

Indice socioéconomique Explication de la procédure de calcul

ISE des structures 2019-20

18 janvier 2021

Administration générale de l'Enseignement (AGE)

Direction générale du Pilotage du Système éducatif (DGPSE)

Direction de l'exploitation des données (DED)

1. Contexte

Le gouvernement de la FWB met annuellement à jour l'indice socioéconomique (ISE) des implantations, établissements et secteurs statistiques¹. Cette mise à jour est basée sur une formule de calcul déterminée au moyen d'une analyse statistique dite « en composantes principales » (ACP). Le décret du 30 avril 2009 précise que « les indices socioéconomiques de chaque secteur statistique, de chaque implantation et de chaque établissement sont exprimés à l'aide d'un indice composite dont une valeur faible indique un niveau socioéconomique moins favorisé. La formule de calcul de ces indices est établie par le Gouvernement et fait apparaître les variables déterminant chacun des critères et la méthode de détermination de la part de chacune d'elles dans le cadre de la formule de calcul. »

La formule en cours d'utilisation est fixée dans plusieurs arrêtés :

- l'arrêté du Gouvernement de la Communauté française du 19 juillet 2017 établissant le choix des variables et la formule de calcul de l'indice socioéconomique de chaque secteur statistique dont l'article 1 (annexe 1) précise les variables ;
- l'arrêté du Gouvernement de la Communauté française du 07 avril 2020 établissant le choix des variables et la formule de calcul de l'indice socioéconomique de chaque secteur statistique (ci-après, l'arrêté) dont l'article 1 (annexe 2) précise la formule de calcul et la méthode de détermination de la part de chaque variable dans le cadre de cette formule.

2. Processus général

La production de l'outil statistique qu'est l'ISE est le résultat d'un processus long :

- Extraction des données Stat Élèves sur 6 années consécutives (2013-14 à 2018-19) et du dernier comptage (2019-20) avec données d'identification par le Centre de Compétences en Business Intelligence (CCBI – ETNIC) ;
- Transfert des fichiers à la BCSS ;
- Extraction des variables de la BCSS sur 7 années civiles consécutives (2013 à 2019) avec données d'identification ;
- Création, à la BCSS, du fichier anonymisé combinant les données du comptage, de la Banque Carrefour de la Sécurité Sociale (BCSS) et de la Direction générale Statistique – Statistics Belgium (DGSSB) ;
- Calcul d'une valeur unique par indicateur et identifiant individuel ;
- Agrégation (moyenne) des variables individuelles par implantation, établissement et secteur statistique de résidence, sur base d'une procédure développée en SAS (logiciel statistique) par l'équipe interuniversitaire, corrigée et complétée par le CCBI ;
- Application de la méthode d'analyse en composantes principales sur base d'une procédure développée en SAS par l'équipe interuniversitaire et reproduite par le CCBI :
 - au fichier des implantations de l'enseignement fondamental ordinaire ;
 - au fichier des implantations de l'enseignement secondaire ordinaire de plein exercice ;
 - au fichier des établissements de l'enseignement fondamental ordinaire ;

¹ « Le secteur statistique est l'unité territoriale de base qui résulte de la subdivision du territoire des communes et anciennes communes par l'Institut national de Statistique pour la diffusion de ses statistiques à un niveau plus fin que le niveau communal. » (Jamagne, Lebrun, et Sajoitte 2012) Il s'agit d'un code en 9 positions.

- au fichier des établissements de l'enseignement secondaire ordinaire de plein exercice ;
- au fichier des secteurs statistiques ;
- Validation de la méthode par la Direction de l'Exploitation des Données (DED – DGPSE - AGE) ;

Calcul des ISE sur base des coefficients extraits de l'analyse en composante principale.

3. Les variables extraits de la BCSS

Sept variables sont extraites à la BCSS :

Revenu par habitant

1. Revenu en euro par unité de consommation (**rev_med**) : Le revenu est la moyenne de toutes les composantes des revenus du ménage de l'élève (tenant compte de la pondération suivante : 1 pour le 1^{er} membre de plus de 13 ans, 0,5 pour les membres de plus de 13 ans suivants et 0,3 pour les moins de 13 ans). Les élèves des ménages comportant au moins un travailleur frontalier sont exclus de l'extraction et reçoivent une valeur manquante ;

Niveau des diplômés

2. Appartenance du plus haut diplôme du ménage aux CITE 5 ou 6 (**etudes_haut**), à savoir un diplôme certifiant la réussite d'un programme d'enseignement supérieur (universitaire ou non-universitaire) ;
3. Appartenance du plus haut diplôme du ménage aux CITE 1 ou 2 (**etudes_bas**), à savoir, un diplôme certifiant la réussite de l'enseignement maternel ou primaire ;

Taux de chômage, le taux d'activité et le taux de bénéficiaires d'une aide sociale

4. Part des personnes de plus de 18 ans ayant un travail dans les ménages (**prop_trav**)
5. Obtention d'une aide sociale (**aide_sociale**) ;

Activités professionnelles

6. Part des ouvriers parmi les personnes du ménage qui ont un travail (**prop_ouv**) : Les élèves des ménages comportant au moins un travailleur frontalier sont exclus de l'extraction et reçoivent une valeur manquante ;
7. Part des personnes travaillant dans le secteur tertiaire bas parmi les personnes du ménage qui ont un travail tertiaire (**prop_ter**) : Les codes NACE sont convertis en 3 niveaux de secteur d'activité tertiaire (A=bas, B=intermédiaire et C=haut) sur base desquels l'indicateur peut être calculé. Les élèves des ménages comportant au moins un travailleur frontalier sont exclus de l'extraction et reçoivent donc une valeur manquante ;

Pour chaque élève, les variables relatives à son ménage sont extraites de la BCSS sur 7 années civiles successives (des valeurs pouvant toutefois être manquante pour une ou plusieurs années). Un calcul est ensuite effectué afin d'obtenir une valeur unique par variable pour chaque élève :

1. **rev_med** : Médiane des données disponibles sur 7 ans ;
2. **etudes_haut** : Diplôme le plus haut (maximum) sur 7 ans ;
3. **etudes_bas** : Diplôme le plus haut (maximum) sur 7 ans ;
4. **prop_trav** : Moyenne de la part sur 7 ans ;
5. **aide_sociale** : Obtention d'une aide sociale au moins une des 7 ;
6. **prop_ouv** : Moyenne des parts sur 7 ans ;

7. **prop_tert** : Moyenne des parts sur 7 ans.

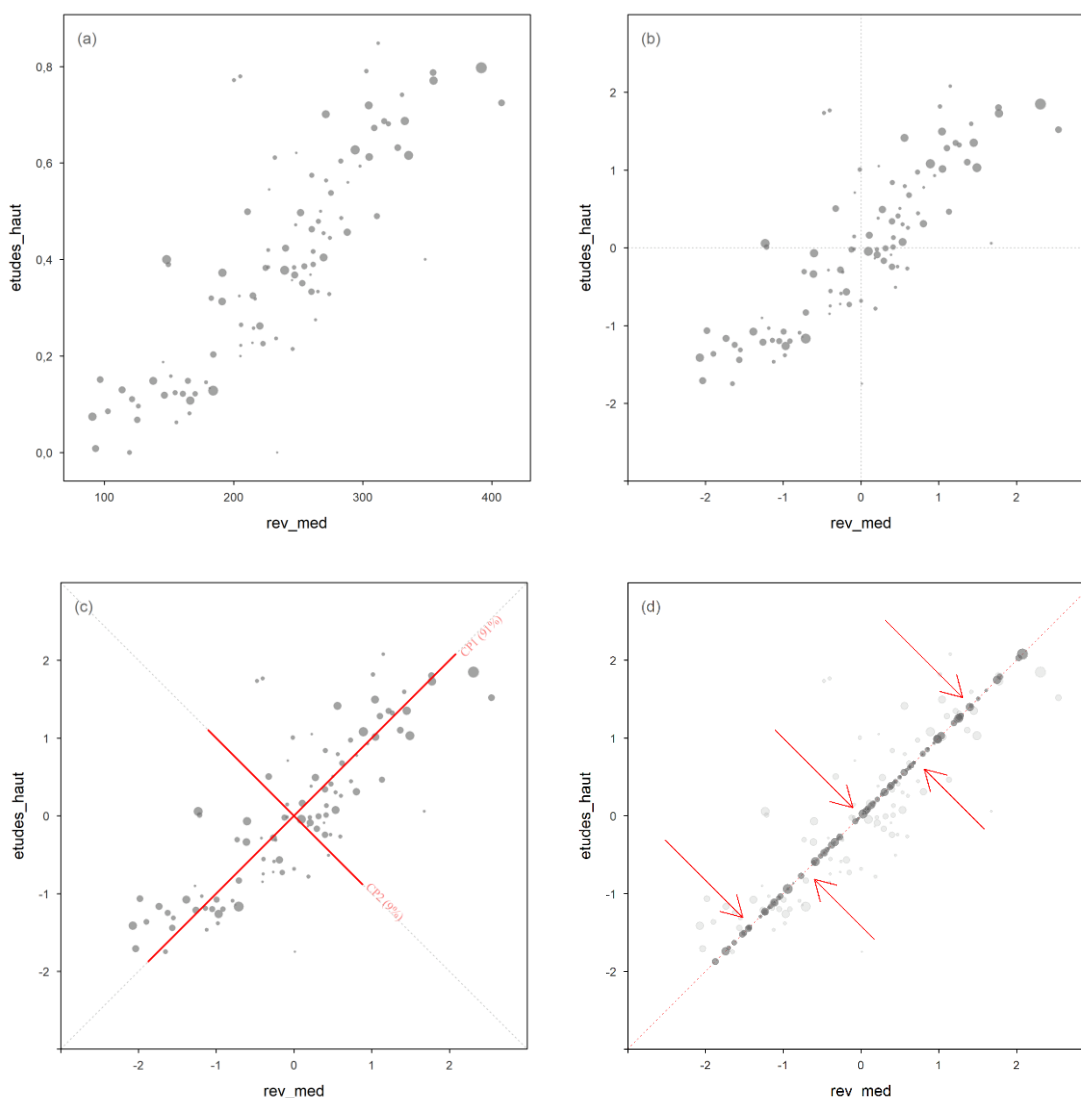
En pratique toutefois, les dernières données n'étaient pas disponibles. Les données précédentes ont alors été copiées. Pour l'exercice courant, les données de 2018 ont dès lors été copiées pour 2019. Cela signifie concrètement que nous disposons des données de 2013 à 2018. Ce choix a des conséquences diverses lors du calcul de la variable individuelle selon la méthode utilisée et les unités concernées : lorsqu'il s'agit d'une moyenne, cela équivaut à faire une moyenne pondérée où la donnée 2018 reçoit une pondération 2 fois plus importante que les années précédentes; pour le maximum cela équivaut à utiliser une fenêtre de 6 ans au lieu de 7 ans théoriques. Une exception demeure : pour le revenu médian, nous ne disposons que des 4 premières années.

Une fois qu'une valeur a été affectée à chaque élève, la moyenne des valeurs individuelles disponibles est calculée pour chaque implantation, établissement ou secteur. Dans le cas des indicateurs dichotomiques (2, 3, 5), calculer la moyenne est équivalent au calcul de la proportion des élèves ayant la caractéristique concernée sur la variable. Trois fichiers agrégés sont donc disponibles pour la suite de l'analyse.

4. L'analyse en composante principale

Une méthode appelée « analyse en composante principale » (ACP) est utilisée pour identifier les coefficients de la formule de calcul de l'indice socioéconomique.

Figure 1 : Exemple illustré de l'analyse en composante principale



La figure 1 fournit un exemple simplifié de la méthode de calcul :

- Considérons, pour les implantations, deux variables parmi les variables présentées ci-dessus : le revenu moyen des ménages des élèves de l'implantation (*rev_med*) et la part d'élève venant d'un ménage où un membre possède un diplôme élevé (*etudes_haut*) ;
- Le panneau (a) expose un nuage de points où chaque point représente une implantation. Chaque implantation peut être positionnée précisément dans ce repère grâce aux valeurs qu'elles ont sur les deux variables (ce qui constitue ainsi leurs coordonnées). Les variables étant corrélées positivement, les implantations qui présentent une valeur élevée sur la variable *rev_med* en abscisse (axe horizontal) ont tendance à également présenter une valeur élevée sur la variable *etudes_haut* en ordonnée (axe vertical). Aucun classement univoque n'est toutefois possible, car certaines implantations présentent à la fois une valeur plus élevée que d'autres implantations, sur la première variable et une valeur plus faible sur la seconde ;
- Les variables ont des échelles très différentes : *rev_med* est supérieur à 100 alors qu'*etudes_haut* est inférieur à 1.

- Avant d'appliquer l'ACP, ces variables sont transformées afin qu'elles varient sur une échelle similaire. Pratiquement, les variables sont « centrées » autour de 0 (leur moyenne est fixée à 0) et « réduites » (rendues « sans dimension » avec une variance de 1). Le panneau (b) montre les 2 variables centrées-réduites, variant à présent entre -3 et 3 ;
- Deux « composantes » sont calculées de telle sorte que la 1^{re} composante résume le maximum de l'information contenue dans les 2 variables initiales. Ces composantes sont dessinées en rouge et traversent littéralement le nuage de points du panneau (c).
- Ces composantes représentent 2 nouveaux axes qui forment un nouveau repère dans lequel les implantations ont de nouvelles coordonnées.
- Les coordonnées des implantations sur la 1^{re} composante correspondent à la valeur de l'indice dit « socioéconomique ». Un classement univoque est ainsi possible sur la 1^{re} composante (panneau (d)) si l'on accepte la perte (limitée) de l'information fournie par la seconde composante.

Dans un langage plus technique, l'ACP est une méthode statistique robuste qui permet de synthétiser l'information fournie par n variables initiales corrélées entre elles en situant les unités dans un nouvel espace dont les axes sont n composantes principales non corrélées entre elles. En maximisant la variation des données sur les premières composantes principales, la méthode permet la sélection d'un sous-ensemble de composantes résumant une part suffisante de ladite variation. En d'autres termes, la méthode vise à créer un ensemble non redondant de nouvelles variables (les composantes) résumant l'entièreté de l'information contenue dans les variables initiales, mais dont un sous-ensemble est susceptible de contenir la majeure partie de cette information.

La méthode est appliquée, en FWB, aux unités « implantation », « établissement » des enseignements fondamental ordinaire et secondaire ordinaire de plein exercice ainsi qu'aux unités « secteur statistique ». Seule la première composante est retenue et nommée ISE. Cette méthode permet dans ce contexte d'obtenir un indice unique qui permet un **classement univoque des unités à partir d'une multitude de variables mesurant (7 dans le calcul actuel) imparfaitement ou partiellement le niveau socioéconomique**. Dans ce classement, une valeur positive signifie que le niveau socioéconomique moyen est plus favorisé tandis qu'une valeur négative indique que le niveau socioéconomique moyen est plus défavorisé.

Le tableau 1 montre que l'ISE explique une part systématiquement élevée de l'information contenue dans les variables initiales (entre 78,2 et 87,5 % pour les structures 2019-20), quelle que soit l'ACP considérée. En d'autres termes, l'outil fournit un bon résumé des variables initiales.

Tableau 1 : Variance expliquée par la 1^{re} composante

Structures	Fondamental		Secondaire		Secteur
	Implantation	Établissement	Implantation	Établissement	
2015-16	74,5		81,3		75,5
2017-18	78,5	80,2	86,4	87,1	79,5
2018-19	78,3	79,9	86,4	87,1	79,7
2019-20	78,1	79,8	86,7	87,2	79,4

5. La construction de l'indice socioéconomique

L'ISE est la combinaison linéaire des variables initiales. Sa formule est la suivante :

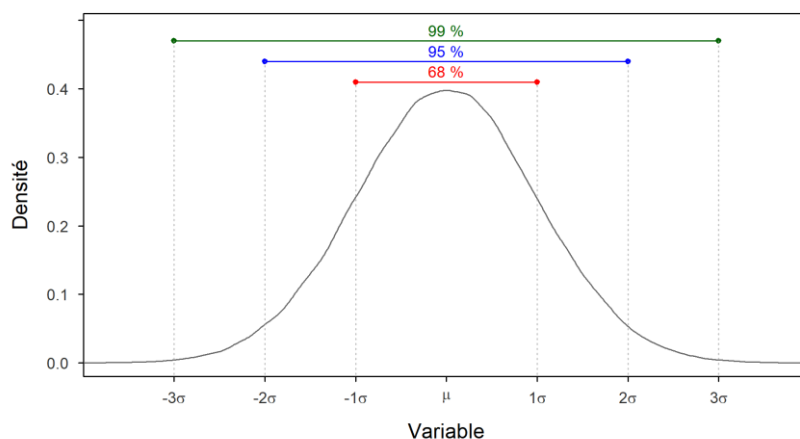
$$ISE_i = \sum_{j=1}^7 C_j \frac{(x_{ij} - A_j)}{B_j} = \sum_{j=1}^7 x'_j * C_j$$

Où i est l'unité statistique considérée (implantation, établissement ou secteur) ; j est le numéro de la variable incluse dans l'ACP ; C_j est le coefficient issu de l'ACP qui permet de pondérer la contribution de chaque variable pour le calcul de l'indice ; A_j est la moyenne et B_j l'écart-type de cette variable. Techniquement, la soustraction de la moyenne A_j permet de centrer les valeurs tandis que la division par l'écart-type B_j permet de réduire à 1 l'écart-type des valeurs. Aussi, l'ISE est-il la somme pondérée des valeurs centrées réduites de chaque variable pour l'unité considérée. Dit autrement, c'est une **moyenne des variables initiales sur une échelle standardisée**.

Cette formule et ses éléments ont été reportés dans les AGCF successifs, mais leur présentation a été simplifiée depuis l'AGCF du 20 mars 2019. Étant donné que la méthode implique d'appliquer le coefficient C_j à des variables centrées réduites « x », il n'est pas nécessaire de préciser les coefficients A_j et B_j pour cette transformation usuelle. La formule simplifiée est écrite dans l'AGCF du 20 mars 2019. Notons que ce n'est pas tant les coefficients qui sont fixés par les textes légaux qu'une méthode statistiquement cohérente. Par ailleurs, l'identification des coefficients et leur application aux données consistent en une analyse continue s'appliquant aux mêmes données. Bien que ces étapes soient distinguées dans les textes légaux, elles sont fondamentalement intriquées du point de vue statistique. Afin de simplifier le processus et de diminuer le temps requis par les publications des arrêtés successifs, le tableau reprenant les coefficients C a été retiré de l'AGCF du 7 avril 2020. Ces coefficients sont toutefois annexés au présent document par souci de transparence.

L'ISE se distribue «normalement», ce qui signifie que ses valeurs n'ont pas d'interprétation directe si ce n'est une interprétation statistique qui permet la comparaison des implantations, établissements et secteurs. Théoriquement, ce type de distribution (figure 2) est notamment caractérisée par le fait que 68 % des valeurs se trouvent, en théorie, à + ou - 1 écart-type de la moyenne et 95 % des valeurs à + ou - 2 écarts-types. En d'autres termes la valeur d'une unité dit quelque chose sur sa position et la proportion d'unités plus défavorisées ou plus favorisées qu'elle. Par exemple, une implantation d'ISE inférieur à -2 fera approximativement partie des 2,5 % d'implantations les plus défavorisées.

Figure 2 : Distribution normale



Deux éléments complémentaires doivent être explicités :

- Les propriétés de l'ISE en matière de classement des unités sur une même échelle normalisée s'appliquent au sein d'une ACP uniquement et non entre les cinq ACP calculées dans cette procédure. En d'autres termes, les implantations du fondamental ordinaire sont comparables avec les autres implantations du fondamental ordinaire. La combinaison de deux classements, par exemple les implantations du fondamental ordinaire et celles du secondaire ordinaire n'est pas statistiquement garantie. Il est, en effet, raisonnablement possible de penser que l'ISE moyen des établissements et que la dispersion des écoles sont différents: le niveau socioéconomique des ménages peut être différents entre le fondamental et le secondaire (le niveau socioéconomique étant liée par exemple à l'âge des membres des ménages) et la stratification sociale peut évoluer différemment au cours du temps. La seule manière de combiner ces unités est de le faire au moyen d'une ACP commune à toutes ces unités. En d'autres termes, **tout classement implique une ACP sur base des unités spécifiques à ce classement** ;
- Un changement d'ISE peut traduire des situations multiples : un changement de population dans l'unité, un changement de population dans les autres unités ou un changement de formule.

6. Évolution de la méthode

6.1. Mesure indirecte agrégée de l'ISE

L'ISE tel qu'il a été calculé initialement a consisté en une mesure indirecte de la situation socioéconomique des familles et synthétise l'information provenant de 12 (ISE 1999 et 2001) ou 11 (ISE 2006) variables couvrant 5 dimensions : le revenu, le niveau de diplomation, l'activité professionnelle et le type d'activité ainsi que le confort du logement. Cet indice prend sa source dans la mise en place du dispositif des discriminations positives (Décret du 30 juin 1998) qui se traduit par la distinction et l'attribution de moyens supplémentaires à certaines implantations sur base de critères socioéconomiques et non scolaires. Plusieurs éléments permettent de comprendre l'implémentation spécifique de ce dispositif. Premièrement, il s'agissait de compenser un désavantage non scolaire dont la science a montré qu'il se traduit en désavantage scolaire. Aussi, le dispositif compensait-il un désavantage lié à la situation socioéconomique de certains groupes (mesurée sur base du niveau socioéconomique de leur quartier) en favorisant les implantations qui les accueillait et ne sanctionnait-il pas les résultats des implantations. Deuxièmement, l'utilisation de critères non scolaires permettait d'éviter certains effets pervers du dispositif, comme l'allocation de moyens supplémentaires à l'implantation socioéconomiquement favorisée, mais aux résultats insatisfaisants et non à une implantation socioéconomiquement défavorisée, mais obtenant des résultats satisfaisants ou la suspension de l'aide aux implantations ayant réussi à améliorer la réussite de ses élèves et la favorisation d'implantations n'ayant pas réussi à mettre à contribution ces moyens supplémentaires. (Demeuse 2003) Troisièmement, l'implémentation technique s'inscrivait dans la réalité des données de l'époque, à savoir le développement du fichier de la carte élève permettant de lier les lieux de scolarisation et de résidence et la disponibilité de données par quartier ou par commune (recensements décennaux, mais également données publiées par l'INS, l'ONEM, ou le Ministère des Affaires Sociales) (Demeuse et al. 1999).

Trois ACP ont été successivement calculées par une équipe interuniversitaire (EIU) et appliquées à tous les secteurs statistiques belges² sur base des données des ménages résidents dans le secteur :

- ISE 1999 pour tous les secteurs statistiques de Belgique (19331 secteurs en nomenclature 1991) résultant de l'ACP de l'AGCF 26924 du 2 mai 2002, sur base des données datant de 1991 à 1999 ;
- ISE 2001 pour tous les secteurs statistiques de Belgique (19331 secteurs en nomenclature 1991) résultant de l'ACP de l'AGCF 30023 du 20 juillet 2005, sur base des données datant de 1991 à 2001 ;
- ISE 2006 pour tous les secteurs statistiques de Belgique (19781 secteurs en nomenclature 2001) résultant de l'ACP de l'AGCF 35651 du 17 juin 2010, sur base des données datant de 2001 à 2006.

L'ISE était calculé sur base des données de tous les résidents qu'ils aient ou non des enfants scolarisés ou en âge d'être scolarisés dans l'enseignement de la FWB. Afin de ne pas donner trop de poids lors la détermination de la formule à des secteurs peu peuplés d'enfants en âge de scolarisation, mais également afin de ne pas dépendre trop d'un découpage arbitraire (Demeuse et al. 1999), l'analyse était pondérée par le nombre de jeunes de 0 à 20 ans qui habitent chaque secteur statistique. En d'autres termes, si la manière dont les variables sont liées entre elles, diffère selon la composition du secteur, ce sont les secteurs davantage peuplés d'élèves en âge d'être scolarisés qui déterminent le plus le résultat de l'analyse.

Une fois les ISE des secteurs calculés, un ISE implantation était dérivé en faisant la moyenne des ISE des secteurs de résidences des élèves, les élèves sans ISE (soit parce que leur secteur statistique n'était pas identifié, soit parce qu'ils ne résidaient pas en Belgique) n'étant pas pris en compte et les élèves primoarrivant recevant un ISE égal à la moyenne arithmétique des indices des 50 secteurs statistiques les plus bas.

Compte tenu de la méthode exposée ci-dessus, ce mode de calcul avait quelques limites notables :

- L'utilisation d'une mesure indirecte agrégée plutôt qu'individuelle ne pouvait rendre compte du niveau socioéconomique des élèves dans les secteurs hétérogènes que de manière imparfaite. Ainsi, les ISE de deux écoles recrutant le même nombre d'élèves dans les mêmes secteurs seront-ils identiques, même si l'une recrute les élèves favorisés et l'autre les élèves défavorisés de ces secteurs ;
- Un ISE non régulièrement mis à jour peut être en décalage avec le niveau socioéconomique de la population actuellement accueillie ;
- L'ISE d'une implantation peut présenter des écarts importants entre mises à jour successives ce qui limite fortement la prévisibilité des ressources disponibles ;
- Les mises à jour peu fréquentes posaient des problèmes spécifiques pour les nouvelles implantations et les implantations ayant connu une restructuration ;
- L'ISE n'est pas renouvelable à l'identique puisqu'une partie des variables sont issues des recensements décennaux qui ne sont plus reconduits après 2001.

L'ISE 2006 a été en application jusqu'en 2016-17

² Les coefficients ont toutefois été calculés sur les secteurs de la région bilingue de Bruxelles-Capitale et de la région de langue française uniquement, bien qu'un ISE soit disponible pour l'entièreté des secteurs pour lesquels les données sont disponibles.

6.2. Mesure directe agrégée de l'ISE

La méthode a connu un changement majeur en 2017. Cinq éléments de ce changement sont particulièrement saillants :

- Une utilisation directe de données individuelles (se rapportant aux caractéristiques socioéconomiques du ménage de l'élève) agrégée au niveau de l'implantation, de l'établissement et du secteur statistique. Les ACP (ou les coefficients de l'AGCF en vigueur) sont appliqués à ces données agrégées ;
- Une mise à jour régulière des données et des coefficients de la formule pour coller le plus possible à la réalité et son évolution. La mise à jour des données permet de tenir compte de la population actuellement scolarisée ou résidente tandis que la mise à jour des coefficients permet de tenir compte des évolutions du caractère discriminant³ des variables mobilisées (et de maximiser l'information synthétisée par l'ISE) ;
- Un changement de variables. L'ISE synthétise à présent l'information provenant de 7 variables couvrant 4 dimensions : le revenu, le niveau de diplomation, l'activité professionnelle et le type d'activité ;
- Un changement d'unité d'analyse. Les ACP sont à présent appliquées aux implantations, (établissements) et secteurs statistiques (où réside au moins un élève avec RN et scolarisé en FWB) ;
- Pour les implantations et les établissements, les valeurs des primoarrivants (plus précisément, les élèves qui ont été qualifiés de primoarrivant une des 7 dernières années de comptage) reçoivent la moyenne des 2000 valeurs les plus faibles.

Il en découle un ISE qui reflète le niveau socioéconomique de la population de l'implantation, de l'établissement ou du secteur avec davantage de précision et de fiabilité. Il solutionne les limites précédemment identifiées. L'ISE n'est toutefois pas individuel en ce sens qu'il ne permet pas de distinguer les variations de niveau socioéconomique au sein des implantations et secteurs.

Trois AGCF ont à ce jour été promulgués et plusieurs ISE ont été calculés⁴ :

- AGCF 44433 du 19 juillet 2017 : ACP sur base des données individuelles agrégées (années civiles 2008 à 2014) des élèves (populations scolaires des années scolaires 2008-09 à 2014-15) pour toutes les unités (implantation et secteur) avec au moins 1 élève scolarisé dans une implantation de la FWB ouverte en 2014-15. Les coefficients ont été appliqués pour le calcul de 2 ISE :
 - ISE des structures 2014-15 sur base des données individuelles agrégées datant des années civiles de 2008 à 2014 produit par l'EIU (chargé le 16 janvier 2018 dans ACISE) ;
 - ISE des structures 2015-16 sur base des données individuelles agrégées datant des années civiles de 2009 à 2015 produit par le CCBI (chargé le 20 novembre 2018 dans ACISE) ;
- AGCF 46854 du 20 mars 2019 : ACP sur base des données individuelles agrégées (années civiles 2011 à 2017) des élèves (populations scolaires des années scolaires 2011-12 à

³ Un exemple classique d'évolution du caractère discriminant d'une variable concerne la disponibilité d'une connexion internet à domicile qui ne discrimine plus aujourd'hui les mêmes populations ou ne les discrimine plus de la même manière que dans les années 90.

⁴ Ces ISE ont été respectivement mobilisés pour les classements « encadrement différencié » pour les années scolaires 2017-18, 2018-19 et 2019-20.

2017-18) pour toutes les unités (implantation, établissement et secteur) avec au moins 1 élève scolarisé dans une implantation de la FWB ouverte en 2017-18. L'ISE des structures 2017-18 a été produit par le CCBI ;

- AGCF 48054 du 7 avril 2020 :
 - ISE des structures 2018-19 sur base des données individuelles agrégées (années civiles 2012 à 2018) des élèves (populations scolaires des années scolaires 2012-13 à 2018-19) pour toutes les unités (implantation, établissement et secteur) avec au moins 1 élève scolarisé dans une implantation de la FWB ouverte en 2018-19. L'ISE a été produit par le CCBI ;
 - ISE des structures 2019-20 sur base des données individuelles agrégées (années civiles 2013 à 2019) des élèves (populations scolaires des années scolaires 2013-14 à 2019-20) pour toutes les unités (implantation, établissement et secteur) avec au moins 1 élève scolarisé dans une implantation de la FWB ouverte en 2019-20. L'ISE a été produit par le CCBI.

Plusieurs modifications mineures (n'impactant pas le mode de calcul général de l'indice) ont été apportées lors du calcul de l'ISE des structures 2017-18 :

- L'ISE des établissements était calculé en faisant la moyenne des ISE des implantations (pondérée par les effectifs). L'ISE des établissements est, depuis l'ISE des structures 2017-18, calculé sur base d'une ACP propre afin d'assurer une cohérence des différentes méthodes de classements et de permettre à des variables de se comporter de manière légèrement différente lorsqu'elles discriminent les implantations et les établissements. Cela crée toutefois une confusion pour la lecture des valeurs des ISE des établissements à implantation unique puisque la valeur de l'établissement sera différente de celle de l'implantation⁵.
- Des contraintes de temps (délai de livraison de la base de données Stat Élèves et des données BCSS ainsi qu'une fenêtre d'observation évoluant annuellement) avaient été identifiées lors de la production précédente avec pour conséquence une utilisation de données datant d'au mieux 2 ans et demi. Le choix a été fait de réduire le délai de 2 ans et demi à 1 an et demi, ce qui a pour conséquence :
 - L'impossibilité d'utiliser la base de données Stat Élèves (desanonymisée) dont le délai de livraison est trop tardif pour respecter les délais de livraison des ISE : est dès lors utilisée une version intermédiaire de la base de données (comptage dit « ++ ») qui intègre le nettoyage des primoarrivants (SCE) et l'attribution des clés de pilotage, la suppression supplémentaire des doublons ainsi que le nettoyage des champs requis par la procédure SAS (CCBI). Le risque qu'une erreur soit identifiée au cours de la procédure de nettoyage de la base de données et qu'elle impacte les ISE existe, mais est minime.
 - Un saut de 2 ans dans la fenêtre d'observation qui devait évoluer d'un an chaque année pour assurer la cohérence.
- Le mode de calcul du revenu médian a été modifié. Pour l'ISE des structures 2015-16, la procédure SAS prévoyait d'extraire les données de 4 années successives (2012 à 2015), ce qui revenait à ne sélectionner que 2012 étant donné qu'aucune donnée plus récente n'était disponible. Il a donc été décidé d'étendre la fenêtre de 4 ans aux 7 années utilisées pour la majorité des autres variables.

⁵ Pour l'ISE des structures 2017-18, les ISE obtenus en appliquant les résultats de l'ACP et en faisant la moyenne pondérée des ISE des implantations sont très proches et sont corrélés à hauteur de 0,9998.

- Des erreurs ont été corrigées par le CCBI depuis le calcul des coefficients de l'arrêté du 19 juillet 2017 :
 - Une boucle incorrecte qui n'extrait qu'une année de données pour les variables relatives à la Diplomation (variables 2 et 3, ci-dessous) ;
 - La structure conditionnelle de la variable 6 (part des ouvriers) prévoyait d'interdire les 0 au numérateur, mais pas au dénominateur, ce qui générerait des valeurs manquantes.

La nouvelle méthode de calcul présente également des limites :

- Aucune ACP n'est appliquée aux implantations et établissements de l'enseignement spécialisé. Ces derniers se voient appliquer les coefficients du fondamental ordinaire (implantations organisant un enseignement spécialisé fondamental ou organisant un enseignement spécialisé fondamental et secondaire) et du secondaire ordinaire (implantations organisant un enseignement spécialisé secondaire). Une ACP directement appliquée aux implantations et établissements de l'enseignement spécialisé pourrait produire des résultats différents dans la mesure où les variables n'auraient pas le même pouvoir de discrimination que dans l'enseignement ordinaire ;
- Dans le cadre de la différenciation lors du calcul des dotations et subventions (Décret du 28 avril 2004), des implantations de l'enseignement fondamental ordinaire, de l'enseignement secondaire ordinaire et de l'enseignement spécialisé sont ordonnées ensemble. Or, ces listes différentes sont des classements sur des échelles distinctes. Certes les valeurs varient sur une même fourchette, mais ni la moyenne ni la distribution de ces listes n'est comparable. Plus concrètement, l'ISE moyen et la dispersion des implantations de l'enseignement ordinaire et spécialisé sont considérés comme comparables alors que les élèves de l'enseignement spécialisé ont tendance à présenter un profil socioéconomique plus défavorisé. Les analyses menées à ce jour ne permettent toutefois pas d'identifier une conséquence sur la répartition du financement, le mécanisme de répartition prévoyant des balises à la redistribution ;
- Des valeurs manquantes existent au niveau des secteurs statistiques et doivent être traitées. En effet, les secteurs pour lesquels certaines données sont absentes ne reçoivent pas d'ISE au terme de la procédure de calcul. Les valeurs manquantes sont imputées en héritant des données issues de l'agrégation des données à un niveau plus large (7 premiers caractères du code secteur, ou au niveau de la commune, à savoir les 5 premiers caractères).

7. Catégorisations dérivées des ISE

Sur base des ISE précédemment calculés, une série de catégorisations sont produites à des niveaux divers et à des fins diverses. Ce chapitre en précise brièvement certaines.

7.1. Classes « encadrement différencié »

Des moyens d'encadrement et financiers complémentaires sont attribués aux implantations identifiées selon le décret du 30/04/2009 organisant un encadrement différencié au sein des établissements scolaires de la Communauté française afin d'assurer à chaque élève des chances égales d'émancipation sociale dans un environnement pédagogique de qualité. Cette identification est basée sur une catégorisation de l'ISE en 20 classes dites « encadrement différencié ».

Pour être bénéficiaire de l'encadrement différencié, l'implantation doit répondre à une des deux conditions suivantes :

- Avoir relevé totalement ou partiellement des classes numérotées de 1 à 5 à trois reprises au moins lors des six derniers classements, en ce compris le classement de l'année scolaire concernée.
- Avoir toujours relevé totalement ou partiellement des classes numérotées de 1 à 5 lors de chaque classement où l'implantation a été classée à partir de l'année scolaire 2017-2018.

Pour l'attribution des classes « ED » 2021-22, les étapes suivantes sont effectuées en mars 2021 :

- Toutes les implantations disposant d'un ISE des structures 2019-20 et comptabilisant des élèves régulièrement inscrits dans SIEL en janvier 2020 sont ordonnées de manière croissante selon leur ISE, en commençant donc par l'implantation à l'ISE le plus faible. Deux listes sont ainsi produites, l'une pour l'enseignement fondamental ordinaire, l'autre pour l'enseignement secondaire ordinaire ;
- Les implantations sont réparties, en fonction de leur population scolaire cumulée, en 20 catégories scolarisant chacune exactement 5 % de la population scolaire considérée. La première classe rassemble ainsi les implantations scolarisant les 5 premiers points de pourcentage de cette population et ainsi de suite ;
- En cas d'ex aequo (situation peu probable au vu de la précision de l'ISE), l'implantation ayant le nombre d'élèves le plus élevé est considérée comme la plus défavorisée (elle sera ainsi prioritaire pour l'attribution des moyens de l'encadrement différencié) ;
- Lorsqu'une implantation est à cheval sur 2 classes, les populations de ces implantations sont considérées comme appartenant à la classe ED la plus défavorisée ;
- La classe 3 est scindée en 2 sous-classes (a et b) avec la proportion 2.5 %/2.5 % pour le fondamental et 3.5 %/1.5 % pour le secondaire ;
- Des exceptions sont prévues pour les implantations restructurées :
 - Dans le cas des moyens pour « l'encadrement différencié » : Aucun ISE n'est recalculé, mais les restructurations sont prises en compte pour déterminer les moyens théoriques du passé (moyennes sur 6 ans) ;
 - Pour le calcul des dotations et subventions de fonctionnement : l'ISE est recalculé pour une implantation restante issue d'une fusion par absorption (moyenne pondérée des ISE des implantations avant fusion) et les nouvelles implantations issues d'une scission se voient attribuer les ISE des implantations dont elles sont issues. Les dites implantations se voient affecter une classe ED relative ment à leur nouvel ISE.

7.2. Implantations ISEF

Les implantations de l'enseignement fondamental ordinaire sont qualifiées d'implantations à indice socioéconomique faible (ISEF) conformément au décret du 30 avril 2009 organisant un encadrement différencié au sein des établissements scolaires de la Communauté française afin d'assurer à chaque élève des chances égales d'émancipation sociale dans un environnement pédagogique de qualité. Dans le cadre de la procédure d'inscription des élèves en 1^{re} secondaire, un quota de places (20,4 % dans chaque établissement) est en effet réservé pour des élèves issus d'implantations ISEF afin de soutenir la mixité sociale, culturelle et académique.

Pour l'attribution de la qualification « ISEF » pour les inscriptions en 2022, les étapes suivantes sont effectuées :

- Toutes les implantations disposant d'un ISE des structures 2019-20 et comptabilisant des élèves régulièrement inscrits dans une 6^e primaire ordinaire dans SIEL en octobre 2021 sont ordonnées de manière croissante selon leur ISE, en commençant donc par l'implantation à l'ISE le plus faible ;

- Les implantations scolarisant 40 % de la population scolaire de janvier 2020 sont qualifiées d'ISEF.

7.3. Catégories des plans de Pilotage

Dans le cadre des plans de pilotage, des profils et des catégories sont produits afin de permettre à chaque établissement de se comparer à un ensemble d'établissements similaires. Les profils d'écoles sont des regroupements d'écoles en fonction du type d'études organisées dans l'école. Les catégories sont des regroupements d'écoles en fonction du niveau socioéconomique des élèves qui y sont scolarisés.

Pour l'attribution des catégories pour données « plan de pilotage » mises à disposition en 2021, les étapes suivantes sont effectuées :

- Tous les établissements disposant d'un ISE des structures 2019-20 et comptabilisant des élèves inscrits dans la version statistique du comptage en janvier 2020 sont ordonnés, par profils, de manière croissante selon leur ISE, en commençant donc par l'établissement à l'ISE le plus faible ;
- Les établissements sont répartis, en fonction de leur population scolaire cumulée, en 4 catégories scolarisant chacune exactement 25 % de la population scolaire considérée. La première catégorie rassemble ainsi les établissements scolarisant les 25 premiers points de pourcentage de cette population et ainsi de suite ;
- Des conditions minimales pour la constitution d'une catégorie ont toutefois été prévues pour garantir le calcul d'indicateurs « catégorie » pertinents. Pour chaque profil, un maximum de 4 catégories « ISE », regroupant chacune au minimum 25 établissements et 1000 élèves sont générés ;
- Lorsqu'un établissement est à cheval sur 2 catégories, l'établissement est considéré comme appartenant à la catégorie la plus défavorisée.

8. Annexe : coefficients C⁶

8.1. AGCF 44433 du 19 juillet 2017

n	Implantation (fondamental)	Implantation (Secondaire)	Secteur statistique
1 rev_med	+0,18255	+0,16474	+0,17579
2 etudes_haut	+0,17436	+0,15795	+0,16475
3 etudes_bas	-0,1754	-0,16237	-0,17391
4 prop_trav	+0,17908	+0,16074	+0,17715
5 aide_sociale	-0,09268	-0,14323	-0,14811
6 prop_ouv	-0,17214	-0,15153	-0,16256
7 prop_tert	-0,16251	-0,15398	-0,1521

8.2. AGCF 46854 du 20 mars 2019

n	Implantation (fondamental)	Implantation (Secondaire)	Etablissement (fondamental)	Etablissement (Secondaire)	Secteur statistique
1	0,17538	+0,16006	+0,17221	+0,15878	+0,16903
2	0,17226	+0,15587	+0,16903	+0,15453	+0,16325
3	-0,16975	-0,15681	-0,16714	-0,15581	-0,16611
4	0,16746	+0,15330	+0,16448	+0,15221	+0,16354
5	-0,08499	-0,14134	-0,09530	-0,14217	-0,13474
6	-0,17199	-0,15524	-0,16903	-0,15453	-0,16624
7	-0,16682	-0,15285	-0,16535	-0,15283	-0,15896

8.3. AGCF 48054 du 7 avril 2020

8.3.1. ISE des structures 2018-19

n	Implantation (fondamental)	Implantation (Secondaire)	Etablissement (fondamental)	Etablissement (Secondaire)	Secteur statistique
1	+0,17582	+0,17268	+0,15965	+0,15855	+0,16824
2	+0,17262	+0,16956	+0,15610	+0,15490	+0,16177
3	-0,17016	-0,16759	-0,15647	-0,15547	-0,16481
4	+0,16761	+0,16478	+0,15305	+0,15201	+0,16345
5	-0,08494	-0,09474	-0,14255	-0,14329	-0,13544
6	-0,17222	-0,16930	-0,15522	-0,15467	-0,16523
7	-0,16690	-0,16534	-0,15244	-0,15246	-0,15823

8.3.2. ISE des structures 2019-20

n	Implantation (fondamental)	Implantation (Secondaire)	Etablissement (fondamental)	Etablissement (Secondaire)	Secteur statistique
1	+0,17599	+0,17268	+0,15888	+0,15813	+0,16835
2	+0,17262	+0,16953	+0,15546	+0,15462	+0,16167
3	-0,16993	-0,16761	-0,15601	-0,15524	-0,16494
4	+0,16803	+0,16492	+0,15282	+0,15206	+0,16320

⁶ Pour rappel, la numérotation n correspondant à l'ordre donné chapitre 3 et est rappelé dans le 1^{er} tableau.

5	-0,08658	-0,09554	-0,14254	-0,14305	-0,13757
6	-0,17258	-0,16939	-0,15510	-0,15472	-0,16576
7	-0,16693	-0,16520	-0,15267	-0,15264	-0,15855