</b>	<b>1,77</b>

Le Tableau 4 ci-après regroupe pour chaque critère les pays pour lesquels une augmentation ou une diminution d'au moins 5% a été enregistrée. Un même critère peut augmenter dans certains pays et diminuer dans d'autres pays.

**Tableau 4 - Augmentation ou diminution d'au moins 5% des 5 critères de 2015 à 2018**

Critères	Augmentation	Diminution
<i>Efficacité</i>	Danemark, Estonie, États-Unis, Hongrie, Pologne, Rép. slovaque, Rép. tchèque, Royaume-Uni, Suède, Turquie	Allemagne, Grèce, Japon, Luxembourg, Norvège, Pays-Bas, Slovénie
<b>Total (nombre de pays)</b>	<b>10</b>	<b>7</b>
<i>Efficience</i>	Danemark, Estonie, Grèce, Hongrie, Irlande, Pologne, Rép. slovaque, Royaume-Uni, Turquie	Chili, France, Islande, Israël, Italie, Norvège, Pays-Bas, Portugal
<b>Total (nombre de pays)</b>	<b>9</b>	<b>8</b>
<i>Équité</i>	Autriche, Canada, Corée, France, Grèce, Irlande, Italie, Pays-Bas,	Allemagne, Espagne, États-Unis, Israël, Italie, Mexique, Suisse,

<b>Total (nombre de pays)</b>	Rép. tchèque <b>9</b>	Turquie <b>8</b>
<i>Engagement des élèves</i>	Autriche, Chili, Corée, Danemark, Hongrie, Luxembourg, Mexique, Rép. slovaque, Rép. tchèque, Slovénie, Suède, Suisse, Turquie	Estonie, France, Grèce, Irlande, Italie, Japon, Lettonie, Nouvelle-Zélande, Portugal, Royaume-Uni
<b>Total (nombre de pays)</b>	<b>13</b>	<b>10</b>
<i>Engagement des enseignant·e·s</i>	Autriche, Belgique, Chili, Danemark, États-Unis, France, Japon, Lettonie, Luxembourg, Norvège	Estonie, Israël, Turquie
<b>Total (nombre de pays)</b>	<b>10</b>	<b>3</b>

Grâce à ce tableau, on relève que sur la somme des variations possibles (positives et négatives) des 5 critères, soit 175 (35 pays x 5 critères), 51 sont des augmentations et 36 des diminutions, soit en tout 87 variations de 5% ou plus, c'est-à-dire la moitié des variations possibles. Dans l'ordre décroissant, les critères qui sont le plus souvent en augmentation sont l'engagement des élèves, l'engagement des enseignant·e·s, l'efficacité, l'efficience et l'équité. À l'inverse, les critères le plus souvent en diminution sont l'engagement des élèves, l'efficience, l'équité, l'efficacité et l'engagement des enseignant·e·s. Les critères présentant le solde positif le plus élevé sont l'engagement des enseignant·e·s (+7), l'efficacité et l'engagement des élèves (+3), puis l'équité et l'efficience (+1).

Une analyse détaillée des politiques éducatives des pays concernés par ces évolutions serait nécessaire afin d'être en mesure d'identifier celles ayant pu favoriser ces évolutions, mais – à ce stade – on peut conclure qu'il n'y a pas un modèle suivant lequel l'ISQ d'un pays s'améliore ou pas au sens où il existerait une combinaison optimale des six indicateurs. En revanche, il s'avère que certains critères contribuent plus que d'autres, d'une part à l'augmentation de l'ISQ et, d'autre part, à sa diminution. Dans le premier cas, il s'agit de l'engagement des élèves, de celui des enseignants et l'efficacité ; et dans le second, de l'engagement des élèves, de l'équité et de l'efficience. Ces constats sont affinés dans notre article sur l'utilisation des réseaux bayésiens, en particulier le DAG (directed acyclic graph), dans l'analyse des causalités (Gerard, Hugonnier & Varin, 2023), ainsi que dans un texte en préparation sur la mise en relation de nos résultats et des mesures politiques prises par les pays de l'OCDE (OECD, 2019).

#### **4.1.2 Synthèse**

L'objectif est ici d'évaluer la mesure dans laquelle la qualité des systèmes éducatifs des pays de l'OCDE – mesurée par un score moyen – a évolué entre 2015 et 2018 et d'identifier parmi les cinq critères ceux expliquant cette évolution.

En utilisant une courbe de régression non paramétrique, nous avons établi qu'il y a davantage de pays qui ont vu leur score ISQ augmenter de 2015 à 2018 que l'inverse, cette évolution étant cependant de faible ampleur. Cette conclusion est confirmée par le coefficient tau de Kendall. Il convient malgré tout de relever que 9 pays sur 35 présentent une évolution de leur ISQ positive ou négative, supérieure à 5% (sans qu'on puisse identifier de traits communs entre les pays concernés).

Si les cinq critères contribuant à l'évaluation de la qualité des systèmes éducatifs présentent tous des augmentations qui, de 2015 à 2018, sont supérieures en nombre à celui des diminutions, ceux présentant le solde positif le plus élevé sont l'engagement des enseignant·e·s (+7), l'efficacité et l'engagement des élèves (+3), puis l'équité et l'efficience (+1). Sur la base de ce résultat, on peut inférer que l'engagement des enseignant·e·s est le critère ayant le plus contribué à l'amélioration des scores de certains pays entre 2015 et 2018.

## 4.2 Classement des 35 pays avec et sans le critère d'engagement des parents

### 4.2.1 Résultats et analyses

Nous présentons dans le Tableau 5 le classement et le score des 35 pays pour les données 2018 avec et sans le critère d'engagement des parents. Nous souhaitons ainsi étudier l'influence de ce sixième critère sur le classement et le score des 35 pays.

**Tableau 5 - Classement des 35 pays avec et sans le critère d'engagement des parents**

Pays	Classement ISQ 2018 avec l'engagement des parents	Classement ISQ 2018 sans l'engagement des parents	Score ISQ 2018 avec l'engagement des parents	Score ISQ 2018 sans l'engagement des parents
Allemagne	20	17	490	499
Australie	16	12	506	514
Autriche	20	23	490	488
Belgique	34	31	425	432
Canada	4	3	554	567
Chili	27	32	467	430
Corée	2	4	560	555
Danemark	15	14	507	509
Espagne	6	7	533	528
Estonie	1	1	604	615
Etats-Unis	6	9	533	522
Finlande	3	2	559	580
France	33	30	437	438
Grèce	25	28	483	452
Hongrie	20	17	490	499
Irlande	9	5	526	548
Islande	26	25	476	477
Israël	35	35	422	394
<b>Italie</b>	<b>20</b>	<b>27</b>	<b>490</b>	<b>459</b>
Japon	8	6	529	547
Lettonie	18	19	500	497
Luxembourg	32	29	446	450
<b>Mexique</b>	<b>27</b>	<b>34</b>	<b>467</b>	<b>421</b>
Norvège	12	11	508	516
Nouvelle-Zélande	12	12	508	514
Pays-Bas	29	26	465	471
Pologne	5	7	539	528
<b>Portugal</b>	<b>12</b>	<b>19</b>	<b>508</b>	<b>497</b>
République slovaque	31	33	450	426
République tchèque	24	21	485	491
Royaume Uni	19	15	494	507
Slovénie	17	16	503	505
Suède	11	10	513	517

<i>Suisse</i>	<i>30</i>	<i>24</i>	<i>461</i>	<i>479</i>
<i>Turquie</i>	<i>10</i>	<i>21</i>	<i>520</i>	<i>491</i>
Moyennes			498,4	496,1

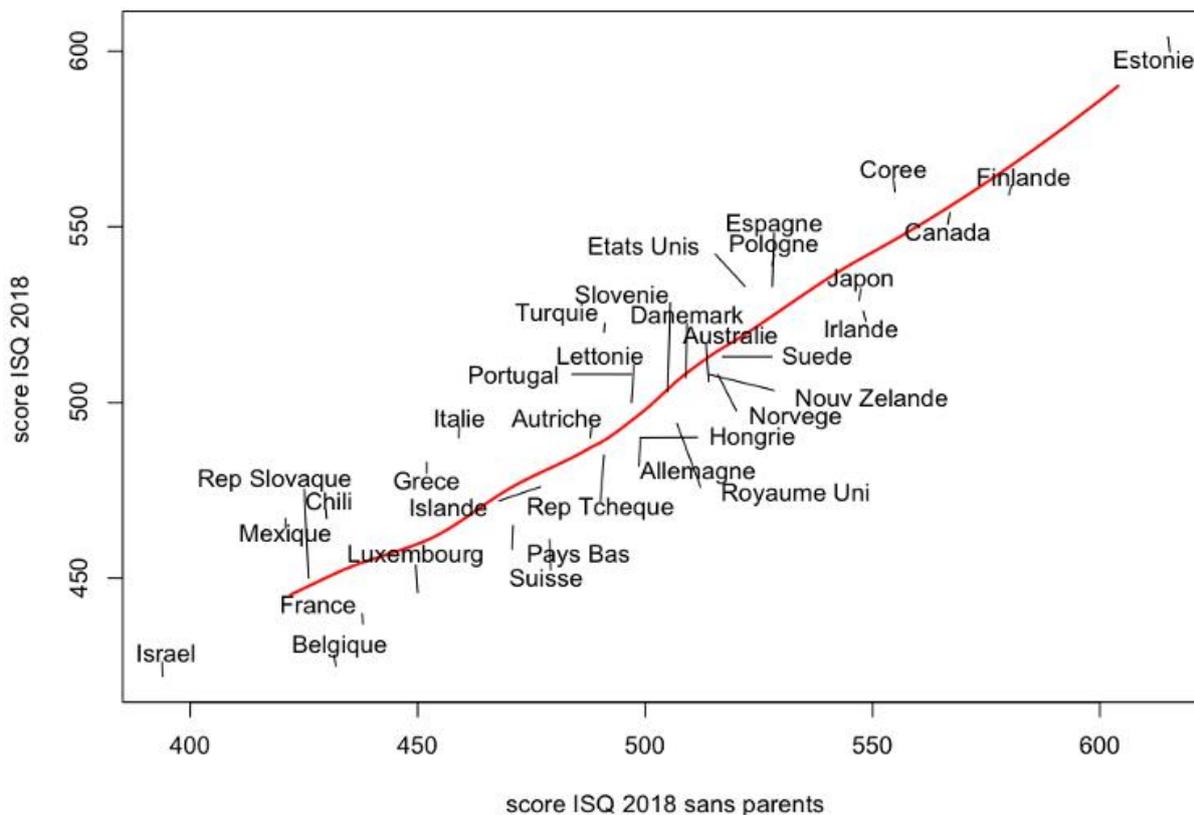
Nous avons considéré les pays dont le classement dans l'ISQ 2018 diffère de plus de cinq places/rangs selon qu'on prend ou pas en compte le critère d'engagement des parents. Figurent en gras italique, les cinq pays concernés. Sur cette base, l'Italie, le Mexique, le Portugal et la Turquie rétrogradent dans le classement si l'on prend en compte ce critère ; alors qu'un seul pays, la Suisse, améliore son rang dans ce cas. On peut donc inférer sur cette base que le critère d'engagement des parents a très peu influencé dans son ensemble l'évolution du classement des pays entre 2015 et 2018.

Cette conclusion est confirmée par le coefficient W de Kendall qui analyse la concordance entre les rangs des deux classements. Ce coefficient est en effet très élevé ( $W = 0,97$ ) ce qui confirme la forte concordance entre les deux classements et que le critère d'engagement des parents ne modifie pas le classement au niveau de l'ISQ.

Qu'en est-il des scores ISQ ? L'analyse statistique des différences des scores entre le classement avec le critère d'engagement des parents et celui sans ce critère a également été menée grâce à la Figure 2 (nuage de points et droite de régression non paramétrique s'y référant) et au calcul du coefficient tau de corrélation de Kendall. La valeur de celui-ci est de 0,80, ce qui établit une relation positive élevée entre les scores ISQ 2018 avec ou sans ce critère.

La Figure 2 permet également de visualiser le positionnement des pays en fonction de la courbe de régression.

Figure 2 - Courbe de régression non paramétrique entre ISQ 2018 et ISQ 2018 sans parents



Sur ce graphique, les pays se trouvant au-dessus de la courbe rouge sont ceux dont le score ISQ sans l'engagement des parents est inférieur à leur score ISQ en tenant compte de ce critère. En dessous de cette courbe, se trouvent les pays dont le score ISQ sans engagement des parents est plus élevé que leur score avec. Ici, aucun pays n'est inscrit en rouge, car aucun d'eux ne se trouve sur la courbe. Plus les pays sont éloignés de la courbe, plus la différence entre leurs deux scores est élevée. Les pays ne sont pas très éloignés de la courbe ce qui démontre une liaison assez forte, comme l'atteste le tau de Kendall.

#### 4.2.2 Influence des autres critères

Si, selon nos analyses statistiques, l'introduction du critère d'engagement des parents modifie peu la qualité des systèmes éducatifs en termes de scores et de rangs pour 2018, on peut se demander si le même résultat serait obtenu en retirant les cinq autres critères un à un. Nous l'avons vérifié, ce qui a permis d'établir (voir Tableau 6) que quatre critères ont un faible impact sur les classements et les scores. En effet, considérons le critère d'équité. Si on retire ce critère de l'ISQ, on s'aperçoit que tant le coefficient de concordance « W » que le tau de Kendall sont très élevés ( $W=0,98$  et  $\tau=0,85$ ). Cela signifie donc que sans le critère d'équité, les classements ( $W=0,98$ ) et les scores ( $\tau=0,85$ ) changent peu. Donc, tout comme pour l'engagement des parents, l'impact du critère d'équité est faible sur l'ISQ 2018. Il en va de même pour les critères d'efficacité, d'engagement des élèves et d'engagement des enseignants. Le seul critère qui a un impact relativement fort sur l'ISQ 2018 est l'efficacité, notamment sur les scores. En effet, en retirant ce critère, les scores sont moins corrélés ( $\tau=0,69$ ) alors que le coefficient de concordance « W » reste lui relativement élevé ( $W=0,92$ ). Il est difficile à ce stade d'élaborer une hypothèse étayée à propos de ce constat de l'impact de l'efficacité. Cela pourrait refléter le fait que les performances des élèves sont encore souvent considérées comme le critère principal de la qualité des systèmes éducatifs. C'est par exemple le cas du site *Worldtop20* qui propose un classement de 201 pays, en se

basant uniquement sur des indicateurs d'efficacité<sup>1</sup>. Nos recherches sur les réseaux bayésiens et les relations causales semblent indiquer par ailleurs un rôle particulier de l'efficacité, elle-même influencée directement par l'engagement des parents, par l'efficience et par l'équité (Gerard, Hugonnier et Varin, 2023).

Enfin, au vu des résultats des coefficients de concordance « W » – tous assez proches les uns des autres et proches de 1 – et des taux de Kendall, nous relevons que l'impact des critères sur l'ISQ 2018 est plus fort sur les scores que sur les classements.

**Tableau 6 - W de Kendall et coefficient de corrélation tau de Kendall sur cinq critères**

	sans efficacité	sans efficience	sans équité	sans engagement élèves	sans engagement enseignant·e·s
W de Kendall (classements)	0,92	0,97	0,98	0,98	0,98
Tau de Kendall (scores)	0,69	0,80	0,85	0,82	0,86

#### 4.2.3 Synthèse

L'objet de cette section était d'analyser l'impact de l'introduction dans l'ISQ 2018 d'un sixième critère, l'engagement des parents, sur les rangs et les scores respectifs des pays. À la suite de l'introduction de ce nouveau critère, cinq pays (14,2%) gagnent ou perdent plus de cinq places/rangs entre les deux classements. Cela ne suffit cependant pas pour établir une différence notable entre les deux classements, ce que confirme le coefficient W de Kendall très élevé qui analyse la concordance entre les rangs des deux classements.

L'analyse de la modification des scores de chacun des pays montre, grâce au coefficient tau de Kendall, que l'introduction du sixième critère ne modifie pas non plus le score des pays. Il en ressort que l'engagement des parents n'influence pas de manière notable l'évaluation de la qualité des systèmes éducatifs des pays de l'OCDE.

Cela ne signifie pas que ce sixième critère n'a pas d'impact. Il peut en avoir dans certains pays. Nos travaux sur l'analyse des causalités à partir des réseaux bayésiens, en particulier le DAG, montrent d'ailleurs que si l'engagement des parents a peu d'influence directe sur la qualité globale des systèmes éducatifs, il en a de manière indirecte à travers le critère d'efficacité (Gerard, Hugonnier & Varin, 2023).

Globalement, la deuxième hypothèse n'est donc pas confirmée. Cependant, l'analyse de l'influence de chaque critère pris un à un sur le score ISQ montre qu'aucun des six critères n'influence de manière significative le score ou le classement ISQ final, si ce n'est légèrement le critère d'efficacité au niveau du score. Cela confirme que l'ISQ est un « indicateur synthétique » qui mesure la variation de la valeur d'une grandeur complexe – dans notre cas, la qualité des systèmes éducatifs – définie comme l'agrégation d'un ensemble de grandeurs élémentaires, à savoir les six critères. Même si ceux-ci sont indépendants l'un de l'autre, comme l'analyse de la validité interne (voir ci-dessous, 4.3.1) permet de le montrer, ils constituent chacun une dimension importante et spécifique de la qualité globale de la qualité des systèmes éducatifs.

### 4.3 Études de la validité et de la fiabilité de l'ISQ

La disponibilité des données et des résultats tant pour 2015 que pour 2018 nous permet d'étudier deux dimensions importantes de tout outil d'évaluation : sa validité et sa fiabilité.

<sup>1</sup> <https://worldtop20.org/worldbesteducationsystem>

### 4.3.1 Validité interne

Afin d'étudier la validité interne de l'ISQ, c'est-à-dire de vérifier la cohérence conceptuelle de l'indice synthétique lui-même, nous avons calculé les corrélations entre les statistiques relatives à un critère. Si elles sont élevées, cela signifierait qu'elles mesurent bien le même critère. Par contre, les corrélations entre les statistiques de deux critères devraient être moins élevées, car elles sont censées évaluer des dimensions différentes et indépendantes (De Ketele et Gerard, 2005).

Dans un premier temps, nous avons calculé les coefficients de corrélations partielles de Kendall pour les trois statistiques de chacun des cinq critères (voir Tableau 7) ; puis les corrélations partielles entre les statistiques liées à des critères (Tableau 8). Pour chacun des deux tableaux, les « moyennes » de ces corrélations sont données (Corey, Dunlap et Burke, 1998).

Les statistiques A, B et C correspondent aux trois statistiques relatives à chaque critère, énoncées dans le Tableau 1. La corrélation partielle A vs B (et C) est celle qui reste entre la variable A et B une fois enlevées toutes les corrélations dues à leurs relations communes avec la 3<sup>e</sup> variable (C dans notre exemple). Cela permet d'éviter qu'une corrélation masque (ou non) une 3<sup>e</sup> variable (l'effet « masque »).

**Tableau 7 - Coefficients de corrélation partielle de Kendall au sein de chaque critère**

Critère	ISQ de référence	Corrélation stat. A vs stat. B	Corrélation stat. A vs stat. C	Corrélation stat. B vs stat. C	Moyenne
Efficacité	2015	0,75	0,54	-0,12	<b>0,39</b>
	2018	0,75	0,59	-0,12	<b>0,46</b>
Efficience	2015	0,61	0,19	-0,02	<b>0,24</b>
	2018	0,65	0,15	0,003	<b>0,25</b>
Équité	2015	0,55	-0,09	0,27	<b>0,24</b>
	2018	0,61	-0,20	0,35	<b>0,34</b>
Engagement des élèves	2015	0,17	-0,04	0,08	<b>0,07</b>
	2018	-0,004	-0,22	0,16	<b>-0,02</b>
Engagement des enseignants	2015	-0,13	-0,44	0,24	<b>-0,11</b>
	2018	-0,29	-0,31	0,14	<b>-0,16</b>
Engagement des parents	2015	/	/	/	/
	2018	0,24	0,24	0,39	<b>0,29</b>

**Tableau 8 - Coefficients de corrélation partielle de Kendall sur les six critères**

	ISQ	Efficacité	Efficience	Équité	Engag. élèves	Engag. enseignant·e·s	Engag. parents	Moyenne
Efficacité	2015		0,45	0,18	-0,02	0,03	/	<b>0,15</b>
	2018		0,34	0,32	-0,04	0,1	-0,2	<b>0,1</b>
Efficience	2015	0,45		0,04	-0,04	-0,003	/	<b>0,11</b>
	2018	0,34		0,11	-0,14	0,04	0,03	<b>0,07</b>
Équité	2015	0,18	0,04		0,31	0,13	/	<b>0,16</b>
	2018	0,32	0,11		0,1	0,01	0,13	<b>0,13</b>
Engagement des élèves	2015	-0,02	-0,04	0,31		-0,05	/	<b>0,05</b>
	2018	-0,04	-0,14	0,1		0,12	0,15	<b>0,04</b>
Engagement enseignant·e·s	2015	0,03	-0,003	0,13	-0,05		/	<b>0,03</b>
	2018	0,1	0,04	0,01	-0,12		-0,12	<b>-0,02</b>
Engagement des parents	2015	/	/	/	/	/		/
	2018	-0,2	0,03	0,13	0,15	-0,12		<b>0,00</b>

Les coefficients de corrélation entre les critères d'efficacité et d'efficience sont relativement élevés, ce qui est logique dans la mesure où les scores d'efficience résultent du rapport entre

les scores d'efficacité et les moyens mis en œuvre. Ces deux critères ne sont donc pas totalement indépendants.

La comparaison des résultats donnés dans la dernière colonne des deux tableaux établit, tant pour 2015 que pour 2018, la validité interne de l'ISQ ; en effet, les moyennes des coefficients de corrélation des statistiques propres à un seul critère sont à chaque fois supérieures à celles des corrélations des statistiques relatives à plusieurs critères, et cela avec une constance remarquable entre les résultats de 2015 et de 2018. Les analyses confirment donc la validité interne de l'ISQ.

#### ***4.3.2 Fiabilité de l'ISQ***

Enfin, la comparaison des résultats de l'ISQ 2015 à ceux de l'ISQ 2018 apporte des informations sur la fiabilité de l'outil. D'une part, les corrélations partielles étudiées dans la section précédente révèlent une grande constance des mesures effectuées, ce qui atteste de la fiabilité des statistiques utilisées et de l'usage que nous en réalisons. D'autre part, l'absence de différences notoires entre les résultats 2015 et 2018, tant pour l'ISQ final que pour chacun des cinq critères communs, montre également la stabilité des mesures effectuées. Celle-ci peut être liée à la standardisation des statistiques, basée sur la technique PISA (valeur centrale de 500, indice de dispersion de 100). L'utilisation d'estimateurs robustes pour cette standardisation contribue néanmoins à la fiabilité de l'ISQ. Les données sont elles-mêmes issues des études PISA dans lesquelles la question de la fiabilité des outils de recueil d'informations et des résultats est traitée avec soin.

#### ***4.3.3 Synthèse***

La comparaison entre les résultats de 2015 et 2018 montre la fiabilité de l'ISQ. La validité interne de l'ISQ a pu également être confirmée : les statistiques liées à un critère sont bien corrélées et mesurent la même dimension, alors que les corrélations entre les critères sont faibles, ce qui indique l'indépendance entre ceux-ci.

## **5. Discussions**

Dans cette partie, nous rappelons et discutons les trois hypothèses de recherche l'une après l'autre.

### ***5.1 Évolution de l'ISQ de 2015 à 2018***

Les 35 pays de l'OCDE ont-ils vu la qualité – mesurée par un score moyen – de leur système éducatif évoluer de 2015 à 2018 ? Le score ISQ moyen qui permet de mesurer la qualité des systèmes éducatifs des pays de l'OCDE a légèrement augmenté de 2015 à 2018 (4 points ou 0,8%). Par ailleurs, le score moyen de chacun des cinq critères est relativement stable, il n'y a pas vraiment de variabilité entre 2015 et 2018. Toutefois, comme nous l'avons décrit, des différences notoires propres à certains pays ont été relevées et analysées, notamment par rapport à l'augmentation ou à la diminution de leurs scores entre 2015 et 2018, ce qui était l'objet de la deuxième partie de notre première hypothèse.

Deux explications peuvent être avancées :

- (1) alors que les scores globaux ou par critère de certains pays varient de manière notable (plus de 5%), on peut supposer qu'en trois ans, la qualité des systèmes éducatifs ne change pas fondamentalement, même si des mesures politiques sont prises pour les faire évoluer, car les effets des politiques ne sont jamais immédiats. De plus, les données PISA portent sur les résultats scolaires des élèves à la fin de leur scolarité scolaire (à l'âge de 15-16 ans). Ces résultats dépendent en grande partie de ce que les élèves auront vécu durant au moins une dizaine d'années de scolarité. Il est donc vraisemblable que les effets des modifications apportées aux politiques éducatives ne puissent se percevoir en termes de résultats (même sur les variables processus) que dans une diachronie beaucoup plus

longue. Il conviendra donc à cet égard d'observer si les données issues de la prochaine enquête PISA en 2023 mettront en avant de plus grandes variations ou confirment les tendances passées ;

- (2) la faible variabilité s'explique en partie par l'indispensable standardisation des données. Lorsqu'on calcule la somme ou la moyenne de statistiques, qui ont des unités différentes pour créer un score composite comme l'ISQ, il est indispensable de standardiser les données afin d'éviter qu'une statistique puisse avoir davantage d'impact sur la somme ou la moyenne uniquement en raison de son unité. Nous avons utilisé des estimateurs robustes de tendance centrale et de dispersion afin de permettre de mieux tenir compte des réelles distributions asymétriques des dix-huit statistiques. Cette manière de procéder permet de faire des analyses valides tout en pouvant mettre en évidence les variations des scores des ISQ.

## 5.2 Introduction de l'engagement des parents

Notre deuxième hypothèse de recherche consistait à vérifier si l'introduction du critère « engagement des parents » avait une influence sur l'ISQ. Cette hypothèse se fonde sur des travaux antérieurs de l'OCDE (Borgonovi et Montt, 2012 ; PISA, 2011) ainsi que sur les mesures qui sont prises actuellement dans de nombreux pays pour accroître l'implication des parents dans le processus éducatif de leurs enfants. Les résultats de nos analyses, à savoir l'absence d'influence significative du critère d'engagement des parents dans le score agrégé de l'ISQ, nous ont donc surpris. Une explication possible peut être liée à l'âge des élèves concerné·e·s par les études PISA : 15 ans. À cet âge, en effet, les élèves, on le sait, sont moins attentifs et sensibles aux conseils de leurs parents. De plus, d'autres travaux (Fan et Chen, 2001 ; Hoover-Dempsey, Battiato, Walker, Reed, DeJong and Jones, 2001 ; Organisation de coopération et de développement économiques [OCDE], 2019) semblent indiquer que l'implication des parents de jeunes de cet âge peut avoir un effet négatif sur eux, contrairement à ce qui semble se passer avec des enfants au début des apprentissages scolaires.

Cependant, nous avons aussi observé qu'aucun des autres critères constitutifs de l'indicateur synthétique de qualité des systèmes éducatifs, à savoir l'efficacité, l'efficience, l'équité et les engagements des élèves ou des enseignant·e·s, n'avait une influence spécifique sur l'ISQ. Celui-ci représente bien une valeur agglomérée des différents critères qui – ensemble – conduisent à des variations de l'ISQ, car ils synthétisent ce que nous considérons comme une représentation de la qualité d'un système éducatif. L'engagement des parents est un critère qui a son sens, comme le montrent les différents travaux en la matière. Il nous semble donc important de continuer à le prendre en compte lors de l'utilisation future des données PISA. En réalité, dans nos travaux, ce qui compte, ce ne sont pas tant les résultats obtenus à un instant donné, que l'évolution sur plusieurs années. Pour le moment, nous n'avons réellement pu comparer que les résultats issus des données 2015 et 2018. C'est évidemment une période trop courte pour pouvoir identifier de réelles évolutions que seules de nouvelles analyses sur de nouvelles données permettront ou non de mettre en évidence.

Il est enfin à noter que les six critères utilisés portent sur des réalités internes aux systèmes éducatifs. C'est donc bien la qualité intrinsèque de ces systèmes qui est mesurée par l'ISQ, alors même que d'autres variables pourraient se concentrer sur leur valeur extrinsèque ; mentionnons l'efficacité externe des systèmes éducatifs, c'est-à-dire les effets générés par les systèmes éducatifs observés hors de ceux-ci comme évaluer la productivité sociale et économique des individus sortis du système éducatif.

## 5.3 Validité et fiabilité de l'outil ISQ

L'indicateur synthétique de qualité a une validité interne élevée, c'est-à-dire que l'ISQ reflète vraiment la « qualité » d'un système éducatif tel que nous définissons celle-ci, à savoir la combinaison des cinq ou six critères. L'ISQ proposé reste néanmoins perfectible comme tout indicateur synthétique. Nous pouvons légitimement penser que d'autres critères, par exemple l'équilibre du système éducatif (prise en compte de divers types d'objectifs, diversité des programmes et des options d'apprentissage proposées...), sa stabilité (sa capacité à maintenir son fonctionnement et sa qualité de manière constante dans le temps), sa cohérence (manière dont tous les éléments du système fonctionnent ensemble de manière harmonieuse), ou encore la formation initiale et permanente des enseignants peuvent influencer sur la « qualité » du système éducatif. Cependant, par manque de données comparables pour tous les pays de l'OCDE, ces critères ne peuvent être pris en compte.

## 6. Conclusion

Après les calculs issus de données collectées en 2012 et en 2015 dans le cadre des enquêtes PISA, cette recherche présente les résultats du troisième calcul de l'indicateur synthétique de qualité des systèmes éducatifs (ISQ). On ne peut réellement comparer que les versions 2015 et 2018, car, les données de 2012 ne sont pas totalement compatibles avec celles de 2015 et 2018.

Le fait d'exister est en soi l'intérêt de l'indice synthétique de la qualité des systèmes éducatifs, car il comble un vide dans la recherche internationale. Intégrer de nouvelles dimensions à évaluer permettrait certainement d'accroître la portée et la signification de l'ISQ, mais il convient – au-delà de la pertinence d'une telle extension – de prendre en compte sa faisabilité. Notamment, il faut interroger la disponibilité et la fiabilité des données relatives à une dimension qu'on souhaiterait ajouter. C'est le cas avec l'insertion d'un sixième critère, l'engagement des parents, en espérant que les statistiques utilisées pour ce critère resteront disponibles dans les futures enquêtes PISA. La pertinence de ce nouveau critère nous semble évidente dans la mesure où il permet de prendre en compte un des acteurs importants des systèmes éducatifs. Néanmoins, on constate qu'il ne modifie fondamentalement ni le score global de l'ISQ, ni le classement des 35 pays sur la période 2015-2018.

On ne constate pas plus d'évolution notable dans les scores globaux et le classement entre l'ISQ 2015 et 2018. Cela ne signifie pas que la qualité des systèmes éducatifs n'a pas évolué dans un sens ou dans un autre. En effet, pour certains pays, on peut déceler des variations notoires, c'est-à-dire supérieures à 5%, positives ou négatives. Ces variations peuvent être observées tant au niveau global que pour certains critères, L'analyse de ces variations ne permet pas, à ce stade, de formuler des hypothèses. Il faudrait notamment pour cela pouvoir relier de manière qualitative les variations observées de l'ISQ avec les politiques éducatives de chacun des pays. C'est ce que nous tenterons de faire dans un prochain article.

## 7. Références

- Asselin de Beauville J.-P. & Varin, S. (2014). *Initiation au logiciel R : premiers pas en analyse de données. Survol théorique et applications avec Rcmdr et FactoMineR*. Academia-L'Harmattan.
- Bautier, É. et Marchand, J. (2020). Peut-on concilier qualité et équité dans les réformes en éducation ? *Revue internationale d'éducation de Sèvres*, 83, 85-93. <https://doi.org/10.4000/ries.9237>

- Behrens, M. (dir.) (2007). *La qualité en éducation : pour réfléchir à la formation de demain*. Presses de l'Université du Québec.
- Borgonovi, F. et Montt G. (2012). « Parental Involvement in Selected PISA Countries and Economies », *Documents de travail de l'OCDE sur l'éducation*, n° 73, Éditions OCDE, Paris, <https://doi.org/10.1787/5k990rk0jsjj-en>.
- Bouchard, C. et Plante, J. (2000). La qualité : sa définition et sa mesure. *Service social*, 47(1/2), 27-62.
- Bouchard, C. et Plante, J. (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, 219-236.
- Bushman, B. J., et Wang, M. C. (1995). A procedure for combining sample correlation coefficients and vote counts to obtain an estimate and a confidence interval for the population correlation coefficient. *Psychological Bulletin*, 117(3), 530–546. <https://doi.org/10.1037/0033-2909.117.3.530>
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297–312. <https://doi.org/10.1037/h0040950>
- 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*, 2(2), 125-140. <https://doi.org/10.3917/es.018.0125>
- Conférence des ministres de l'éducation des pays ayant le français en partage (2011). *Qualité de l'éducation: un enjeu pour tous: constats et perspectives: document de réflexion et d'orientation*. <https://www.confemen.org/wp-content/uploads/2019/06/DRO-SUR-LA-QUALITE-DE-L-EDUCATION.pdf>
- Corey, D. M., Dunlap, W. P. et Burke, M. J. (1998). Averaging Correlations: Expected Values and Bias in Combined Pearson rs and Fisher's z Transformations. *The Journal of General Psychology*, 125(3), 245-261. <http://dx.doi.org/10.1080/00221309809595548>
- Corriveau, L., Boyer, M. et Fernandez, N. (2009). La qualité en éducation: un enjeu de collaboration à cerner. *La revue de l'innovation : La Revue de l'innovation dans le secteur public*, Vol. 14(3), 1-24.
- Croux, C. et Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods and Applications*, 19(4), 497-515. <https://doi.org/10.1007/s10260-010-0142-z>
- De Ketele, J.-M. & Gerard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences, *Mesure et Évaluation en Éducation*, Volume 28, n°3, 1-26.
- 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, pp. 19-38.
- De Ketele, J.-M. et Roegiers, X. (1993). *Méthodologie du recueil d'informations*. De Boeck Université.
- De Landsheere, G. (1982). *Introduction à la recherche en éducation*. Georges Thone.
- Dominguez-Gil, C., Segovia-Gonzalez, M.M. & Contreras, I. (2022). A multiplicative composite indicator to evaluate educational systems in OECD countries, *Compare: A Journal of Comparative and International Education*, 52:8, 1296-1313, DOI : 10.1080/03057925.2020.1865791
- Fan, X. et Chen, M. (2001). Parental Involvement and Students' Academic Achievement: A Meta-Analysis. *Educational Psychology Review*, 13, 1–22. <https://doi.org/10.1023/A:1009048817385>

- Foin, S. et Gauthier, R.-F. (2020). Quelle mesure de la qualité dans les réformes éducatives ? *Revue internationale d'éducation de Sèvres*, 83, 95-104.
- Fortin, F. (1994). Propriétés métrologiques des instruments de mesure (fidélité-validité). *Recherche en Soins Infirmiers*, 39, 58-62.
- 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. (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, pp. 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, pp. 61-73.
- 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.
- Gerard, F.-M., Hugonnier, B. & Varin, S. (2023). Bayesian networks and causal discovery: what lessons for the synthetic indicator of the quality of education systems in OECD countries?, *International Journal of Research in Social Sciences*, Volume 13 Issue 1, January 2023, ISSN: 2249-2496, Thomson Reuters ResearcherID: L-5180-2015, 73-81.
- Hanushek, E. A. et Wößmann, L. (2012). *Education Quality and Economic Growth*. The World Bank. [https://www.inesss.qc.ca/fileadmin/doc/INESSS/Rapports/ServicesSociaux/INESSS\\_Outils\\_evaluation\\_risque\\_abus\\_physique\\_enfants.pdf](https://www.inesss.qc.ca/fileadmin/doc/INESSS/Rapports/ServicesSociaux/INESSS_Outils_evaluation_risque_abus_physique_enfants.pdf).
- Hoover-Dempsey, K.V., Battiato, A.C., Walker, J.M.T., Reed, R.P., DeJong, J.M. and Jones, K.P. (2001). Parental Involvement in Homework. *Educational Psychologist*, 36(3), 195–209.
- Institut national de la statistique et des études économiques (2016). *Indice synthétique – Définition*. <https://www.insee.fr/fr/metadonnees/definition/c1205>
- Institut national d'excellence en santé et en services sociaux (2015). *L'efficacité d'outils d'évaluation du risque d'abus physique envers les enfants* (J. Damasse, rédac. ; I. Beaumier et d'É.-L. Bussièrès, colla. ; ETMIS, vol. 11 no 3).
- Laveault, D. et Grégoire, J. (1997). *Introduction aux théories des tests en psychologie et en sciences de l'éducation* (2<sup>e</sup> édition). De Boeck Université.
- Murafa, C. Ikonomi, E., Todorakov, L., Mickovksa-Raleva, A., Brankovic, N., Georgiana Radu R. Ivan, C., Stankovic, D., Radisic, J. et Klemencic, E. (2012). *Quality assurance in Basic education in South Eastern Europe: What Works in Our Region?* Global Development Network, Open Society Institute Romanian Academic Society (RAS).
- Organisation de coopération et de développement économiques (2018). Résultats du PISA 2015 (Volume III) : *Le Bien être des élèves. Chapitre 9, L'implication des parents, la performance des élèves, et leur satisfaction à l'égard de leur vie*. Éditions OCDE.
- Organisation de coopération et de développement économiques (2019). *Résultats du PISA 2018 (Volume I) : Savoirs et savoir-faire des élèves*. Éditions OCDE. <https://doi.org/10.1787/ec30bc50-fr>
- Organisation for Economic Cooperation and Development (2019). *Education Policy Outlook 2019 – Working Together to Help Students Achieve their Potential*. OECD Publishing, Paris. <https://doi.org/10.1787/2b8ad56e-en>
- Programme international pour le suivi des acquis des élèves (2011). Réussite scolaire : comment les parents peuvent-ils aider leurs enfants ? *PISA à la loupe*, 2011-10, OCDE.

- R Core Team (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. <https://www.R-project.org/>.
- Roegiers, X. (1997). *Analyser une action d'éducation ou de formation*. De Boeck Université.
- Sall, H. N. et De Ketele, J.-M. (1997). L'évaluation du rendement des systèmes éducatifs : apports des concepts d'efficacité, d'efficience et d'équité. *Mesure et évaluation en éducation*, 19(3), 119-142.
- Stufflebeam D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O. et Provus, M. M. (1980). *L'évaluation et la prise de décision en éducation*. N.H.P.
- Van Driel, M. (2004). *Minerva Glossaire des termes utilisés en Evidence-Based Medicine*. Minerva.
- 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). « La qualité du collège enfin mesurée : onze cantons suisses sous la loupe », *L'Éducateur*, pp. 33-36.
- Varin, S. (2020). "Comparing the performances of Generalized additive models, Multivariate adaptive regression splines and polynomial linear models on a real and simulated datasets". *International Journal of Multidisciplinary Sciences and Advanced Technology (IJMSAT)* Vol 1 No 6: 10-35.
- Varin, S. (2021). « Classement de la qualité des systèmes éducatifs européens », in : *Variations critiques à propos du néolibéralisme : aspects philosophiques, politiques, économiques et éducatifs*. (2021). Sous la dir. de Varin, S. & Chancerel, J.-L. Thélème-L'Harmattan- Bruxelles.
- Varin, S. (2021a). Comment et pour quelles raisons évaluer la qualité des systèmes éducatifs ? Dans S. Varin & J.-L. Chancerel, *Variations critiques à propos du néo-libéralisme – Aspects philosophiques, politiques, économiques et éducatifs* (p. 51-101). Academia L'Harmattan.
- Varin, S. (2021b). La qualité des systèmes éducatifs européens comme déterminant de la croissance économique. Dans S. Varin & J.-L. Chancerel, *Variations critiques à propos du néo-libéralisme – Aspects philosophiques, politiques, économiques et éducatifs* (p. 201-223). Academia L'Harmattan.
- Varin, S. & Panagiotakos, DB. (2020). A review of robust regression in biomedical sciences' research, *Archives of Medical Sciences (AMS)* 6 (5) : 1267–1269.