RAPPORTS

Aigul Alieva

Thiago Brant

Eugenio Peluso

Philippe Van Kerm

An empirical and methodological review of the “Indicateur social”

October 2021

Technical report

www.liser.lu

An empirical and methodological review of the “Indicateur social”

This research is part of the project “Évaluation de l'application de l’indicateur social dans le cadre de l’attribution du contingent par commune dans l’enseignement fondamental” supported by the Ministry of National Education, Children and Youth of Luxembourg. Technical report not for dissemination.

Table of Contents

List of figures

Introduction

The report is written in the framework of the collaborative project “Évaluation de l'application de l’indicateur social dans le cadre de l’attribution du contingent par commune dans l’enseignement fondamental” [Evaluation of the application of the social index in the framework of contingent allocation by the municipality in primary education]. The partners in this project are the Ministry of National Education, Children and Youth (MENJE), the National Observatory on School Quality (ONQS), Luxembourg Institute of Socio-Economic Research (LISER) and the University of Luxembourg (UniLU). The goals of the project are 1) to review the existing methodology in light of the scientific advancement in this domain; 2) to provide the recommendations for modifications in the methodology for the forthcoming indices, as well as discuss the issues regarding the data quality and availability; 3) to evaluate the effect of contingent mechanism on the reduction of socio-economic inequalities in primary schools.

This report addresses the first goal of the project and evaluates the methodology employed for the construction of the social index between 2009 and 2019. Part I of the report focuses on the longitudinal development of the index as part of the robustness analysis: both the changes in the ranking of the municipalities and the changes in the index values attributed to the municipalities over time are analyzed. Additional important issues, such as the size of a commune are investigated. In part we also review the changes in the four dimensions that generate the final index and issues pertinent to data availability are discussed accordingly. Additionally, it touches briefly on the comparability of the social index calculated by LISER with the social index, similarly calculated at the municipality level, by the National Institute of Statistics and Economic Studies of the Grand Duchy of Luxembourg (STATEC). It also offers some insight on the comparability of the LISER social index with the index derived from the national school monitoring data Épreuves Standardisées (ÉpStan) collected by the Luxembourg Centre for Educational Testing (LUCET) at the University of Luxembourg.

Part II develops a comparison of the social index calculated in Luxembourg by LISER with four selected indices that have a similar policy goal of addressing socio-economic inequalities in compulsory education. Two of the indices are developed by the cities of Hamburg (Germany) and Zurich (Switzerland), one index is developed at the regional level in Wallonia-Brussels in Belgium, and the fourth index is computed at the national level in France.

Part III is dedicated to additional empirical questions that were raised during the first phase of the project to better understand other features of inequality within the school population in Luxembourg that go beyond local differences among communes, which are the center of the composite indicators (CIs). Among them is the heterogeneity of school population between international and “standard” primary schools, among the population accessing the “précoce”, which is part of early education, and finally, the inequality within large municipalities containing several primary schools, such as Luxembourg City and Esch-sur-Alzette. These findings will contribute to the forthcoming report on strengthening the methodological approach for the future calculation of the social index.

Part IV of the report provides a discussion of academic literature on relevant themes for CIs: data selection and data normalization, an overview of relevant methods' strengths and weaknesses, as well as an overview of weights and weighting strategies.

The main empirical findings and questions that require additional attention, both at the policy and methodological levels are summarized in the concluding part of the report.

PART 1 - Overview of the social index

1.1. The policy framework of the index and the contingency mechanism

Educational inequalities driven by ascriptive characteristics of an individual such as social and educational background, gender, disability, language skills and immigration background remain persistent across many European education systems. For example, the impact of the economic, social and cultural status (ESCS) of the families on variation in reading performance in Luxembourg is 17.8% and is one of the largest among countries participating in PISA study in 2018 and is nearly three times higher than in Estonia (6.2%) (OECD, 2019). Inequalities in the education system evolve through a complex interplay between the contextual level (societal macro-level, institutional meso-level) and the individual micro-level, resulting in significant between-country differences, as seen in PISA results mentioned above. They are also subject to temporal development and are detected already in the early stages of life and gradually accumulate over the lifespan of an individual (DiPrete and Eirich, 2006). Without proper identification of vulnerable groups and targeted intervention, the gap will continue to grow with each subsequent educational stage leading to educational poverty, both in terms of certificates and in terms of minimum skills and competences (Allmendinger 1999; Allmendinger and Leibfried, 2003; Checchi, 1998). Research on educational interventions concludes that the window of intervention is relatively narrow and the most malleable stage that leads to best outcomes is in the early educational stages (Heckman, 2006).

The persistence, in particular, of socio-economic inequalities in the context of schools led the educational policy makers to devise the compensatory measures based on the allocation of resources in order to ease the social disparities among the schools and schoolchildren (OECD, 2010). Composite indicators are then used to access socioeconomic inequalities in the educational system. The main goal of composite indicators is to translate a complex phenomenon, which generally would need to be represented by different dimensions, into a single number, this leads to an advantage for policymakers to efficient convey a message which has an easy interpretability for the general public (Saltelli, 2007). [1]

The national law on the organization of primary education of February 6, 2009 introduced the contingent mechanism of teaching lessons available per municipality and school district based on a number of parameters (article 38).[2] The objective of this mechanism is, among others, to promote academic success among the students and ensure a more equitable allocation of resources given the pronounced socio-

[1] See Section 4 for more details.
2 http://data.legilux.public.lu/eli/etat/leg/loi/2009/02/06/n3/jo

economic heterogeneity between the municipalities. The law stipulates a gradual introduction of the quota over the duration of ten years (article 38).

The following Grand-Ducal Regulation (RGD) of February 18, 2010 specifies the mechanisms and conditions for allocating the contingency lessons with an emphasis on providing extra resources to the municipalities with more vulnerable socio-economic and socio-cultural composition (article 2).[3] The distribution of these resources is based on an index calculated for every municipality and school district that takes into account school population characteristics. The RGD delegated the task of calculating the index to the Luxembourg Institute of Socio-Economic Research (LISER)[4] in partnership with the Ministry of National Education, Children and Youth (MENJE) and the General Inspectorate of Social Security (IGSS). The index (composite indicator – CI) was calculated for the following school years: 2009-2010, 2011-2012, 2014-2015 and 2018-2019. According to the RGD, the index varies between 100 and 120 points for each municipality, and those municipalities closer to 120 are considered socio-economically disadvantaged in comparison to more affluent communes with a score closer to 100 points. The distribution of the teaching quota is based on this final index, where a commune with a hypothetical score of 120 points will receive 20% more resources, while a commune with a score of 100 points will receive no *additional* support. In other words, based on the index, a commune can receive between 0% and 20% of extra resources. It is important to note that the social index is a complementary tool for resource allocation to schools and is one of the four existing instruments alongside the linear allocation (contingent de base), transitional allocation of ten years to even out the situation among under- and over-staffed schools, and finally the special allocation.

The index consists of four dimensions (see the following section) and its construction requires the combination of data from the IGSS and from the MENJE. The IGSS data provides information on the socio-economic background of the students' families attending the public school(s) within a respective municipality, the MENJE data provides information on their socio-cultural background. In other words, the index does not take into account the characteristics of children that live in a given municipality, but attend other schools (private, across the borders) or are home-schooled, or have special needs and are in charge by specific developmental centres (*Centres de compétences*) supported by MENJE. Another category of children that is excluded from the index is children in state care, such as orphans (living in *Maisons d'Enfants de l'Etat*). While these children might be attending the local municipal school where such care home is located, given the absence of information in the IGSS database about the parental background, they are dropped from the calculation of the index. Finally, the availability of the variables and the

3 http://data.legilux.public.lu/eli/etat/leg/rgd/2010/02/18/n9/jo

4 Former CEPS/INSTEAD

methodological changes in data collection between the years have affected the evolution of the index across the years.

1.2. The socio-economic index for Luxembourg

The indices from 2010, 2012, 2015 were obtained through the means of the Principal Component Analysis (PCA), a popular statistical technique used to reduce the number of relevant dimensions of a dataset. The main goal of the PCA is to simplify the variables so that data can be expressed in a simple way. However, the reduction of the number of dimensions comes with a cost in terms of information loss, generating a trade-off between accuracy and simplicity. The result provided by the calculation of a PCA aims to provide the maximum amount of variance of the data, simplifying the dimensions and minimizing the loss of information, in terms of variance between the variables. This technique is often applied in the construction of CIs, either to reduce the number of variables in a dimension of a CI, (for example, if a CI has two variables under the dimension of 'Culture' such as 'Number of books at home', 'Frequency of visiting the museum together with parents' a PCA can be applied to reduce this two variables to single number) or a PCA can be applied on all dimensions of a CI so it is possible to have a single number for the final index.

The Luxembourg index takes into account four dimensions: household income (indice du revenu), working condition (indice de la précarité professionnelle), family structure (indice de la structure familiale), and linguistic background (indice linguistique).[5] Between 2010 and 2019 there were some changes in the dimensions, mainly due to data availability and quality. For instance, in 2010 the dimension originally called 'indice de nationalité' (nationality indicator) was changed in 2012 to 'indice linguistique' (language indicator) in the following editions of the index.[6] Using the nationality information can be problematic in cases of naturalized Luxembourgish citizens with two or more nationalities: in the administrative records, Luxembourgish nationality becomes the first one, while the second (and third) nationality might not be recorded or no longer reported. As language skills play a more important role for school results than the nationality as such, using information about language background in the following editions is well justified.

Below is a brief overview of the dimensions and variables that were used to construct them across the years:

- **Household income -** the data used is from the IGSS database, the calculation is based on the methodology used by OECD. It considerers income from all sources of a household weighted by the number of household members and their age, throughout standard equivalence scales. This

[5] An overview of similar indices in Hamburg, Zurich, Wallonia and in France (see Part 2 for more details) shows that a number of dimensions range from two (e.g. income and migration in Zurich; income and education in Wallonia) to four dimensions.

[6] In 2010 both nationality and language background were available and calculated for the index, but only the results based on nationality were reported.

indicator is computed as the average for each municipality for families with children at primary education age. In the reports referring to the years 2010, 2012, 2015 and 2019 the calculation for 'Household income' remained the same, and data used from the IGSS data was respective from the years 2007, 2008, 2012 and 2017.

- **Working condition -** the data used is from the IGSS database and the indicator, pointed out at the municipality level, has changed over time. In 2010 and 2012, the calculation was based on five variables: "percentage of the blue-collar workers in the private sector", "percentage of the white-collars in the private sector", "percentage of the white-collar, the blue-collar, and the state employees in the public sector", "percentage of couples where one person works", "percentage of couples where both people work". For 2015 two new variables were additionally introduced: "if a person receives unemployment benefit" and "if a person receives the guaranteed minimum income". For 2019, there was a simplification and the number of available variables was limited to only two variables, such that "if a person receives unemployment benefit" and "if a person receives the guaranteed minimum income", were used to summarize the "work condition dimension".
- **Family structure -** the data used is from the IGSS database, the calculation is based on the average of the commune, divided into three different factors, "households with single parents", "households with two parents with less than 35 years", and "households with two parents with more than 35 years" for the years 2010, 2012, 2015. The age of parents was dropped in 2019 due to concerns about the quality of the data, as well as over the doubts about its relevance. If age is indicative of the economic position of a family, this factor is already being measured by income dimension. As a result, in 2019 there was only a distinction between households with single vs two parents. Apart from the family structure, no other family characteristics are taken into account, for example whether there are children with disabilities, children in foster care families, refugee and asylum seeking families, and etc.
- **Linguistic background -** is calculated based on data from the MENJE. In the first edition of the index in 2010 this dimension was reported based on the nationality of the student.[7] In 2012 and 2015 index, the language dimension was classified in 5 categories: "percentage of children speaking either Luxembourgish only or speaking Luxembourgish as their first language and another language as their second language", "percentage of children speaking either only one Latin language or one Latin language as first language and any other language as second language (except Luxembourgish)", "percentage of children speaking either only a language other than

7 categories: "percentage of individuals of Luxembourg nationality", "percentage of individuals of French nationality", "percentage of individuals of Belgian nationality", "percentage of individuals of Portuguese nationality", "percentage of individuals of Italian nationality", "percentage of individuals of other European nationality (EU 15)", "percentage of individuals of other nationality".

Luxembourgish, German or a Latin language, or a language other than the languages mentioned as first language and any language other than Luxembourgish as a second language", "percentage of children speaking either only German or German as first language and any other language as second language (except Luxembourgish)", and "percentage of children who speak any language other than Luxembourgish as a first language and Luxembourgish as a second language". In 2019 the index adopted a different classification with two categories: "percentage of children that speaks either Luxembourgish or German as first or second language" and "percentage of the students that do not speak neither Luxembourgish nor German as first or second language".

2. Analysis of the robustness of the index

2.1 Stability of the ranking of the municipalities over time

To assess the performance of the social index we start with the overall stability of the results. All four indices calculated in 2010, 2012, 2015, and 2019 are analyzed with respect to 1) the changes of position in the rank of communes, and 2) the changes in the final index score, which defines the distribution of the contingent for the communes. At first glance, the communes ranked at the top of the list (with values closer to one) have the most advantaged socio-economic composition and the smallest share of the non-Luxembourgish and non-German speaking population. At the bottom of ranking (with large values) are the communes with a high share of a vulnerable population – lower income, in a precarious working situation, with many households not speaking neither Luxembourgish neither German as their first or second language. However, more precise features can be described by looking at the overall ranking over time.

Table 1 starts by summarizing the position of a commune or school district in the overall ranking for each year. The columns "min" and "max" indicate the highest and the lowest position that a particular commune occupied in the different years under exam. The final column "variation" shows how many positions a commune shifted between the years (by a subtraction of "min" from the "max" column) and represents a simple measure of fluctuation. Results are pretty heterogeneous, pointing out a substantial stability for some communes and high variability for others. For instance, Esch-sur-Alzette was frequently at the bottom of the ranking given the city's socio-economically and linguistically disadvantaged population composition, and its position remained stable throughout these years (between 97 and 99th position). Similarly, Reckange-sur-Mess commune with favorable population composition remained at the top of ranking each of these years (between 2nd and 5th position) and had very minimal variation. On the other side, the commune of Saeul moved from the 8th position in the 2010 edition to the 75th position in the 2019 edition, resulting in a change of 67 positions, with the most considerable shift taking place between rank position in

2015 and 2019. A closer look at other communes with high variation tends to point to a similar finding that the most noticeable changes took place somewhere between the two last editions of the index with communes moving both upwards and downwards. For example, communes Bertrange, Strassen, Kopstal, Leudelange, Ell, and Préizerdaul moved down more than 30 places in 2019. Contrary to that, communes Fischbach, Grosbous, Tandel, Wahl, Waldbillig, and Vichten moved more than 30 places upward in 2019.

Part of these changes can be attributed to data availability and the subsequent changes in the measurement of some of the dimensions in the 2019 calculation. For example, as explained above, until 2015 the dimension “family structure” was defined as a combination of two sources of information: age of the head of household and distinction between single- vs. two-parents households. In other words, single-parent families where a parent was under 35 years old were considered more vulnerable in comparison to all other types of families. In 2019 this dimension was defined based solely on information about the share of single vs. two-parents families within a commune. In this regard, Saeul is a good example: according to the 2015 definition the commune was ranked in the middle (*commune sans tendance*), while according to the 2019 definition Saeul was identified as one among four communes with the highest share of single-parent households, thus drastically changing its ranking from 2015. This and other similar examples help to explain the fluctuation of the communes over time, as well as reveal the sensitivity of the social index to the data-related changes. We also suggest that the changes in the size of communes due to merging might have affected the results: in 2010 there were 102 municipalities, while in 2019 there were 98 municipalities and school districts (syndicates).

Table 1. Rank position of the communes within and across the years

Commune	Rank 2010	Rank 2012	Rank 2015	Rank 2019	Min	Max	Variation
Äerenzdall	88	71	65	48	48	88	40
Bascharage	40	NA	NA	NA	40	40	0
Beaufort	91	85	84	86	84	91	7
Beckerich	39	62	61	40	39	62	23
Berdorf	57	66	58	58	57	66	9
Bertrange	26	20	23	53	20	53	33
Bettembourg	73	73	66	70	66	73	7
Bettendorf	78	74	79	78	74	79	5
Betzdorf	30	14	10	10	10	30	20
Bissen	69	57	59	69	57	69	12
Biwer	22	4	9	6	4	22	18
Boevange/Attert	33	31	19	NA	19	33	14

Bourscheid	55	54	30	64	30	64	34
Bous	28	39	38	13	13	39	26
Burmerange	12	NA	NA	NA	12	12	0
Clemency	17	NA	NA	NA	17	17	0
Clervaux	66	65	73	74	65	74	9
Colmar-Berg	77	84	82	73	73	84	11
Consdorf	53	38	41	47	38	53	15
Contern	5	1	1	12	1	12	11
Dalheim	35	37	42	36	35	42	7
Diekirch	94	89	93	88	88	94	6
Differdange	101	99	98	92	92	101	9
Dippach	21	23	21	16	16	23	7
Dudelange	87	81	80	82	80	87	7
Echternach	95	90	94	95	90	95	5
Eis schoul	NA	NA	NA	14	14	14	0
Ell	19	26	29	60	19	60	41
Erpeldange	37	34	51	21	21	51	30
Esch-sur-Alzette	99	97	99	97	97	99	2
Esch-sur-Sûre	75	77	74	71	71	77	6
Ettelbruck	98	96	97	93	93	98	5
Feulen	76	67	57	54	54	76	22
Fischbach	56	58	24	17	17	58	41
Frisange	29	18	25	49	18	49	31
Garnich	13	7	2	4	2	13	11
Goesdorf	36	51	44	50	36	51	15
Grevenmacher	80	76	69	65	65	80	15
Grosbous	72	61	53	22	22	72	50
Habscht	NA	NA	NA	31	31	31	0
Heffingen	23	27	14	1	1	27	26
Helperknapp	NA	NA	NA	34	34	34	0
Hesperange	63	63	50	79	50	79	29
Hobscheid	31	45	60	NA	31	60	29
Junglinster	10	6	7	20	6	20	14
Käerjeng	NA	35	45	62	35	62	27
Kayl	89	83	78	85	78	89	11
Kehlen	1	9	6	15	1	15	14
Koerich	42	30	11	25	11	42	31
Kopstal	11	22	13	46	11	46	35
Larochette	97	94	89	81	81	97	16
Lenningen	7	16	27	18	7	27	20

Leudelange	6	11	16	43	6	43	37
Lintgen	86	75	77	83	75	86	11
Lorentzweiler	61	48	52	44	44	61	17
Luxembourg	83	82	75	87	75	87	12
Mamer	15	24	12	26	12	26	14
Mersch	79	79	72	61	61	79	18
Mertert	70	68	62	67	62	70	8
Mertzig	65	60	67	41	41	67	26
Mompach	49	10	34	NA	10	49	39
Mondercange	27	25	32	52	25	52	27
Mondorf-les-Bains	67	72	76	76	67	76	9
Niederanven	4	8	3	5	3	8	5
Nommern	18	15	20	3	3	20	17
Pétange	96	93	95	90	90	96	6
Préizerdaul	14	5	17	57	5	57	52
Rambrouch	48	56	63	51	48	63	15
Rambrouch-Neunhausen	45	NA	NA	NA	45	45	0
Reckange-sur-Mess	3	3	5	2	2	5	3
Redange-sur-Attert	41	50	46	38	38	50	12
Reisdorf	81	92	83	96	81	96	15
Remich	90	87	85	98	85	98	13
Roeser	51	41	47	55	41	55	14
Rosport	50	40	56	NA	40	56	16
Rosport-Mompach	NA	NA	NA	37	37	37	0
Rumelange	92	88	88	89	88	92	4
Saeul	8	21	37	75	8	75	67
Sandweiler	25	28	15	27	15	28	13
Sanem	71	70	70	72	70	72	2
Schengen	68	47	31	33	31	68	37
Schieren	85	80	91	80	80	91	11
Schifflange	93	91	86	84	84	93	9
Schuttrange	16	17	8	19	8	19	11
Stadtbredimus	34	49	49	24	24	49	25
Steinfort	44	52	43	45	43	52	9
Steinsel	38	44	40	39	38	44	6
Strassen	20	33	36	66	20	66	46
Syndicat Billek Dreiborn	46	19	33	35	19	46	27
Syndicat SCHOULKAUZ	64	59	87	63	59	87	28
Syndicat Harlange	NA	43	54	56	43	56	13

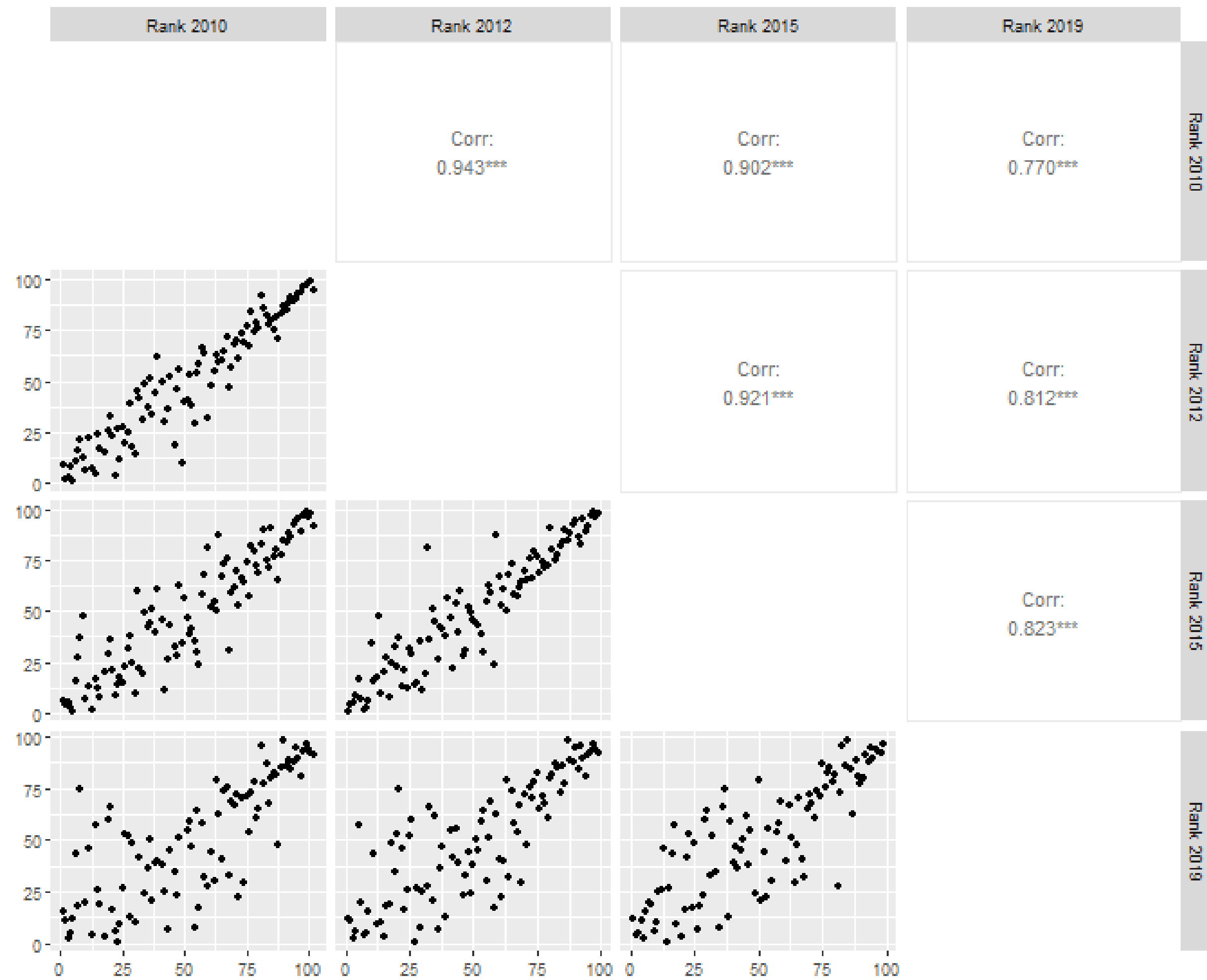
Syndicat SISPOLO	62	55	55	30	30	62	32
Syndicat SYNECOSPORT	24	12	18	9	9	24	15
Tandel	74	69	64	29	29	74	45
Troisvierges	82	86	90	77	77	90	13
Tuntange	9	13	48	NA	9	48	39
Useldange	47	46	28	23	23	47	24
Vianden	102	95	92	91	91	102	11
Vichten	59	32	81	28	28	81	53
Wahl	43	36	26	7	7	43	36
Waldbillig	54	29	35	8	8	54	46
Waldbredimus	32	42	22	42	22	42	20
Walferdange	52	53	39	59	39	59	20
Weiler-la-Tour	2	2	4	11	2	11	9
Weiswampach	84	78	71	68	68	84	16
Wellenstein	60	NA	NA	NA	60	60	0
Wiltz	100	98	96	94	94	100	6
Wincrange	58	64	68	32	32	68	36

2.2. Ranking order analysis

The previous analysis of Table 1 pointed out some instability in the ranking of the communes between the years, with some of them experiencing substantial movements (both upward and downward). An additional way to look at the correspondence between the indices calculated in different years is by focusing on the degree of correlation in the ranking order of a commune. Figure 1 depicts the pairwise rank comparison between the indices from 2010, 2012, 2015 and 2019. The top left of Figure 1 contains the rank distribution of the communes in 2010. Under the first square, we can see the scatter plot between the rank represented at the top of the figure, in this case 'rank_2010' with the 'rank_2012'. The white squares in the figure represent coefficients of correlation between two years, higher correlation means smaller changes in the rank position between the communes over the years. The squares with a curve show the density plot, which tells us how the communes are distributed along with the rank. The shape of the curve is a reflection of the changes in communes, due to communes been merged or the creation of syndicats, otherwise it would be a perfect Gaussian distribution.

The cases below the main diagonal visualize the different ranking of each municipality over two different years. Ideally, values on a 45 ° line would indicate exactly the same rank over the two years for each municipality. This feature seems to describe what mostly happens at the top of scatter charts. Looking more closely at Figure 1, the shape of its scatterplots shows that those communes presenting the highest values of the index (i.e. with vulnerable socio-economic compositions) generally remain in their position (tight grouping of communes in the top right corner of scatterplots). Communes with a more affluent population show moderate variability (more dispersed points appear in the left bottom corner of scatterplots). This is particularly the case for the years 2010, 2012, and 2015. The communes in the middle of the distribution are the ones less consistent thought the years, in particular for 2019 (middle part of scatterplot between the tails). Overall, a greater instability appears in their central part and remains in the lower left part. Furthermore, by comparing more and more distant years, instability increases. This is due to changes both in the socio-economic conditions of the municipalities and in the variables selected. These general trends are also confirmed by looking at the values of the correlation coefficients between the rank of each municipality in different years, contained in the cases at the top right of Figure 1, above the main diagonal. Correlation coefficients generally present high values, decreasing with the length of the time interval considered.

Figure 1. Pairwise comparison of the ranking order of communes



2.3. Stability of the social index over time

Table 2 shows the impact of the changes in the rank on the allocation of resources for the different communes. The final index is listed for all the communes for each year. The columns “min” and “max” represent the lowest and the highest value of the index that a commune was given in any of those years, while the column “variation” is the difference between the highest and the lowest index score (subtraction of “min” from “max” column). As the index is standardized between 100 and 120 points, the variation is smaller in magnitude, in contrast to the change in the ranking order in the section above. However, the final index score offers crucial information regarding the magnitude of the additional resources allocated to each commune or school district. As mentioned in Section 1.1., the extra support to communes ranges from 0% to 20% based on the index result. Communes that benefited the most from the contingency policy between 2010 and 2019 by receiving on average 15% to 20% more resources are: Beaufort, Diekirch, Differdange, Echternach, Esch-sur-Alzette, Ettelbruck, Pétange, Remich, Rumelange, Schifflange, Vianden and Wiltz. There is a number of communes with favorable social index scores that remained stable over time, hence they received under 5% of additional funding: Biwer, Contern, Garnich, Junglinster, Kehlen, Niederanven, Nommern, Reckange-sur-Mess, Schuttrange, and Weiler-la-Tour. Finally, there are communes that experienced dramatic changes over time. For example, the commune of Saeul experienced the largest increase in additional funding in comparison to other communes, moving from an extra 2.22% of resources allocated in the commune in (2010) to an additional 11.87% of resources in 2019, with an increase of 9.65 percentage points in the extra teaching quota. Other communes that benefited from the increasing support over time are: Préizerdaul (7.72 p.p.), Strassen (6.94 p.p.), Leudelange (6.56 p.p.), Ell (6.1 p.p.), Reisdorf (5.97 p.p.), Kopstal (5.45 p.p.), Remich (5.39 p.p.) However, there are communes that experienced a reduction in teaching quota: commune’s index stood at 105.86 in 2012 and dropped to 100 in 2019 that resulted in the loss of 5.86 p.p. Similarly, communes of Wahl (5.45 p.p.), Waldbillig (5.37 p.p.), Grosbous (5.14 p.p.), Fischbach (4.83 p.p.) underwent a similar path. Finally, some communes, like Vichten experienced an irregular pattern: between 2010 and 2015 the index value grew from 108.36 to 114.39 points that led to 6.03 percentage points increase during those five years, and followed a drop to 106.98 points in 2019 leading to a decrease by 7.41 percentage points.

The situation of the largest municipalities in Luxembourg, such as Luxembourg-city, Esch-sur-Alzette, Differdange, Dudelange, Ettelbrück, Petange, Wiltz has been stable throughout all the years in constrast to cases discussed above. They are the largest recipients of complementary funding since the first edition of the index in 2010 – e.g. Esch-sur-Alzette was eligible for an additional 19.3% of resources. Between 2010 and 2019 we observe a marginal change (1 to 2 p.p.) An important take from this is that these large communes well-known for their relatively unfavourable socio-economic and linguistic composition were not affected by the fluctuations in the ranking between the years that would have otherwise significantly impacted their complementary school resources.

Table 2. Index scores of the communes within and across the years

Commune	Index 2010	Index 2012	Index 2015	Index 2019	Min	Max	Variation
Äerenzdall	113.84	111.63	111.18	108.82	108.82	113.84	5.02
Bascharage	106.26	NA	NA	NA	106.26	106.26	0
Beaufort	115.21	114.91	115.17	114.9	114.9	115.21	0.31
Beckerich	106.23	109.97	110.26	107.7	106.23	110.26	4.03
Berdorf	108.26	110.71	109.77	109.67	108.26	110.71	2.45
Bertrange	104.56	105.21	105.72	109.14	104.56	109.14	4.58
Bettembourg	110.98	111.89	111.21	111.38	110.98	111.89	0.91
Bettendorf	111.92	112.12	114.03	112.56	111.92	114.03	2.11
Betzdorf	105.01	104.09	102.97	103.06	102.97	105.01	2.04
Bissen	110.49	109.33	109.93	111.31	109.33	111.31	1.98
Biwer	104.1	101.33	102.94	102.03	101.33	104.1	2.77
Boevange/Attert	105.28	106.48	104.85	NA	104.85	106.48	1.63
Bourscheid	108.03	108.56	106.77	110.56	106.77	110.56	3.79
Bous	104.8	107.64	107.26	103.95	103.95	107.64	3.69
Burmerange	102.63	NA	NA	NA	102.63	102.63	0
Clemency	103.33	NA	NA	NA	103.33	103.33	0
Clervaux	110.27	110.7	112.47	111.82	110.27	112.47	2.2
Colmar-Berg	111.74	114.88	114.63	111.79	111.74	114.88	3.14
Consdorf	107.62	107.61	107.71	108.1	107.61	108.1	0.49
Contern	100.99	100	100	103.63	100	103.63	3.63
Dalheim	105.58	107.6	107.77	107.52	105.58	107.77	2.19
Diekirch	116.31	115.95	117.32	115.62	115.62	117.32	1.7
Differdange	119.85	120	119.78	117.88	117.88	120	2.12
Dippach	104.04	105.48	105.6	104.56	104.04	105.6	1.56
Dudelange	113.79	114.18	114.22	113.71	113.71	114.22	0.51
Echternach	116.42	116.1	117.34	118.36	116.1	118.36	2.26
Eis schoul	NA	NA	NA	104.5	104.5	104.5	0
Ell	103.72	105.67	106.34	109.82	103.72	109.82	6.1
Erpeldange	106.06	106.87	108.63	105.3	105.3	108.63	3.33
Esch-sur-Alzette	119.3	119.05	120	119.3	119.05	120	0.95
Esch-sur-Sûre	111.45	112.32	112.52	111.48	111.45	112.52	1.07
Ettelbruck	117.8	118.71	119.1	117.92	117.8	119.1	1.3
Feulen	111.63	111.09	109.75	109.19	109.19	111.63	2.44
Fischbach	108.18	109.42	105.74	104.59	104.59	109.42	4.83
Frisange	104.85	104.96	105.74	108.85	104.85	108.85	4
Garnich	102.72	102.87	100.13	101.17	100.13	102.87	2.74

Goesdorf	105.93	108.46	107.81	108.85	105.93	108.85	2.92
Grevenmacher	112.54	112.18	111.45	110.57	110.57	112.54	1.97
Grosbous	110.92	109.57	108.81	105.78	105.78	110.92	5.14
Habscht	NA	NA	NA	107.15	107.15	107.15	0
Heffingen	104.24	105.86	103.98	100	100	105.86	5.86
Helperknapp	NA	NA	NA	107.37	107.37	107.37	0
Hesperange	109.1	109.98	108.54	112.72	108.54	112.72	4.18
Hobscheid	105.13	107.85	110.18	NA	105.13	110.18	5.05
Junglinster	102.5	102.28	102.54	105.1	102.28	105.1	2.82
Käerjeng	NA	107.28	107.94	110.33	107.28	110.33	3.05
Kayl	114.48	114.54	113.82	114.09	113.82	114.54	0.72
Kehlen	100	103.4	102.31	104.55	100	104.55	4.55
Koerich	106.46	106.29	103.35	106.7	103.35	106.7	3.35
Kopstal	102.57	105.39	103.74	108.02	102.57	108.02	5.45
Larochette	117.29	117.36	116.31	113.69	113.69	117.36	3.67
Lenningen	101.84	104.44	105.85	104.97	101.84	105.85	4.01
Leudelange	101.37	103.43	104.37	107.93	101.37	107.93	6.56
Lintgen	113.75	112.14	112.85	114.01	112.14	114.01	1.87
Lorentzweiler	109.05	108.07	108.63	107.93	107.93	109.05	1.12
Luxembourg	113.63	114.42	112.57	115.45	112.57	115.45	2.88
Mamer	103.04	105.51	103.67	106.83	103.04	106.83	3.79
Mersch	112.41	112.9	112.25	109.95	109.95	112.9	2.95
Mertert	110.58	111.29	110.32	110.77	110.32	111.29	0.97
Mertzig	109.58	109.52	111.37	107.76	107.76	111.37	3.61
Mompach	107.07	103.43	107.12	NA	103.43	107.12	3.69
Mondercange	104.78	105.54	106.87	109.05	104.78	109.05	4.27
Mondorf-les-Bains	110.34	111.7	112.6	112.35	110.34	112.6	2.26
Niederanven	100.6	103	100.29	101.17	100.29	103	2.71
Nommern	103.55	104.35	105.09	100.92	100.92	105.09	4.17
Pétange	116.47	116.42	118.2	116.88	116.42	118.2	1.78
Préizerdaul	102.77	101.78	104.51	109.5	101.78	109.5	7.72
Rambrouch	107.06	108.93	110.32	109.04	107.06	110.32	3.26
Rambrouch-Neunhausen	106.76	NA	NA	NA	106.76	106.76	0
Reckange-sur-Mess	100.33	101.26	101.04	100.48	100.33	101.26	0.93
Redange-sur-Attert	106.3	108.19	108	107.54	106.3	108.19	1.89
Reisdorf	112.82	116.41	114.7	118.79	112.82	118.79	5.97
Remich	114.61	115.37	115.28	120	114.61	120	5.39
Roeser	107.54	107.67	108.29	109.25	107.54	109.25	1.71
Rosport	107.07	107.65	109.28	NA	107.07	109.28	2.21

Rosport-Mompach	NA	NA	NA	107.53	107.53	107.53	0
Rumelange	115.54	115.82	115.78	116.02	115.54	116.02	0.48
Saeul	102.22	105.21	107.25	111.87	102.22	111.87	9.65
Sandweiler	104.46	106	104.21	106.92	104.21	106.92	2.71
Sanem	110.67	111.47	111.59	111.71	110.67	111.71	1.04
Schengen	110.39	108.01	106.85	107.29	106.85	110.39	3.54
Schieren	113.67	113.98	117.02	113.1	113.1	117.02	3.92
Schifflange	115.62	116.16	115.4	114.05	114.05	116.16	2.11
Schuttrange	103.28	104.86	102.82	105	102.82	105	2.18
Stadtbredimus	105.55	108.15	108.38	106.44	105.55	108.38	2.83
Steinfort	106.61	108.51	107.78	107.95	106.61	108.51	1.9
Steinsel	106.12	107.75	107.56	107.67	106.12	107.75	1.63
Strassen	103.76	106.74	107.24	110.7	103.76	110.7	6.94
Syndicat Billek Dreiborn	106.85	105.1	107	107.48	105.1	107.48	2.38
Syndicat SCHOULKAUZ	109.14	109.48	115.41	110.39	109.14	115.41	6.27
Syndicat Harlange	NA	107.73	108.89	109.26	107.73	109.26	1.53
Syndicat SISPOLO	109.09	108.82	109.01	107.04	107.04	109.09	2.05
Syndicat SYNECOSPORT	104.44	103.62	104.61	102.89	102.89	104.61	1.72
Tandel	111.02	111.32	110.85	107.01	107.01	111.32	4.31
Troisvierges	113.59	115.18	116.96	112.54	112.54	116.96	4.42
Tuntange	102.47	103.69	108.3	NA	102.47	108.3	5.83
Useldange	106.96	107.98	106.3	105.86	105.86	107.98	2.12
Vianden	120	118.11	117.29	117.18	117.18	120	2.82
Vichten	108.36	106.48	114.39	106.98	106.48	114.39	7.91
Wahl	106.58	107.5	105.76	102.05	102.05	107.5	5.45
Waldbillig	107.82	106.13	107.13	102.45	102.45	107.82	5.37
Waldbredimus	105.15	107.69	105.7	107.86	105.15	107.86	2.71
Walferdange	107.58	108.52	107.54	109.79	107.54	109.79	2.25
Weiler-la-Tour	100.16	100.49	100.49	103.41	100.16	103.41	3.25
Weiswampach	113.64	112.71	111.82	111.12	111.12	113.64	2.52
Wellenstein	108.8	NA	NA	NA	108.8	108.8	0
Wiltz	119.58	119.85	118.66	118.16	118.16	119.85	1.69
Wincrange	108.31	110.08	111.44	107.24	107.24	111.44	4.2

Due to the merging of some of the municipalities and the changes within the school districts (syndicates), there are different names and numbers of municipalities in each of the reports, there were 102 municipalities and syndicates in 2010, 99 in 2012, 99 in 2015, and 98 in 2019.

2.4. An overview of the dimensions over time

To get a better insight into the performance of the social index over time we perform three empirical exercises. First, we look at the rank correlation within each of the four dimensions across different editions of the index. Second, we look at the correlations between the four dimensions within each of the years. Finally, we re-calculate the index by excluding one of the dimensions in order to understand whether it affects the overall rank.

Figure 2 below captures the ranking of the communes between years with respect to the income component. Scatterplots reveal particularly high stability at the top and bottom of the distribution (upper right and lower left corners of each scatterplot), with moderate fluctuation in the middle. Overall, this is the most stable dimension.

Figure 2. Household income rank correlation

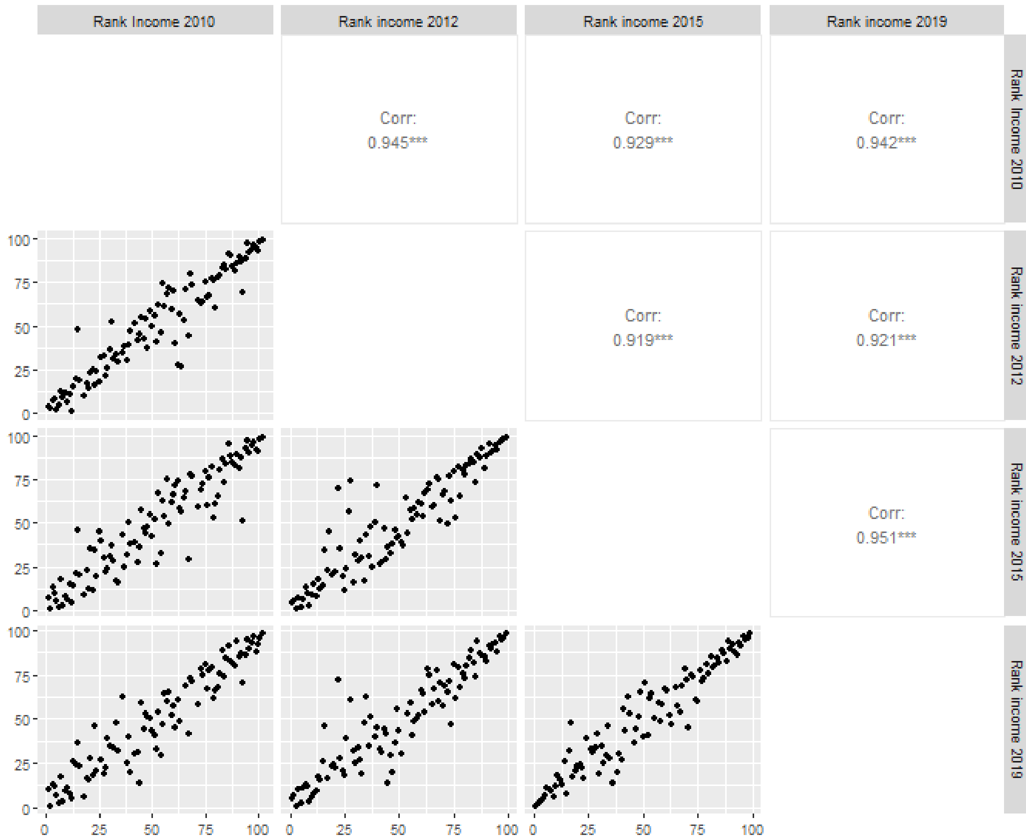


Figure 3 depicts the rank correlations in the second dimension of the social index related to the working conditions. In contrast to the income dimension that remains stable across all the years, this dimension has a high correlation in rank order of the communes between 2010 and 2012 and a fairly high correlation with results in 2015. However, the position of the communes in 2019 is rather different from the previous years, arguably as a result of significant data reduction and of the subsequent change in the methodology.

Figure 3. Working conditions rank correlation

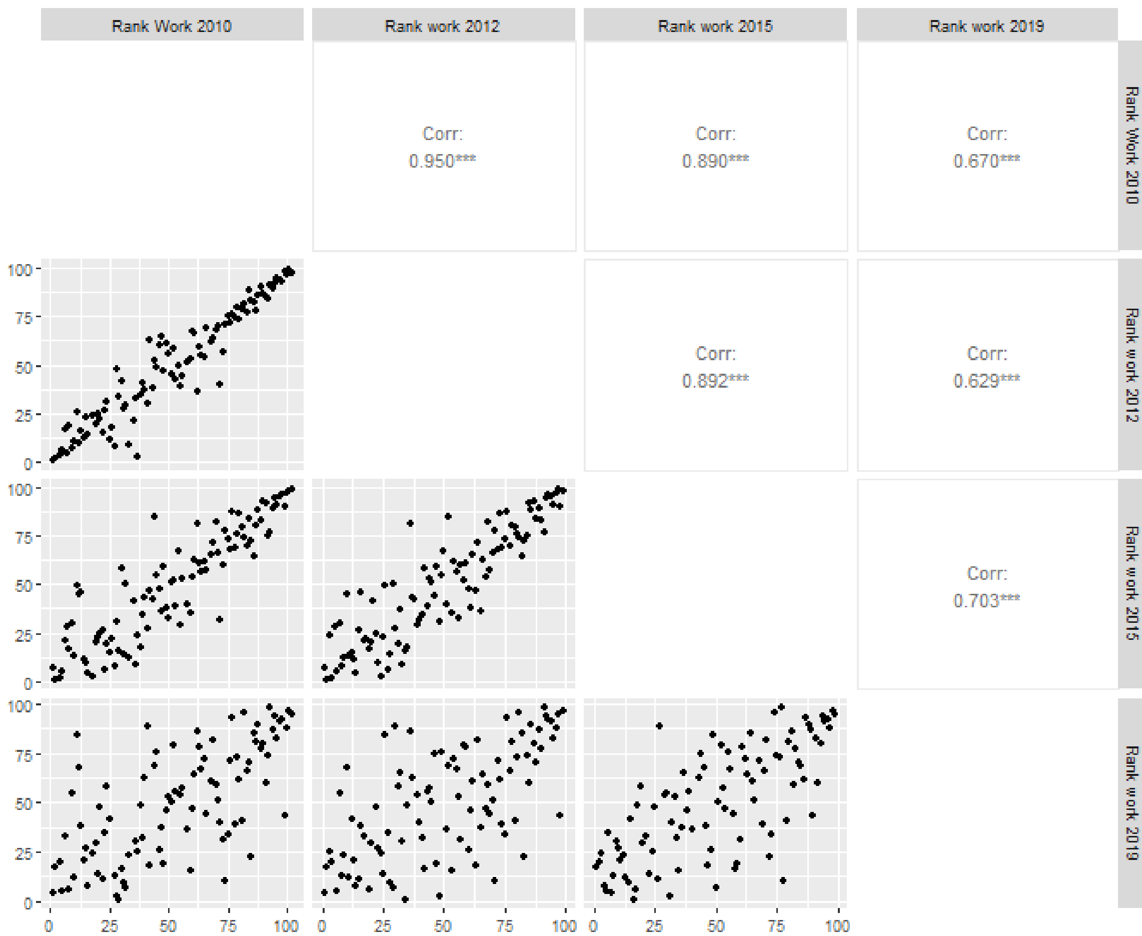
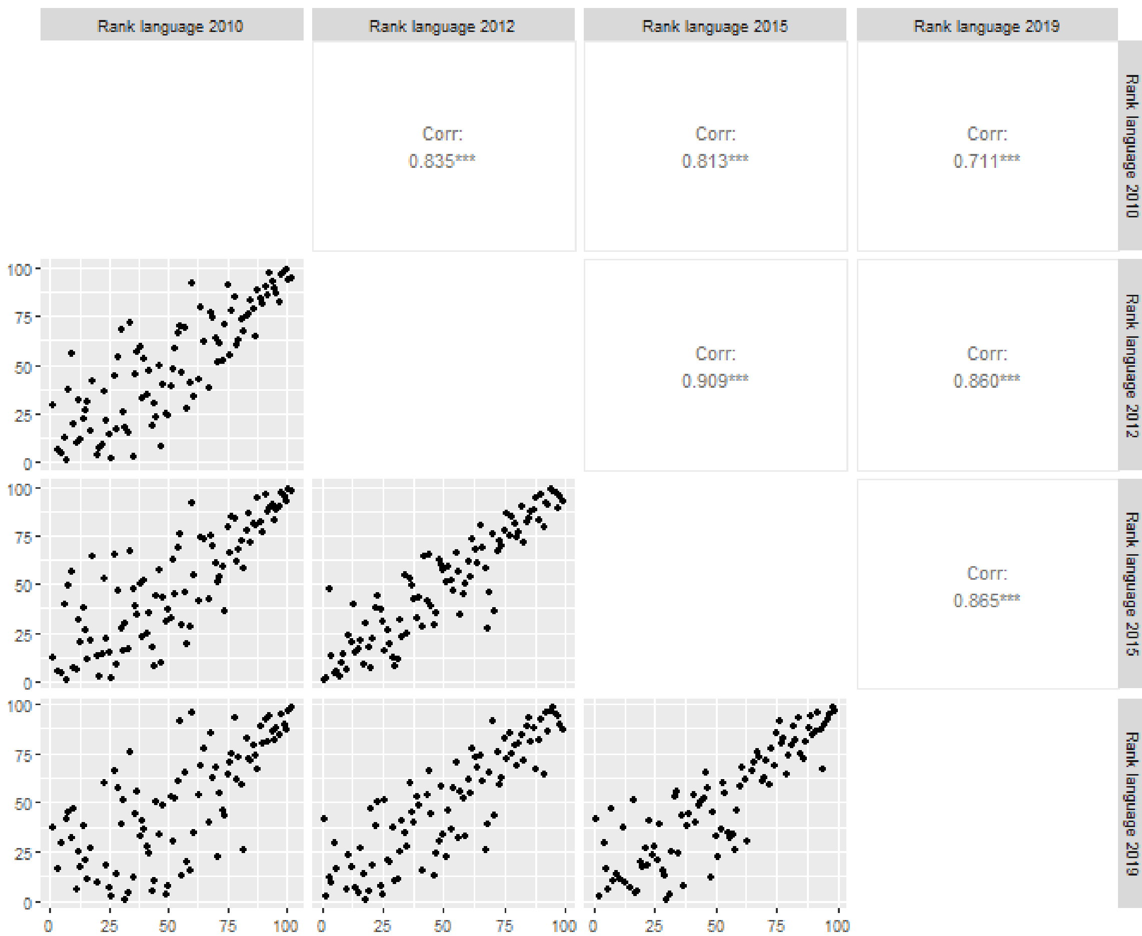


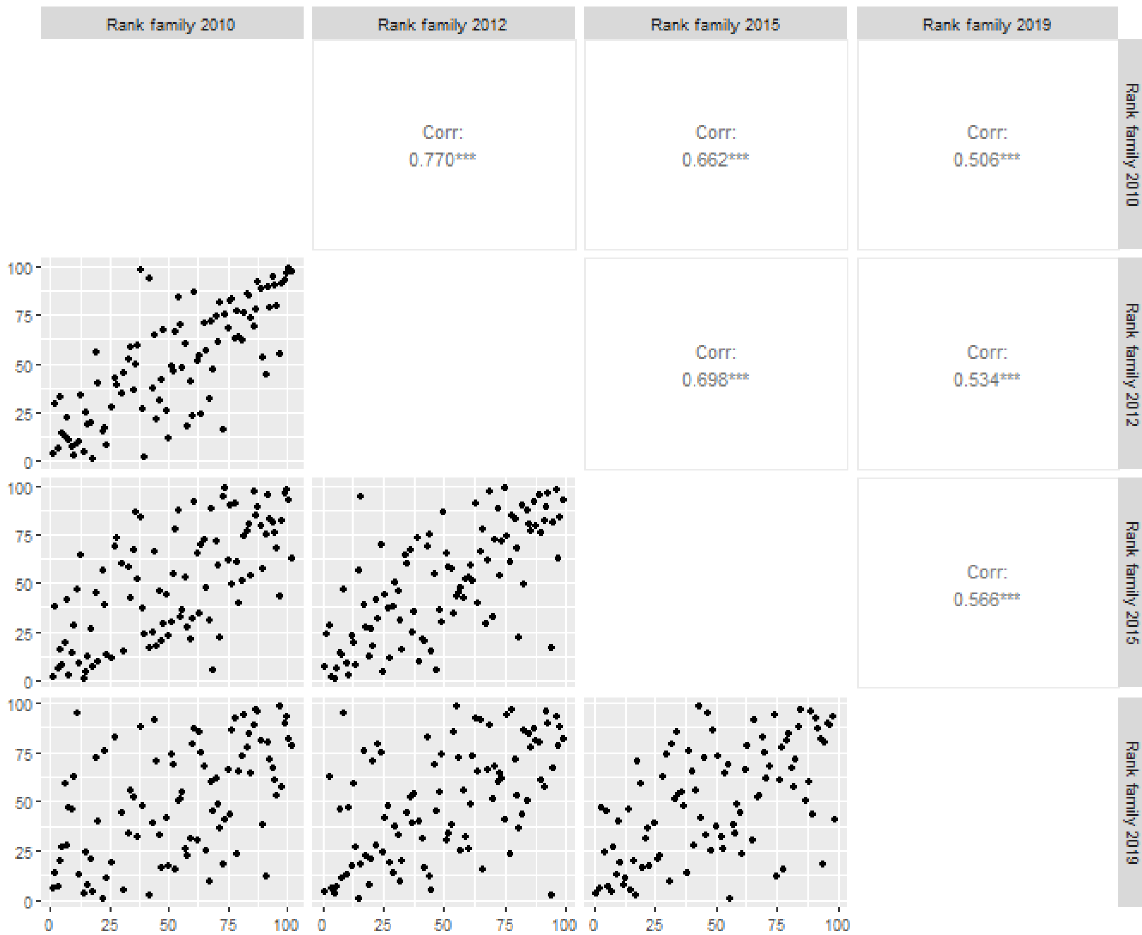
Figure 4 provides an interesting insight into the language dimension in the index. Overall, the correlation in the ranking of the communes between years is strong. There is a potentially important trend that we can see in the scatterplots: communes that are at the bottom of distribution with respect to language capital (upper right corner in each scatterplot) maintain their position over time. However, the middle and the top part of the distribution show a lot of movement. Interestingly, while the 2010 index was based on nationality and indices from 2012 onwards were based on the language spoken by the student and family, the correlation in the rank position between these two years is relatively high.

Figure 4. Language rank correlation



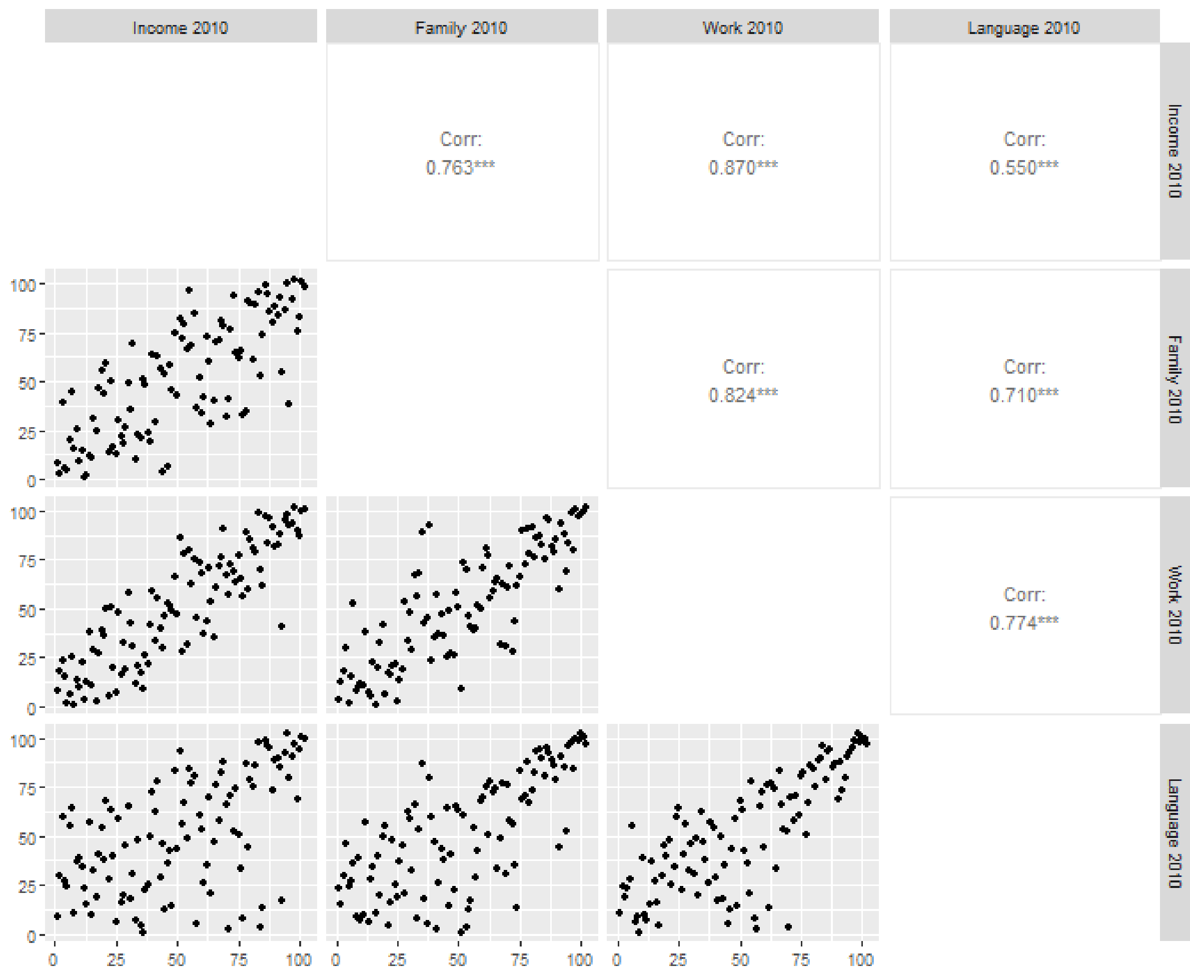
In comparison to the other three dimensions, the family composition is overall less consistent. The rank correlation is weak, in particular when compared to the 2019 results. As a reminder, in 2010, 2012 and 2015, the dimension was measured by two variables: age of the head of household and single- vs. two-parent family. In the 2019 index only the second variable was used. As mentioned above, the main reasons were the concerns over the data quality and the scientific relevance of the age information.

Figure 5. Family composition rank correlation



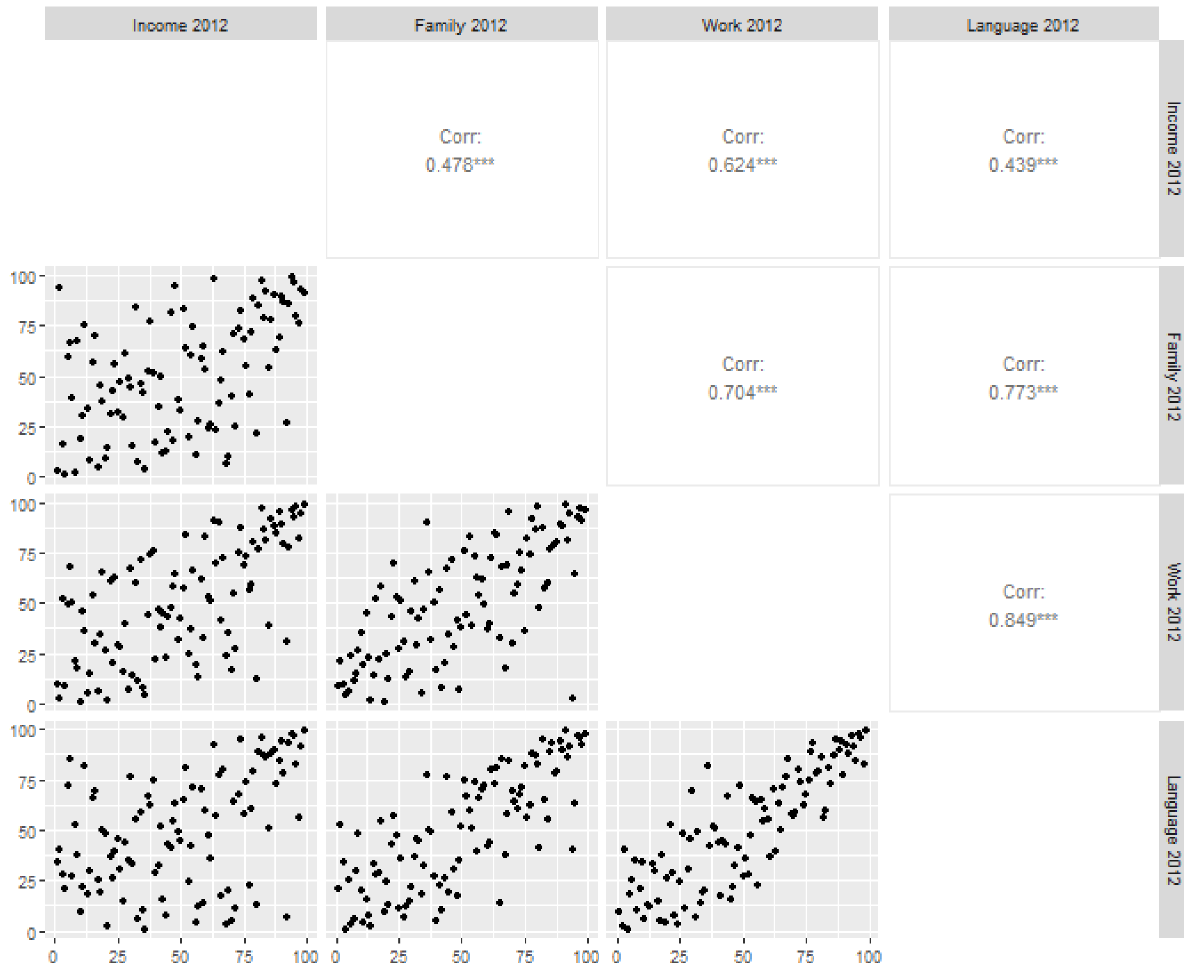
The second analysis, based on the correlations between the dimensions within each of the years reveals some mixed results. In 2010, dimensions “work” and “income” and “work” and “family” were correlated stronger than the other dimensions, with “language” and “income” having the weakest correlation. In general, in 2010 the correlation coefficients were the highest in comparison to the following years.

Figure 6. Correlation between the dimensions in 2010



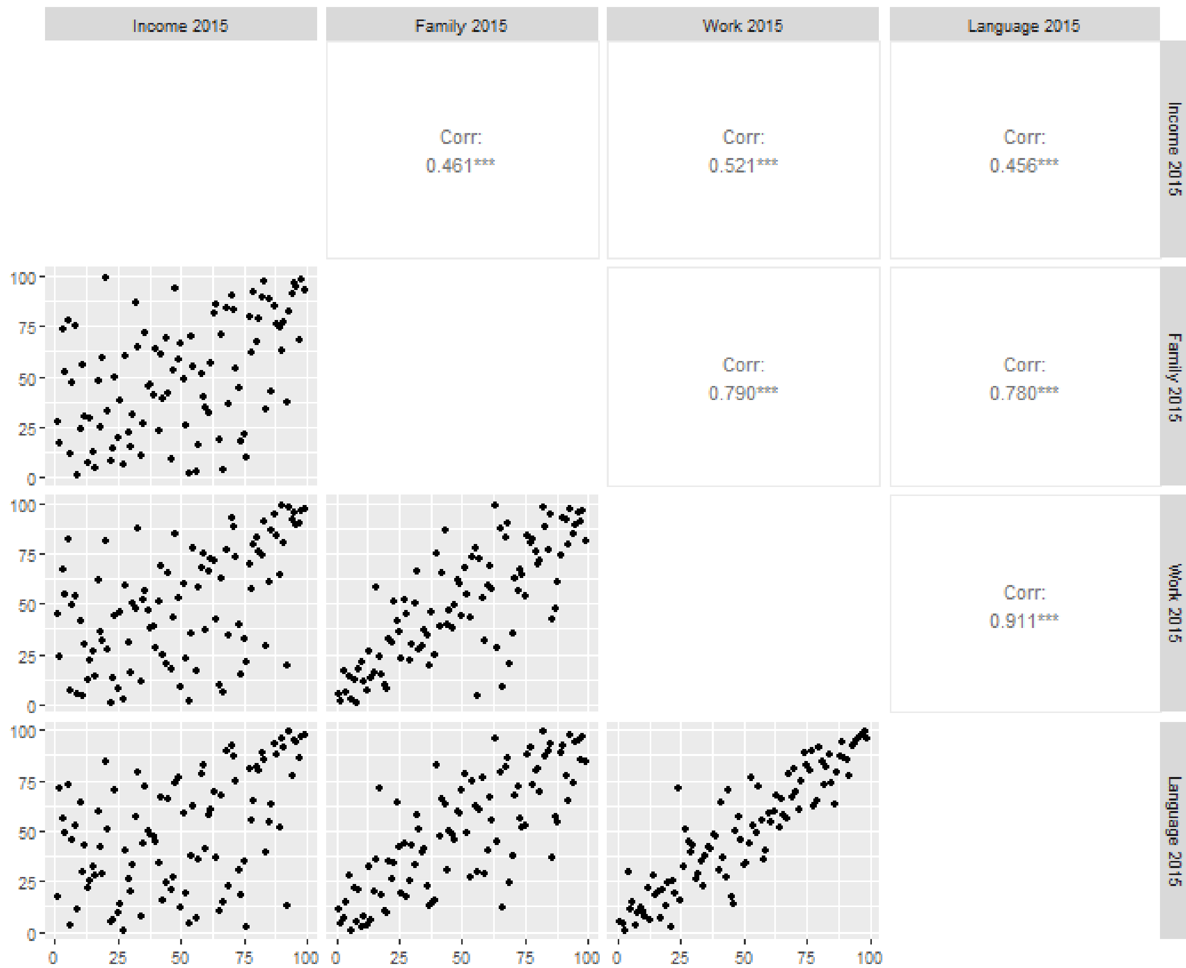
In 2012 associations between the dimensions were somewhat different. The highest correlation coefficients were observed between “language” and “work” and “language” and “family” dimensions, “Family” and “income” and “language” and “income” had the lowest correlation coefficients.

Figure 7. Correlation between the dimensions in 2012



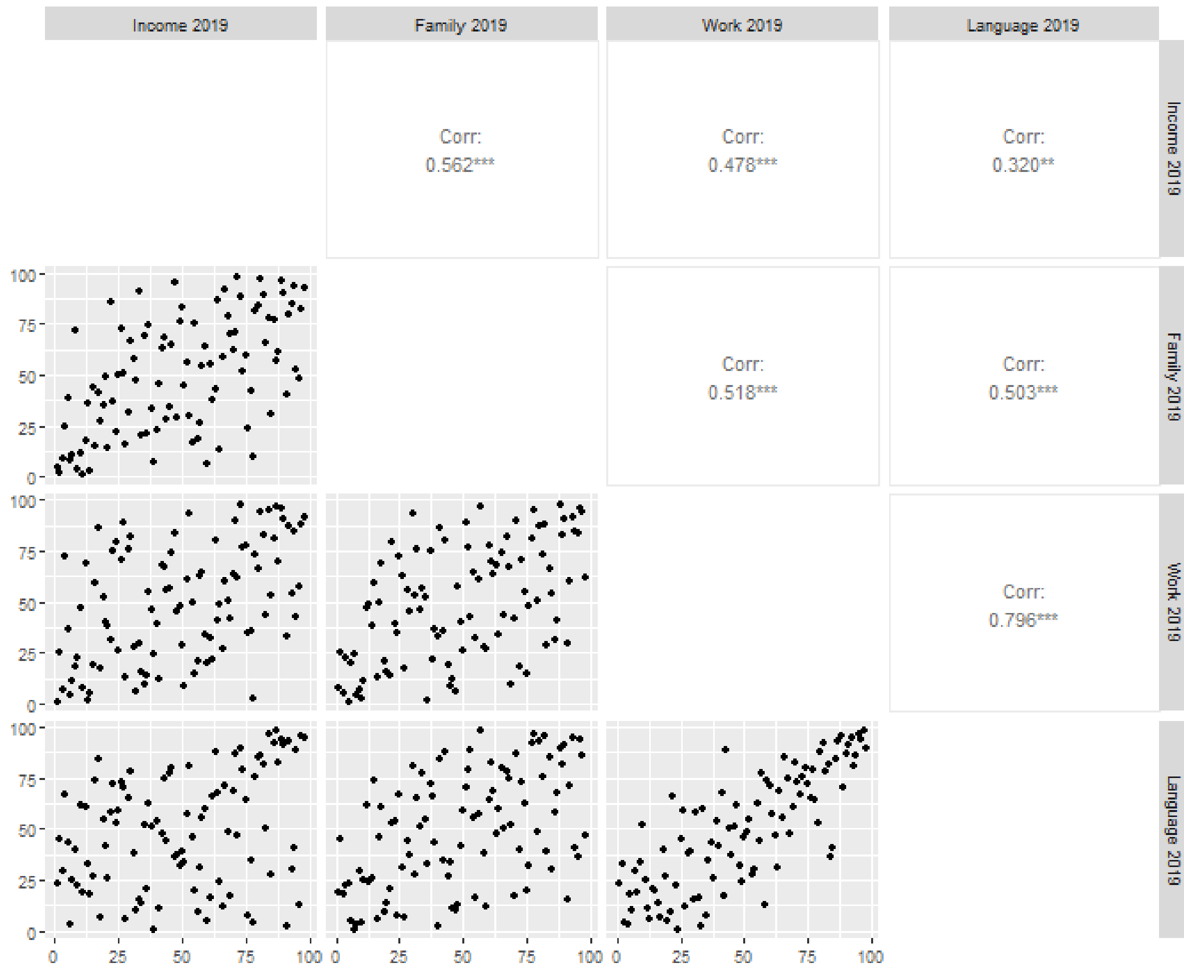
In the 2015 edition, correlations between “income” and three other dimensions were the weakest. The highest correlation was observed between “language” and “work” dimensions.

Figure 8. Correlation between the dimensions in 2015



Finally, in 2019, all correlation coefficients were in general lower in comparison to the previous years, with "language" and "work" being the only highly correlated dimensions. "Language" and "income" had the weakest correlation, similarly to the previous years.

Figure 9. Correlation between the dimensions in 2019



As mentioned at the start of the section, our third exercise tests the stability of the ranking of each commune when one of the dimensions is excluded. In this analysis, the ranking of each commune was recalculated four times omitting one of the dimensions. This has been repeated for each edition of the index (2010, 2012, 2015, 2019). Results (shared in a separate file due to a large number of columns and rows) frequently show large differences in rank changes in 2010, 2012, 2015. In 2019, the rank variation following the exclusion of each of the dimensions was significantly smaller.

This section concludes with a short note on the weights that each of these four dimensions has in the final index. Based on a multivariate linear regression, the weights for each of the dimensions can be extracted. Results suggest that each of these dimensions had different weights in 2010, 2012 and 2015, while in 2019 they received equal weight. The largest difference is observed in the year 2012, between the “working conditions” and “language” dimensions. Two potential explanations can be put forward to explain these differences. First, the calculation of the weights in the principal component analysis is sensitive to small changes in the values in data. Second, the weights also depend on the correlation between the dimensions, when the variables with high correlation are given less weight.

Table 3. Weights assigned to dimensions

Dimensions	2010	2012	2015	2019
Household income	24%	22%	27%	25%
Family structure	25%	27%	25%	25%
Language	24%	18%	20%	25%
Working conditions	27%	34%	28%	25%

2.5. Three commune-level indices in Luxembourg – how do they compare?

The reform on the financing of the communes in Luxembourg that started by the Ministry of Interior based on the law of 14 of December 2016 included, among other measures, the distribution mechanism of resources based on a socio-economic index calculated by the STATEC at the commune level.8 At the basis of the index there are five variables:

- The share of the guaranteed minimum income recipients;
- Unemployment rate;
- The median salary;
- The proportion of workers in “low-level” occupations;
- The share of single-parent households.

The indices calculated by LISER and the STATEC are aggregated at the communal level and cover similar socio-economic characteristics of the population, except the language background. Below we provide a

8 https://legilux.public.lu/eli/etat/leg/loi/2016/12/14/n1/jo

brief comparison based on rank correlations of the communes in two indices. Overall, the STATEC 2017 commune ranking and LISER ranking in 2010, 2012, and 2015 are highly correlated, particularly for the communes at the top and the bottom of the distribution. The correlation between STATEC 2017 and the LISER index of 2019 is on a weaker side. This goes in line with the correlation between the LISER 2019 index and indices produced in previous years in the framework of the teaching contingency mechanism.

As an empirical exercise, we compute a new index based on the 2017 data from the national school monitoring data Épreuves Standardisées (ÉpStan) collected by the LUCET. This index, also calculated at the municipality level, is based on information about parents' highest socio-economic status (HISEI) and language background. Similar to the STATEC 2017 index, the ÉpStan 2017 index is highly correlated with the LISER indices of 2010, 2012 and 2015 mainly for the communes at the bottom of rank, and a weaker correlation with the LISER index of 2019. The ÉpStan and STATEC municipal indices are also highly correlated with each other. To conclude, all three indices provide a very similar outlook on the socio-economic composition of the municipalities and rank order that results from it, albeit the latest LISER index for 2019 appears to be somewhat different from others.

Figure 10. Correlations between the social indices – STATEC, LISER and ÉpStan

2.6. Sensitivity analysis based on the exclusion of one commune at a time.

Another approach for testing the stability of the rank of a commune is to re-calculate its rank by excluding other communes, one at a time in an iterative manner. Following that, we register the highest and lowest rank order and compute the variation by subtracting the lowest from the highest position (results are shared in a separate file given the size of the table). A low variation value would then suggest that the rank position of a commune is robust. Results show the opposite: the rank positions of communes were unstable (with very few exceptions) with very large fluctuations in 2010, 2012 and 2015. In contrast to earlier years, the 2019 rank position was significantly more stable among all communes. Two main changes were implemented in the calculation of the social index in 2019: methodologically, PCA was no longer applied in comparison to earlier years; with regard to the data, dimensions were measured using a much smaller number of variables. Our findings indicate that any future change on the number of communes/syndicats can have a significant impact in the final rank when applying a PCA.

2.7. Stability of individual variables from 2010 to 2019.

One of the plausible hypotheses that needed to be tested to access what are the driven factors of instability in the rank was to look at the demographical changes that might have occurred between 2010 and 2019 in the communes, this could help to explain the rank variability. First we look at the individual variables based on the methodological approach applied in 2015 (the methodologies applied in 2010 and 2012 are very similar to the one applied in 2015) and 2019 to analyze if there is a significant variation between years, which would otherwise indicate that part of the instability is driven by natural changes in the demographics of the communes. For this we calculated the index, using both methodologies, for all the years available (2010-2019) using the raw data from IGSS and MENJE, looking at figures 19 to 32 in the annex, we can see that most of the variables are very stable through time. The exceptions are 'unemployment', 'monoactive' which summarize the number of households where just one of the parents works, and the 'fourth language condition', which deals with the percentage of children speaking either only German or German as the first language and any language other than Luxembourgish as a second language.

The variable 'unemployment', which is equally applied in 2015 and 2019, and the variable 'monoactive', which was only applied in 2015 (and in the previous years) have a very low variance, this means that small changes in one commune can have a significant impact in the rank of the commune. While in the case of the instability of the 'fourth language condition', it was only applied in 2015 (and in the previous years), we believe that this is driven primarily due to the small number of observations that fit in this criteria. This makes the variable sensitive to small changes.

We concluded that the demographics are stable, and that the variables that bring forward some level of instability do not reflect major changes in demographics but deal with a high level of susceptibility to changes due to the way they are measured. Some dimensions do not differ much across municipalities. Most notably the "family" variable. Therefore, an index that implies 'normalizing' differences (as implicit in PCA) and/or using commune's ranks in each dimension will mechanically have a lot of variability: small percentage point changes in a commune's raw 'family' variable can lead to big changes in "normalized" differences (or rank differences) and this automatically will reflect in variability the total index for that municipality.

2.8. Correlation between commune size and rank variation.

In order to better understand whether the observed fluctuation in rank order and social index score, among some of the municipalities, is dependent on their size, we performed three exercises.

First, using the existing information from previous indices (2010, 2012, 2015, 2019), we compute the correlation between the size of the commune (measured by the number of students per commune in IGSS data) and the variation of the rank position across the years to test the basic hypothesis whether small communes' ranking varies stronger than the rank of larger communes. Below are the graphs that depict changes across the years. We start by looking at a total variation across all the previous years, using both non-scaled and scaled (with mean equal to 0 and the standard deviation equal to 1) data (see Figure 11 and Figure 12). Results indicate that the municipality size does not affect its variation in ranking order.

Figure 11. Correlation between rank variation from 2010 to 2019 and number of students per commune

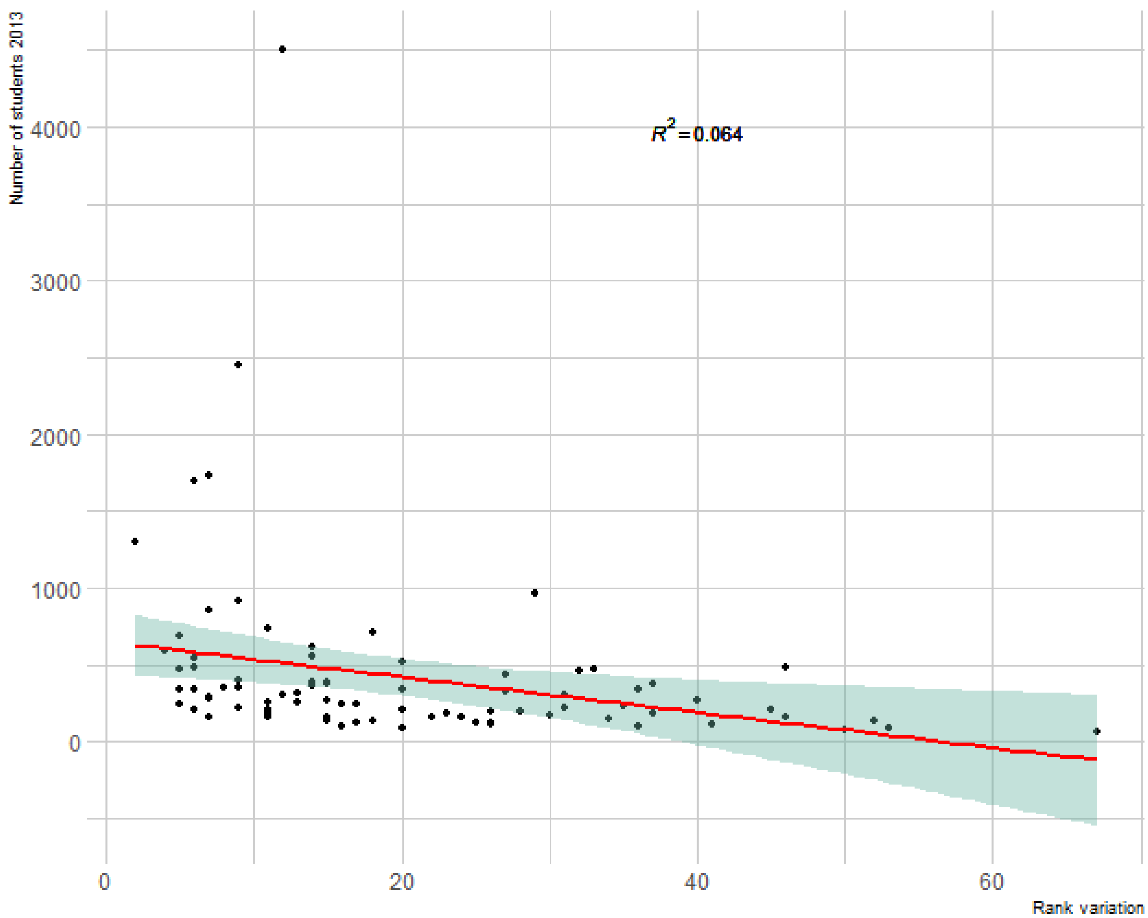


Figure 12. Correlation between rank variation from 2010 tp 2019 and number of students per commune (scaled)

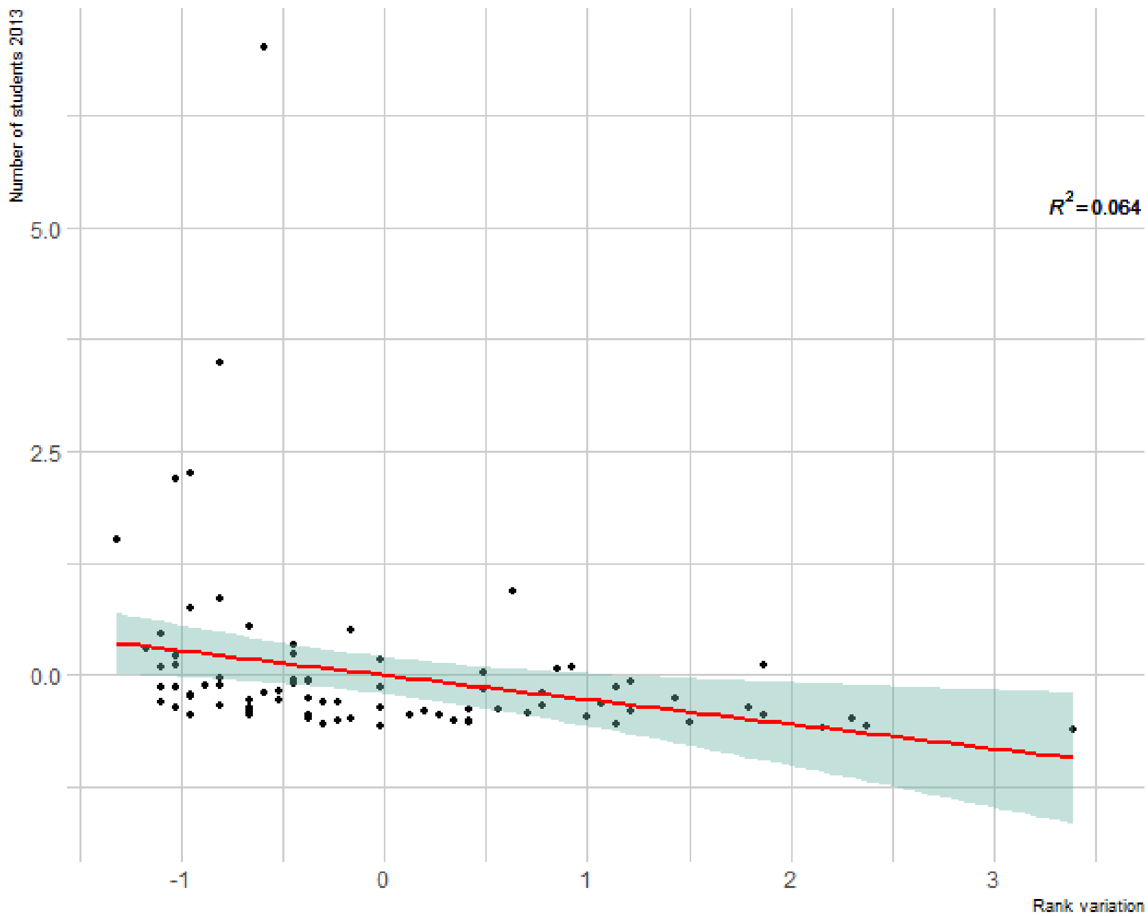


Figure 34 in the annex depicts the results from the same exercise, but excluding the year 2019, as the latest index had a lower correlation with the indices of earlier years. Similar to the results above, there appears to be a very weak association between the commune size and the rank variation for the years 2010, 2012 and 2015.

Finally, the Figure 35, Figure 36 and Figure 37 in annex show the correlation between two indices from one year to another, instead of the total variation. Regardless of what pairs of years are compared, the ranking fluctuation and the municipality size are not correlated to any considerable extent.

In the second exercise with respect to the effect of the communal size, we correlate the size with the variation in the social index score based on scaled data. Figure 38 (variation across 2010, 2012, 2015,

2019) and Figure 39 (variation across 2010, 2012, and 2015) reveal a very weak correlation with the size of a municipality.

Correlation between the variation of the social index from one to the following year, also, regardless of the compared years, does not show meaningful association with the commune size.

Finally, as a third exercise, we compare the change in the rank order as well as the social index score between larger and smaller communes. Using the student population data from 2013 from 94 communes, we identify the median population as equal to 254 students. Based on this number, the communes are split into two groups: those with more than 254 students (larger communes) and those with less than the median number (smaller communes). For each of the two groups, we sum the total variation in rank and compute the group mean. Smaller size communes have a total rank variation of 1181 positions for all 10 years with a mean value of 25. Larger size communes have a total rank variation of 755 positions for all 10 years with a mean value of 16. While smaller communes have a larger group mean value, larger communes' mean value is not too far from the first group, which means that no statistically significant effect can be found. When the same exercise is performed on index score variation, the results between the two groups are even closer to one another: total index variation in smaller communes is 182 with a group mean of 4, while in larger communes the total variation is 123 with group mean of 3.

To sum up, none of the three analyses carried out in this part indicate that the size of the commune is the main factor driving instability in the index, actually the impact of size on communes seems to be very small in the overall instability.

2.9. The impact of Principal Component Analysis (PCA) on the stability of the rank.

The most complex part of the methodology applied in 2010, 2012 and 2015 is the use of the Principal Component Analysis (PCA) for dimensionality reduction. The PCA, as mentioned before, is applied twice, first within the dimensions with more than one variable and then between dimensions. We applied the same methodology applied in 2015, which makes use of the PCA and calculated the index for the years 2010 to 2019 using the raw data from IGSS and MENJE. Looking at Figure 41 in the annex, we can see that the rank is highly stable through the years even if compared with the calculation of the index that does not use a PCA as in Figure 42 in the annex.

Second with repeat the same exercise, but this time we compare the stability of the rank when the second PCA is not applied, in this case instead of applying a PCA between the dimensions to have a final index number, we just sum the four dimensions and divide by four, similar with the methodology applied in

2019. Figure 43 in the annex shows that the application of the second PCA does not produce a significant effect on the stability of the rank.

These both exercises lead us to believe that although some qualitative assessment of the use of a PCA can be made in terms of loss of information and accuracy of the representation of the four dimensions, there is no indication that the use of a PCA in the calculation of the indices of 2010, 2012 and 2015 is responsible for the rank instability over the years.

2.10. The impact of methodological changes.

Finally, we investigate methodological changes that occurred between editions of the index, more specifically we look at the methodological approach used for the calculation of the index in 2015 and 2019. Although there were some minor changes in the selected variables between the index of 2010 and 2012 and between 2012 and 2015, the most significant methodological change happened from the index of 2015 to the one in 2019, were not only the PCA is not applied but a number of variables were not included for the calculation of the index of 2019. There were also some changes in the variables that remained in the index, for example in the Family dimension the consideration with the age of the parents does not take in account anymore in the edition of 2019, another drastic change was in the Language dimension where only German and Luxembourgish are measured for 2019.

While in the previous section we compared the two methodologies to see if one or another could produce a more stable rank, we now look at how the transition of one methodology to another can influence the rank stability. If we look at Figure 44 in the annex, we can see the correlation between method 1(the one used in 2015) and method 2(the one used in 2019), we check the correlation for the two methods comparing how they would have correlated with each other if we applied the two methodologies on the raw data in the years 2010, 2015 and 2019. Initially they seem to have a very good correlation, however if we look at the correlation between the rank of 2015 using ‘method 1’ and the rank of 2019 using ‘method 2’, we can see a considerable increase of entropy in their correlation, this reflects exactly what happened when we analyze the correlation between the index calculated in 2015 and the one from 2019 presented in the report from 2015 and 2019.

PART 2 – Overview of existing composite indicators

The CIs that are compared to the socio-economic and cultural index calculated in the framework of educational contingency in Luxembourg are:

- "Der Sozialindex für Hamburger Schulen" [Social index for Hamburg schools] from city of Hamburg in Germany (Schulte et al., 2014),
- « L'indice de position sociale des élèves » [Index of student's social position] with the national coverage from France (Rocher, 2016),
- "Der Sozialindex" [Social index] from the city of Zürich in Switzerland,9
- "L'indice socioéconomique dans l'enseignement fondamental et secondaire" [Socioeconomic index in primary and secondary education] from the Federation Wallonia-Brussels10

These CIs have the common goal of measuring the socio-economic status of students' families, with the aim of assisting the policy-makers to strengthen the allocation mechanisms of different resources between public schools. The comparison of these CIs reveals different choices made for their construction, both with respect to data sources, as well as methodological procedures.

Table 4 below summarizes the main **dimensions** taken into account for building these CIs. Although it is possible to notice some similarities between countries, several differences appear. The economic situation is one of the main dimensions in all five analyzed CIs and is measured through the *income of the households* of the students. Wallonia-Brussels, Hamburg and Luxembourg (2010, 2012, 2015) take into account additionally the type of job of parents, as well as the *unemployment* experience, which affects the economic dimension and potentially also the psychological dimension (Ruiz-Valenzuela, 2020). In Luxembourg, unemployment is integrated into 2015 and 2019 and measured through the share of individuals receiving unemployment benefits within a given commune. Reliance on *social assistance* is also included in the economic dimension in Zurich, Wallonia-Brussels, Hamburg, as well as in Luxembourg (in 2015 and 2019). Two CIs, in Hamburg and in France consider additionally the *living conditions* of a family by including information on whether a child has own room, the total number of rooms in the house, and availability of a computer and internet connection.

9 More information: https://www.zh.ch/de/bildung/informationen-fuer-schulen/informationen-volksschule/volksschule-fuehrung/volksschule-klassen-stellen-planen/volksschule-stellen-berechnen.html#-858642922

10 More information: https://statistiques.cfwb.be/enseignement/fondamental-et-secondaire/indice-socioeconomique/ and https://www.gallilex.cfwb.be/document/pdf/44433_002.pdf

Other important characteristics of parents that greatly affect the student's achievement, namely their *education level* (OECD, 2012; OECD, 2019) is used only in three of the five CIs, in Wallonia-Brussels, Hamburg, and France (see the right column "Cultural Dimension" in Table 1). Participation and exposure to *culture and cultural activities*, another important dimension for academic achievement (Bourdieu, 1986) are explicitly included in two of the five CIs – in Hamburg and in France.

The migratory background is associated with vulnerability in the educational context through at least two fundamental mechanisms: language skills and linguistic background of the country of origin, the economic position of the families (OECD, 2019). The overview of five CIs reveals different approaches and choices with regard to the inclusion of this dimension. The CI from France does not consider immigration, because of a limited measured impact of the migrant background on students' performance (Rocher, 2016). This is also the case for Wallonia-Brussels, where migration is not included in the index. Zurich CI takes into account a share of pupils with foreign *nationality*, but not the language skills. Hamburg CI remains the most detailed in this regard: it includes both the country of birth of father and mother, the frequency of speaking German with mother, father, and siblings. Luxembourg has made several changes throughout the years: in 2010 it started with the nationality of students, from 2012 and onwards the *language background* of students was taken for the calculation of this dimension, but the measurement varied from one edition to the next (see Part 1, section 1.1.).

Some of the CIs include the social dimension that consists of two main sub-dimensions: *family structure*, namely differentiation between single- and two-parents households, as implemented in Luxembourg; and *parent-child interaction*, as implemented in Hamburg and France. Hamburg CI includes additional information about the interaction of a child with classmates. Also, Hamburg CI is the only one among five that includes *political participation* information measured as voter turnout.

To summarize, even when there are similar dimensions in CIs, the **specific variables** chosen to asses them can differ substantially. For example, the migration dimension is introduced in different ways: Luxembourg, which is a multilingual country with the highest percentage of migrants in the EU, considers the language spoken at home, Hamburg exploits more detailed information on the background of each of the parents and the student, including several different measures for this dimension, while in the Zurich index only the nationality of the student is accounted for. Arguably, the choice of indicators used to capture each dimension is related to the availability and nature of the data, but it can also reflect the social and cultural context of the country where the index was created. Some of the indices appear to be inspired by the contextual variables used in the PISA test, which seek to capture not only the socio-economic background of students but also other aspects such as pupils' ambitions and aspirations (OECD, 2019).

Table 4. Summary of indicators for the selected CIs

CI	Economic Dimension	Social Dimension	Migration Dimension	Cultural Dimension
Luxembourg	Equivalised income (2010, 2012, 2015, 2019), Share of recipients of unemployment benefits and the minimum income guarantee (2010, 2012, 2015, 2019), Share of blue-collar workers vs. white-collar and state employees (2010, 2012, 2015), Share of families with one economically active adult vs. two adults (2010, 2012, 2015)	Family structure - share of single- vs. two-parent families (2010, 2012, 2015, 2019). Age of the head of household (2010, 2012, 2015, 2019)	Nationality (2010), Languages spoke at home (2012, 2015, 2019)	
Zurich	Share of taxpayers, with at least one eligible child, with income below the cantonal median, Proportion of children and adolescents aged 5 to 14 years with social assistance		Proportion of pupils with foreign nationality	
Wallonia	Average household income, Unemployment rate, Share of manual workers, Share of people working in the third sector, Share of the household that obtained social assistance			Highest educational qualification in the household, Lowest educational qualification in the household
Hamburg	Income, Work class of the father, Work class of the mother, Child has his room, Unemployment rate, Share of people in need of assistance who are unable to work	Child spends his free time with his classmates, Child spends his free time with parents, The parents praise the child for a good school grade, The parents are proud of the child, Voter turnout	Country of birth father, Country of birth Mother, Frequency of speaking German with mother, Frequency of speaking German with father, Frequency of speaking German with siblings	Number of books at home, Frequency of visiting the museum together with parents, Education qualification father, Education qualification mother
France	Income, Number of rooms at the house, Room, Computer, Internet	*Aspirations, Most useful diploma, Parent involvement, Conversations (school life), Conversation (school future)		Number of books at home, T.V. at the bedroom, Time spent watching T.V., Attendance to sports, Attendance to concerts, Attendance to theater, Attendance to cinema, Attendance to museum, Extra-curricular activities, **Education qualification father, Education qualification mother

*Ambitions dimension; ** France has a separate dimension for parents

Also the **number of indicators** used to proxy the different dimensions can dramatically differ. Hamburg and France used respectively 24 and 21 indicators for the construction of their index, followed by Wallonia with 7, Luxembourg has 5 indicators and Zurich only 3. In theory, more indicators do not necessarily translate into a better CI. The choice of the number of indicators depends on their ability to correctly capture the effects of the measured dimension on school performance. Following this logic, the inclusion of uncorrelated indicators would only add noise to the data. However, some indicators could be considered essential to describe the socio-economic conditions of families, regardless of their impact on school performance. To understand exactly the nature of the CIs, it would therefore be necessary to have a thorough knowledge of the legislator's aims and normative views, which are not always explained in a way that is easy to conciliate with the adopted statistical methodology for building the CIs.

The **reference unit** for the index is also different between the compared CIs. Luxembourg's calculation is historically based on the municipality-level. Given the population size, the majority of municipalities have only one primary school, hence the social index is calculated *de facto*, at the school level. However, there are traditionally bigger municipalities e.g., Luxembourg city (22 schools), Esch-sur-Alzette (9 schools), Differdange, Dudelange, Sanem (6 schools each) with more heterogeneity in their population (see Part 3 for discussion). Additionally, there have been changes in the sizes of communes following their merging since 2004 that affected at least 30 municipalities. CIs in Hamburg and France calculate the indicator at the school level, while Zurich and Wallonia-Brussels use the district for their base calculation. This has an important consequence for the main goal of CIs, which is to inform policy-makers to implement suitable compensatory measures. In this perspective, the capacity of the policy-maker to identify, as precisely as possible, a unit of intervention (e.g. a specific school or school district) where the resources need to be allocated for the maximum impact, is key for improving the efficiency of such policy tools. However, the socio-economic situation of families can depend on some local characteristics, amenities and policies, which often go beyond the school level, which generate externalities among school districts. Ideally, the optimal statistical identification of the "jurisdiction" should be determined by maximizing the predictive capacity of the adopted methodology, which is a measurable feature of the analysis.

Regarding the statistical **methodology** behind these five CIs, each of them adopted a different approach. Luxembourg (2019) and Zurich did not perform multivariate analysis, Hamburg used a Factorial Analysis and France, in which the data was composed of categorical variables, applied Multiple Correspondence Analysis. Finally, Wallonia-Brussels chose their variables based on a Principle Component Analysis. This implies that had these five CIs been computed by using the same data, they could have reached quite different conclusions about which variable to use in their final model. The choice of variables and the distribution of weights between these variables on the index represent fundamental decisions and a substantial amount of literature on CIs stressed the importance of these two steps in index construction. The most contentious part of this literature deals with the implementation of the **weights**. Different methodological choices can lead to very different final ranks based on the CI, and the same unit of analysis can move from the first to the last place of the rank after a non-dramatic change in the weights (Becker et al., 1987). Luxembourg applied equal weights for the construction of the index in 2019, both between the dimensions and between the indicators (for other editions of the index see Table 3). This implies a logic of substitution behind the aggregation of the index: for example, one unit from the indicator ‘income’ can be compensated by one unit of the indicator ‘family structure’. This leads not only to an assumption of equal importance between indicators, but it also carries implicit assumptions about this trade-off, the independency of the indicators, and the possibility to explain the marginal contribution of each variable separately. Table 2 summarizes the above discussion.

Table 5. Summary of selected CIs on education

CI	Number of dimensions	Total number of variables	Level of aggregation	Multivariate analysis
Luxembourg	4	5	municipality /school syndicat	
Zurich	2	3	district	
Wallonia-Brussels	2	7	district	Principal Component Analysis (PCA)
Hamburg	4	24	school	Factorial Analysis (FA)
France	5	21	school	Multiple Correspondence Analysis (MCA)

Finally, the **data sources** also vary between CIs: France and Hamburg use a combination of data from the official government database and the surveys. In contrast, Luxembourg, Wallonia-Brussels and Zurich rely only on data from the government. The decision of including survey data in a CI has its pro and cons: more relevant information can lead to a more precise CI, however both the quality of the information obtained through the surveys as well as the costs associated with carrying out the surveys need to be taken into consideration. These points are particularly relevant when for keeping the data up to date for the revision of the index.

It is essential to mention that the CIs discussed in this section were created in countries with **different socio-economic, cultural and political backgrounds**. In other words, to fully assess the quality of a CI the context and the precise objectives of the policy should be taken into account. Similarly, from the perspective of data, each country has its own idiosyncrasies and therefore, it is a challenge to develop a model that "fits all". Instead, analyzing the different perspectives underlying these five CIs can help not only to understand their features, but also to identify relevant criteria for assessing their robustness to changes in variables and methodology, as a crucial element of our analysis.

PART 3 – Inequality beyond the municipality level: an exploratory analysis

The CI adopted in Luxembourg focuses on the socio-economic composition of students attending public schools, which is aggregated at the municipal level. However, other potential sources of inequality may affect the education system but do not necessarily reflect differences existing at the municipal level. To justify this conjecture, let us consider the trends of the average household incomes across communes over time. Table 6 shows a slight variation from 2010 to 2019. This is also true for the GINI coefficient of inequality calculated in table 7. Not only the GINI index does have a limited change over time, but it is also pointing to a moderate level of inequality between the communes.

Table 6. Minimum, median, mean and maximum percentage variation of income between communes – 2010 - 2019

Parameters	2010 - 2011	2011 - 2012	2012 - 2013	2013 - 2014	2014 - 2015	2015 - 2016	2016 - 2017	2017 - 2018	2018 - 2019
Min.:	-8	-3	-8.7	-4.6	-8.4	-10.2	-10.3	-4.5	-13.2
Median:	1.8	2.8	1.4	2	1.4	0.4	4.8	3.1	2.4
Mean:	1.8	3	1.4	1.9	1.4	0.5	4.8	3.6	2.1
Max.:	8.7	14.6	7.5	13	8.7	16.8	19.6	14.2	8.2

Table 7. GINI coefficient, school population – 2010 - 2019

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
GINI	0.250	0.246	0.250	0.250	0.252	0.255	0.258	0.259	0.263	0.257

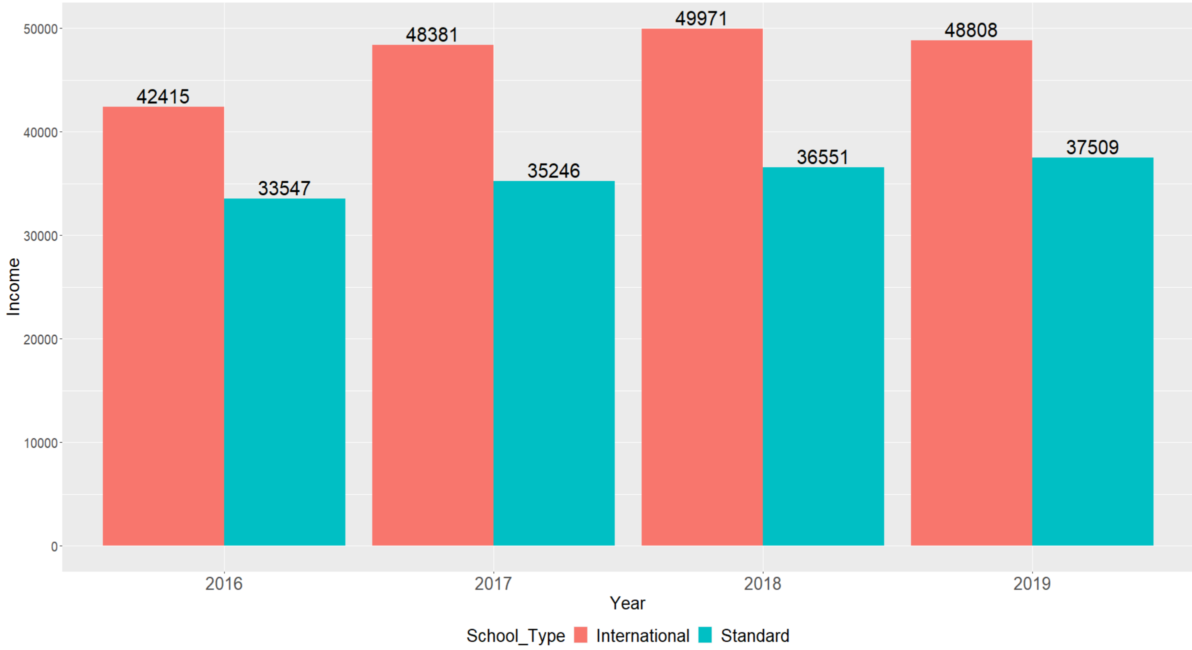
These figures suggest investigating some potential additional differentiation factors in the national school system, which could guide the sorting of families between different schools based on their socio-economic composition and are not captured by the previous analysis.

Based on the data provided by the IGSS and MENJE, this part analyses the social stratification 1) for different types of school, such as public standard and public international; 2) between the student population who attended "Précoce" education and those who did not. Finally, 3) the heterogeneity of the student population among the schools of the same municipality is also analyzed.

3.1. Public standard vs. public international schools

We consider four public internationals schools located in five different communes: Luxembourg, International School of Differdange and Esch-sur-Alzette (EIDE) which have a unified data as a single school, Junglinster, and Mondorf-les-Bains. These schools are publicly funded; however, they differ from the other public schools as they offer different international curriculums with different options regarding the primary language. Figure 13 contrasts the **household income** data of the pupils enrolling in these schools with the school population of other public primary schools. We remark that families of students from the international schools have higher income levels in all the years analyzed. In 2018 and 2019 this difference was more than 30%.

Figure 13. Average household income by school type – 2016 - 2019 (in euros)



Moving to the citizenship/cultural dimension, consistently with the current CI for Luxembourg, we use the first **language spoken** by the students. Figure 14 shows the relationship between income and the first language: German and Luxemburgish languages are associated with higher income and non-European and Portuguese languages are associated with low income households. French and EU languages are in between the two groups. Non-European languages make an interesting case: figure 8 shows that these languages are associated with low income levels. However, this group comprises a plethora of different languages (arguably related to different types of migration in Luxembourg), corresponding to different income levels. In other words, it is not possible to cluster this population as a homogeneous group, however at Figure 45 in the annex it is possible to see an important difference in the average income of students who have the first language non-European languages in standard schools and international schools, the later has an average income much higher that students in standard school. Finally, the language and income association remains stable across all analyzed years.

Figure 14. Average household income by first spoken language – 2010 - 2019 (in euros)

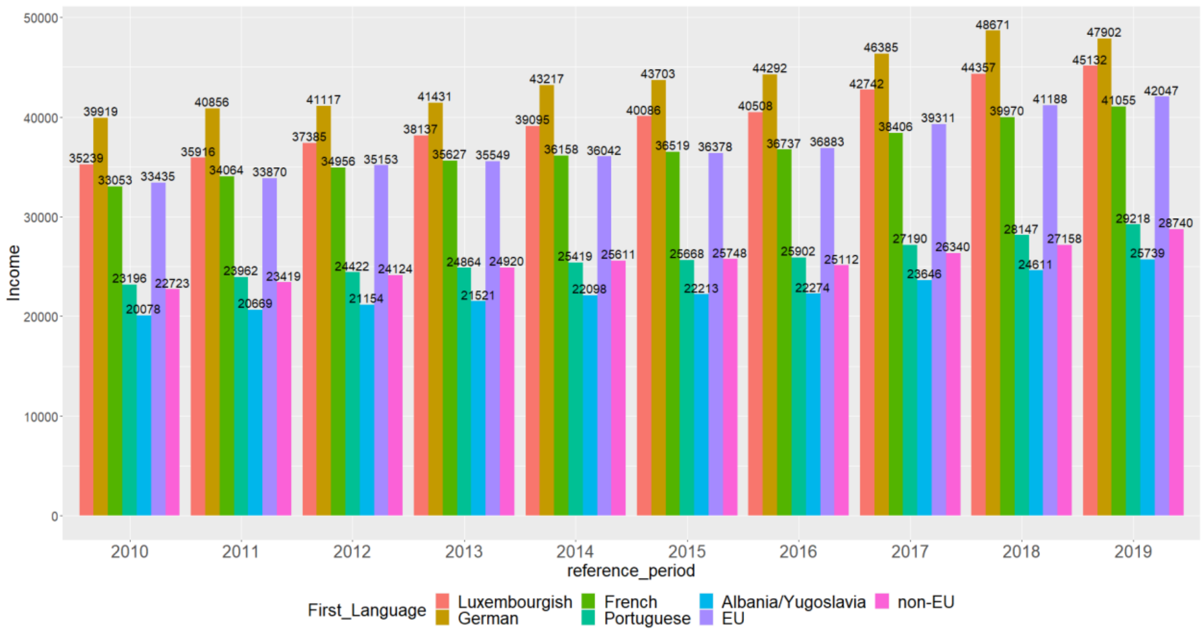
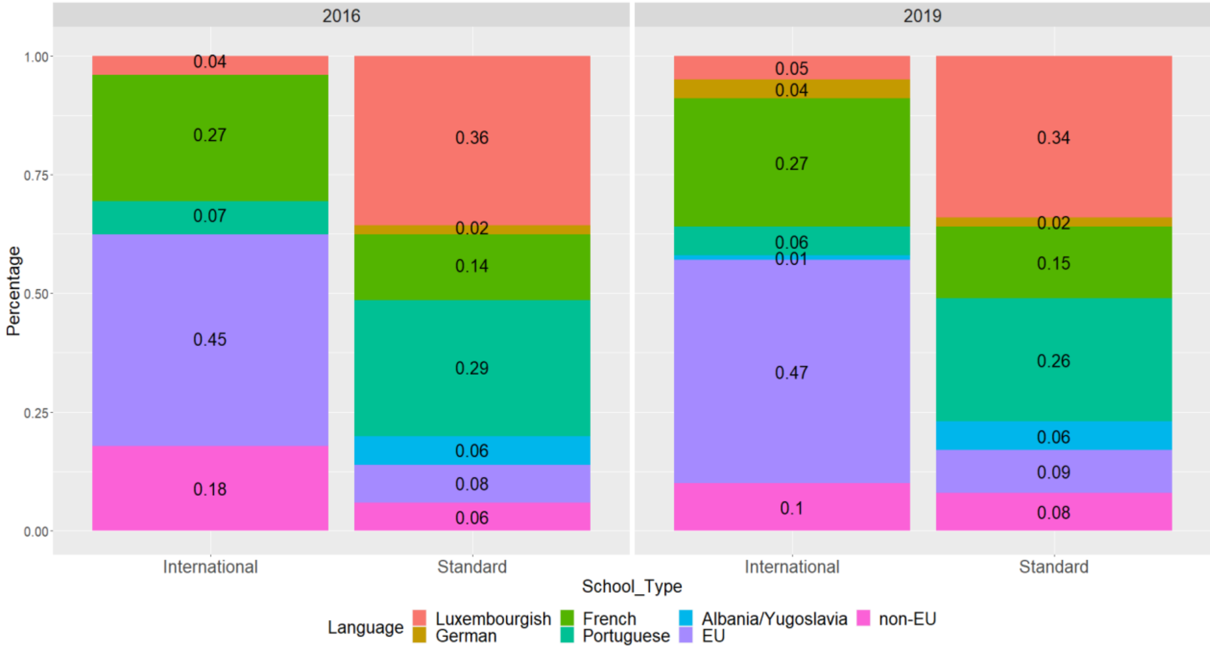


Figure 15 provides a comparison between the population of public international and public standard schools by their first language. Data shows a clear overrepresentation of French and other EU languages in the public international schools when compared to standard schools, while the Luxembourgish and Portuguese languages are much more popular standard public schools. While the corresponding data is available only for the last few years, the distribution in terms of percentages remains very similar.

Figure 15. Proportion of student's first language by type of school – 2016 and 2019



3.2. Participation in early education (éducation précoce)

One of the EU's educational policy targets is to provide regular early childhood education and care (ECEC) to 95% of children by 2020. The legal entitlement to early education and introduction of childcare vouchers in 2009 resulted in a more than three-fold increase of available childcare places for children 0 to 3 y.o. and between 2009 and 2011. Attendance rates among children aged 0 to 5 grew from 28% to nearly 50% between 2009 and 2013 (Bousselin 2019). While recent statistics show a successful inclusion of children aged 4 years and older (95%), there remains a noticeable gap with the participation of children aged 3 (85%).Missing out on quality ECEC leads to different early literacy starting levels among children in primary education, with children whose parents fail to provide compensatory educational input being particularly disadvantaged.

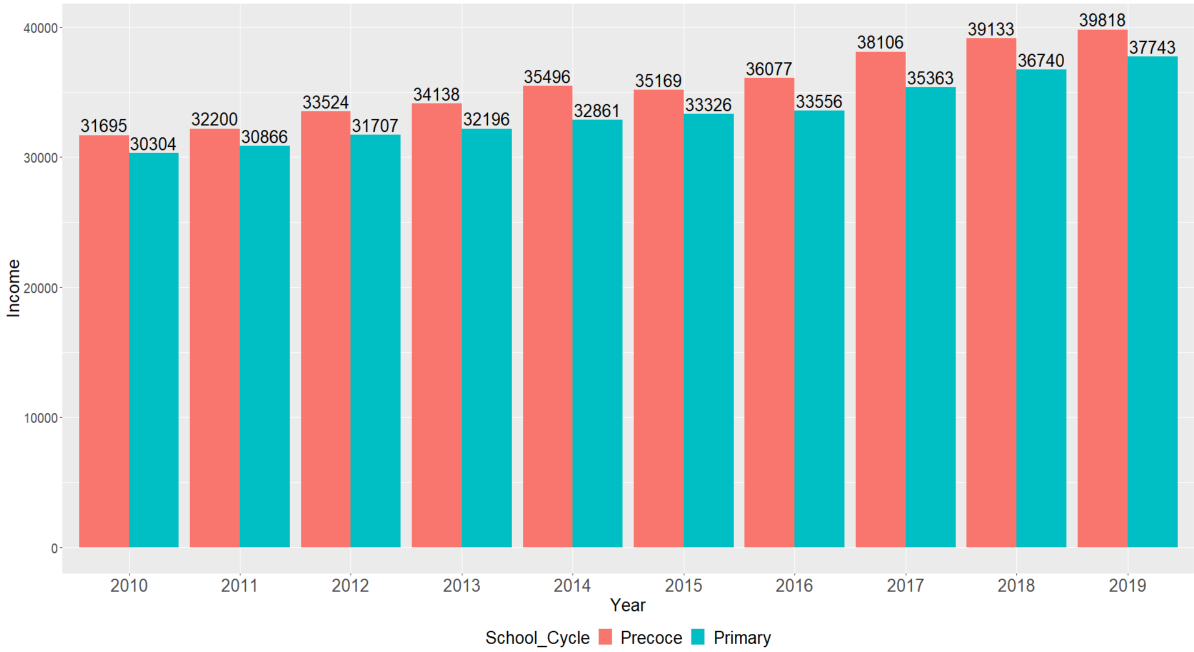
Luxembourg education system offers a possibility to families to enroll the 3-year-olds in one year of early education (éducation précoce) on a voluntary basis, as defined by article 2 of the amended law of 6 February 2009 on compulsory education.11 This year precedes the start of obligatory Cycle 1 attended by children between 4 and 5 years old. Although optional, précoce can help children start learning or improve the Luxembourgish language at an early stage, which can be particularly useful for

11 https://legilux.public.lu/eli/etat/leg/loi/2009/02/06/n2/jo

families who do not speak Luxembourgish as their first language. Hence it is crucial to investigate the characteristics of the student population that attend précoce in Luxembourg.

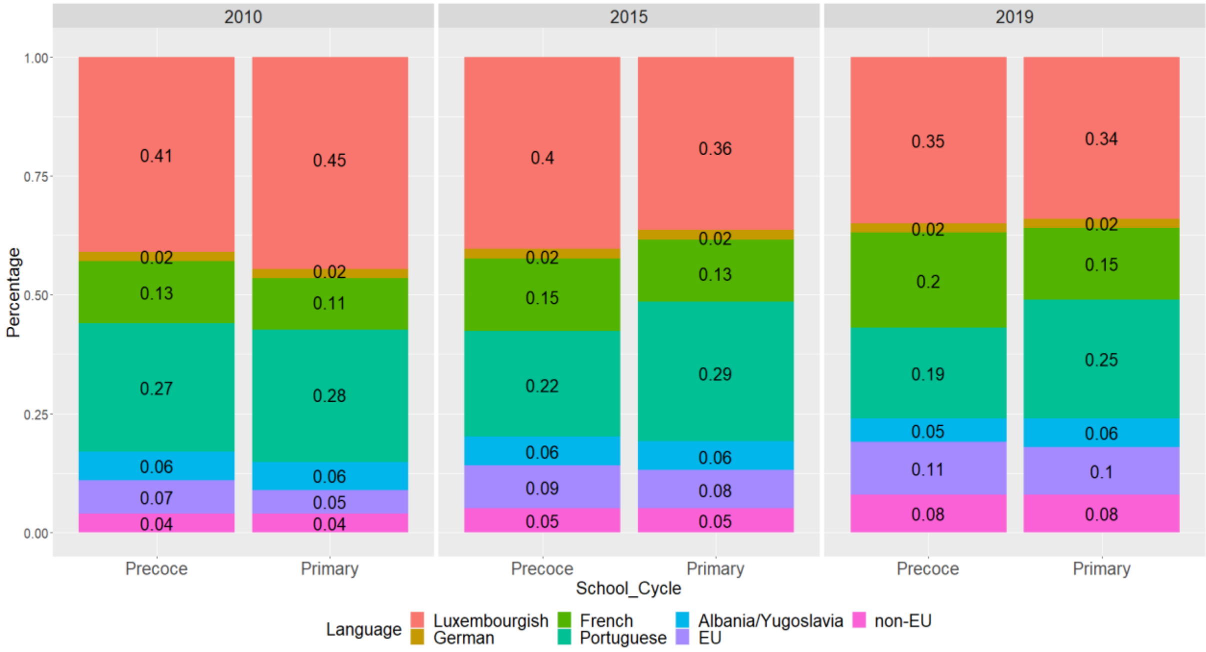
Figure 16 compares the average income of families of children enrolled in précoce with the rest of students in primary education and it shows that families with a higher income are predominant in précoce education, which means that there is a differentiation between the two populations. The trend is also stable across the years 2010 and 2019.

Figure 16. Average household income – précoce vs. primary school – 2010 – 2019 (in euros)



This difference can be seen even more clearly in Figure 17 where the two populations are grouped according to the first language spoken at home. Although the differences are minor, it is possibly to see that Luxembourgish (in 2015, 2019), French, and European languages have been overrepresented in précoce education and Portuguese is underrepresented, when compared to the overall population of primary education. However, it is important to keep in mind that 1) the comparison is made between the children attending non-compulsory précoce education (age 3) and those attending obligatory primary school that includes children from age 4 to12, which, as a result, is a much larger and more heterogeneous group. In other words, these are two different populations and their distributions that we are speaking about. In future analysis, a comparison between précoce and e.g. only the Cycle 1 students could make the comparison more precise. 2) results in the graph should be treated carefully when making conclusions about specific language groups. For instance, while Portuguese-speaking students might appear as less frequently enrolling in early education, Luxembourgish-speaking students had similar results in 2010. Better fitting data to analyze more accurately both the enrollment rate, as well as the reasons behind lower participation in early education should be made available before making the final conclusions.

Figure 17. Proportion of student's first language – Précoce vs. Primary – 2010, 2015 and 2019



3.3. Stratification within the communes

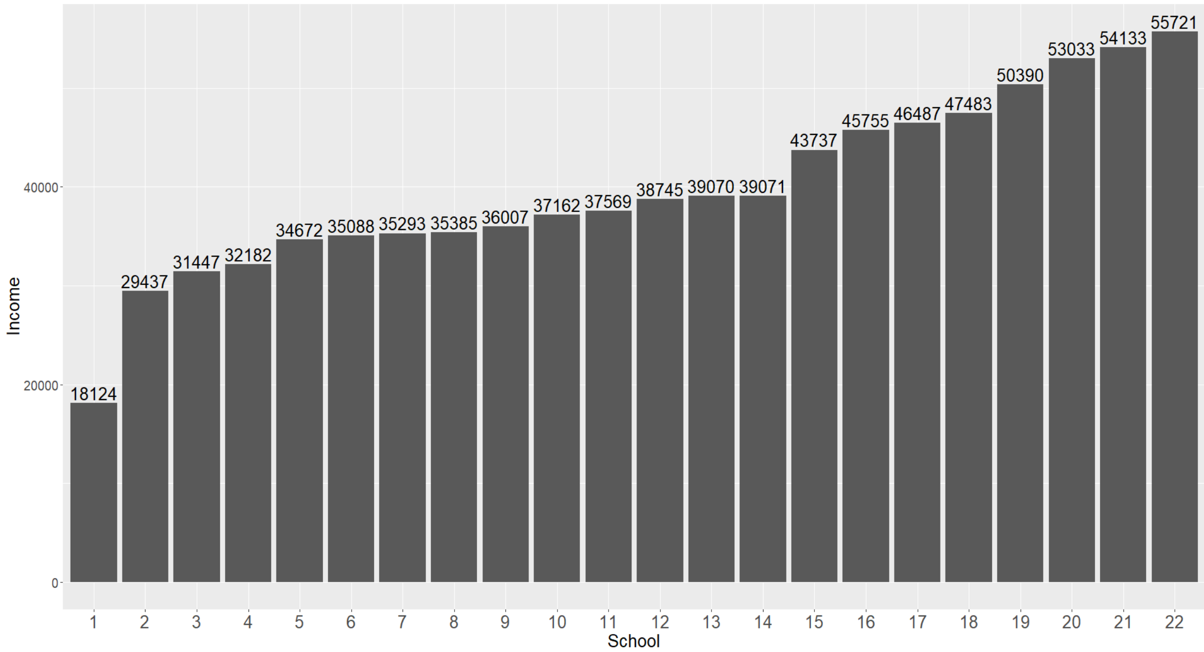
In 2019, 23 communes of 97 (24% of the total) had more than one public primary school within their administrative unit. These communes, with more than one school, account for 55% of all the public primary schools in Luxembourg. This is a crucial factor to take into consideration when comparing different communes, as one of the key dimensions of the social index, the household income is aggregated at the commune level based on the average of the student population within a given commune.

The level of aggregation can have a dramatic impact on the shape of the average income distribution when we contrast schools and municipalities. Figure 46 in the annex shows that the distribution of income per commune is positively skewed, meaning that few communes have a much higher average income than the other communes. However, if we aggregate household income at the school level, we can also see in Figure 46 a different shape of the distribution. In this case, the distribution is negatively skewed meaning that there is a concentration of much lower average income in some of the schools.

The significant difference in average income between schools can be better illustrated if we look closely inside the main cities, as Luxembourg Ville. In 2019, there were 22 public primary schools in the city of

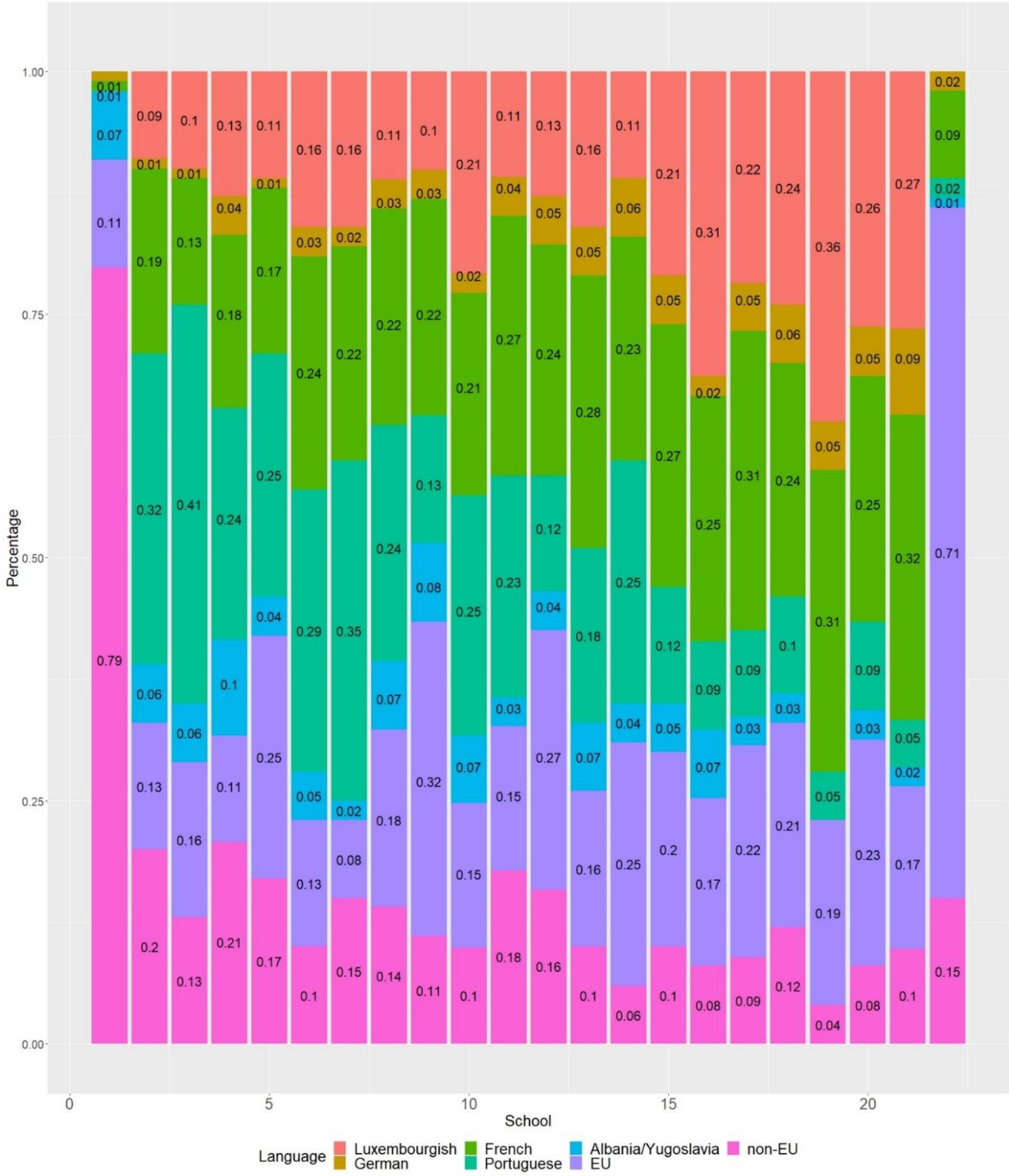
Luxembourg according to the data. The average equivalised household income of the primary school population for the city is 40,665 euros. However, a closer look at Figure 18 shows a considerable difference between the average income of the schools: one school has less than 50% of the mean income of the city and the richest one with a 30% higher average income than the mean of the city. Even if we are to ignore these extreme cases, there still remains a significant heterogeneity between the schools in the city of Luxembourg.

Figure 18. Average household income by school – Luxembourg City – 2019 (in euros)



This difference between schools within the city is evident not only with respect to the economic background of a school but also with respect to the linguistic background that changes remarkably from lowest income schools to the highest income schools (see Figure 19). It follows the tendency that we observed earlier, with schools with higher income counting more students speaking either Luxemburgish, German or French language, or lower proportions of students speaking Portuguese as their first language.

Figure 19. Proportion of student's first language by school – Luxembourg City – 2019



Luxembourg City is a compelling but not the only example: communes of Esch-sur-Alzette, Dudelange, Differdange and Sanem offer similar evidence (see Figure 47 to Figure 54 in Annex). All these communes have more than one school under their administration and show significant differences between the average income of their schools. The same heterogeneity is also true for the linguistic profile of the schools. No doubt, these findings highlight the importance to closely inspect the schools inside the communes so the inequalities can be analyzed at a finer level, in order to correctly assess the effect of transfers between communes.

PART 4 – Methodological review

In the past decades, the availability of educational data has increased exponentially, leading to the challenges of meaningful interpretation of large masses of information. The complexity of societal processes and phenomena require a multidimensional approach with the combination of different indicators into composite indicators (CI). The use of information in data-driven policies follows the principle of “what we measure affects what we do” (Stiglitz et al., 2009). CI is defined by OECD as ‘[…] formed when individual indicators are compiled into a single index on the basis of an underlying model’ (Nardo et al., 2008). There exist more than 400 CIs in different areas such as economy, development, and technology (Bandura, 2011) and they share the same goal of translating a complex evaluation of a system, composed of different indicators and dimensions, into a single number. This has the main advantage of the ease of interpretability by the general public (Saltelli, 2007).

At the national and supranational levels, these indicators have been gaining increased importance in helping to shape various policies. The OECD and the European Union even created a handbook and a tool to facilitate the creation and implementation of CIs for local level governments (Nardo et al., 2008). Some countries such as Belgium, France, Germany, Luxembourg and Switzerland developed CIs to monitor the socio-economic inequalities between schools based on different dimensions. In this case, CI helps optimizing its distribution policy while striving to minimize the socio-economic inequalities in the school population.

The construction of CIs, regardless of their objective, share some common steps. The choice made in each of these steps affects subsequent steps in its creation, as well as the final indicator and its quality (Mazziotta and Pareto, 2013). Based on the ‘Handbook on Constructing Composite Indicators’ (Nardo et al., 2008) we identified six main steps for the creation of a CI and discussed them throughout this part of the report:

- The first step in creating a CI is **defining what phenomenon will be measured**, and to stablish what is the goal of the CI. An indicator is only important if there is a clear view of its application (Dewey, 1938); otherwise, without a clear goal, the CI becomes an end for itself (Kuc-Czarnecka et al., 2020).
- The second step is the process of **data selection**, which relies on the preview step of defining the goal of the CI but also on the availability and quality of the relevant data. The quality of the chosen data will have a lasting impact through the entire process of constricting the CI.
- The third step is the **multivariate analysis** of the data, this stage is used to analyze the data structure and how the available variables are related to each other, and if the selected variables are useful for the CI. It plays a crucial role in the interpretation and the aggregation of dimensions and their weights (Nardo et al., 2008).

- The fourth step is the **normalization of the data**, at this stage the selected data is transformed, so different variables can have the same scale. As, it is the case, very often the data can have variables with different units of measurement, for instance income measured in the currency of a country while the unemployment rate in percentages, they need to have a common basis before any comparisons can be made.
- The fifth step is the **weighting** of the indicators and dimensions. This step is often subject to criticism given that sometimes developers of CIs opted for normative decisions regarding the implementation of weights. Without proper care at this stage, the final index can have an inaccurate measure of the target phenomenon.
- Finally, at the sixth step, **uncertainty and sensitivity analysis** is applied to the final model. It is an essential step, where potential problems can be corrected and the accuracy of the choices made at previous steps can be verified. This final analysis can help to ensure that a rank based on the CI is sufficiently robust to guide policy makers.

4.1. Theoretical framework

While defining the phenomenon that will be taken into account by the CI, it is necessary to consider the current knowledge on the subject, both theoretical and empirical. As Atkinson (2019) noted “the role of social science is to provide answers to questions posed by society and to convey the limits to the answers that can be given in the present state of knowledge”. With a clear theoretical background, it is easier to explain the choice of the different indicators in a CI.

There is a vast field of multidisciplinary research created for studying the impact of different characteristics of the students on their academic achievement. The main dimensions used to analyze pupils’ achievements at school are social, economic and cultural (Bourdieu, 1986), and pedagogical (Heck, 2009). However, within each of these dimensions, there are several variables that can impact the student performance at school, and for the construction of a CI for education it is necessary to combine some of these dimensions. The main examples of CIs for education rely mostly on the impact of socio-economic status (SES) of the students on their academic performance at school (BMBF, 2010; NCES, 2012; Rocher, 2016). The relationship between student’s SES and achievement has been studied extensively, both from the theoretical and empirical perspectives. However, the definition and measurement of SES are still subject to discussion (Bornstein and Bradley, 2003), and different countries apply different definitions. Even with some variability between countries regarding which variables to select for a CI aiming to capture the multidimensionality of the aspects that will impact students achievement, very often the background of the student’s family based on education, income, and occupation, are common indicators used to measure differences of outcome between students

(Brese and Mirazchiyski, 2013). Education funding and education outcomes have been an object of scrutiny by researchers for a long time (Jackson, 2020). A new wave of researchers from the beginning of the 21st century reported positive relationship between education spending and student achievement. This relationship not only showed that an increase of spending per student could increase his grades at school, but it also found that more resources could mitigate school dropout and increase the educational attainment of students (Hyman, 2017).

4.2. Data selection

Although several variables have been found to be relevant for the assessment of educational achievement, two persistent factors will impose constraints on the construction of a CI for schools.

Firstly, the **availability** of data for the construction of a CI, is in itself a potentially limiting factor on the amount of information that can be captured by the indicator. The data availability affects not only the current CI but is also relevant when considering about how often the CI needs to be updated with new data. For example, if the allocation of resources for schools is based on a CI, and if there are dynamic indicators that change frequently, the data needs to be available for constant updates of the CI.

Secondly, the **quality** of the data is an essential factor in the creation of CI. Some of the CIs rely on governmental data sources, while others use, additionally, the survey data (see overview in Part 2 as examples). Survey data can feed relevant for CI information, but there might be accompanying issues related to representativeness of data, or other issues, such as mismatch (e.g. survey data in the state of Hesse in Germany, where data collected from the parents of the students and the data supplied by the school's directors did not correspond each other (Makles and Weishaupt, 2010).

4.3. Multivariate analysis

There have been considerable advances in the field of multivariate analysis. Additionally, recent developments in the field of CIs have provided new perspectives on how to look at the non-linear relations between the indicators (Becker et al., 2017). Other new developments are the use of Machine Learning by creating a tree model to check whether the variables capture the phenomenon in focus and decompose the relative importance of selected variables (Oțoiu and Țițan, 2020). There are several ways of performing a multivariate analysis, we list some of the more common options with their advantages and disadvantages. We also included a more recently developed approach for feature selection and aggregation for CIs from the field of Machine Learning with supervised and unsupervised techniques.

Table 8. Summary of selected multivariate analysis methods

Method	Advantages	Disadvantages
Principal Components Analysis (PCA) – Tries to explain the variance of the data with less dimensionality.	● Easy to apply, relies upon few assumptions, can significantly reduce the number of variables when they are highly correlated.	● Data with nonlinear dependencies requires higher dimensionality. ● Can mask the real underline structure of the data.
Factor Analysis (FA) – Similar to PCA but can help to find underlying relations based on a latent variable.	● Factor loadings can help to distinguish between variables that have more or less impact between schools/cities, helping to explain variability on the unit of analysis. ● It can help to define the weight of the different variables in the final index.	● Requires some degree of intercorrelation to work. ● Highly pulverized opinion on how to apply 'stopping rules' when choosing factors to be retained.
Bayesian Factor Analysis – Similar to FA but takes advantage of Bayes' Theorem (Conti et al., 2014).	● Helps to avoid the arbitrariness of applying 'stopping rules' and the decision regarding the rotation method.	● As FA, it requires some degree of intercorrelation to work.
Multiple Correspondence Analysis – Similar to PCA and FA with the aim of reducing dimensionality but used on categorical data.	● This allows the correspondence of categorical data in multiple dimensions.	● Difficult to interpret, limited to categorical data.
Cronbach Coefficient Alpha – It measures the internal consistency by analyzing the indicators' proximity as a group.	● Useful to cluster similar variables.	● It can be hard to understand or misleading when dimensions are correlated. ● Only useful if CIs are computed as a scale.
Decisions trees – Aims to reduce dimensionality by selecting the most important variables that help in the process of the classification of the output (Oțoiu and Țițan, 2020).	● It works with nonlinear data; it is not sensitive to outliers or missing data.	● It is harder to explain how the algorithm works when compared with other methods. ● It relies on the existence of an output or proxies of the phenomenon that has been measuring.

4.4. Normalization of data

There are considerable implications at this stage for the remaining steps of the CI construction, not only the variance of the data will be affected by the normalization process, but also future decisions for the development of the CI will be limited by the choices made at this step.

Outliers and skewness of the data need to be studied at this stage, so the relationship between extreme values and their relative importance can be better understood (Decancq and Lugo, 2012). Qualitative decisions are also crucial at this point; for example, in the CI from France there was a concern not to have negative numbers as they could be misinterpreted by the general public (Rocher, 2016). We list some of the options for normalization and their advantages and disadvantages.

Table 9. Summary of selected methods for normalization of data

Method	Advantages	Disadvantages
Standardization (or z-scores) – Returns a normalized value based on the mean and the standard deviation.	• It is the most common procedure used for the normalization of data.	• Z Scores assumes a normal distribution. Extreme values have a greater impact on the indicator.
Min-Max – Transform the indicators to be in an identical range between 0 and 1.	• It produces a more homogenous scale when compared with Z scores.	• It does not work well with outliers.
Distance to a reference – Based on the position of an indicator for a given unit compared to the position to another unit or any artificially selected poinst of reference such as the average unit or the lower or higher unit in a specific point in time.	• Easy implementation.	• Loss of the disaggregated level of the data.
Categorical scales – Gives categorical values to the indicators.	• Small changes in the measures of indicators do not affect the variable.	• Hides the variance between units.
Indicators above or below the mean – Creates a classification based on a threshold relative to the mean.	• Good to show the evolution of an indicator in a time series. It is not affected by outliers.	• The definition of thresholds is arbitrary. Loss of information on the absolute level.

4.5. Weighting and Aggregation

Weighting is the most discussed part of the creation of a CI and has amassed a significant amount of literature on the topic. Not only several technical methods can be applied to define the weights of the indicators and dimensions, but there are also distinct *philosophical* views on the subject. Based on Decancq and Lugo (2013), we can classify the different methods into three major categories: data-driven, normative, and hybrid. However, there is no clear-cut answer on what should be considered best practice when building a new CI.

This stage's complexity means that even opting for what it seems simplistic methodological choices such as not applying any weight to the CI, in reality can have consequences on the final model, as this decision would mean to give equal weights to different indicators or different dimensions. There is no single solution when choosing a weighting method. Instead, the theoretical framework, the available data, and the phenomenon been measured should guide the decision over the choice of the method in this stage. More recently, developments on the creation of CIs have been focusing on the improvement of the data-driven approach (Becker et al., 2017) thus diminishing, when possible, normative decisions for weighting and aggregation. In Table 3, we present some of the options for weighting and their advantages and disadvantages, including the most recent literature on the construction of CIs that addresses the problem of nonlinear effects for adjusting the weight of indicators (Becker et al., 2017).

The decision made regarding the approach for weighting the indicators and the dimensions will affect the indicators' aggregation. The quality of a CI is based on how well it can measure a phenomenon but it is also based on the principle of the minimum 'information loss' (Zhou et al., 2010). Conceptually, the aggregation step deals with some fundamental questions on a creation of a CI. Firstly, it imposes a tradeoff between disaggregated information, where indicators could be presented individually, or a single number combing all the indicators, which will lead to the final CI with a loss in the level of information. Secondly, based on the weighting scheme applied to the indicators it also leads to a question of substitution between indicators or dimensions, where one unit of the indicator 'a' could compensate for one unit of the indicator 'b'. Without the correct application of weights, the marginal contribution of each indicator to the final index can derive from incorrect assumptions about the implicit marginal rate of substitution between indicators or dimensions (Greco et al., 2019). Finally, the different aggregation methods are based on the assumption of underlying independency of indicators, which in practice is rarely the case, much less when dealing with education where there is a degree of collinearity between common social-economic determinants.

Table 10. Summary of selected methods for weighting

Method	Advantages	Disadvantages
Data-driven – Based on different models to extract the weight of the indicators from the data.		
PCA/Factor analysis – PCA and FA can be used to compute the overlap variance between the correlated indicators.	● It is straight forward, and the data speaks for itself.	●The data need to be correlated to work. ● Dimensions with lower correlation will receive lower weights, which is not necessarily a true reflection of the phenomenon.
Benefit of the doubt (BOD) – Introduces the idea that specific indicators in which the unity of analysis performs well are more important for this unit and thus should be weighted accordingly with this importance. It uses each indicator of each unit to benchmark against indicators of other units.	● It gives transparency to the weighting process. ● It does not need a theoretical background for implementation.	● Harder to find a unique solution for ranking, as many units can be on the top. ● The index can reward the status quo.
Unobserved Components Models (UCM) – Similar to a multiple regression model, but the dependent variable is unknown. It considers that the measure of a given indicator is imperfect and thus includes an error term in the final value of the weight.	● The estimation of the weight does not depend on any random cut.	● Can be affected by outliers. ● High correlation between indicators can impose difficulties in extracting the correct weights.
SHapley Additive exPlanations (SHAP) – It is a unified framework to explain predictions of a model assigning each feature an importance value (Lundberg and Lee, 2017).	● Highly flexible in applying on different complex models. ● It extracts only what was strictly used to make the prediction.	● Interpretability has a high level of complexity when compared to other traditional ways of extracting weights of data.
Hybrid – Combines a data-driven approach with a normative approach, which means that not only expert opinions are factored for the extraction of the weigh, but it is also used statistical methods to validate these opinions.		
Analytic hierarchy process (AHP) – Uses expert opinions to compare a given pair of indicators where the expert decides which indicator is more important and by which factor it is more important. Then a pairwise matrix is assembled, and the relative weights of the indicator are calculated using an eigenvector.	● It benefits from having an expert opinion supporting the choices of the weights.	● Results depend on the set of evaluators chosen and the setting of the experiment. ●The magnitude of importance of an indicator in relation to another can vary significantly depending on a person.
Conjoint analysis (CA) – It uses a survey to get information from a sample of people on their opinion of how they would rate different scenarios of weights distribution. After the survey is applied, the collected data is used to desegregate the information at the indicator level, and a function of	● It values different opinions thus is more democratic.	● Weights can vary accordingly with the sample of respondents.

preference is applied to define the weights.		
Normative – It relies specifically on an opinion.		
Arbitrary – Weights are applied based on researchers, experts, or policymakers' opinions or some combination of different stake holders.	● Transparent and straight forward.	● Can end up reflecting what should be instead of what it is.
Budget allocation process (BAP) – Experts are asked to distribute an n number of points between all indicators. Afterward, the weights are calculated based on the average number of points of each indicator.	● Specialist of a specific area selects the weights.	● Instead of the weight of individual indicators in the allocation of the points, the experts can express what they believe to be the priority of what needs to be changed.

4.6. Uncertainty and sensitivity analysis

The creation of a CI is just like a creation of an empirical model, and thus it requires a final step to carry out the stability of the model. This step is important, for it brings transparency to the model: as a series of decisions were made in the previous steps, it is important to demonstrate the impact of these decisions in the final model (Saisana et al., 2005). CIs are commonly used for the construction of ranks, the robustness of these ranks based on the CI needs to be tested in order to convey how precise it can guide policy makers. Uncertainty analysis examines the behaviour of the variables used in the model, and how this uncertainty propagates throughout the model.

Sensitivity analysis is used to understand how the different indicators affect the overall variance of the final score of the index. Although, often both types of analysis had been used separately from each other, there is a clear gain in using them together (Saisana, et al., 2005b)

Common questions that are addressed at this point based on (Nardo et al., 2008) are:

- Which are the most volatile cities/schools and why?
- What if a measurement error is incorporated?
- What if skewed distributions are not treated?
- What if we change the method of normalization of the data?
- What is the impact of alternative weighting schemes?
- What if the aggregation function is geometric instead of arithmetic?

The current literature provides several tools to help answer these questions, depending on the complexity of the aggregation of indicators/dimensions and the number of layers in the final model different techniques can be explored for uncertainty and sensitivity analysis. Due to the changes made into the data during the different stages in a creation of a CI, the data itself can have its proprieties changed, which can lead to loss of linearity and possibly becoming a non-additive model (Saltelli et al., 2008). A common approach is to use variance-based techniques to explore different properties of the model. Nevertheless, the goal of this step is to move backward in the creation of the CI and verify if the more appropriated choices were made, and if necessary to correct the final model.

5. Concluding remarks

The aim of this technical report was to review the results and methodology behind the social index used to allocate part of the 'contingent scolaire' across municipalities, an essential idea of the policy for the distribution of resources among the schools across Luxembourg.

The social index was first calculated in 2010 and was later calculated in 2012, 2015, and 2019. Part 1 of the report provided details on the evolution of the index across municipalities, examining the (in-)stability of the indices, the underlying forces for changes therefore and the sensitivity of results to a range of factors. The main observations pointed out in Part 1 are the following:

- There is stability in the social index across years for the communes at the bottom of the ranking (communes with unfavorable socio-economic school population composition), and -to some extent- also for those at the top of ranking (with a few exceptions). The more sizeable fluctuations occur among the communes in the middle of the distribution.
- Both rank order and social index scores were generally similar to each other in 2010, 2012, and 2015, but the index varied more in the 2019 release. Variations in 2019 can be linked to the change in the measurement of the underlying dimensions, following a drastic decrease in available data, in part due to GDPR's data minimization principles. While the income component remained stable, the family composition dimension was most affected by the change in underlying data.

Examination of the details of the construction of the index reveals that part of the instability of the social index (notably for communes in the middle of the distribution) is inherent to the methodology applied. One key element is that the four independent constituent dimensions are scaled to have common mean and variance across municipalities before being aggregated into a composite index. This mechanically introduces some risk of instability for dimensions that do not vary much across municipalities, such as employment and family composition dimensions. For such variables, small changes in the raw data can have relatively large effects on the aggregate index. More substantively, the scaling of each dimension to a common mean and variance gives equal "importance" to each dimension, although one may argue that dimensions which do not vary much across municipalities (and therefore do not discriminate clearly 'advantaged' from disadvantaged' areas) should not be given the same importance as variables that discriminate communes more (such as income).

Comparison of the Luxembourg CI with four other CIs used in other countries revealed in Part 2 differences in the choices made with respect to principal dimensions, data sources, level of aggregation, and weighting strategies. To sum up, the Hamburg and France CI appear as more detailed, covering four to five dimensions with an extensive list of variables. Interestingly, France and Wallonia-Brussels CIs do not include the migration background dimension. Zurich CI consists of only two dimensions and is based on 3 variables. Hamburg, France and Wallonia-Brussels CIs aggregate data at the school-level, Zurich at the district level.

Additional analysis into questions that are pertinent to the goal of the effectiveness of the distribution policy in Luxembourg was performed in in Part 3:

- The standard public schools and the international public schools are different in their socio-economic and linguistic composition. International schools are attended by children from higher income families and with a higher proportion of French and other EU-language speaking students. Standard schools are more frequently attended by students with partially lower income and a higher proportion of Luxembourgish and Portuguese speaking students. Given their important differences both in funding policy (international public schools are funded directly by the Ministry of education, and do not receive additional subsidies from the municipalities) and in admission policy (residence based in standard schools vs. open registration with a number of other criteria in international schools), a future discussion on whether a municipality-level social index is adjusted to the growing diversity of school population in Luxembourg and whether the new schools should also be considered would need to take place.
- Our attempt to better understand which population groups make more extensive use of early education (education précoce) led to supplementary analysis. Our preliminary results suggest that higher income families are more likely to enroll their children in early education at age 3. While data suggest potentially different behaviour by Portuguese-speaking families, we recommend to cross-validate these findings with other data sources that are better suited for this purpose.
- Other pertinent findings that could help to calibrate the social index are the heterogeneity within communes with multiple schools. Review of schools in Luxembourg city, Esch-sur-Alzette, Dudelange, Differdange, Sanem point to a diversity of schools both with respect to the income level and language composition. These differences within communes were not taken into account in previous versions of the social index.

Overall, results presented in the current report will serve as a ground for the following part of the project, where suggestions for potential improvement of the social index and its methodology will be elaborated.

Several directions of potential development seem relevant at this stage, including ways to address the “common variance” issue (and the instability it creates), to introduce a multi-level perspective allowing for school heterogeneity within municipalities, and to handle the concentration of cumulative disadvantage at the household-level and inequality thereof within municipalities.

Bibliography

Allmendinger, J., 1999. 'Bildungsarmut. Zur Verschränkung von Bildungs-und Sozialpolitik' [Educational poverty. On the Interconnection between Education and Social Policy], Soziale Welt, 50(1), 35-50.

Allmendinger, J.; Leibfried, S., 2003. 'Education and the welfare state: the four worlds of competence production', Journal of European Social Policy, 13(1), 63-81.

Atkinson, A.B., 2019. Measuring Poverty around the World, Measuring Poverty around the World. Princeton University Press. https://doi.org/10.1515/9780691191898

Bandura, R., 2011. Composite indicators and rankings: Inventory 2011.

Becker, R.A., Denby, L., Mcgill, R., Wilks, A.R., 1987. Analysis of Data from the Places Rated Almanac.

Becker, W., Saisana, M., Paruolo, P., Vandecasteele, I., 2017. Weights and importance in composite indicators: Closing the gap. Ecol. Indic. 80, 12–22. https://doi.org/10.1016/j.ecolind.2017.03.056

Bornstein, M.H., Bradley, R.H., 2003. Socioeconomic status, parenting, and child development. Lawrence Erlbaum Associates.

Bourdieu, P., 1986. Pierre Bourdieu 1986 - The forms of capital. Handb. Theory Res. Sociol. Educ. 241–258.

Bousselin, A. 2019. Expanding access to universal childcare: Effects on childcare arrangements and maternal employment. LISER Working Papers N 2019-11.

Brese, F., Mirazchiyski, P., 2013. Measuring students' family background in large-scale international education studies. Issues Methodol. large-scale assessments.

Bundesministerium für Bildung und Forschung (BMBF), 2010. Zur Konstruktion von Sozialindizes 31.

Checchi, D., 1998. 'Povertà ed istruzione: alcune riflessioni ed una proposta di indicatori' [Poverty and Education: Some Reflections and a Proposal of Indicators], Politica economica, 14(2), 245-282.

Conti, G., Frühwirth-Schnatter, S., Heckman, J.J., Piatek, R., 2014. Bayesian exploratory factor analysis. J. Econom. 183, 31–57. https://doi.org/10.1016/j.jeconom.2014.06.008

Decancq, K., Lugo, M.A., 2012. Weights in Multidimensional Indices of Wellbeing: An Overview. Econom. Rev. 32, 7–34. https://doi.org/10.1080/07474938.2012.690641

Dewey, J., 1938. Experience and education. Macmillan, New York.

DiPrete, T.A. & Eirich, G. M., 2006. Cumulative advantage as a mechanism for inequality. Annual Review of Sociology 32, 271--297.

Education and Training Monitor, 2019. https://ec.europa.eu/education/policy/strategic-framework/et-monitor_en. Brussels: European Commission.

Greco, S., Ishizaka, A., Tasiou, M., Torrisi, G., 2019. On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. Soc. Indic. Res. https://doi.org/10.1007/s11205-017-1832-9

Heck, R.H., 2009. Teacher effectiveness and student achievement: Investigating a multilevel cross-classified model. J. Educ. Adm. 47, 227–249. https://doi.org/10.1108/09578230910941066

Heckman, J.J., 2006. Skill formation and the economics of investing in disadvantaged children. Science 312 (5782): 1900–1902.

Hyman, J., 2017. Does money matter in the long run? Effects of school spending on educational attainment. Am. Econ. J. Econ. Policy 9, 256–280. https://doi.org/10.1257/pol.20150249

Jackson, C.K., 2020. Does school spending matter? The new literature on an old question. Confronting Inequal. How policies Pract. shape Child. Oppor. 165–186. https://doi.org/10.1037/0000187-008

Kuc-Czarnecka, M., Lo Piano, S., Saltelli, A., 2020. Quantitative Storytelling in the Making of a Composite Indicator. Soc. Indic. Res. 149, 775–802. https://doi.org/10.1007/s11205-020-02276-0

LISER, 2019. Etablissement d'un indice socioéconomico-culturel communal sur la base des élèves fréquentant l'enseignement fondamental en 2018-2019.

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. Adv. Neural Inf. Process. Syst. 2017-December, 4766–4775.

Makles, A., Weishaupt, H., 2010. Sozialindex für Schulen – Möglichkeiten und Probleme der Konstruktion am Beispiel einer Untersuchung in Nordrhein-Westfalen. R. der Jugend und des Bild. 58, 196–211. https://doi.org/10.5771/0034-1312-2010-2-196

Mazziotta, M., Pareto, A., 2013. Methods for Constructing Composite Indices: One for All or All for One? Riv. Ital. di Econ. Demogr. e Stat. 67, 67–80.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E., 2008. Handbook on constructing composite indicators, OECD Statistics Working Papers. https://doi.org/10.1787/533411815016

NCES, 2012. Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation. Recommendations to the National Center for Education Statistics. Washington, DC.

OECD, 2019. PISA 2018 Results (Volume II), PISA. OECD. https://doi.org/10.1787/b5fd1b8f-en

OECD, 2012. Equity and quality in education: Supporting disadvantaged students and schools, Equity and quality in education: Supporting disadvantaged students and schools. Organisation for Economic Cooperation and Development (OECD). https://doi.org/10.1787/9789264130852-en

OECD, 2010. Overcoming School Failure: Policies That Work OECD Project Description.

Oțoiu, A., Țițan, E., 2020. Using Decision Trees to Improve Variable Selection for Building Composite Indicators. Stat. Stat. Econ. J. 100, 296–208.

Rocher, T., 2016. Construction d'un indice de position sociale des élèves. Educ. Form. Ministère l'éducation Natl. l'enseignement supérieur la Rech. 5–27.

Ruiz-Valenzuela, J., 2020. Job loss at home: children's school performance during the Great Recession. SERIEs 11, 243–286. https://doi.org/10.1007/s13209-020-00217-1

Saisana, M., Saltelli, A., Tarantola, S., 2005. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. J. R. Stat. Soc. A 168, 307–323.

Saltelli, A.., Ratto, M.., Andres, T.., Campolongo, F.., Cariboni, J.., Gatelli, D., Saisana, M.., Tarantola., S., 2008. Global Sensitivity Analysis: The Primer. John Wiley & Sons.

Saltelli, A., 2007. Composite Indicators Between Analysis and Advocacy. Soc. Indic. Res. 81, 65–77. https://doi.org/10.1007/s11205-006-0024-9

Schulte, K., Hartig, J., Pietsch, M., 2014. Der Sozialindex für Hamburger Schulen.

Stiglitz, J.E., Sen, A.K., Fitoussi, J.-P., 2009. Report by the Commission on the Measurement of Economic Performance and Social Progress. Paris.

Taussig, F.W., 1920. Principles of Economics. Cambridge Scholars, Newcastle.

Zhou, P., Fan, L.W., Zhou, D.Q., 2010. Data aggregation in constructing composite indicators: A perspective of information loss. Expert Syst. Appl. 37, 360–365. https://doi.org/10.1016/j.eswa.2009.05.039

Annex

Figure 20. Comparison between years of income variation (2010-2019)

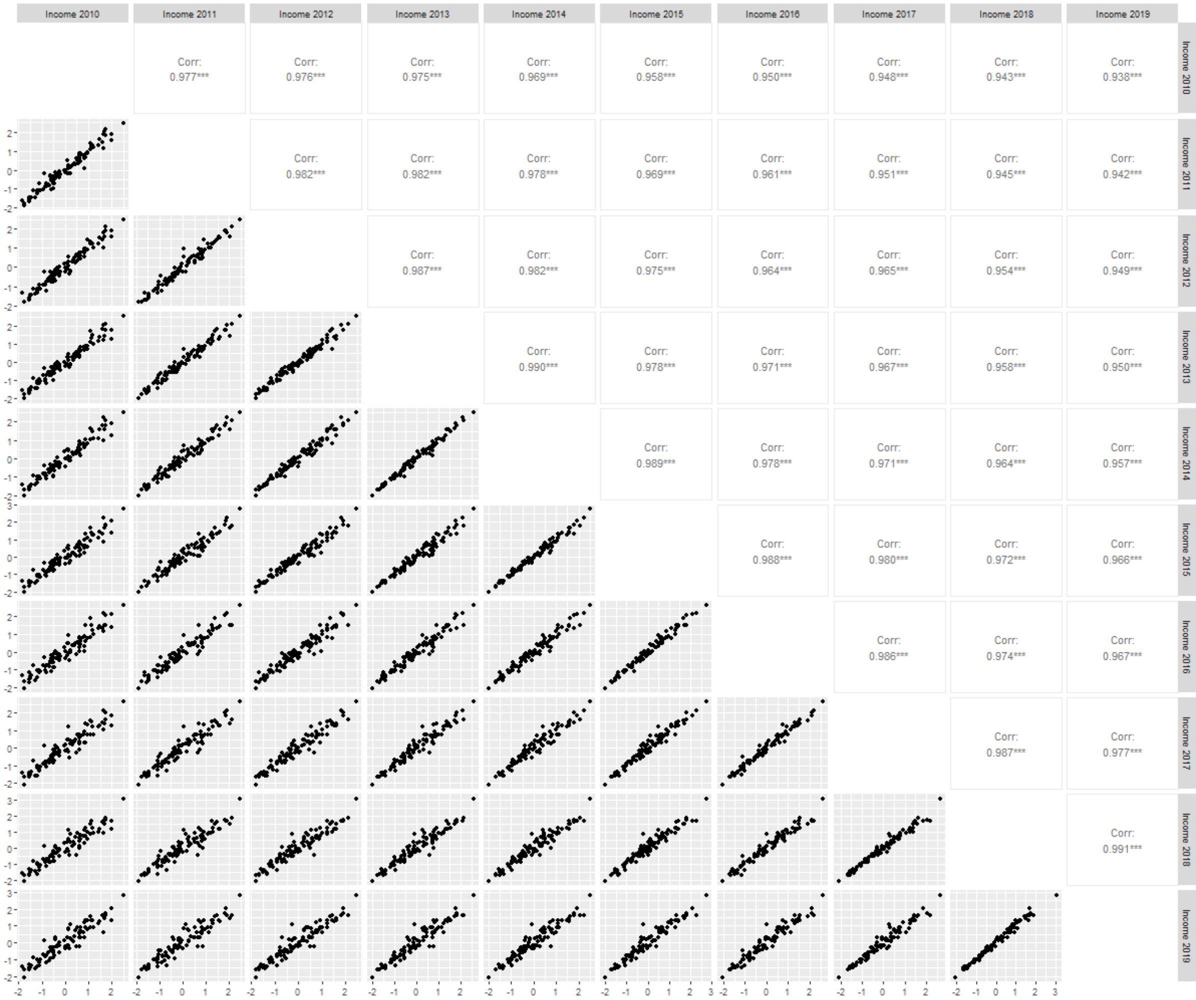


Figure 21. Comparison between years of Blue Collar variation (2010-2019)

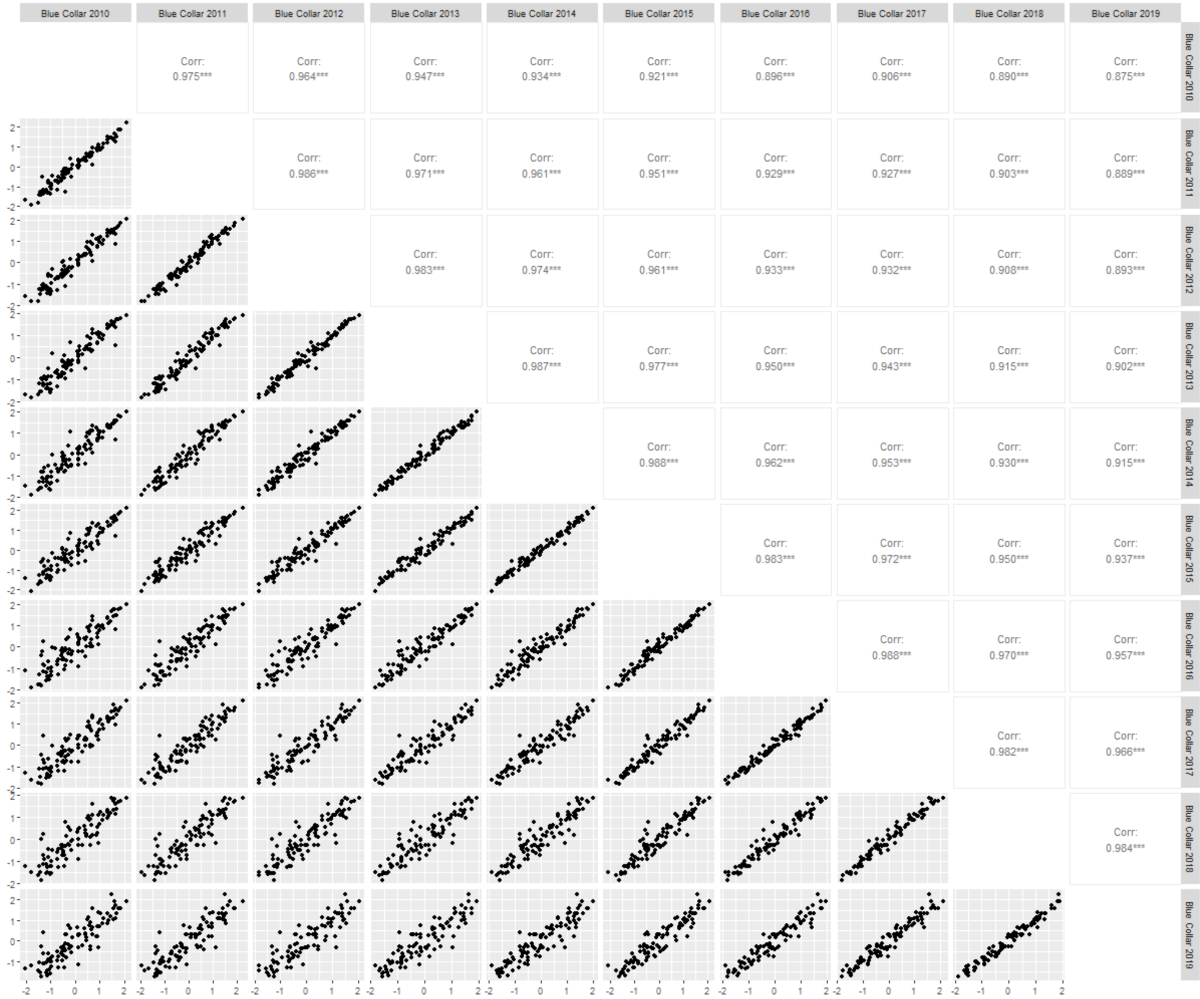


Figure 22. Comparison between years of Civil Servant variation (2010-2019)

Figure 23. Comparison between years of Unemployment variation (2010-2019)

Figure 24. Comparison between years of RMG variation (2010-2019)

Figure 25. Comparison between years of Monoactive variation (2010-2019)

Figure 26. Comparison between years of Biparental younger variation (2010-2019)

Figure 27. Comparison between years of Biparental Older variation (2010-2019)

Figure 28. Comparison between years of First Language Condition variation (2010-2019)

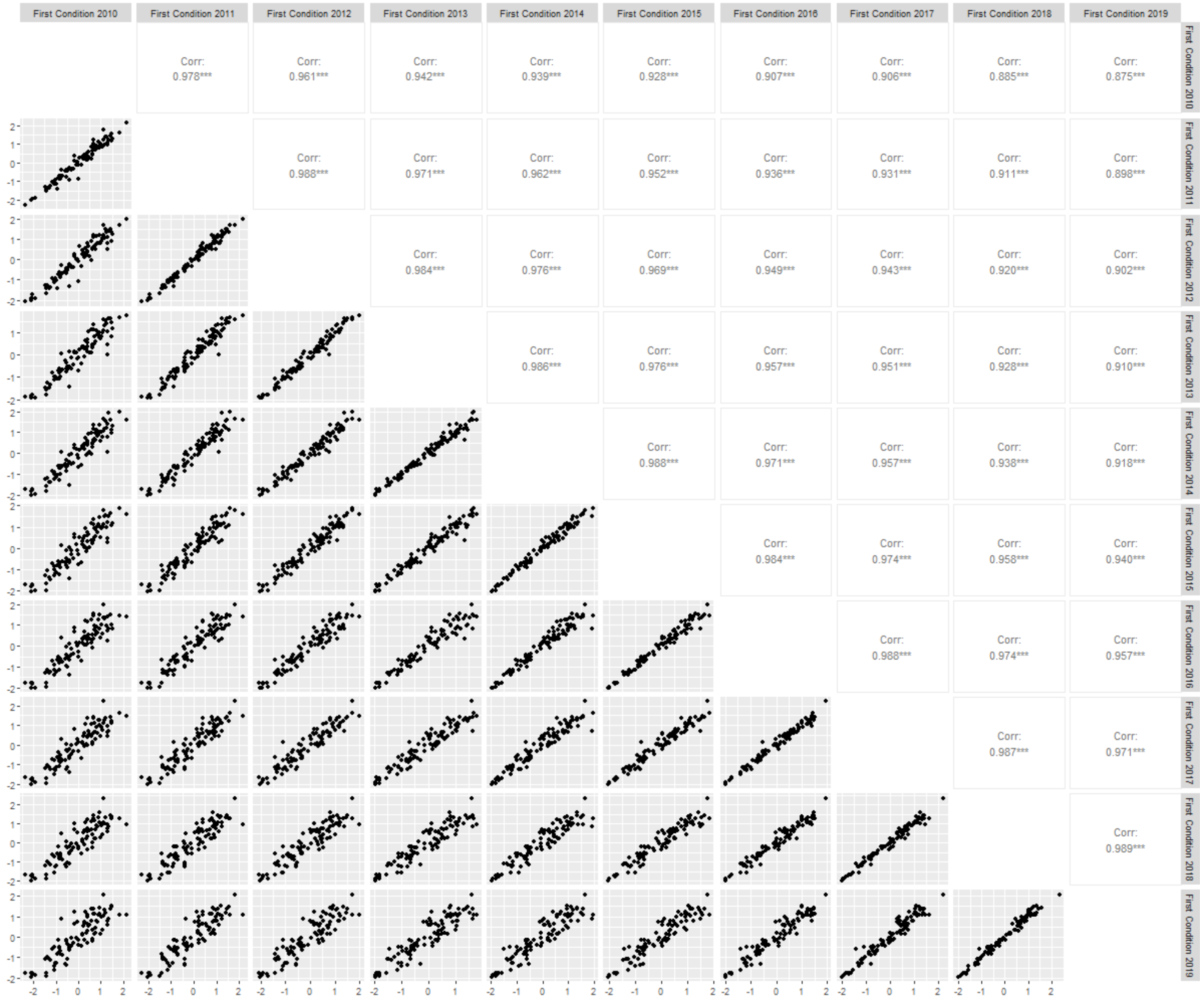


Figure 29. Comparison between years of Second Language Condition variation (2010-2019)

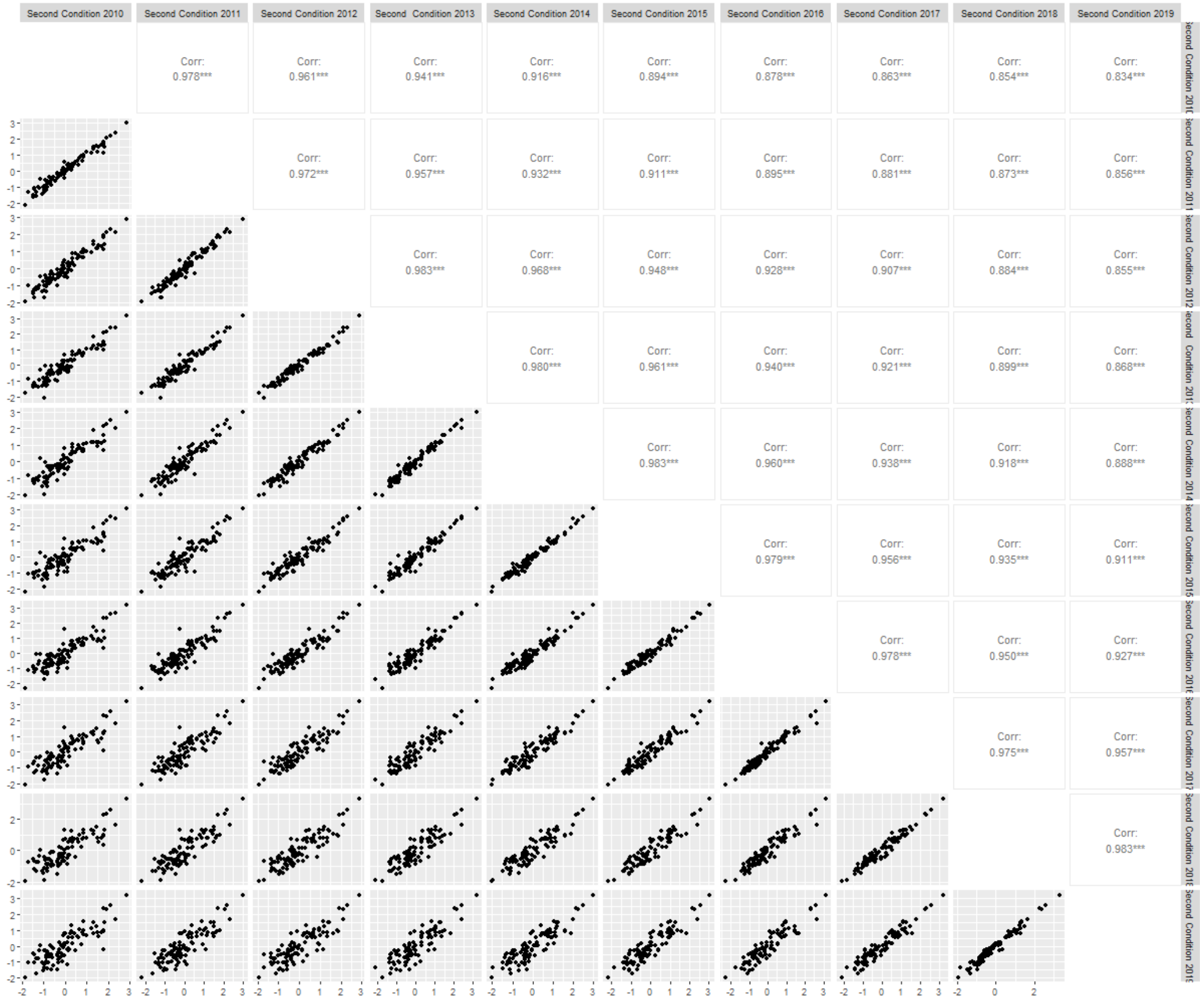


Figure 30. Comparison between years of Third Language Condition variation (2010-2019)

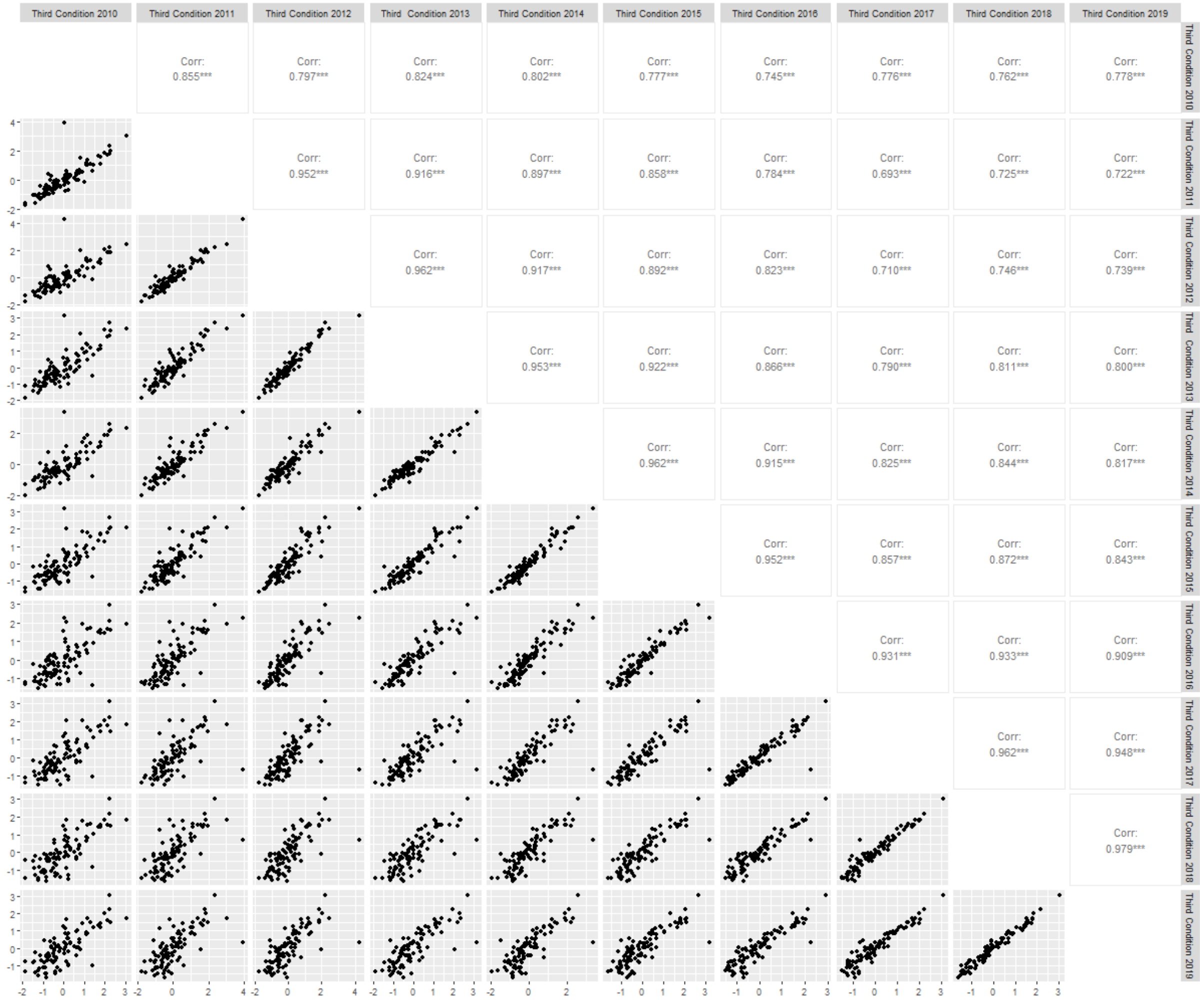


Figure 31. Comparison between years of Fourth Language Condition variation (2010-2019)

Figure 32. Comparison between years of Fifth Language Condition variation (2010-2019)

Figure 33. Comparison between years of Language Condition variation (2010-2019) (methodology 2019)

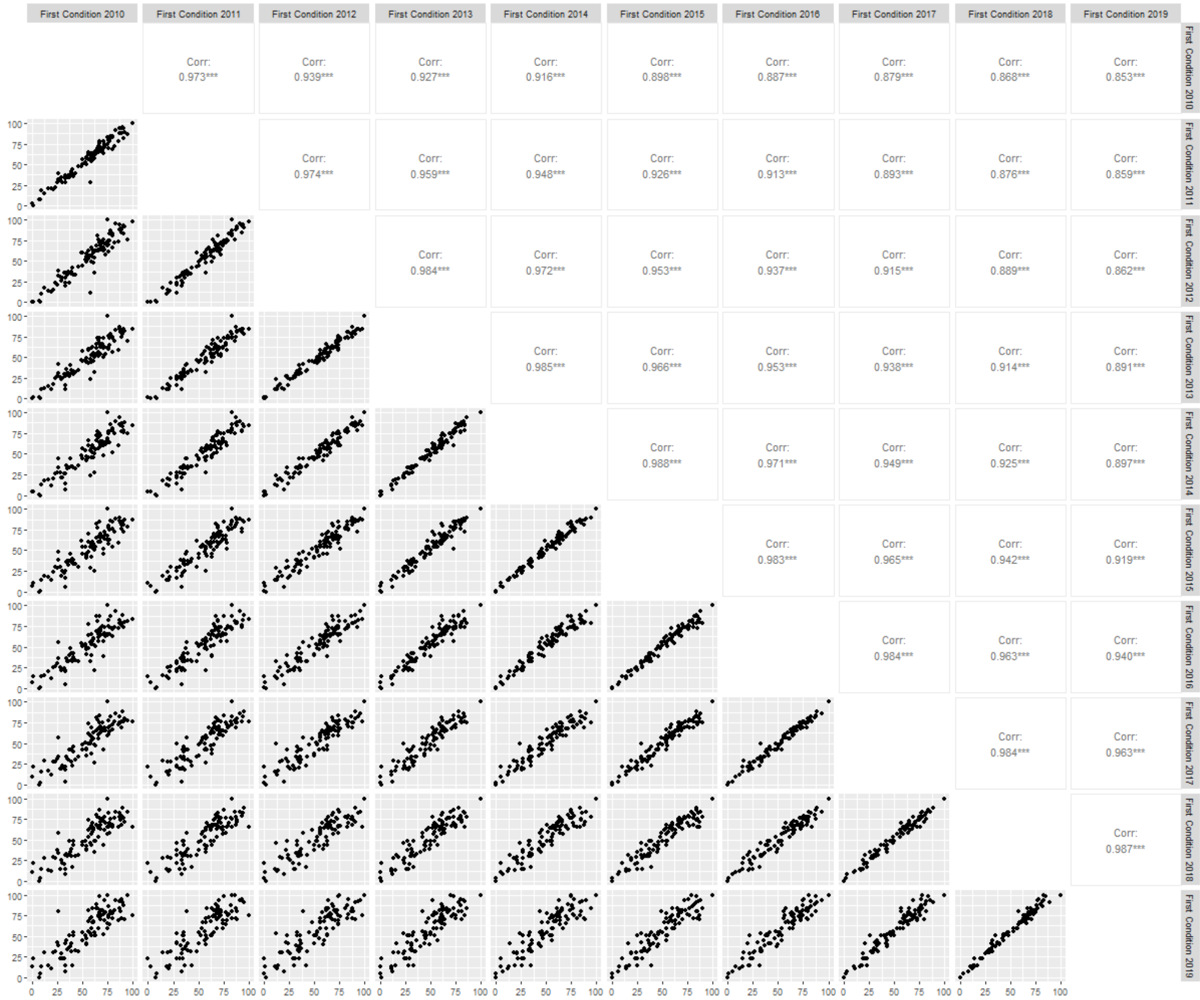


Figure 34. Correlation between rank variation across years 2010, 2012, 2015 and the number of students per commune (scaled)

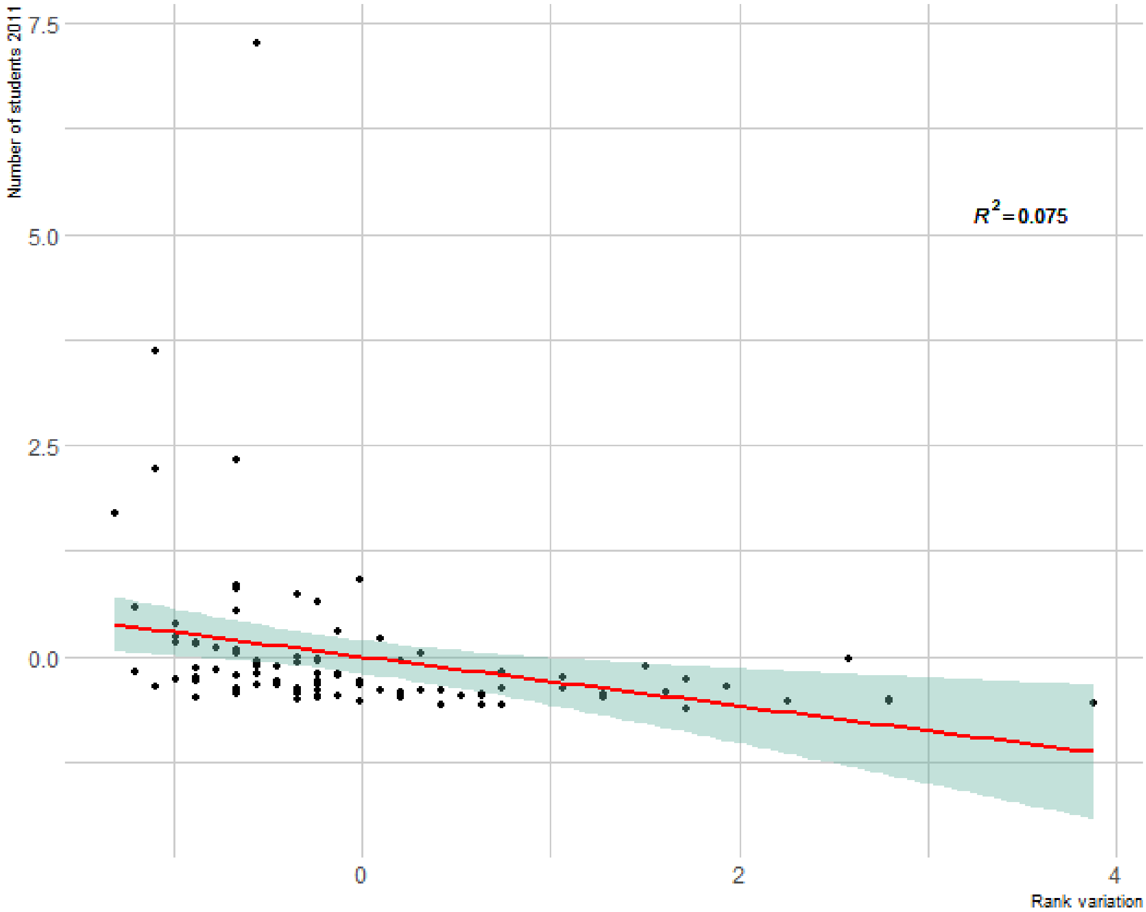


Figure 35. Correlation between rank variation between 2010 and 2012 and the number of students per commune (scaled)

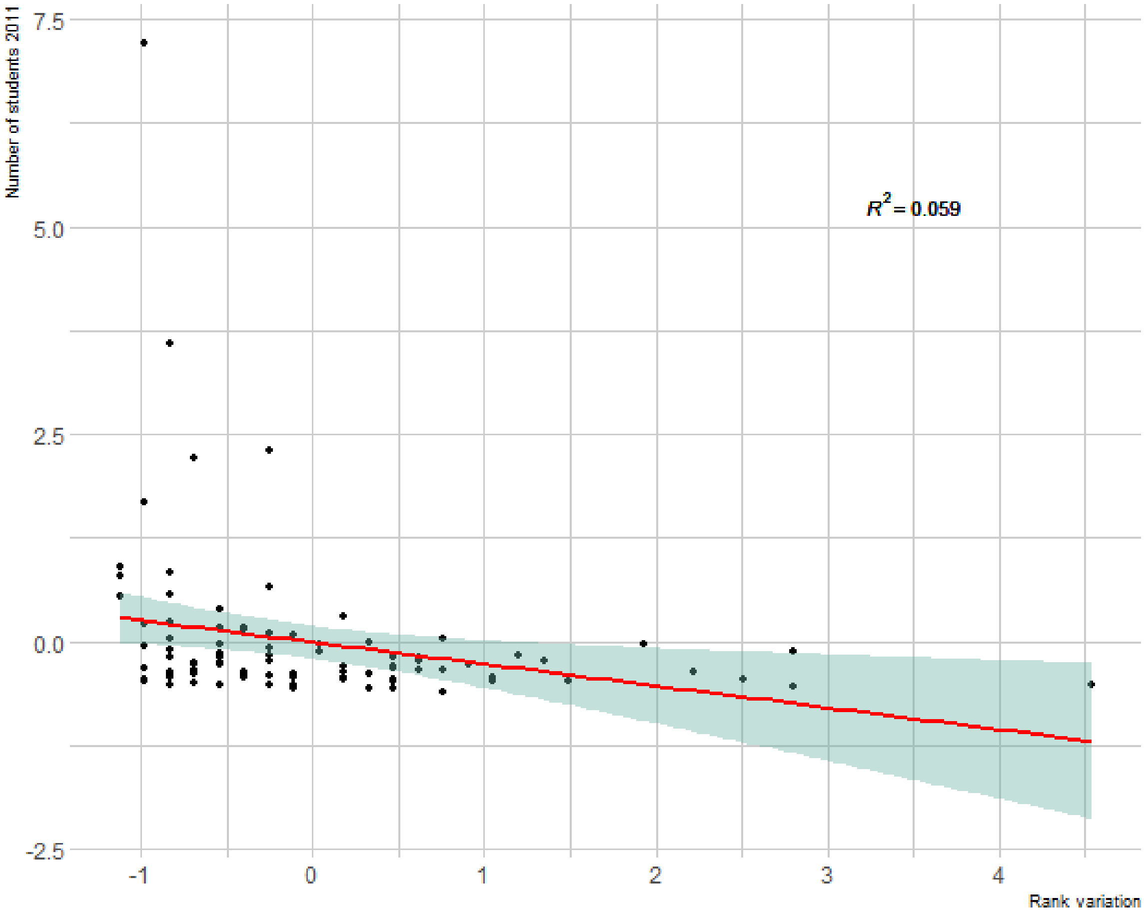


Figure 36. Correlation between rank variation between 2012 and 2015 and the number of students per commune (scaled)

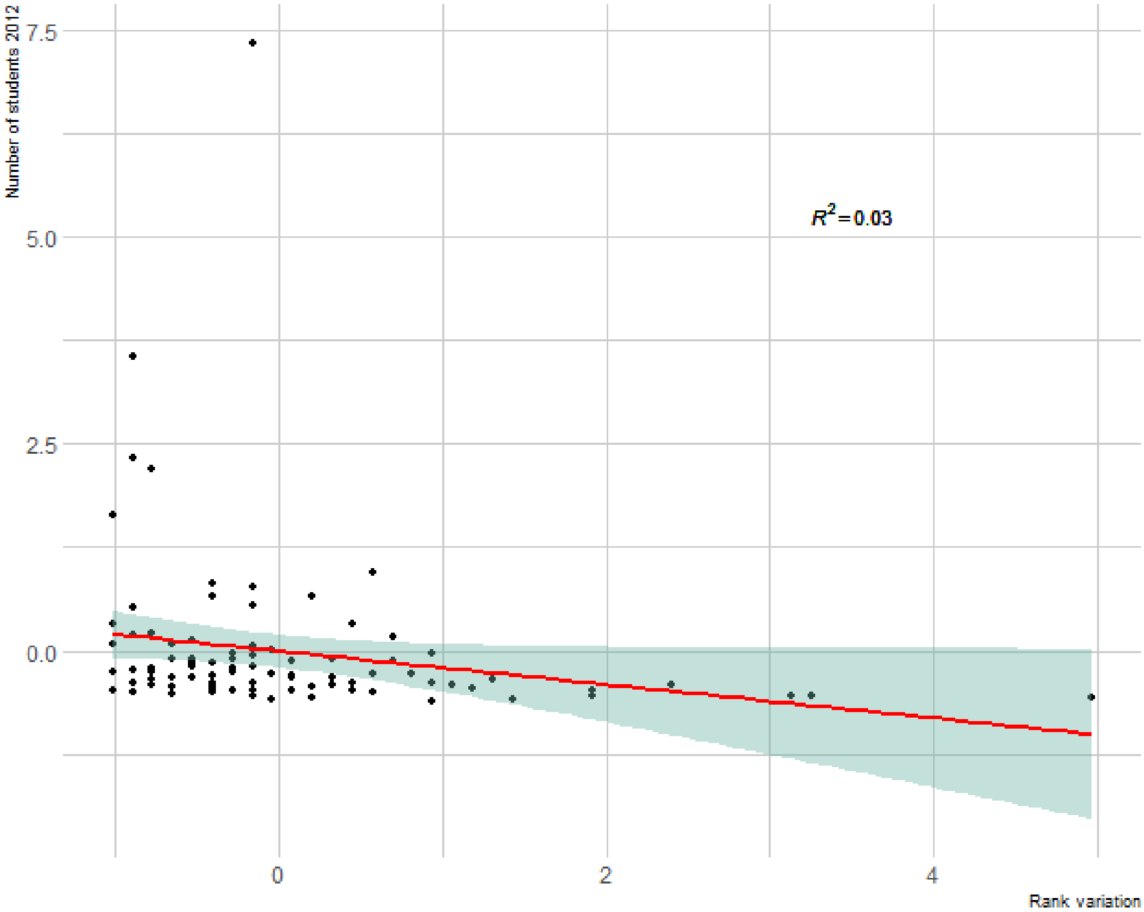

Number of students 2012
7.5
5.0
2.5
0.0
$R^2 = 0.03$
0
2
4
Rank variation

Figure 37. Correlation between rank variation between 2015 and 2019 and the number of students per commune (scaled)

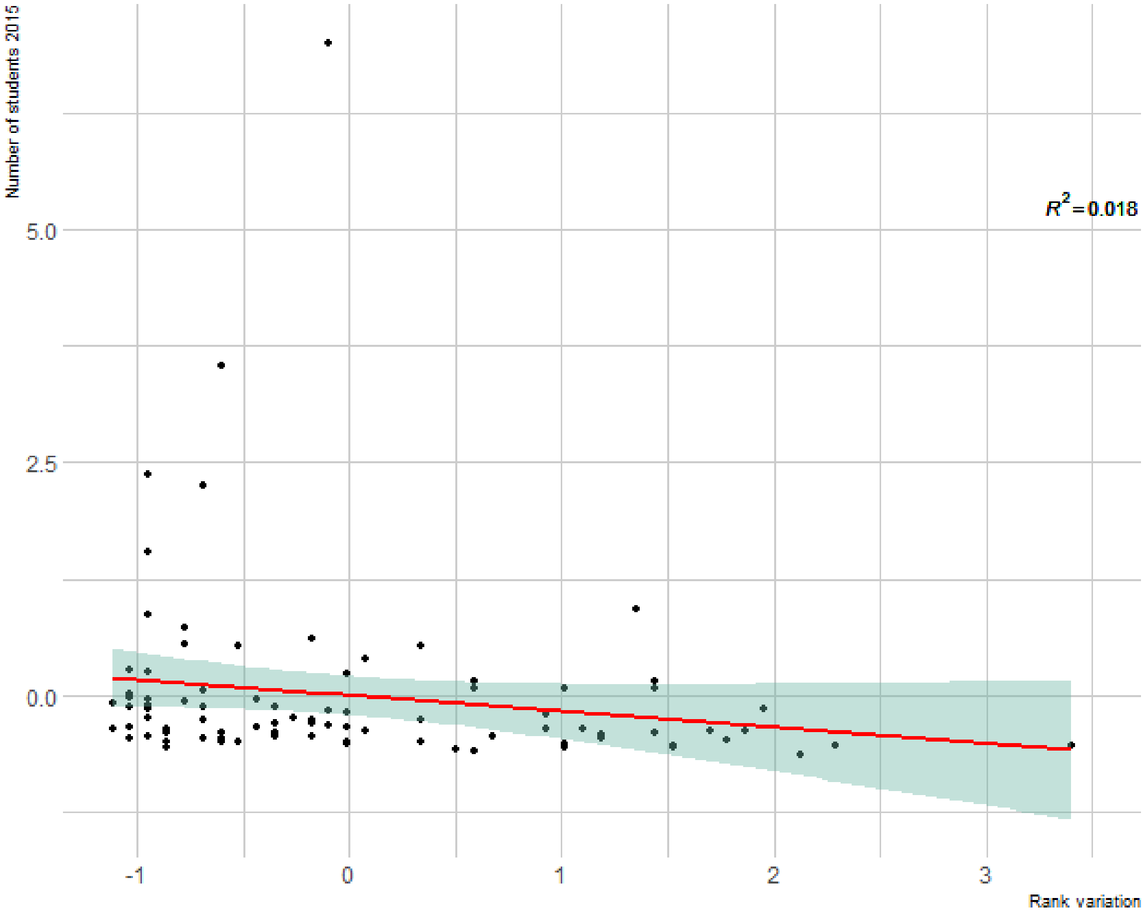


Figure 38. Correlation between index variation across 2010, 2012, 2015, 2019 and the number of students per commune (scaled)

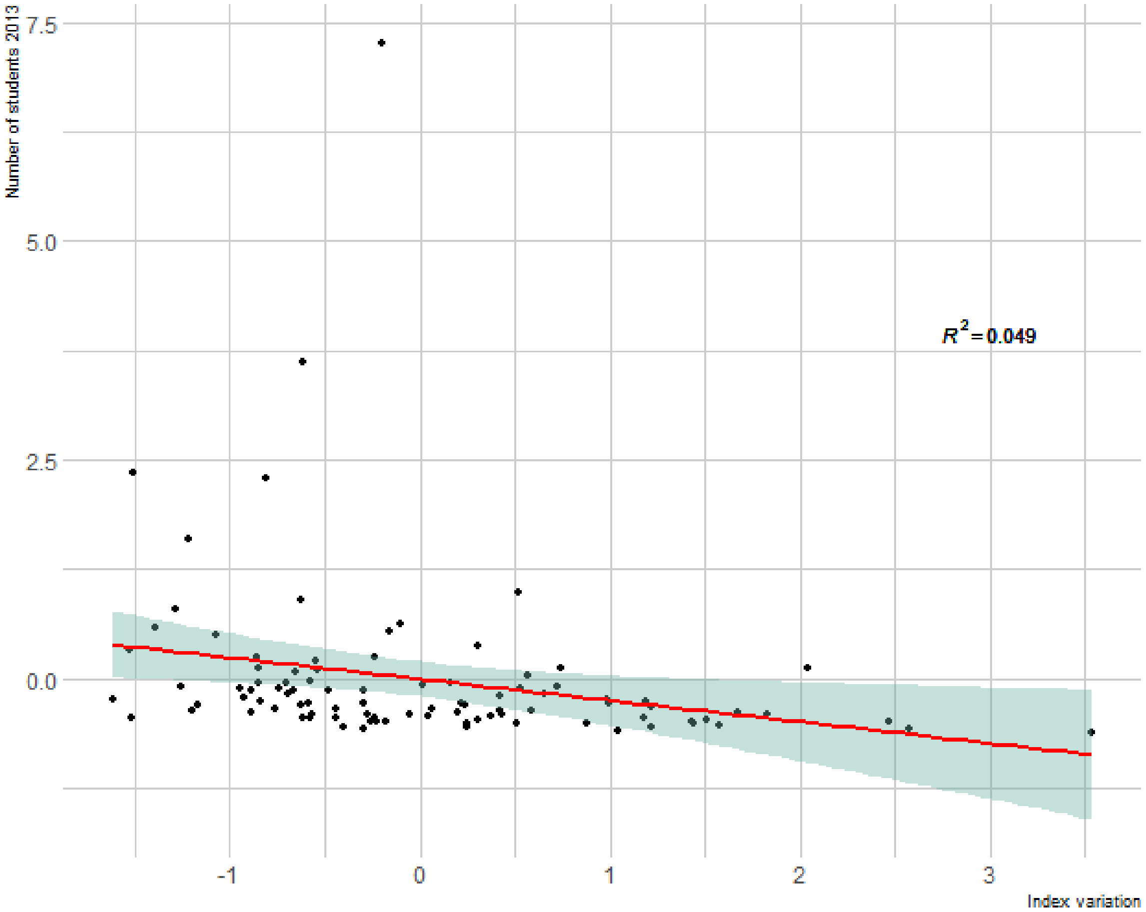


Figure 39. Correlation between index variation across 2010, 2012, 2015 and the number of students per commune (scaled)

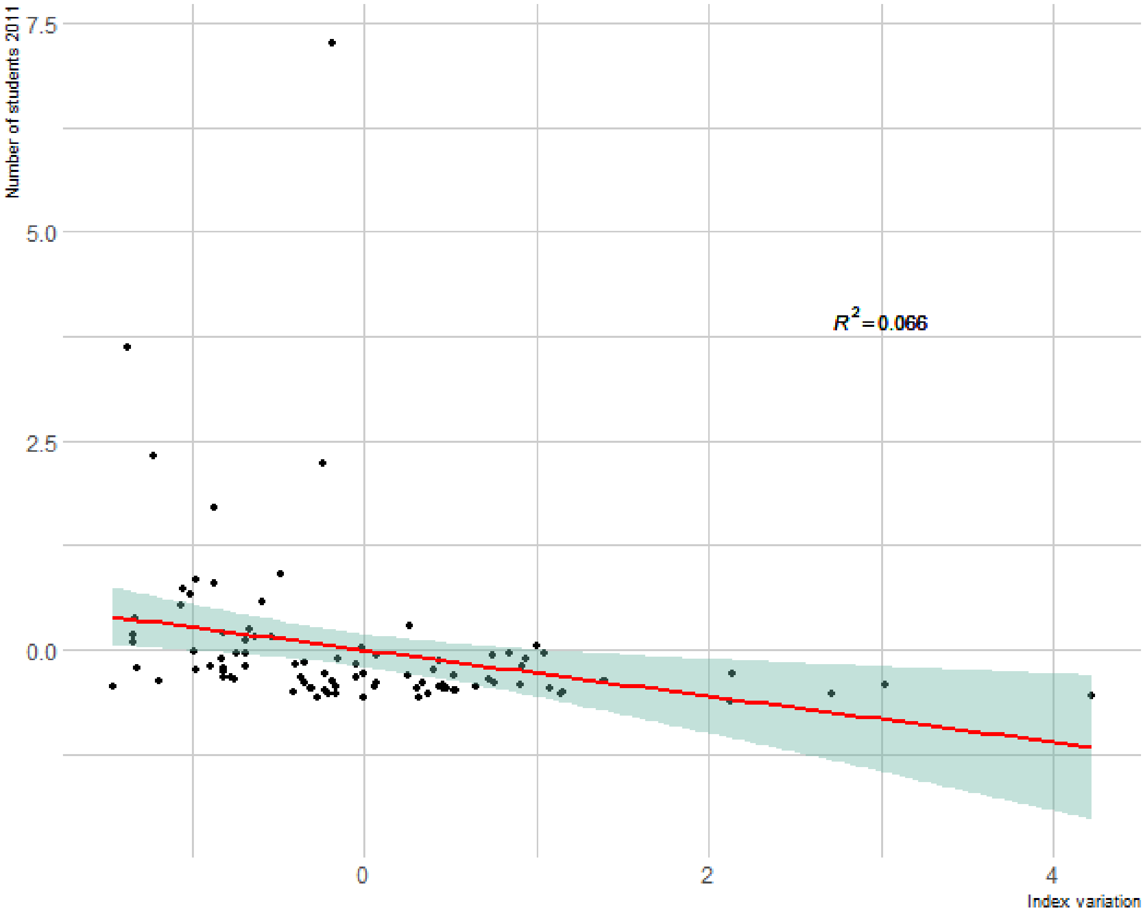


Figure 40. Correlation between index variation between 2010 and 2012 and the number of students per commune (scaled)

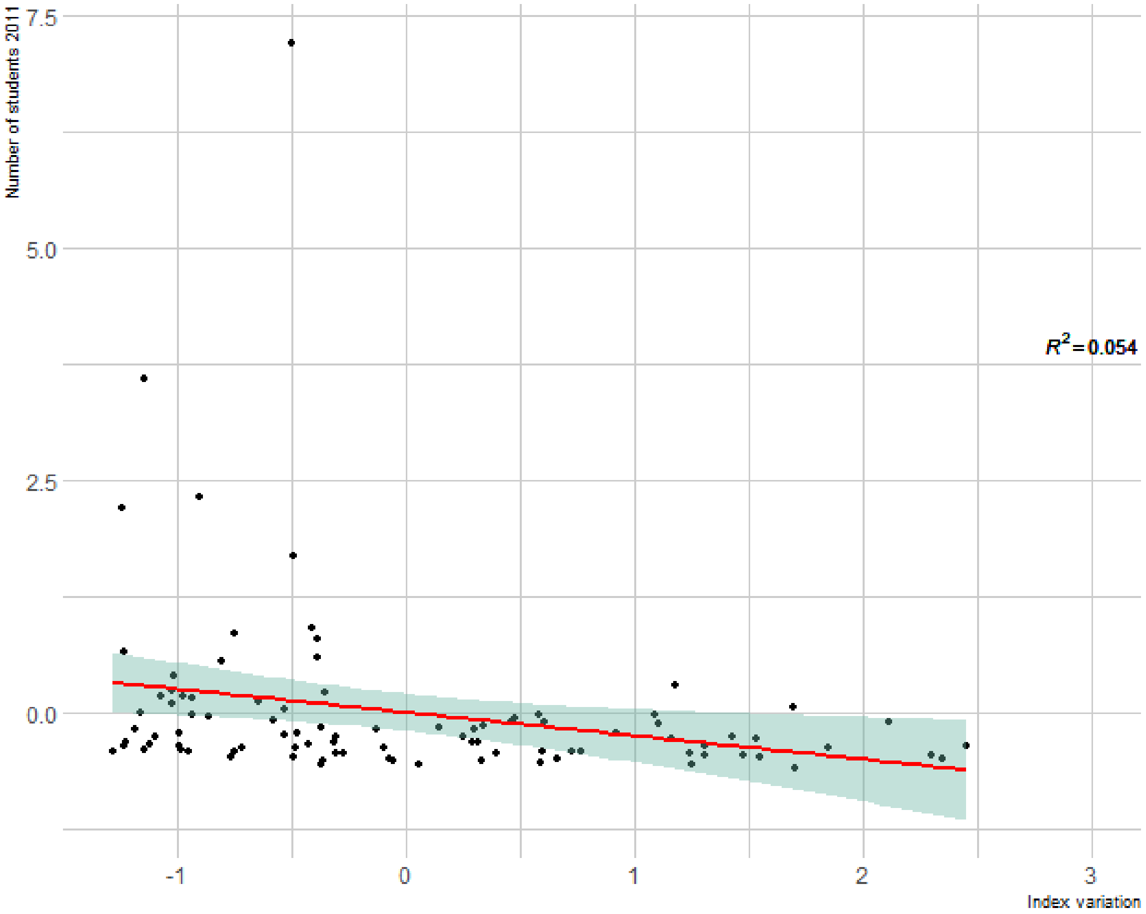


Figure 41. Rank comparison between years, calculation using a PCA (2010-2019)

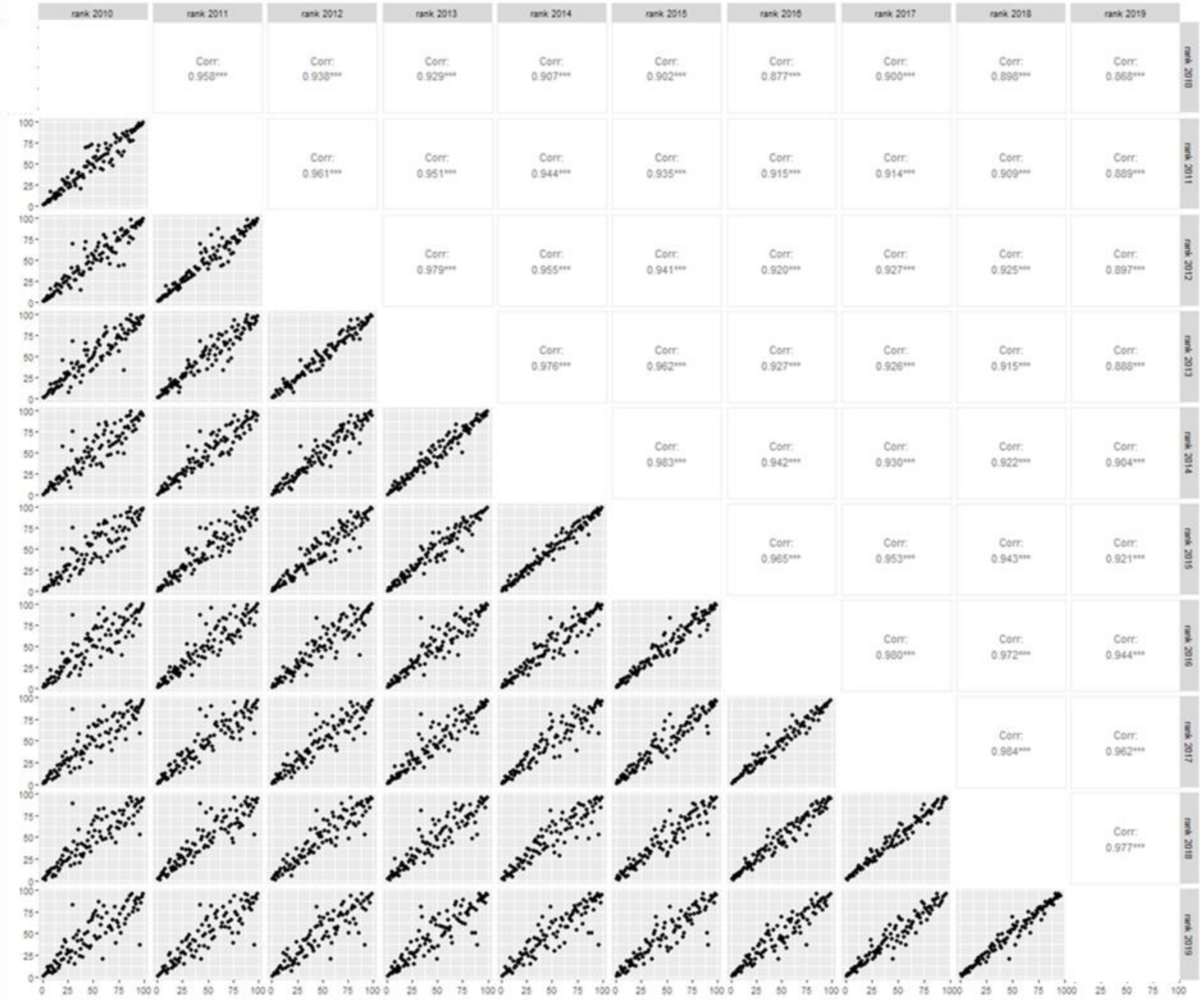


Figure 42. Rank comparison between years, calculation without PCA (2010-2019)

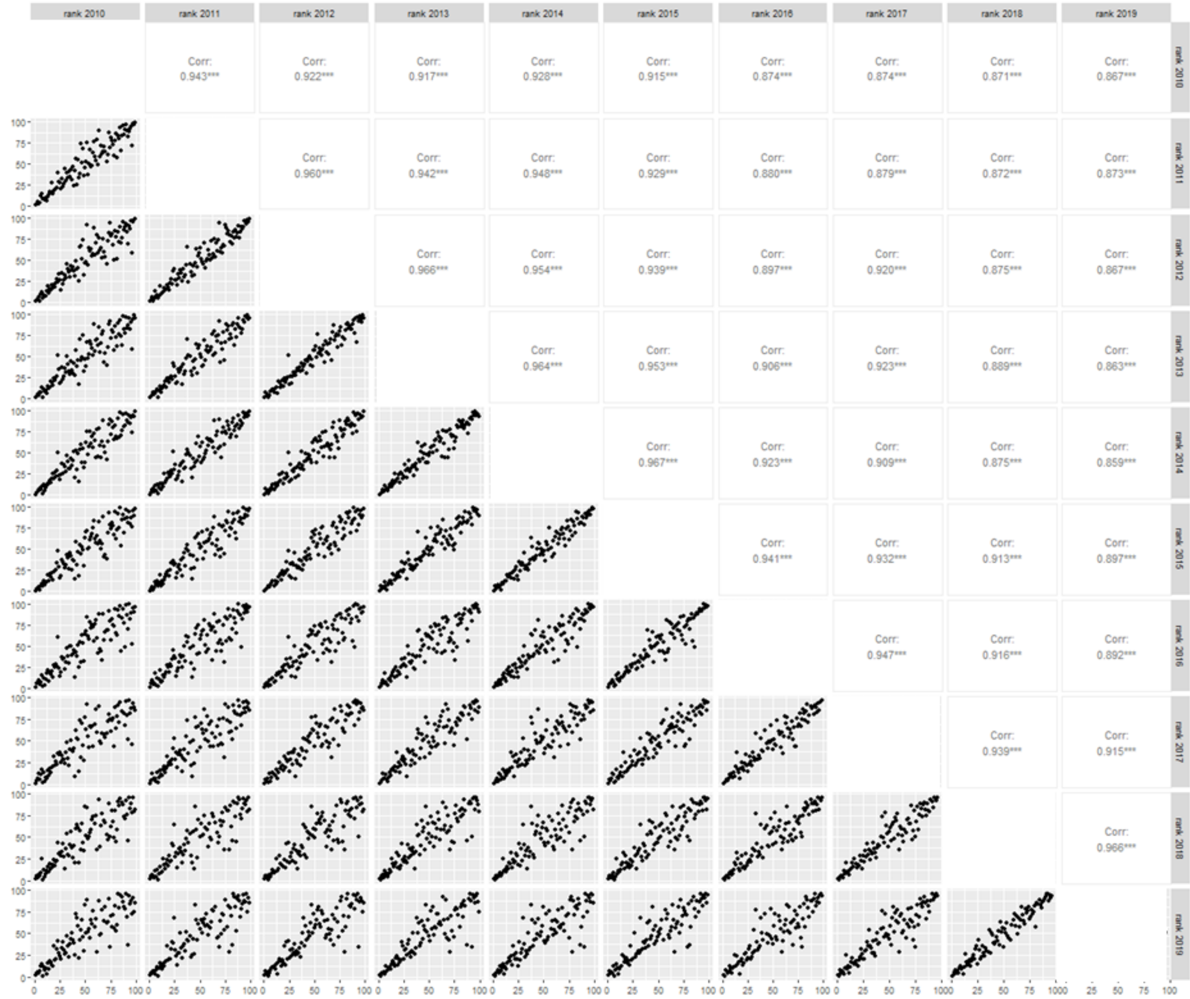


Figure 43. Comparison of rank correlation between calculation using PCA and not using it.

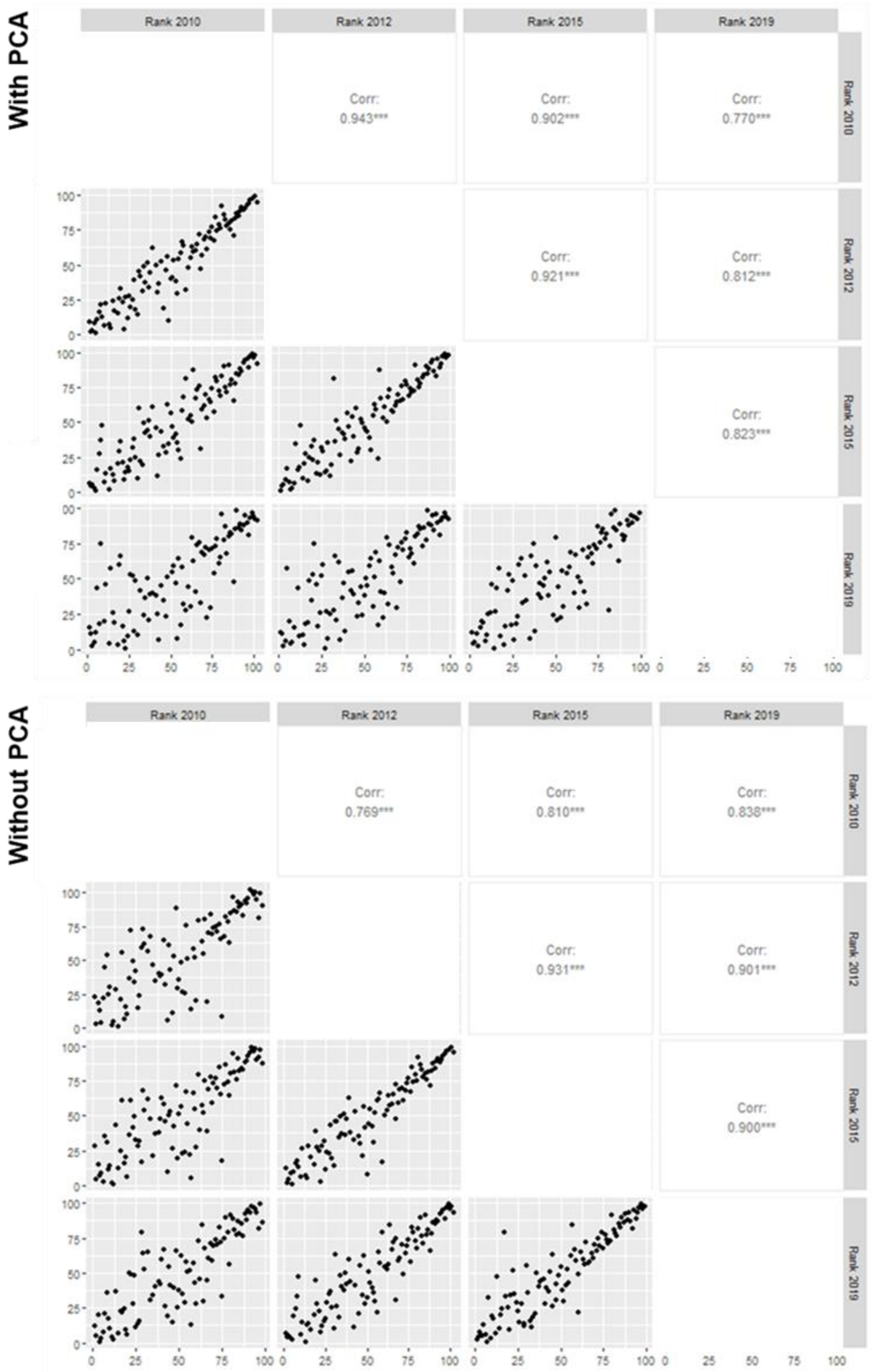


Figure 44. Comparison of rank correlation between the two methods (2010, 2015, 2019)

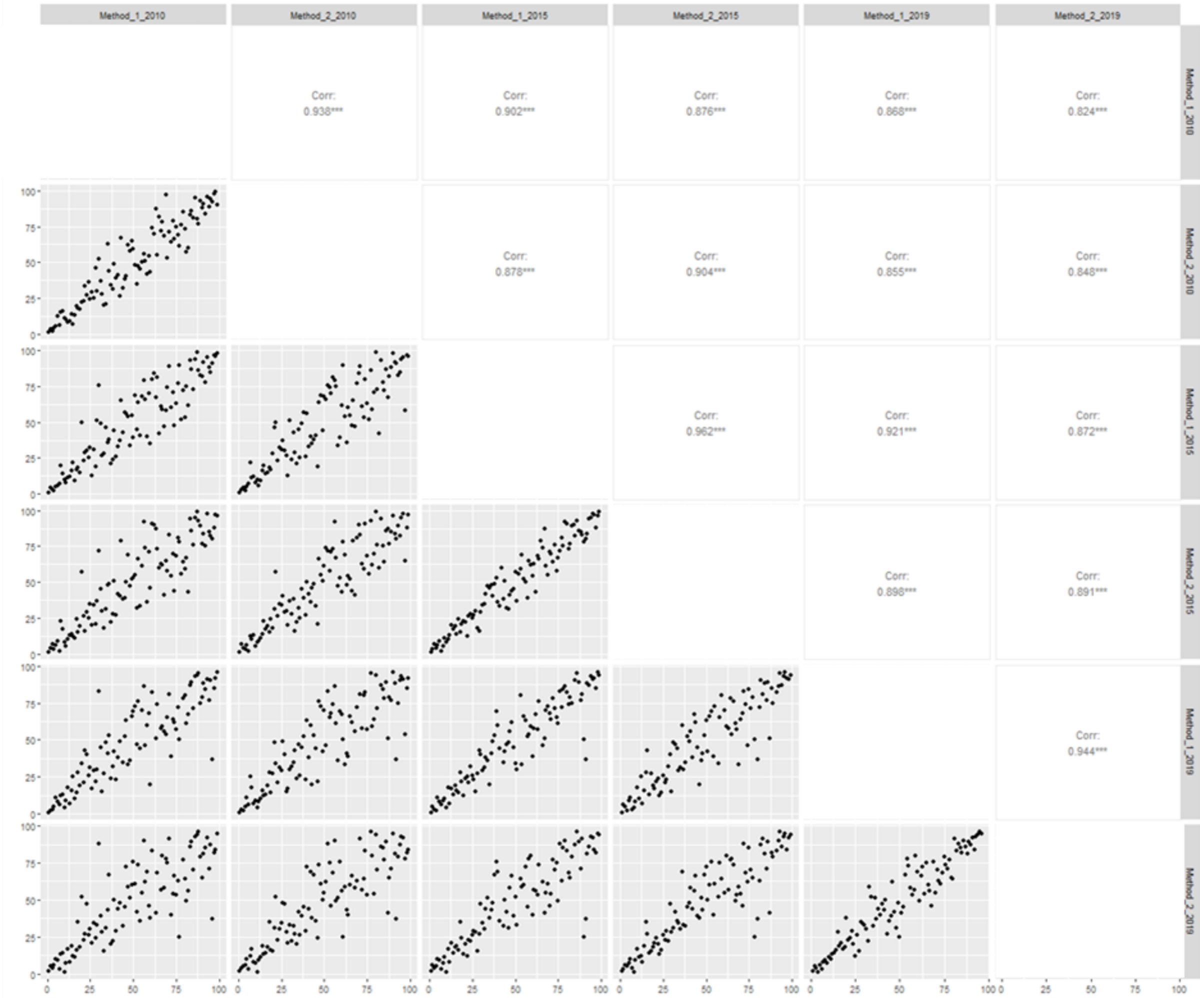


Figure 45. Average household income by first spoken language by type of school – 2019 (in euros)

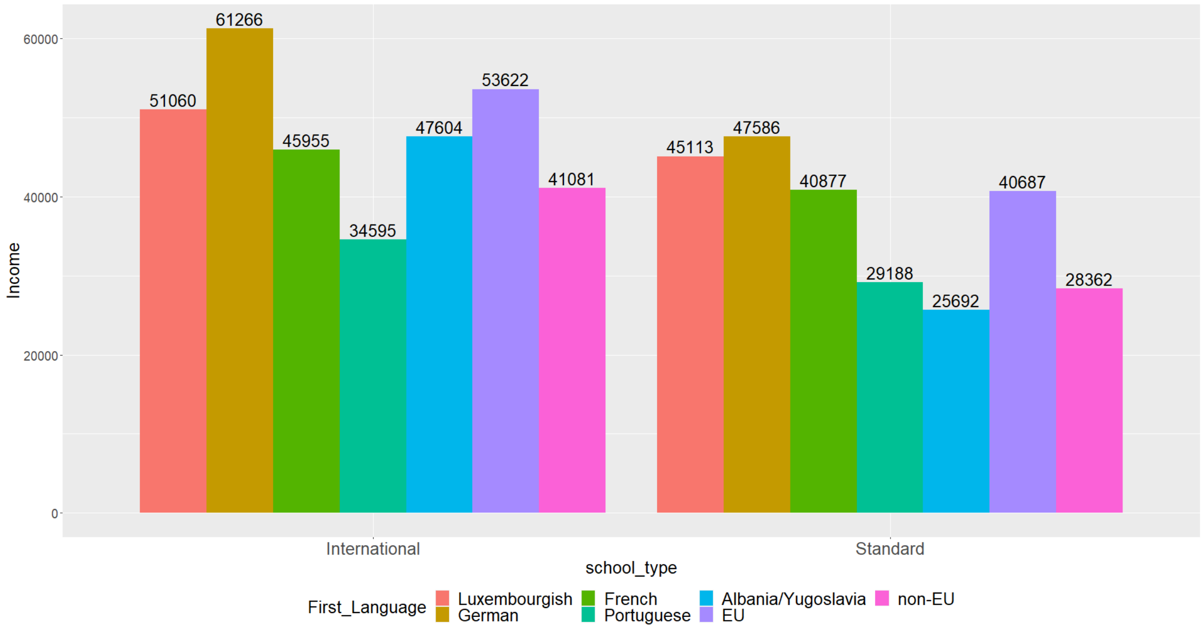


Figure 46. Average household income density distribution aggregated at school level and commune level – 2019 (in euros)

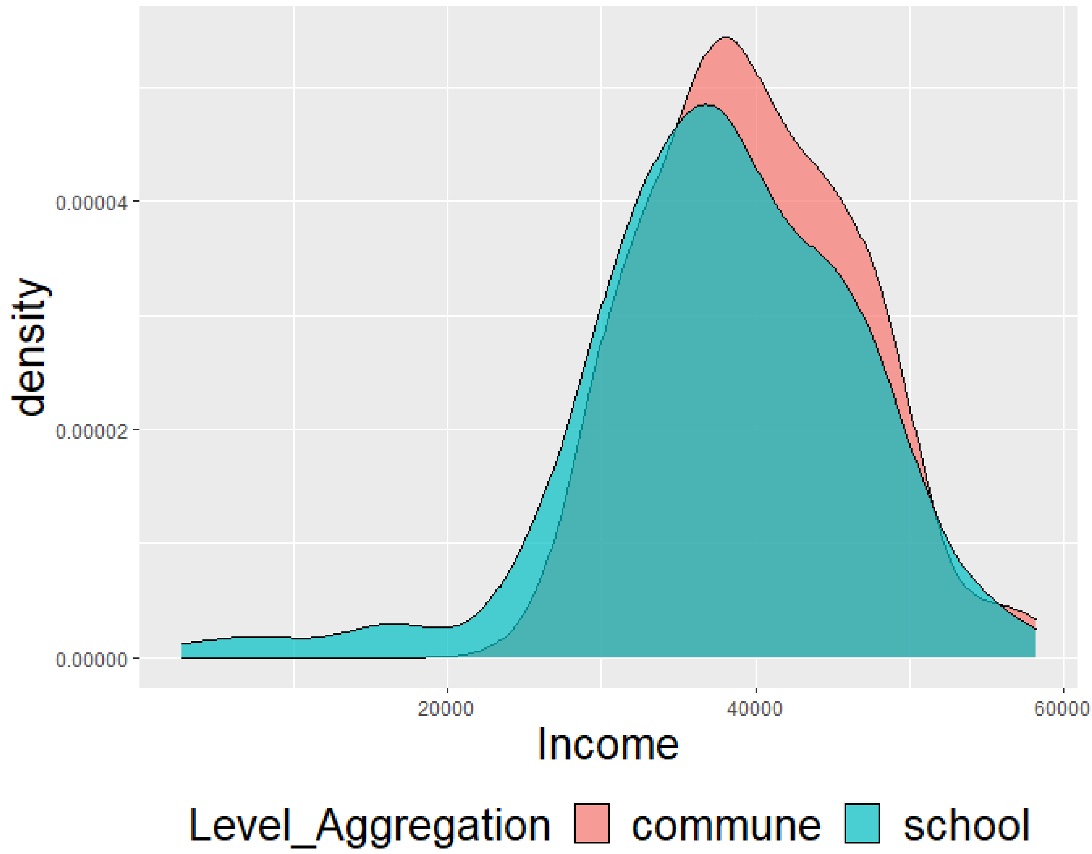


Figure 47. Average household Income by school – Esch-sur-Alzette – 2019 (in euros)

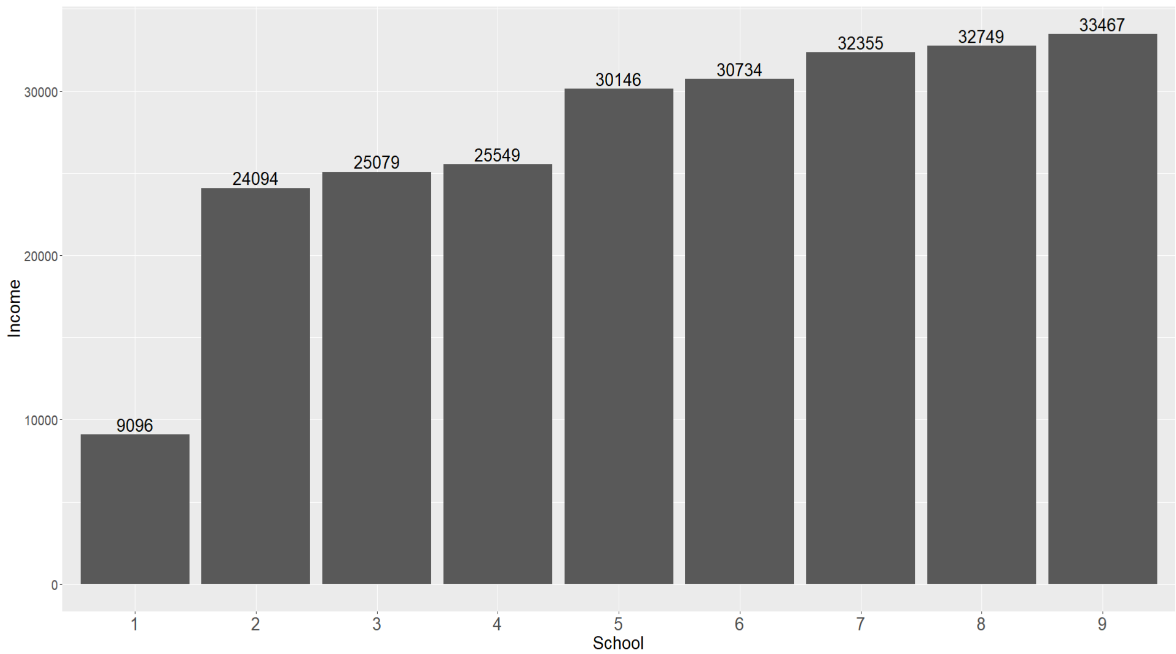


Figure 48. Proportion of student's first language by school – Esch-sur-Alzette – 2019

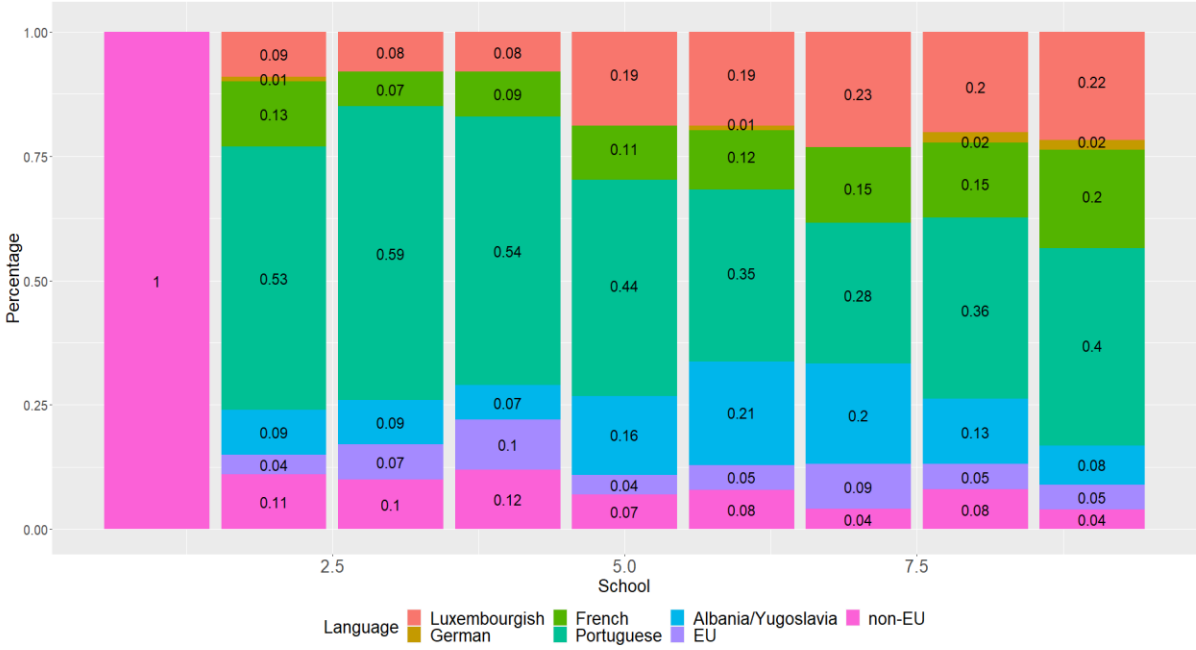


Figure 49. Average household Income by school – Dudelange – 2019 (in euros)

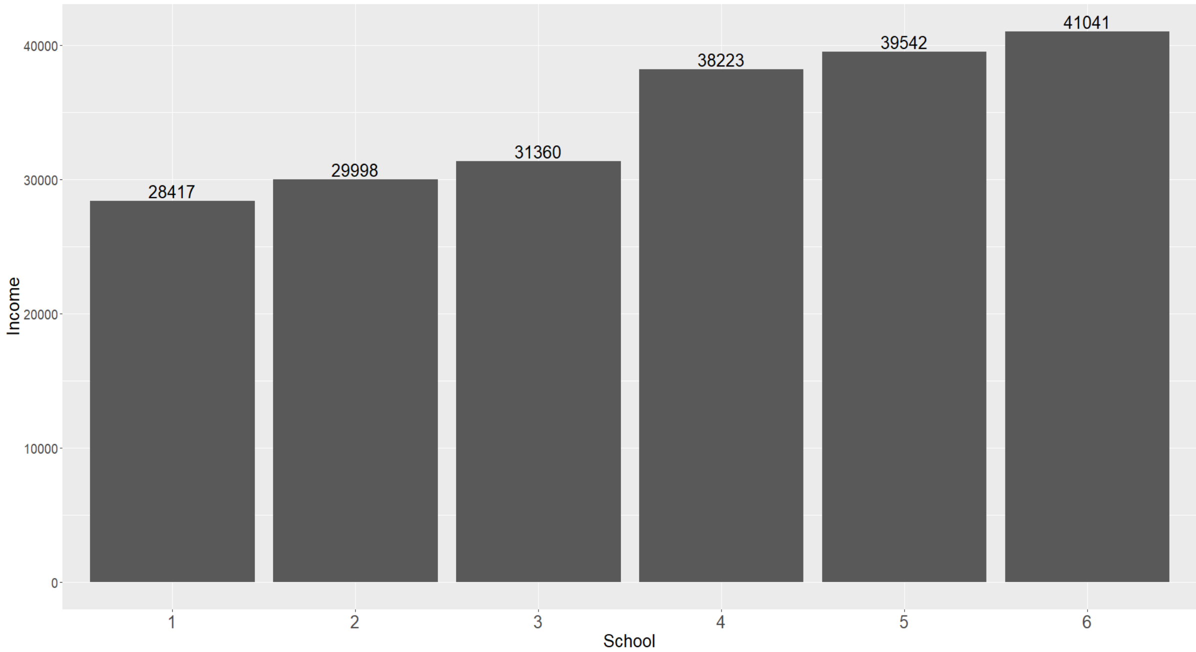


Figure 50. Proportion of student's first language by school – Dudelange – 2019

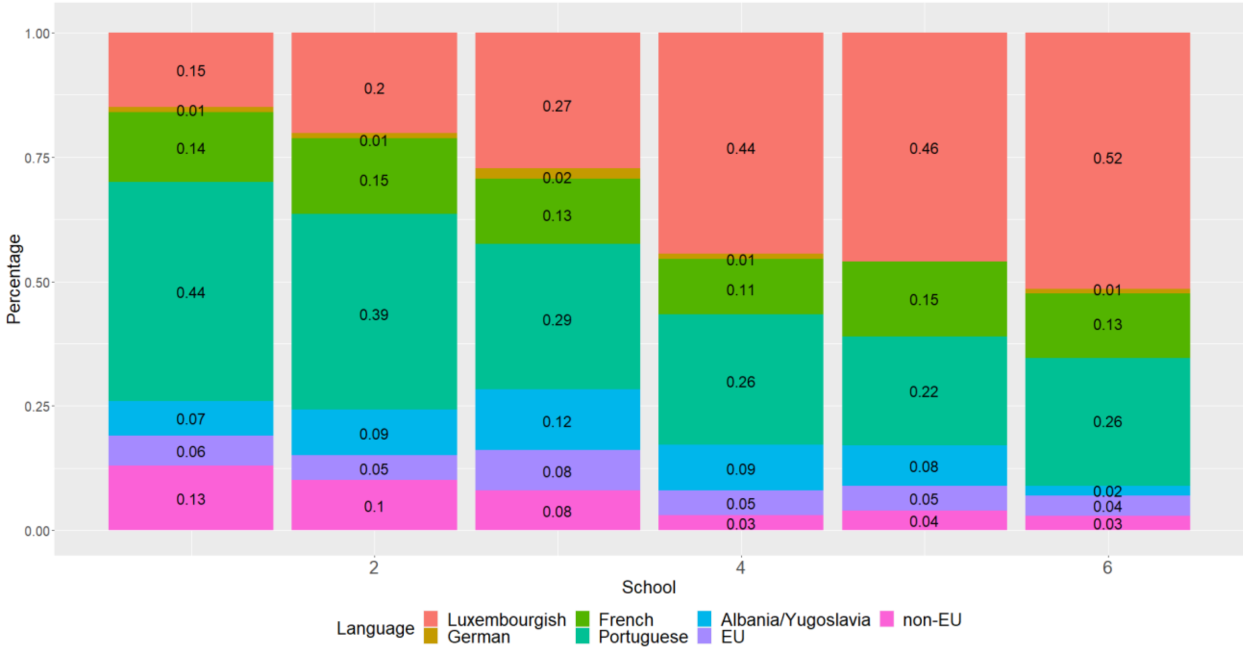


Figure 51. Average household income by school – Differdange – 2019 (in euros)

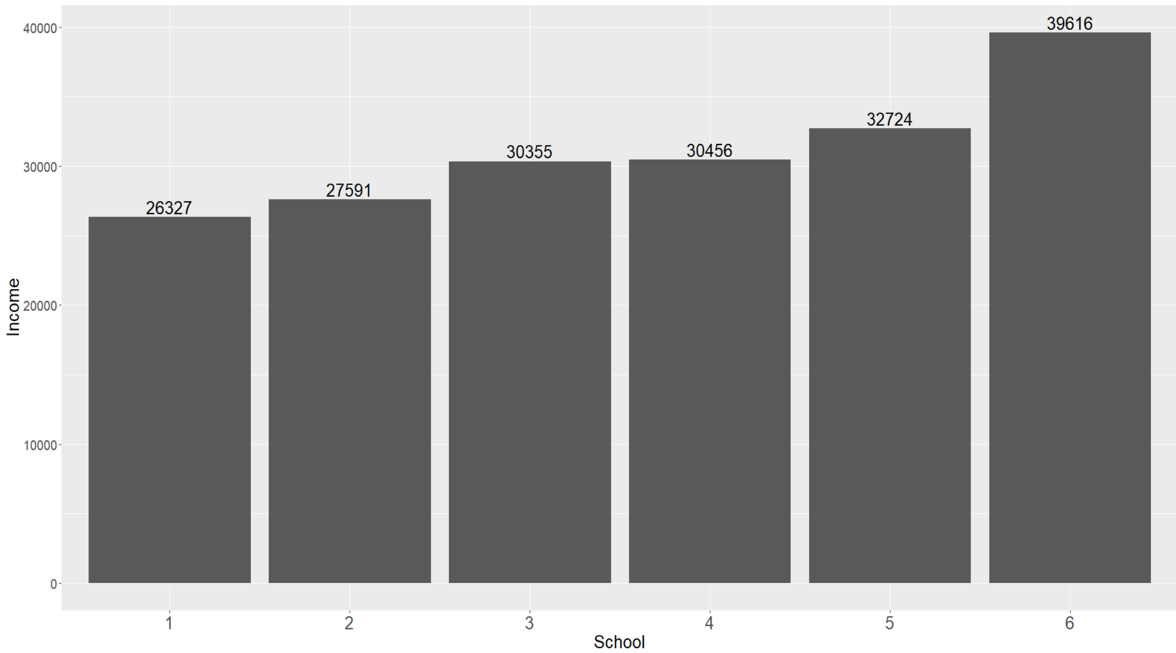


Figure 52. Proportion of student's first language by school – Differdange – 2019

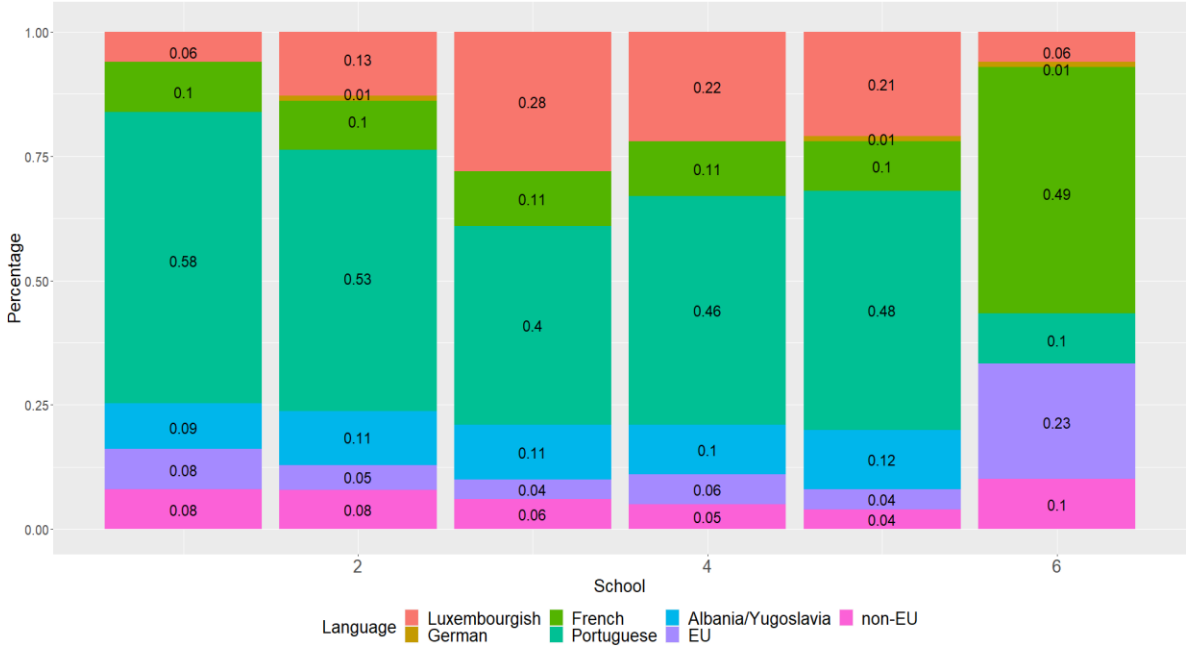


Figure 53. Average household income by school – Sanem – 2019 (in euros)

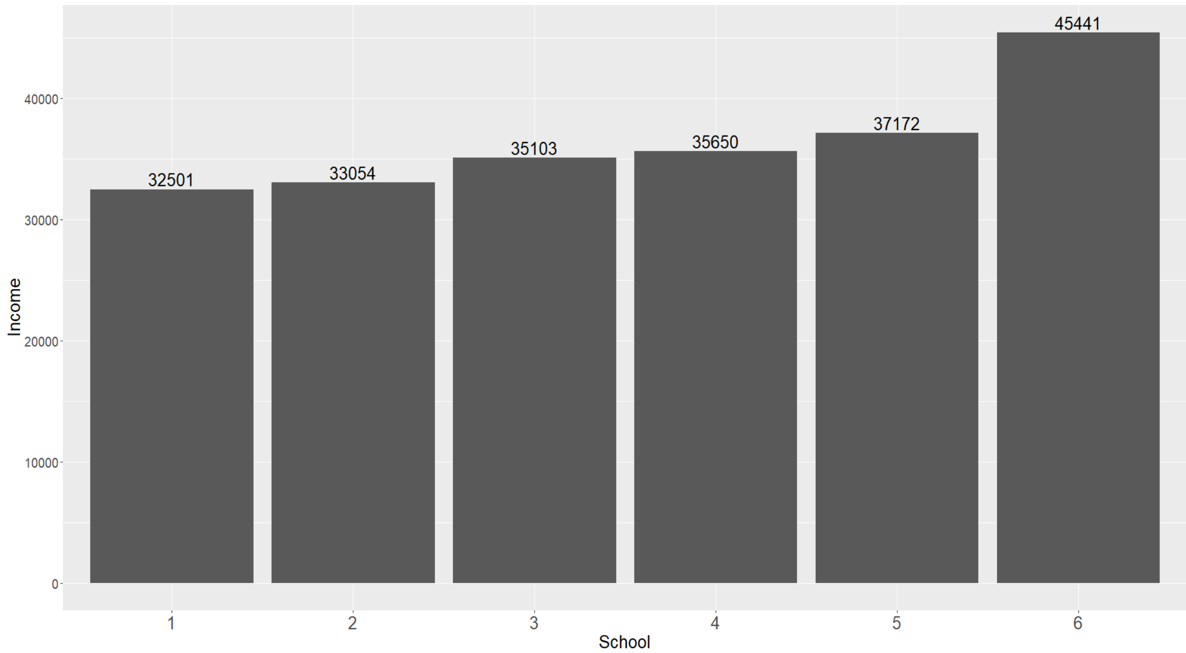
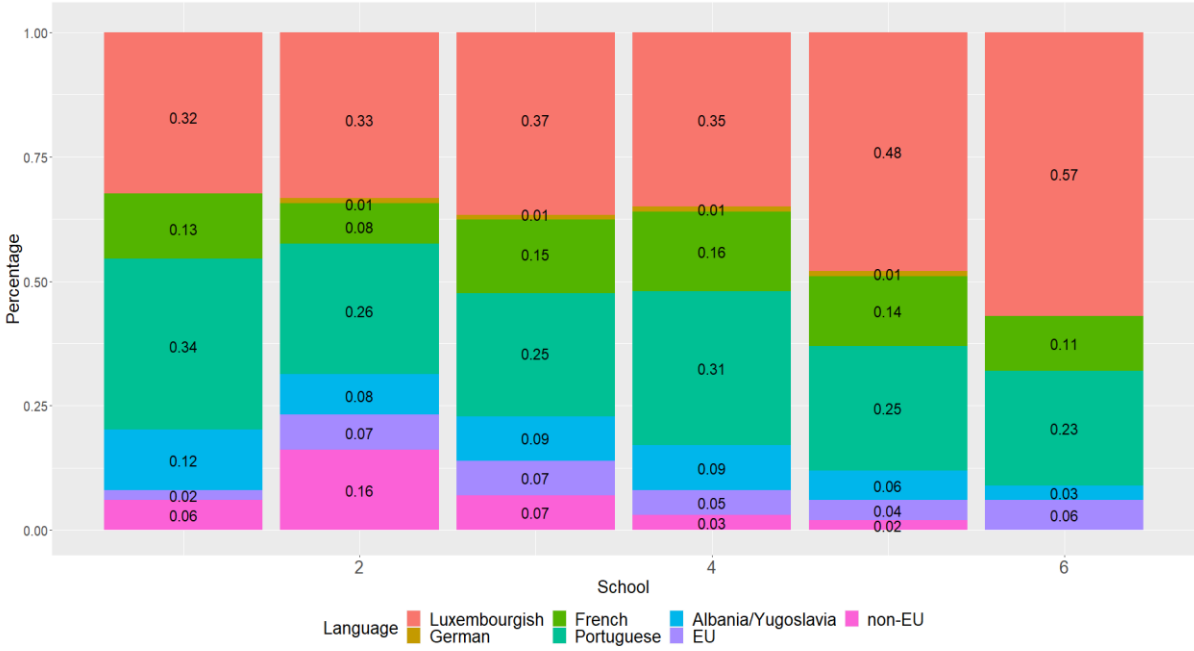


Figure 54. Proportion of student's first language by school – Sanem – 2019



LUXEMBOURG INSTITUTE OF SOCIO-ECONOMIC RESEARCH
11, Porte des Sciences
Campus Belval
L-4366 Esch-sur-Alzette
T +352 58 58 55-1
www.liser.lu