THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

UKRAINE
Education Policy Note:

Introducing the New Ukrainian School in a Fiscally Sustainable Manner

UKRAINE
Education Policy Note:

Introducing the New Ukrainian School in a Fiscally Sustainable Manner

2018

ISBN 978-617-7402-21-2

Contents

List of Tables

List of Figures

List of Boxes

Acknowledgements

This policy note was prepared as a background paper for Ukraine Public Finance Review 2 (P166106). A shorter version of this policy note will appear as a chapter in that report. The main author of the policy note is Lars Sondergaard, drawing on inputs from Irina Capita, Lucia Casap, James Gresham, and Simon Thacker and with useful guidance from Christoph T F Ungerer and Anastasia Golovach. The report was prepared under the guidance of Cristian Aedo (Practice Manager).

The policy note draws on analysis of student learning conducted by Dilaka Lathapipat (included as Annex 3) and a technical paper exploring the scope for network optimization by Kasper Brandt and Thomas Sohnesen (included as Annex 4).

The authors would like to thank the Institute of Educational Analytics and staff at the MoES, including S. Londar; B. Kudrenko; T. Zaporozhets; O. Denysiuk; K. Kuznetsov; A. Lytvynchuk; N. Tytarenko; M. Lavrentiev, Valeriya Izhyk and Oleksandra Husak for their invaluable help in preparing this policy note, including providing school level data, detailed information about how teachers and non-teaching staff are paid, reviewing and commenting on draft terms of references and draft findings. The World Bank team would also like to thank First Deputy Minister V.Kovtunets and Deputy Minister P. Khobzey for their helpful advice and guidance on draft findings of this policy note.

This work would not have been possible without the generous support of the EU Programme for the Reform of Public Administration and Finances (EURoPAF)

Peer reviewers for this note were Igor Kheyfets, Anna Olefir, and Dina N. Abu-Ghaida.

Abbreviation

GDP	Gross Domestic Product
GSE	General Secondary Education
IEA	Institute of Educational Analytics
ISCED	International Standard Classification of Education
LU	Law of Ukraine
OECD	Organization for Economic Cooperation and Development
MIS	Management Information System
MOES	Ministry of Education and Science of Ukraine
MOF	Ministry of Finance of Ukraine
NUs	New Ukrainian School
PIAAC	Programme for the International Assessment of Adult Competencies
PISA	Program for International Student Assessment
TIMSS	Trends in International Mathematics and Science Study
UAH	Ukrainian Hryvnia
UN	United Nations
UNDP	United Nations Development Program
UNESCO	United Nations Educational, Scientific and Cultural Organization
WB	World Bank

Summary

The education sector in Ukraine is in the middle of ambitious – and long overdue – reforms that hold great promise to fundamentally transform the sector. New laws have been passed for Higher Education (in 2014), for Research and Scientific Activity (in 2015), and, more recently, the framework law for the education sector, "Law on Education" (in 2017). Alongside the Budget Decentralization Reform (2014) these laws represent a major shift towards devolving authority from central to local government and the expansion of decision-making autonomy by local authorities and education service providers (e.g., schools and universities). Moreover, as part of these reforms, per student financing for schools was introduced in 2017, with the potential to incentivize local actors to use resources more efficiently. Taken together, these changes represent the most ambitious reform agenda for the education system since the collapse of the Soviet Union. However, the costliest feature of the general secondary education law is a promise to increase the starting salary of teachers to four times the minimum living wage by 2023. If not managed carefully, this increase threatens to put the sector on a fiscally unsustainable path which could undermine the broader reform agenda. This note highlights some areas of the reform agenda where more focus will be needed, and presents some options for how to implement the promised wage increase in a fiscally sustainable manner.

Source: www.adm.dp.gov.ua

Why are the reforms being launched and what do they entail?

Triggered by public dissatisfaction with the quality of education provided, a fundamental overhaul of the education system is well underway

The past few years have witnessed the most ambitious attempt at reforming Ukraine's education system since the collapse of the Soviet Union. Laws have been passed for Higher Education (in 2014), for Research and Scientific Activity (in 2015), and, more recently, the framework law for the education sector, "Law on Education" (in 2017). Alongside the Budget Decentralization Reform (2014) these laws represent a major shift towards devolving authority from central to local government and the expansion of decision-making autonomy by local authorities and education service providers (e.g., universities). While the Law on Education which was passed in 2017 is a framework law that covers the entire education sector, this note focuses on the implication of that law on the General Secondary Education for two reasons: first, this subsector consumes the largest proportion of the budget. Second, there is more clarity on what changes the Law on Education will entail for that subsector. [1] For other subsectors (e.g. technical and vocational education), future legislation (for those subsectors) will provide more clarity on the reforms needed, their costs, and the pace with which they will be implemented.

Three years of national dialogue took place with the aim to create a common understanding of the need for change. Building a broad, shared understanding – across the political spectrum – on why reforms are needed is an important ingredient in sustaining reforms. On this front, the new law was underpinned by years of discussion, culminating in a set of reasons explaining why reforms were both needed and long overdue (laid out in the Ministry of Education and Science of Ukraine's pamphlet on "The New Ukrainian School. Conceptual Principals of Secondary School Reform"). These reasons – as presented in this pamphlet – include the following: (1) Pupils are only able to reproduce pieces of unstructured knowledge; however, they often do not know how to use it to solve everyday problems; (2) The way of teaching in contemporary Ukrainian schools does not motivate children to learn: Textbooks are too theoretical and overburdened with secondary factual materials; and teachers use mostly outdated teaching methods; (3) A low social status and low salaries demoralize teachers. Teachers lack real motivation as well as opportunities for personal and professional growth; and (4) Due to chronic underfunding in the sector, today not all Ukrainian citizens enjoy equal access to the quality education that has been guaranteed by the government. Schools reproduce the poverty ladder: children from poor families have worse chances to get a good education and climb the social strata.

The broad dialogue helped identify and create a common understanding of some of the potential drivers for transforming Ukraine's education system. The end result is a framework law which puts in motion some drastic changes to Ukraine's Soviet-era education system, including codifying into law the main elements of the "New Ukrainian School concept". These include modern approaches to: (i) school curricula, focused on 21st century skills and competencies; (ii) teacher professional development, emphasizing student-centered learning; and (iii) system management and school administration, emphasizing greater local decision-making powers and a different role for the central government with a focus on setting and monitoring learning standards. Indeed, the law sets the stage for much-needed structural reforms in the education sector, with the following measures representing the most significant structural changes:

i. De-bureaucratization of the education sector by granting increased professional autonomy to teachers, schools, and local authorities;
ii. Aligning the Ukrainian school system with European norms, including the transition to 12 years of schooling rather than the current 11 years;
iii. Modernizing school curricula to emphasize competencies and 21st century skills, rather than content knowledge;
iv. Introducing a national system of qualifications, including a National Qualifications Framework;
v. Launching a National Agency for Educational Quality Assurance.[2]

[1] In Ukraine, "General Secondary Education" refers to schooling provided at the primary, lower secondary and upper secondary level.

[2] A more detailed summary of the draft law can be found in Annex 1.

Source: www.adm.dp.gov.ua

How well-positioned is Ukraine to launch these reforms? The answer is mixed

In terms of learning outcomes, the reforms are being launched from a relatively strong starting point: Ukraine performs better than countries at the same income level

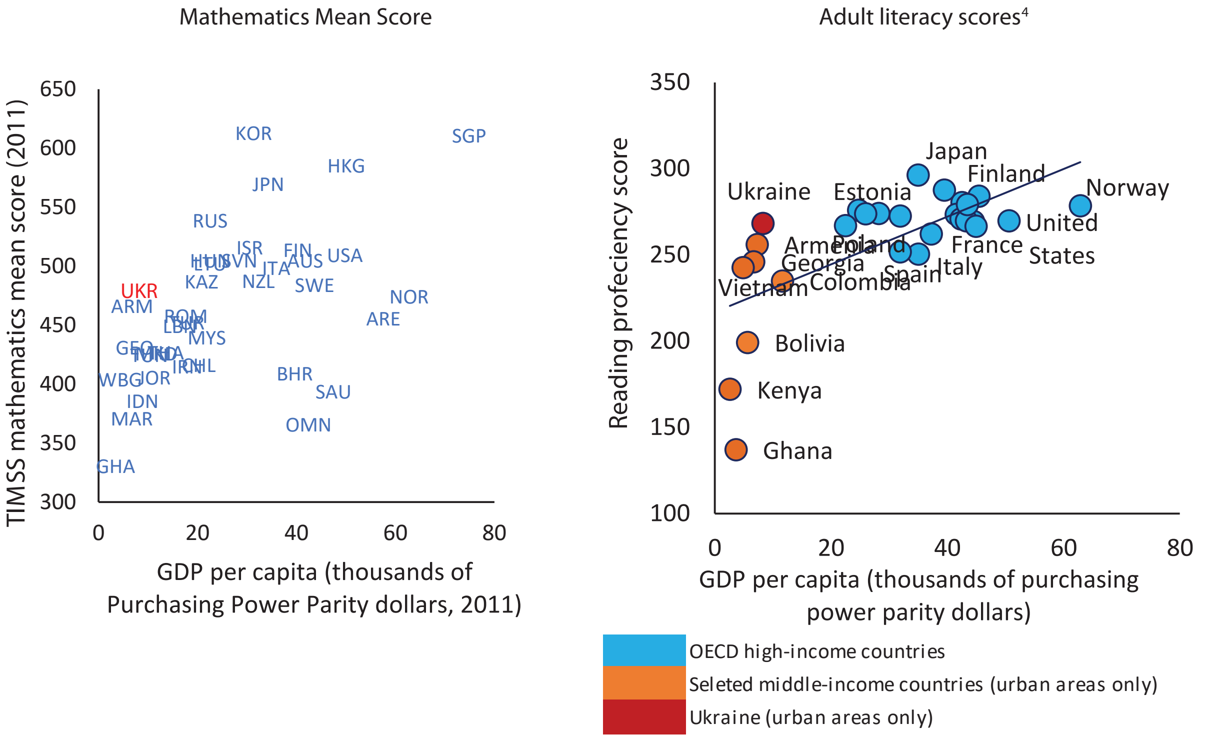
When Ukraine has participated in international assessments of students' and adults' cognitive skills, it has done relatively well. According to 2011 data from the Trends in International Mathematics and Science Study (TIMSS 2011), Ukrainian grade 8 students performed relatively well given the country's level of economic development. Ukraine's mean scores on the mathematics test were significantly higher than those of Armenia, Georgia, and Indonesia (countries with similar GDP per capita). In fact, Ukraine's results were close to those achieved by eighth graders in far richer countries, including Norway, Sweden, and New Zealand. Similarly, when adults' cognitive skills were measured in Ukraine and in a number of OECD countries and other upper middle-income countries in 2011-13, adults in Ukraine outperformed adults in much richer countries, including Poland, France, Italy and Spain (please see Del Carpio et al (2017) and Roseth et al. (2016)) (Figure 1). Ukraine has not yet participated in OECD's Programme for International Student Assessment (PISA), but Ukraine participated for the first time in PISA 2018 (with results due to be released in late 2019).

Although Ukraine performed well on TIMSS 2008, there were significant variations in performance in terms of school location, with students in smaller towns and rural areas scoring significantly lower than students in urban cities. Furthermore, nearly 20 percent of grade 8 students—or about 1 in every 5 students—failed to achieve even the low benchmark in mathematics.[3] This indicates that some areas of the country and some population sub-groups are at a disadvantaged starting point in terms of launching the above-mentioned reforms.

Although Ukraine's general education outcomes look good, firms report having difficulty finding workers with the necessary skills. In particular, a 2014 survey of firms conducted by the World Bank found that 40 percent of firms in four key sectors (agriculture, food processing, information technology, and renewable energy) report a significant gap between the type of skills their employees have and those they need to achieve their business objectives (Del Carpio et al. 2017). This likely reflects an outdated curriculum, outdated teaching equipment/facilities as well as demotivated and poorly trained teachers. Moreover, it likely reflects too little emphasis on teaching pupils to learn independent problem-solving skills.

[3] Mullis et al. (2012). TIMSS 2011 International Results in Mathematics. International Association for the Evaluation of Educational Achievement.

Figure 1: TIMSS 2011 Mathematics Performance and GDP Per Capita (left) and Adult literacy scores and GDP per capita (right)



Source: World Bank EdStats database (left) and Del Carpio et al (2017) (right), using data from ULMS-STEP Household Survey (2012), STEP Household Surveys (2012-13), OECD's calculation for OECD-23 based on PIAAC data (2011-12)

Notes: [1] Qatar, with GDP per capita at PPP of USD134,000, is not shown on the figures.
[2] Panel (b) shows GDP per capita on horizontal axis and adult literacy scores on the vertical axis

Box 1: Ukraine's education system in brief

Ukraine's law on education guarantees every citizen the right to education. Public expenditures go towards maintaining an extensive network of public education institutions covering preschool, general secondary, vocational, and higher education.

- Preschool. Preschool education is mandatory in Ukraine. It can be obtained within the family setting, (until the child reaches age 5) and from qualified providers. When children reach 5 years of age, parents can choose a form of preschool education from among full-time preschool institutions, part-time groups, or special pre-primary groups within primary schools.
- General secondary. General secondary education (GSE) in Ukraine is divided into three levels: primary (level I: grades 1–4), basic general secondary (level II: grades 5–9), and high school/complete general secondary (level III: grades 10–12 (with the 12th year having been introduced by the Law on Education in 2017). A certificate of completion of lower secondary general education is issued after level II, and a certificate of completion of upper secondary general education is issued after level III. GSE services are provided through a network of institutions of various types.

[4] Data for middle-income countries (Armenia, Colombia, Georgia, Ukraine, and Vietnam) are only representative of urban areas. Reading proficiency scores range from 0 (lowest) to 500 (highest). 22-OECD country average is 273. For description of reading scores, see table A.2 in appendix A (in Del Carpio et al 2017)

- Vocational education and training. Vocational education and training (VET) is also offered in several types of institutions. Students can enroll in VET after completing lower secondary (grade 9) or upper secondary (grade 11). Those enrolling in VET after grade 9 can receive an upper secondary education certificate together with a "skilled worker diploma" after two years of study. Those enrolling after grade 11 receive a "skilled worker diploma" after one year of study.
- Higher education. Higher education in Ukraine is provided by colleges, technical colleges, universities, institutes, and academies. Before the Law on Higher Education was passed in 2014, Ukraine had four levels of higher education institutions (HEIs): level I, technical colleges; level II, colleges; level III, institutes and conservatories; and level IV, universities, academies, institutes, and conservatories. Since the 2014 reform, a junior specialist degree is awarded upon completion of 1 – 1.5 years or 3 - 3,5 years of study, depending on the specialty. A bachelor's degree after four years of study. Postgraduate Master's, Candidate of Science (PhD), and Doctor of Science degrees are available.

According to the latest data available, over 32,000 institutions provide educational services in Ukraine. Of these, 14,949 offer preschool education, 16,180 general secondary, 756 vocational, and 661 higher education. These institutions enroll 7,020,804 students and employ 777,335 teaching staff (Table 1).

Table 1 Basic Statistics of the Ukraine Education System, 2017/18[1]

	Institutions	Students	Teachers	Students Per Institution	Students Per Teacher
Preschool Education[2]	14,949	1,291,207	137,881	86	9.4
General Secondary Education	16,180	3,921,673	439,701	242	8.9
Vocational Education[3]	756	269,359	43,416	356	6.2
Higher Education[4]	661	1,538,565	156,337	2328	9.8

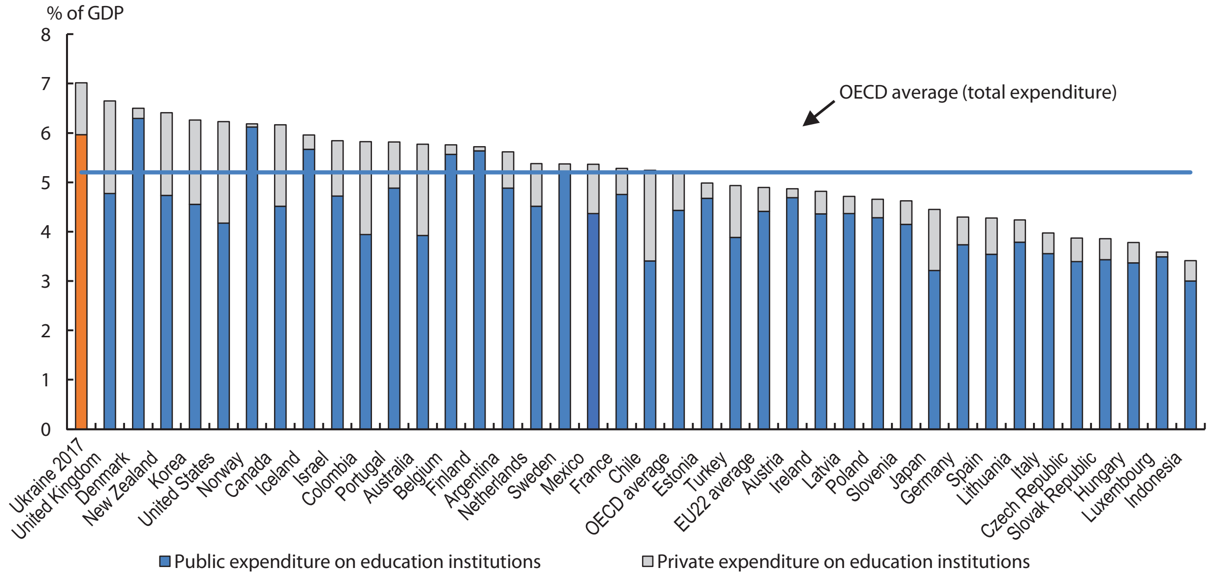
Notes: [1] Excluding the Autonomous Republic of Crimea, the city of Sevastopol, and parts of the regions of Donetsk and Luhansk.
[2] Figures as of the end of 2016 (latest available year).
[3] Number of teachers is from 2015
[4] Includes short programs (accreditation levels I–II), bachelor programs (accreditation levels III–IV), and postgraduate programs.

Source: World Bank (2017), updated with data from State Statistics Service's website and through an email exchange with MOES officials

In terms of spending on education, the starting point is less favorable: reforming the system will require new spending but spending is already high

By most metrics, Ukraine's public spending on education is high, although spending has declined in recent years. Between 2013 and 2017, budget financing shrank from 7.2 to 6.0 percent of GDP (from 21 to 15 percent of total government spending). After the steep devaluation of the hryvnia, public education spending declined by 35 percent in real terms over two years. The decline brought Ukraine closer to international benchmarks in terms of the share of national wealth devoted to education, but spending remains high: with public spending on education at 6.0 percent of GDP and with private spending adding another percentage point of GDP, Ukraine's education spending is amongst the highest in the world (Figure 2). Box 2 provides more details on the composition of spending and its trends.

Figure 2: Public and Private Spending on Education,[1] Percent of GDP, Ukraine (2017) and OECD Countries (2014)



Source: Ukraine BOOST 2017 based on data from State Treasury Service and State Statistics Service's National Education Accounts 2016, and OECD AAG 2017[5].
Note: [1] Pre-primary through tertiary, including expenditures not allocated by level.

Despite being an outlier internationally, the Law of Ukraine "On Education" (Art. 78) commits Ukraine to spend at least 7 percent annually on education, potentially taking off pressures to improve efficiency of spending. The law does not specify the mix between private and public spending but, nevertheless, at 7+ percent of GDP, Ukraine would remain an outlier, internationally. More importantly, with such pre-commitment to spending, there is a risk that there will be less pressures to improve the efficiency of spending. In particular, there is a risk that the key structural cause of Ukraine's inefficiencies in general secondary education – namely, the large network of schools – will remain unaddressed.

Box 2: Most of education spending is allocated for recurrent expenditures
As in many countries, most public education spending finances personnel costs and general secondary education. According to economic classification, 55 percent of all budgeted spending on education went to labor costs; non-personnel recurrent costs accounted for 37 percent, and capital spending made up 8 percent. However, these figures mask an issue in higher education, where more than 90 percent of spending is allocated through a single budget program classified as non-personnel recurrent expenditure.[6] In terms of functions, the largest part of public spending on education is allocated to the general secondary subsector, followed by higher education and preschool education. Of the 6.0 percent of GDP public expenditure on education in 2017, 47 percent went to general secondary schools, 22 percent to higher education institutions, 16 percent to preschools and other preprimary education institutions, and 5 percent to VET schools (Table 2).

[5] The BOOST initiative is a World Bank collaborative effort launched in 2010 to facilitate access to budget data and promote effective use for improved decision-making processes, transparency and accountability. Currently deployed in about 40 countries globally, the appeal of the BOOST approach is that it provides user-friendly platforms where all expenditure data can be easily accessed (often in conjunction with nonfinancial indicators) and used by researchers, government officials and ordinary citizens to examine trends in allocations of public resources, analyze potential sources of inefficiencies, and become better informed about how governments finance the delivery of public services. In Ukraine, the data for the BOOST comes from the State Treasury Service.

[6] Expenditure category 2282, "Special initiatives to implement national/regional programs other than development activity." These block transfers to universities and other HEIs contain a high proportion of personnel spending, thus biasing the categorical distribution of public spending.

Table 2: Distribution of Public Expenditures on Education by Economic and Functional Classification in 2017

	Labor	Non-labor	Capital expenditures	Total
Preschool education system	10%	5%	1%	16%
General secondary education system	35%	8%	4%	47%
Vocational training system	3%	2%	0%	5%
Higher education	1%	19%	1%	22%
Postgraduate education system	1%	0%	0%	1%
Out-of-school education system	4%	1%	0%	4%
Material procurement programs	0%	0%	0%	0%
R&D in education	0%	1%	0%	1%
Other	2%	2%	0%	4%
	55%	37%	8%	100%

Source: Ukraine BOOST 2017 based on data from State Treasury Service

The increased spending on capital in 2016 is a welcome sign, following years of underinvestment. As outlined in discussion on the Ministry's planned reforms, more spending on refurbishing hub schools (including investing in better insulation and better heating systems) and purchasing busses will be needed to support school optimization. By reducing future recurrent expenditures (on staff and on heating), these investments are well worth making.

Figure 3: Public Spending on Education by Category, 2007–17, Percent of total

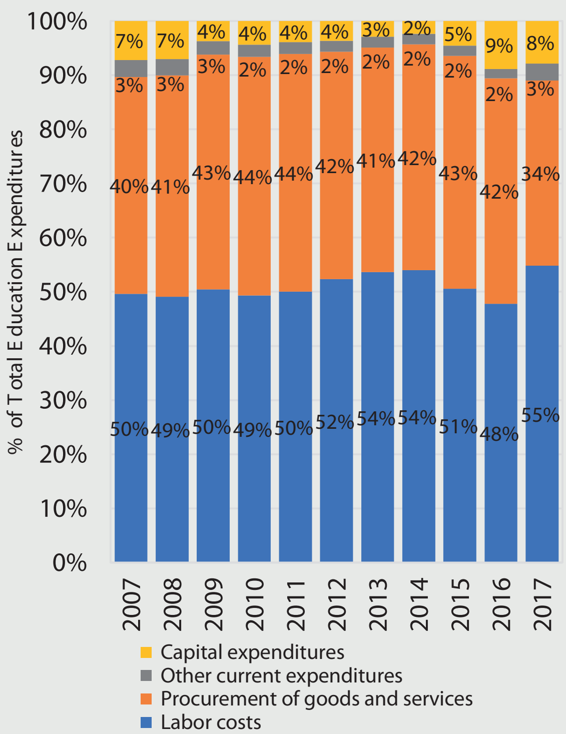
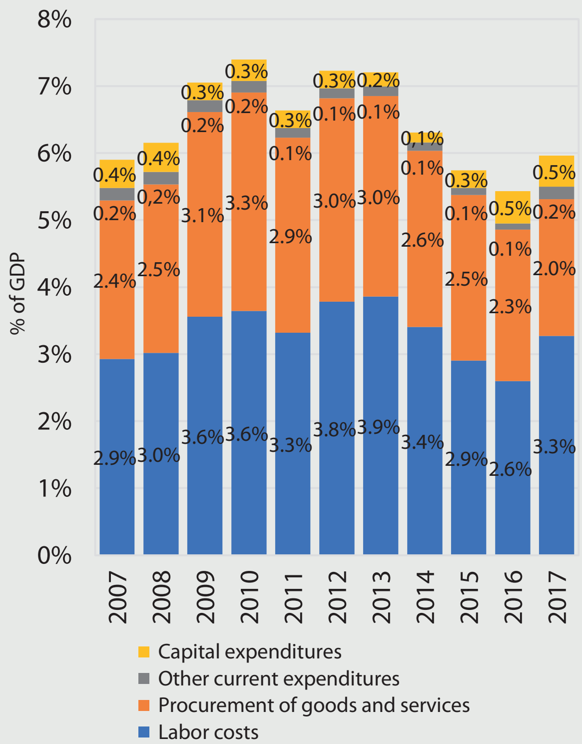


Figure 4: Public Spending on Education by Subsector, 2007–17, Percent of GDP

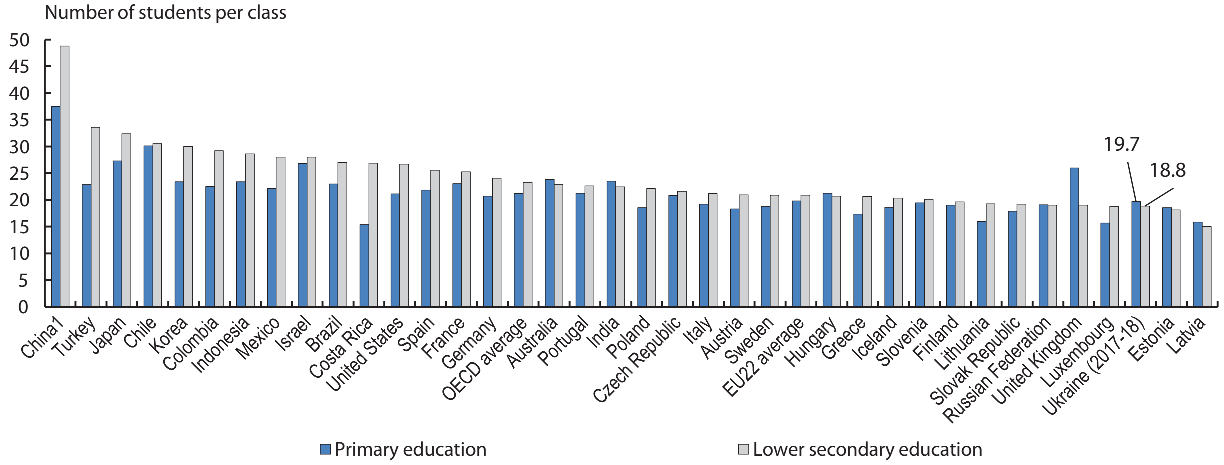


Source: Ukraine BOOST 2017 based on data from State Treasury Service

An oversized school network and the related overstaffing is the primary cause of high levels of spending[7]

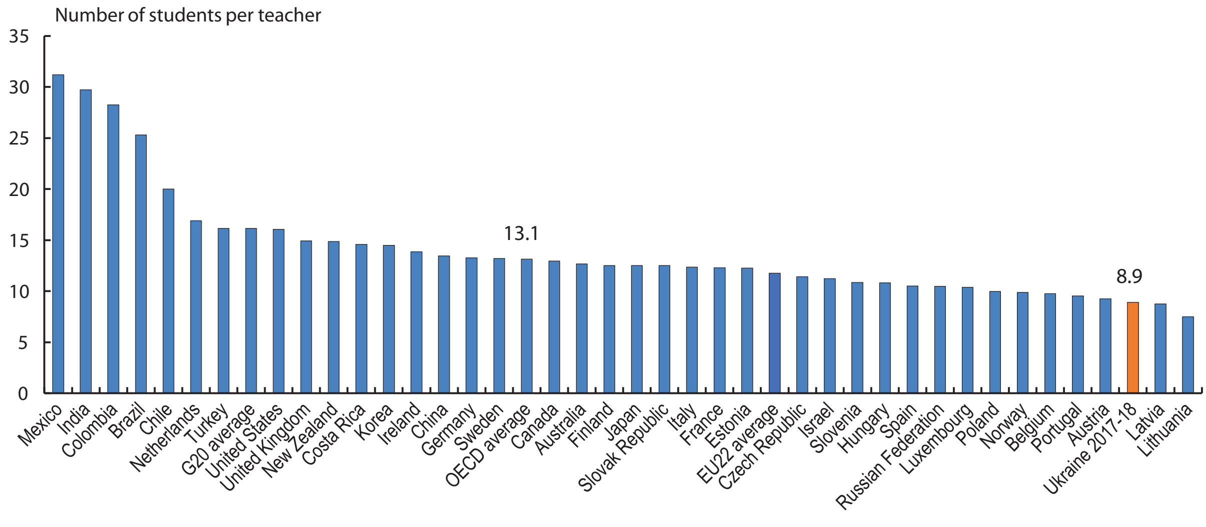
Exceptionally low average class sizes and student-teacher ratios are an important reason for Ukraine's high level of spending. With less than 20 students per class, Ukraine has one of the smallest average class sizes in the world (Figure 5). Maintaining these small classes, in turn, drives the large number of teachers and non-teaching staff working in the system: The education sector employs more than 1.4 million people, with half working in the general secondary schools (439 thousand pedagogical workers, and approximately 280 thousand non-pedagogical workers). With less than 4 million students in the system, this translate into only 9 students for every teacher, with only Latvia and Estonia reporting smaller ratios than this (Figure 6).

Figure 5: Average class size



Source: OECD (2017b), Table D3.2a. (www.oecd.org/education/education-at-a-glance-19991487.htm), except Ukraine (2017-18) which is from Institute of Educational Analytics school level data.

Figure 6: Student-teacher ratio in lower and upper secondary education (except for Ukraine where the figure includes primary education)



Source: OECD (2017b) (Table D2.3). Ukraine is from State Statistics Service's website and covers both primary and secondary education

[7] As noted in World Bank (2017): Ukraine Public Finance Review, Ukraine also spends an unusually large proportion of GDP on higher education (1.5 percent of GDP). Understanding the causes of this, and identifying options for generating savings in that sector will be the focus of future reviews.

The underlying cause of the low average class sizes (and student-teacher ratios) is the inability of the school system to adjust to falling student numbers. The number of students enrolled in school has declined due to demographic changes. Falling birth rates and outward migration have led to an aging of the population and a reduction in the number of school-aged children (Figure 7). Since gaining independence from the U.S.S.R., the number of Ukrainian children aged 5 to 19 years has declined by 41 percent—from 11 million in 1990 to 6.5 million in 2014.[8] The decline was observed in both urban and rural areas of the country. The urban population of 5- to 19-year-olds shrank by 45 percent, from 7.7 to 4.2 million, and the rural population by 33 percent, from 3.3 to 2.2 million. General education institutions, which enrolled 7.1 million students at the beginning of the 1990/91 school year, saw enrollments shrink by 41 percent by 2013/14, to 4.2 million. Over the same period, the number of general education schools declined by only 11 percent (Figure 8), and the number of teachers fell by 5 percent. Meanwhile, the nationwide average school size shrank from 327 to 218 students, and the student-teacher ratio dropped from 13.3 to 8.3. As Figure 6 shows, the student-teacher ratio has improved slightly to 8.9 in 2017/18.

[8] The latest year for which comparable figures are available. After 2014, most data exclude the Autonomous Republic of Crimea, the City of Sevastopol, and parts of the regions of Donetsk and Luhansk.

Figure 7: School-Age Population by Urban-Rural Location, 1990–2016, Millions

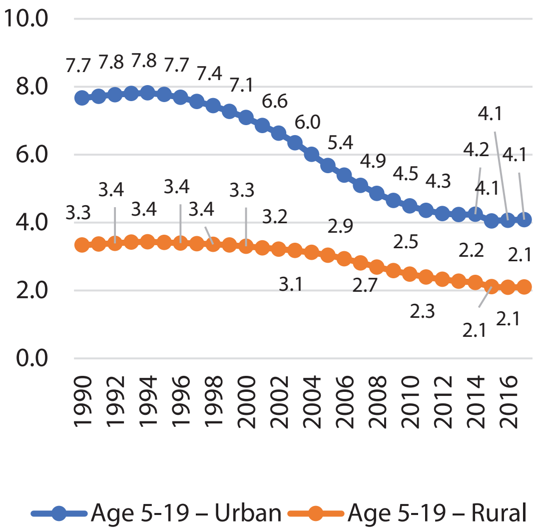
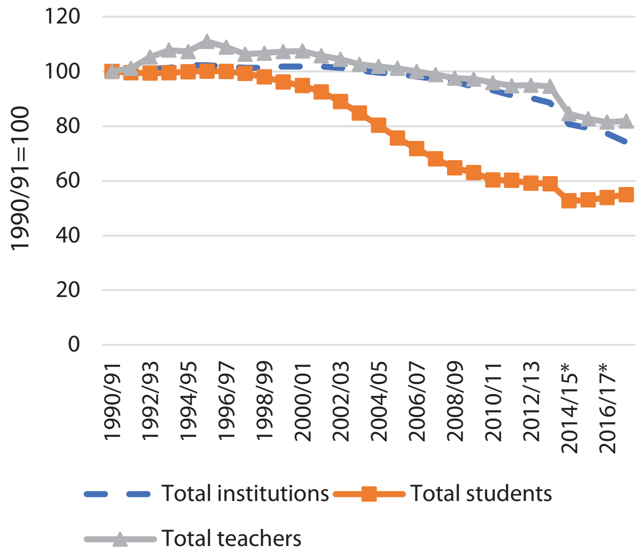


Figure 8: General Education Schools, 1990/91–2016/17



Several factors explain why the oversized network has not been downsized at the same rate as falling student numbers

There are multiple reasons why the school network has been slow to adjust. The principle reason, though, is that closing a school is a politically sensitive maneuver as communities are often vehemently opposed. Schools are seen as the heart of communities, and one of the few reliable providers of jobs. With population numbers falling and with few available jobs in rural areas, the closure of the school is widely seen – whether true or not – as potentially accelerating the demise of the community. Although not formally enshrined in political manifestos, there has been broad political agreement on the need to maintain public sector jobs, especially in rural areas; essentially, the education sector was tasked with providing lifelong livelihoods to those employed in schools, irrespective of the falling number of students.

Moreover, Herczyński (2017) provides two technical reasons why the downsizing moved at a very slow pace: (1) 25 years of confused responsibilities in the education sector and (2) poor sectoral and budget management. First, a very important example of this confusion was, until 2016, the legislative norms which delegated the recurrent financing of schools to *rayons*, but the decision to close the schools was taken by the villages. In other words, a village could vote to keep a school open without taking upon itself the responsibility to finance it. As a result, over two decades very few schools were actually closed, despite the serious decline in birth rates and massive migration to the cities. Second, on poor sectoral and budget management, Herczyński argues that local authorities and schools never faced "hard budget constraints" or any financial incentives to address the problem. Although a per student financing formula was introduced already in 2001, the formula included "adjustment factors" which meant that local authorities did not see declining resources as a result of falling student numbers (which you would normally expect with a per student formula). In fact, Voytov (2003) argues that the formula actually created incentives for local authorities to create *more,* not *fewer* classes in response to falling student numbers. In any case, what is clear is that local authorities and schools could maintain their schools, classes and teachers while enrollments declined without seeing a drop in their available resources.[9] A new funding formula, with better incentives to decrease the number of classes in response to falling student numbers was introduced in 2017.

To summarize, the sweeping reforms are being launched from a strong starting point in terms of learning outcomes, but with significant inequalities in outcomes as well as high levels of spending, stemming from, still, unaddressed inefficiencies. For general secondary education, those inefficiencies are represented by an oversized school network, with approximately twice the number of schools needed, and employing 20-30 percent too many individuals (as the analysis conducted for this note and discussed in more details below suggests).

[9] Voytov (2003) describes the weaknesses of the Soviet-style funding system.

The good news is that the costs of most of the reform initiatives are relatively modest, except for the promised salary increases

Despite the reforms' broad ambition, the costs of most of the reform initiatives are relatively modest, especially in the near term. During the first phase of implementation, annual costs of new reform initiatives are expected at around UAH 2 billion, equivalent to less than 0.05 percent of GDP per year. These costs include financing new textbooks (which will gradually be introduced for all grades), the development and rollout of a new digital e-learning platform, increased spending on teachers' professional development, and new equipment for schools (furniture, learning resources, and computers).

Further out in the future (by 2029), additional reform initiatives are likely to add more costs. In particular, a 12th year of schooling will be added in 2029. Adding an extra year would add an estimated 180,000 additional students to the general secondary system (by keeping 11th graders an extra year in school). In turn, this would result in an estimated 11,000 additional classes (a 5 percent increase), and require an additional 22,000 teachers. However, the Ministry hopes to find savings by simultaneously reducing the number of years required for a bachelor's education (from 4 to 3 years) and, thereby, reducing staffing numbers in higher education. Finally, the law also envisions introducing a voluntary mechanism to allow teachers to become certified, with mechanisms still to be established. On achieving such certification, teachers would receive a 20 percent salary increase. Moreover, as noted earlier, the Law on Education is a framework law, with the details – including on costs – only to be spelled out in subsequent legislation and decrees for subsectors. As such, these additional pieces of legislation could add further costs in the future.

However, by a wide margin, one feature of the new law is the most costly: the law's promise to raise the social status of teachers by raising the salary of the least paid teachers to four times the living wage by 2023. As mentioned above, one of the arguments for change presented by the MoES was that teachers have low social status, and that their low salaries demoralize them. As such, the law promises to increase the starting salary of teachers to four times the living wage by 2023. Already, starting on January 2018, teacher salaries rose, on average, by nearly 25 percent (relative to 2017). If the law's promise were to be introduced as an increase to the starting base salary of a teacher "and without a change to the various bonuses and top-ups which teachers receive (which, on average, amount to nearly half of a teachers' take-home pay, as discussed in more details in Box 3), this promise would increase spending on education to 8.8 percent of GDP (from 6 percent in 2017) (see Figure 7). Since all teachers – irrespective of which sub-sector of education they work in – are on the same wage grid, they would all benefit from that increase (again, please see Box 3 for details). Moreover, that grid also specifies how much more, say, the most senior teachers make as a ratio to the starting salary of a new teacher. As such, unless the grid is modified, the salary increase to a new teacher would be passed on to experienced teachers, resulting, in short, in a 199 percent (nominal) increase in salaries of teachers between 2017 and 2023 (whereas, expected nominal GDP per worker is only expected to increase by only 74 percent during this period). A teacher's salary would increase even further if he or she undergoes the envisioned voluntary certification process.

Box 3: How are teachers in Ukraine paid? The importance of distinguishing between base salary vs. take-home pay

According to Article 8 of the Law of Ukraine "On Labor Remuneration", conditions of payment of work of employees of institutions and organizations, financed from the budget, are determined by the Cabinet of Ministers of Ukraine. This implies that all individuals – teachers, principals, professors, as well as non-teaching staff – have their conditions of pay governed by the Cabinet of Ministers.

Specifically, the size of the employees' salaries is determined by a Resolution of the Cabinet of Ministers of Ukraine which specifies a "Uniform Scale of Wages and Salaries of Categories and Coefficients". This Uniform Scale specifies the types of jobs which exists in the sector, and how much is the **base salary** for each of these position (and, as result, how much wage progression there is across the different levels in the grid).[10] Importantly, the Uniform Scale determines the base salaries, not the take-home pay.

The take-home pay is a combination of the Cabinet resolution on base salary and a Ministerial Order on labor remuneration conditions. In particular, labor remuneration conditions of employees of general educational, preschool, out-of-school, vocational and higher educational institutions of I-II accreditation levels were specified in Order of the Ministry of Education of Ukraine No. 102 of 15 April 1993 (registered with the Ministry of Justice of Ukraine on 27 May 1993 under No. 56) as follows:

A salary of a pedagogical employee (teacher of a general educational institution) consists of

1. a position salary (the base salary set by the Cabinet resolution but defined by the Ministry Order to equal 18 hours of teaching (one "Stavka")),
2. A mark-up for years in service,
3. A mark-up for "occupation prestige",
4. perks (premiums, annual cash remuneration for conscientious work, money towards vacation, etc)

The Minister Order also specifies that teachers' work is paid in accordance with the work load, distributed by a manager of an educational institution in approval with a trade union committee at the beginning of the educational year. The scope of educational work that may be performed by a teacher is not limited by any ceiling, i.e., the load on a teacher may be either lower or higher than a salary rate. The salary rate is established presuming 18 teaching hours of work per week.

Currently, the Institute of Educational Analytics estimates the average take-home pay of a teacher of a general educational institution as follows:

1. a base salary (set to 18 hours of teaching/week). Importantly, a teacher can be paid a fraction of the "Stavka", depending on how many hours of teaching he/she has;
2. markup for years in service (more than 3 years - 10 percent, more than 10 years - 20 percent and upward of 20 years – 30 percent);
3. markup for occupational prestige (up 20 percent);
4. other extra pay and markups (class supervision, checking of notebooks, supervision and oversight of classrooms/workshops, etc., on average 22 percent);
5. remuneration for a bona fide work (1/12 part of the monthly average salary per each month of the year);
6. additional pay to cover part of annual vacation (1/12 part of the monthly average salary per each month of the year).

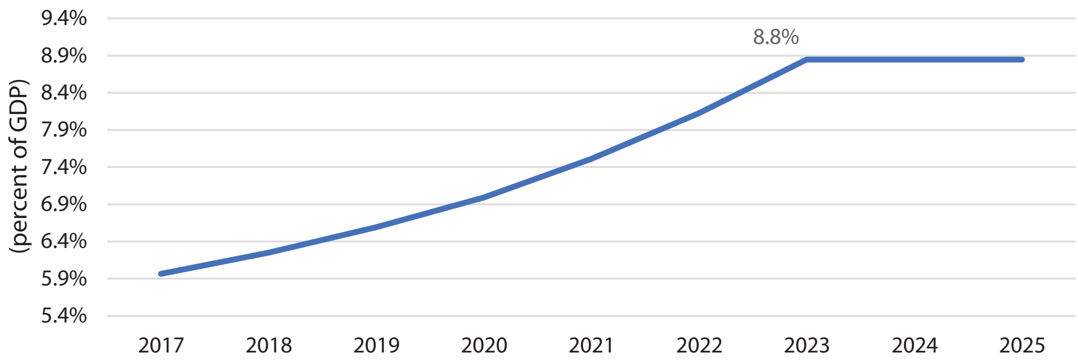
[10] Resolution of the Cabinet of Ministers of Ukraine No. 1298 of 30 August 2002 "On Employees' Labor Remuneration on the Basis of a Uniform Scale of Wages and Salaries of Categories and Coefficients for Employees of Institutions, Establishments and Organizations of Separate Budget Sectors" (as amended), in execution whereof the Ministry of Education and Science of Ukraine issued Order No. 557 of 26 September 2005 "On Streamlining Work Remuneration Conditions and Approval of Salary Levels for Employees of Educational Institutions, Educational Establishments and Scientific Institutions" (as amended).

Nearly half of a teachers' take-home pay can come from the various top-ups. As of January 2018, a person who has just joined the teaching profession is paid a base salary of UAH 3735 (provided they teach 18 hours per week). However, given the various top-ups above, the average take-home pay is UAH 6671. That is, the base salary is only 56 percent of a young teachers' take-home pay. And, for more senior teachers who benefit from a 30 percent mark-up for years in service, the base salary is only 53 percent of take-home pay. Annex table 1 provides more details.

Compared to other countries, the salary progression of a teachers' career in Ukraine is relatively small. In particular, teachers of the "highest rank" who are at the top of the salary scale only makes about 30 percent more than a new teacher. Compared to other countries, this is a relatively small premium for years of service and experience (see Annex figure 1).

Source: Email exchanges with officials from Institute of Educational Analytics

Figure 9: Forecast of education spending as a share of GDP, should the base salary of a new teacher be increased to 4 times the living wage by 2023[11]



Source: World Bank projections[12]

[11] Beyond 2023, it is assumed that wages grow at the rate of nominal GDP

[12] To model the implications of the planned salary increases as well as the fiscal implications of reducing staffing numbers, Microsoft Excel was used to create a model of education sector spending. The model (available upon request) models labor costs based on forecasts of salaries and staffing numbers whereas all other costs are assumed to grow at the rate of nominal GDP. Different scenarios are modelled, including: (i) increasing base salary of teachers with "no rank" to 4 times living wage (and, then, modelling the implication that this would have on take-home pay and the salary of teachers with more experience and seniority); (ii) increasing take-home pay of teachers with no rank to 4 times living wage; and (iii) conducting the same scenario as (ii) but with the number of staff being reduced (by approximately 100,000 in total).

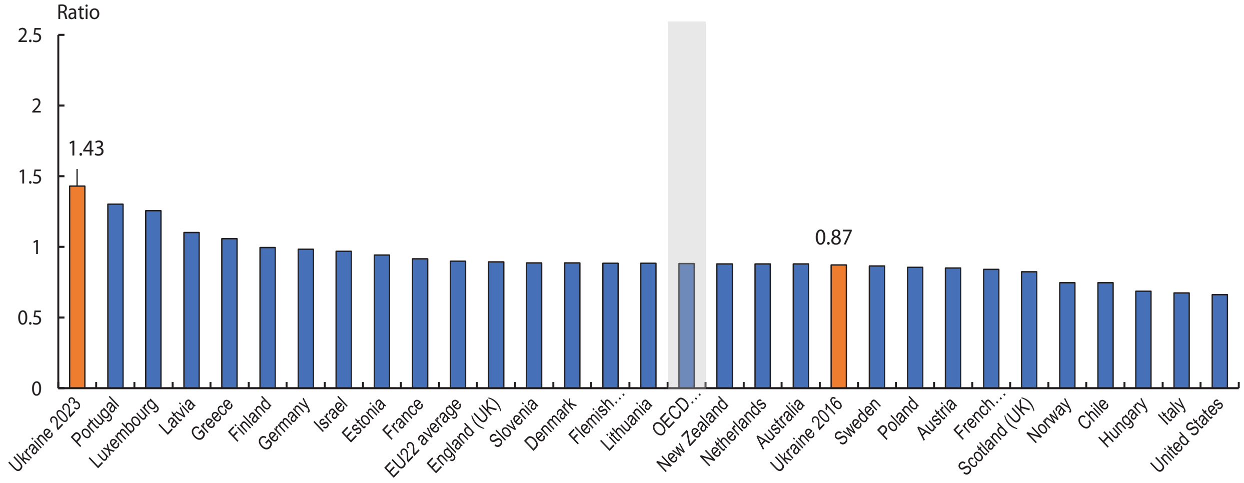
There are at least three reasons why implementing this large increase to teachers' salaries (between now and 2023) is not the best use of public resources. First, it would put the education sector on a fiscally unsustainable path, and limiting the resources available for other much-needed public services (such as health care, pensions, infrastructure etc.); at Ukraine's level of development, the country simply cannot afford wage increases of this magnitude. Second, it would make Ukraine an outlier from an international perspective. It is true that teachers are paid less than teachers in, say, Western Europe when comparing their salaries in EUR-terms (and, even when adjusted for differences in their purchasing power). However, this is not a particularly useful comparison, given that Western European countries have substantially higher overall salaries and higher levels of income (that allow them to afford higher public sector wages). The more meaningful comparison is a comparison of how much teachers in Ukraine are paid relative to individuals in Ukraine with the same level of experience and education. On this dimension, teachers' salaries in Ukraine resemble teachers' salaries in most other countries: teachers – who tend to work fewer hours and enjoy more job security – usually receive 10-30 percent less than peers with similar education and experience (Figure 10). However, if salaries were to be increased to four times the living wage, this would make Ukrainian teachers the highest paid relative to OECD countries (again, see Figure 10). Third, from an education perspective, the putative causal link between increasing teachers' salaries and improving education sector outcomes – including learning outcomes of students – is weak. One potential link between the two is that increasing salaries would make the teaching profession more attractive, attracting more talented individuals into the profession. There is some international evidence that show that higher salaries can attract higher quality public servants (e.g. see Dal Bó et al. 2013). However, in Ukraine's case, as discussed below, the sector needs to downsize the number of teachers (by an estimated 20-30 percent), implying that very few new teachers will need to be recruited in the coming years. As such, it would make better sense to only raise salaries after that downsizing had taken place. Another potential link is that higher pay would improve motivation and, then, performance. But, again, the experience from other countries does not provide much support for this link (see Evans (2018).

A separate but related potential link is that higher pay would discourage teachers from searching for additional paid work, such as fee-based private tutoring, which takes time and attention away from a teacher's primary task of classroom teaching.[13] Private supplementary tutoring is common in Ukraine, especially for teachers whose contracted workload is less than a full working week (UNESCO 2015; OECD 2017a). Because many teachers do not have full workloads and their non-teaching time is not regulated, they would still be permitted to use non-teaching time for private tutoring, which remains in high demand particularly in upper secondary education (OECD 2017a). Therefore, without changing teachers' workloads and the way that teachers are compensated, the incentives to participate in private tutoring would remain strong regardless of the proposed pay increase.

[13] Fee-based private supplementary tutoring – a practice which is not prohibited or regulated in Ukraine – creates problematic incentives for teachers, such as offering preferential treatment in class to students they tutor, or marking poorly those they do not tutor in order to create demand for tutoring services (OECD 2017a).

The more likely scenario is that a large increase in teacher salaries will have a range of unfortunate spill-over effects. As discussed above, there is little evidence to support the idea that a large increase in teacher salaries would improve education sector performance. What is more likely to happen is the following: (a) overall education sector spending will increase substantially, with the growing wage bill squeezing out other priority areas, including the resources needed to support the implementation of the New Ukrainian School and maintain quality during this transition; (b) the salary increase will make it more attractive for older teachers close to their retirement to stay in the system (to earn a few more years of salary at the much higher wage), leading to more pressures to maintain their jobs; (c) the salary increases would also lead to larger-than-necessary future pension expenditures because the teachers who benefit from the increases would retire with higher pensions; and (d) other public sector workers will start demanding increases on the same magnitude, leading to a broader fiscal crisis.

Figure 10: Teachers' salaries relative to earnings for tertiary-educated workers (2015, except for Ukraine 2016 and 2023 (forecast))[14]



Source: OECD, Education at a Glance 2017b, Table D3.2a (www.oecd.org/education/education-at-a-glance-19991487.htm) and World Bank staff calculations for Ukraine using Ukrainian Labor Force Survey and World Bank modelling exercise for 2023.

As will be presented below, there are several other possible options for improving teachers' status while implementing the promised salary increase in a fiscally responsible way.

[14] For Ukraine, the comparison is between teacher salaries and that of "GDP per working age individuals". For 2023, nominal GDP is forecast as in table 3 while the forecast for the number of working age individuals is taking from UN 2017 Revision of World Population Prospects.

What needs more focus in the reform agenda?

A. Ensure that the New Ukrainian School is implemented in a fiscally sustainable way

As outlined above, the passing of the new Law of Ukraine on Education lays the foundation for modernizing virtually all dimensions of primary and secondary education. Yet, as discussed above, the promised wage increases risk placing the entire reform agenda on a fiscally unsustainable path, with spending for other priority areas likely to be squeezed by rapidly rising labor costs.

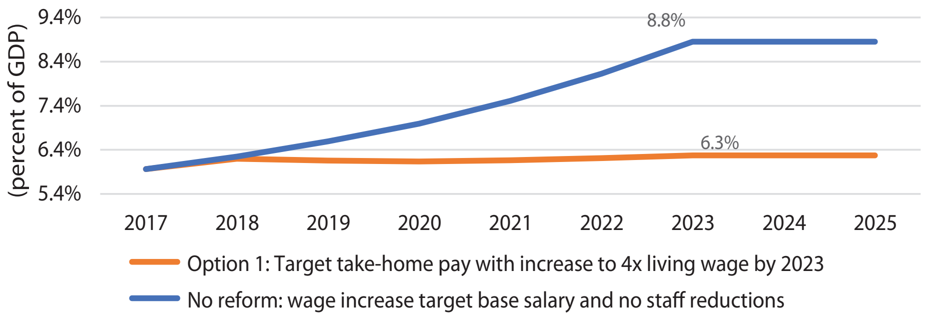
Below, options are presented, first, for how the wage increases could be implemented in a fiscally sustainable way. Second, the case for accelerating optimization is presented, with options on how this could be supported. Finally, some additional suggestions for strengthening the implementation on the new law are presented.

A.1 Target wage increases to teachers' take-home pay (not the base salary)

The language of the new law regarding the promised wage increase is somewhat vague, leaving room for different ways of implementing the increase. With the absence of a centralized payroll database for education staff, attempting to simulate the impact of the promised salary increases is fraught with difficulties (as discussed in Box 4). Notwithstanding these, simulations conducted for this note suggest that if the planned wage increase targets the least paid teachers' *take-home pay*, this would substantially lower the overall cost of such an increase while still increasing teachers' pay. If implemented this way, the promised wage increase would result in education spending rising to 6.3 percent of GDP by 2023 (instead of 8.8 percent of GDP if the base salary were to be targeted).

However, another option for containing salary increases would be to reduce staffing numbers (see discussion below).

Figure 11: Introducing planned salary increases in fiscally sustainable way: target take-home pay as opposed to base salary for increase to 4x living wage



Source: World Bank simulations

Box 4: Challenges to simulating the impact of promised salary increases and details on the assumptions made in this note's simulations

Simulating what the impact of the law's promised salary increase would be is fraught with difficulties, resulting in forecasts with a wide margin of uncertainty. Three main challenges are worth highlighting:

First, there are no centralized payroll information on the workers in the general secondary sector. As such, no information is available on where in the wage grid the current workforce resides, how experienced the workforce is, how many hours they work, and what current take-home salaries are. For the simulation presented here, school level data containing information on teacher rank (for all general secondary teachers, except those working in Kiev (which were missing from the database) were used.

Second, even less information is available about non-teaching staff, and education sector workers employed in preschools, technical and vocational schools, and in higher education. Yet, because all of these workers are on the same wage grid, the planned salary increases will affect all of them. As such, for the purposes of the simulations presented here, it was assumed that:

- Salaries of non-teaching staff would grow at the rate of nominal GDP;
- Salaries of all teaching staff (across all education subsectors) would grow at the same rate as salaries of teachers in general secondary education.

Third, the salary increase is made in terms of what will be the living wage by the year 2023. As such, that wage needs to be forecast. In the simulations presented here, it was assumed that the living wage grows at the rate of the minimum wage, as follows:

Table 3: Assumptions used regarding future growth of minimum, living wages and nominal GDP

	2017	2018	2019	2020	2021	2022	2023	2024	2025
Min wage	3,200	3,723	4,173	4,425	4,868	5,330	5,810	6,332	6,902
living wage	1,625	1,891	2,119	2,247	2,472	2,707	2,950	3,216	3,505
Nominal GDP (bn UAH)	2,983	3,428	3,854	4,314	4,745	5,196	5,664	6,174	6,729

Source: World Bank assumptions for simulations conducted in this policy note

Source: World Bank assumptions

More broadly, the salary increase presents an opportunity to re-design how teachers are remunerated using a model where the take-home pay is not compartmentalized into base-salary vs. various top-ups. This is a much-needed reform, which deserves further study to help guide the Ministry to making this change. In general, though, paying teachers for every task they do equates them with being simply "piece-meal" workers, like workers on a production line. A move to a full undifferentiated salary suggests the move to professionalism. A professional is expected to do many things differently, rather than counting every task.

A.2. Link salary increases to progress on reducing staffing

Accommodating regular salary increases of teachers is important to maintain teachers' social status and help attract high caliber individuals into the profession. However, promising to increase salaries by nearly 200 percent for an oversized workforce (in relation to a declining student population) can undermine the fiscal sustainability of the sector. From an education perspective, scarce public resources would be better used by investing in a better-paid smaller workforce, which is well-prepared and adequately supported (through enhanced pre-service training and in-service training, better school leadership, and a modern learning environment).

Moving forward, education sector budgets could be linked to a multi-year plan where staffing is on a downward trajectory (with an approximately 20 percent reduction as the medium-term target) and **where salary in-**

creases -- that exceed the rate of inflation – are only accommodated when the agreed staff reduction targets are met. The following policy recommendation presents some concrete proposals on how Ukrainian authorities can accelerate the ongoing school optimization reforms that should reduce the number of schools, classes, and staffing levels in the education system to sustainable levels.

A.3. Accelerate school network optimization through the full use of the policy tools available

For a number of reasons, more efforts should be mobilized to accelerate school network optimization as an integral part of the ongoing reforms. First, downsizing staffing numbers would make it less costly to implement the promised salary increase. Second, the reforms envision training all teachers in a new curriculum and with more focus on student-centered pedagogical practices. The logistics and costs involved in re-training and providing mentoring and support to a reduced workforce of, say, an estimated 350-380,000 teachers would be substantially smaller than trying to do the same for 440,000. Third, it should be clear that trying to raise the quality of the facilities, learning materials and the leadership of the existing 15,000+ schools – instead of, say, 8,000 schools is comparably difficult. The investments in materials and school leadership for "New Ukrainian Schools" should take place in, the approximately 8,000 schools that are needed for today's and tomorrow's school population; investing in 15,000 dilapidated buildings which are not needed is a waste of resources. Fourth, one quarter of Ukraine's teachers are 55 years of age, or older, implying that they will retire in the coming years. Unless schools are merged and the number of classes reduced, these teachers will be replaced by new teachers who will be teaching the same small classes. The aging workforce represents an opportunity to downsize through attrition, without having to lay off workers (with the emotional and fiscal costs this entails).

The most important reason for placing more emphasis on school optimization, though, is that students enrolled in Ukraine's small schools are being left behind. Indeed, if there are two features that characterize students attending smaller schools in Ukraine it is this: first, on average, they have the worst examination scores on the school leaving examination (e.g. see Coupe et al 2011). Second, their schools have the lowest teacher-class ratio, a likely tell-tale sign of the difficulties principals and school administrators have in attracting and retaining staff in such places. In Annex 3, Lathapipat shows that the problem is that small schools struggle to attract enough (good) teachers, as reflected in their lower teacher-class ratios (again, please see Annex 3). To provide these children with a quality education, they need to benefit from the same quality of teachers, school leadership and facilities which children enrolled in Ukraine's larger schools enjoy.

The implication for policy makers is an important one: the inability to downsize the network has resulted in a large number of schools that are struggling to attract and retain quality staff. And students in these schools end up learning less than they would have learnt in a better-resourced school.

Detailed analysis conducted as part of this note suggests that the downsizing could be done without impairing access to and quality of education, and largely through teachers' attrition – through retirement as opposed to being laid off. This analysis and its findings are described in more details in Annex 4. First, in terms of access, detailed school mapping analysis conducted using GoogleMap travel distances and school GPS locations, suggest that 66 percent of Ukraine's 15,000 schools are located within 4 km distance of another school (see example below). More astonishingly, 24 percent are located within 1 km (a short walking distance) of another school (Figure 12). For schools located within such close proximity to one another, classes could relatively easily be merged (and staff reduced), without waiting for roads to be improved and busses procured. The analysis suggests that as many as 10 percent of all classes could be eliminated (and 10 percent of the teacher workforce reduced) by creating networks of schools located within close proximity (i.e. maximum 30 minutes' walk for primary students and maximum 60 minutes' walk for secondary students) (see Brandt and Sohnesen 2018).[15] And, where the current quality of

[15] Importantly, these are maximum walking distances and, hence, not the walking distances that every student in Ukraine would have to work to reach his/her school. These maximum values were included for the purposes of simulating the full scope for optimizing. Given the small distances between schools (as shows in Figure 12), there is wide scope for consolidating schools where children would walk much shorter distances. Annex 4 includes robustness simulations that show what the scope for optimizing would be for shorter distances.

roads allows for it – and/or, where, over time, the roads could be improved to allow for it – busses could be procured to allow for even further consolidation (allowing as many as 24 percent of classes (and teachers) to be reduced). Second, in terms of the quality of education: the analysis presented in Annex 3 shows that students from smaller schools -- where teacher-classes are smaller -- underperform their peers from larger schools, even after controlling for other factors that could influence outcomes. The analysis suggests that test score results could potentially be increased by consolidating the school network and, in the process, creating larger classes in better resourced schools. In addition, given that 12 percent of the teacher workforce is above 60 years of age (and 13 percent is above 55), if implemented over a 5-year period, staff reductions could be done by letting teachers retire when they reach retirement age.[16]

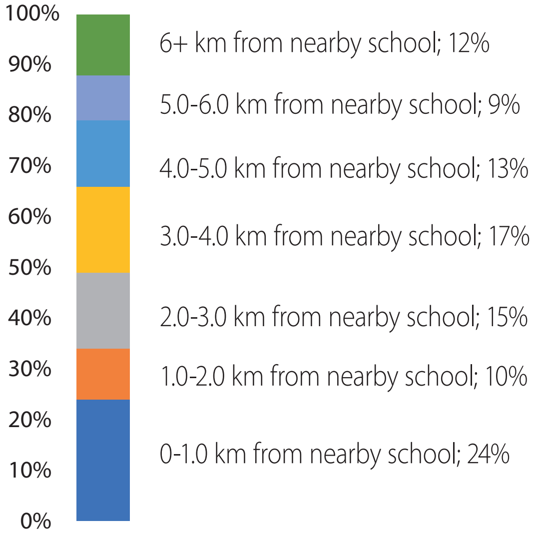
[16]These numbers are from World Bank calculations based on IEA's 2017/18 school level data. However, in that data set, data on teachers in Kiev are missing. In email exchanges with MoES, the number of retired teachers were reported as high as 16 percent of teachers.

Example: In the town of Dolynska, four schools offering primary classes are located within walking distance, and an additional two are located within a short driving distance. Using Google maps and school GPS coordinates, similar analysis was done to identify the scope for school optimization throughout all of Ukraine (see annex 4 for details)

Source: Brandt and Sohnesen (2018) and included in annex 4.

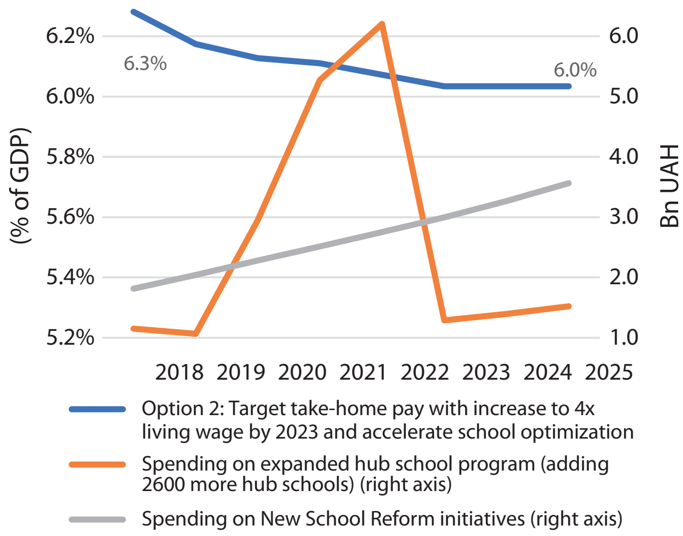
Accelerating school optimization would entail additional costs but many of these additional costs are temporary (e.g. refurbishing schools, or paying severance payments) and they would more than pay for themselves by generating substantial (and permanent) savings. In particular, receiving schools would need to be refurbished, and made more attractive for parents to want to send their children to them. Moreover, in some instances, staff who are not near retirement age would need to be made redundant, resulting in a need for severance payments. Moreover, busses would need to be procured or transportation costs for students covered in other ways. Finally, resources would need to be invested in launching a communication campaign to help explain that optimization is done to ensure that every child benefits from a quality education. Such campaigns will be critically important to counter the opposition that affected communities will likely mobilize (again, fearing that the loss of their school will mean the end of their community, loss of jobs, etc.). Still, even when accounting for such increased spending, the overall savings would be substantial. For instance, an estimated 100,000 staff could be made redundant (60,000 teachers and 40,000 non-teaching staff), reducing overall spending at 6.0 percent of GDP, even while accommodating the promised salary increase (see Figure 11 and Annex table 2)).

Figure 12: Only a small proportion of Ukraine's schools are "isolated"



Source: Analysis conducted by Brandt and Sohnesen in Annex 4

Figure 13: Increasing wages but also spending to support school optimization



Source: World Bank simulations

Below are some options for how the government could accelerate this process.

Policy tool 1: Scale up the hub school program, from 499 hubs (in 2018) to more than 3,100 by 2022

Introduced in 2016 to "provide conditions for equal access to quality education", the School Hub program holds great promise as the main vehicle with which MOES can accelerate school optimization in the coming years. It empowers local authorities to come up with their own plans to (a) identify a hub school and (at least) three schools that will, then, become affiliated schools (subordinated to the hub); and (b) identify their needs to facilitate this re-organization (e.g. busses etc.) (see Box 5 for more details on hub schools).

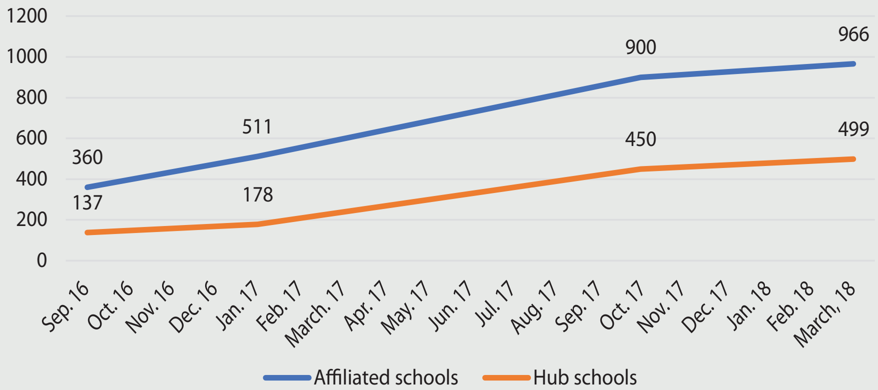
Box 5: Details on the hub schools and some lessons learnt during early years of implementation

The hub schools initiative launched in early 2016 is a major stride toward rationalizing Ukraine's network of general education schools. The MoES concept presents a vision for school consolidation that aims to achieve four main objectives: (1) providing conditions for equal access to quality education; (2) improving the quality of education; (3) encouraging the efficient use of available resources; and (4) enhancing the capacity of local authorities.

Local communities founding such institutions select the prospective hub facilities among their general secondary education institutions on a competitive basis (taking into account several factors aimed at identifying the most suitable candidate for a hub in the community).

Since its launch, the number of hub schools – and their affiliated schools – have grown steadily, upgrading schools to more modern facilities and allowing for classes to be made redundant. By 2018, 499 hubs and 966 affiliated had been created (see Figure 14). These hubs benefitted for investments in their buildings and available learning materials, and allowed these schools to become "magnets" for children enrolled in nearby smaller, and (often) poorer quality schools. As of late 2017, more than 200 classes have been made redundant as a result of the creation of hub schools. Moreover, the investments allowed the hubs to accommodate children with disabilities.

Figure 14: Evolution in the number of hub schools and affiliated schools



Source: Institute of Educational Analytics 2018

To learn from its implementation, the Institute of Educational Analytics has been carefully monitoring the hub school program, including commissioned a survey of school principals. Continuing such careful monitoring and regular surveys will be critically important to learn from implementation challenges, and allow the program to be scaled up.

Early lessons suggest four main implementation challenges which will need to be addressed moving forward:

1. A new interpretation is needed on the state's guarantee to provide education "which is the most accessible and is in the closest proximity to person's place of residence" (enshrined in the Law of Ukraine "On Education", Section II, Art. 13). That new interpretation needs to emphasize the importance of having a quality education provided nearby, not the presence of a crumbling school building. Moreover, that interpretation needs to emphasize that schools located within 4 km (as is the case of 66 percent of Ukrainian schools) is, still, both accessible and neaby (especially if accompanied with new school busses).
2. Being adequately prepared for staff redundancies, both for staff at or near retirement age, as well as younger teachers (who may need to be re-trained).
3. Resistance to the school optimization process by local communities. In a survey of school directors, 42.6 percent of them reported facing resistance during organization of hub schools. Providing training to directors and equipping them with the tools to effectively engage local communities will be critically important moving forward.
4. The requirements of having a minimum of three affiliated schools for each hub have gradually been watered down, limited the scope for creating efficiency savings. In short, communities have been interested in having the investments in a hub school but not all communities have delivered on creating schools whose status would be turned into "affiliated school". For hub schools to be an effective vehicle for optimizing the school network, enforcing this requirement will be key.

Source: World Bank staff based on Institute of Educational Analytics 2018 and MoES 2018

The Ministry might also explore presenting communities with a broader range of options for how, in addition to its function as a school, the community could expand its use: for instance, the building could serve as a school in the day, and a community center on the weekend. Similarly, with increased autonomy provided to the school, its school theater, or sports facilities could be rented by the local population for events. In other words, the Ministry could explore how to present to the community that it is getting more than just a renovated school in the new Hub.

However, to make a dent in the size of the problem, the program needs to be increased substantially from 499 hubs (and 966 affiliated schools) to at least 3,100 hubs and 6,500 affiliated schools. Again, as suggested by the school mapping analysis conducted as part of this note, at least 8,100 schools are located within close proximity to one another, presenting opportunities to have 2,600 of them identified as hubs and 5,500 become affiliated schools quite easily. The Ministry's role should be to provide financing to communities and nudge, incentivize and coach them to take action (see discussion below on making better use of the power of the purse). But a strong feature of the program should be kept: namely, having communities use their knowledge of local circumstances to make the ultimate decisions about how best to organize education in their community. However, experience from elsewhere suggests that they will need both support, convincing and strong incentives to agree to consolidate their schools. Below are ideas for how to provide that support, the convincing and the incentives. **In particular, the following policy tools are ideas for actions which the Ministry could take to support the scale up of the hub school program:** using the power of the purse more actively; improving the legal and regulatory framework to be more accommodating of efforts to optimize; and persuading local actors to take action, and monitor quality and access.

Policy tool 2: Use the power of the purse more actively

Mobilize financing to hubs and require that funds be disbursed against important optimization milestones only when those milestones are met.

Creating hubs involves additional costs to refurbish the hubs. This includes making them capable of catering to a broader set of learners (including students with disabilities or special educational needs) as well as improving their learning environment through investments in better libraries, laboratories, and information and communication technologies (ICTs). By mobilizing and making available such resources to local authorities, MoES and MOF can send a strong signal that they are expecting action, and that they are willing to help address local constraints.

However, for these additional resources to support optimization, MoES and MOF need to remain firm on one point: resources will only be released conditional on important milestones having been met (along the way towards consolidation). For this to work, careful attention will need to be paid to defining those milestones. Experience from the Hub School program already suggests that local authorities are eager to receive money to refurbish one of their schools but will try to water down the requirement that other schools be either made into branches and/or closed (see Box 5).

MoES can also set aside the resources to launch a public communication campaign making the case for fewer but better resourced schools (e.g. "My village's future rests on better education children, not a school building"). This communications campaign could be situated within the wider vision for the New Ukrainian School, including efforts to connect use schools to a digital learning platform granting students access to modern digital textbooks and learning resources. Such a communications campaign could help parents and communities to visualize the improved learning environment that their students will experience in optimized schools.

Another important consideration is to ensure that money is set aside to pay for staff who lose their job as a result of a school closure. To facilitate school consolidation, support may need to be offered to all teachers losing work. The inability to pay such "severance payments" could become a constraint (or an excuse) for local authorities not to proceed with school closures. To ensure that this does not happen, such compensation might be supported through a special grant (subvention) from the national budget to local budgets (Herczyński 2017).

Actively use the per student financing formula to encourage authorities to optimize

The introduction of a per-student financing formula (in 2017) marked an important step towards addressing a root cause of the inability to downsize the network (namely that money used to be tied to inputs, not students). With money now tied to the number of students enrolled, local authorities will start questioning whether they need all their schools, classes and teachers – or whether education could be organized more efficiently. The details of the new formula is described in Herczyński (2018).

However, experience from other countries in ECA who have introduced per-student financing suggests that, by itself, this sort of financing is unlikely to be sufficient to drive consolidation. More particularly, MoES and MOF will need to monitor on a regular, perhaps annual basis to determine whether the formula is putting the right amount of financial pressure on the locations where savings can be found. If not, the formula will need to be tweaked to put more pressure on such locations. MOES counterparts report that they are planning on making further improvements for the 2019 formula to further incentivize local authorities to optimize.

One reason why the per student formula may not work by itself is that, since 2015, local authorities have the authority to use their own resources to make up for short-falls in the education subvention they receive from the central government. As such, in the face of pressures from local communities and teachers, local authorities may simply opt to use their own resources to maintain their oversized school networks.[17]

[17] The mechanism for providing an educational subvention from the state budget to local budgets, approved by the Cabinet of Ministers of Ukraine No. 6 dated January 14, 2015, allows to finance expenditures for the salaries of teaching staff, simultaneously from state and local budgets and / or simultaneously from different local budgets, which allows founders of educational institutions and, if necessary, to co-finance expenditures for the payment of teachers from local budgets.

Enforce hard budget constraints on local governments

A per student formula can help incentivize local authorities and schools to downsize their network in response to falling student numbers but only if backed up by hard budget constraints. According to Herczyński (2017), a perennial problem is that local authorities can obtain mid-year increases to their budgets by claiming that their initial budgets are insufficient to cover their needs. For the per student formula to work, it is critically important that these mid-year demands are not met.

Policy tool 3: Improve the legal and regulatory environment to be more accommodative of optimization efforts

While the hub school program is likely to be the main "vehicle" to drive optimization, MoES leadership in identifying regulations that stand in the way of more efficient resource use will be equally important. For instance, modifications to existing regulations would include the following:

i. Eliminate the norms specifying the required number of non-teaching staff; schools, with local knowledge, are much better placed to determine how many security guards, cafeteria workers, and coat attendants are needed (and can be afforded).

ii. Introduce term contracts for teachers to specify what happens at retirement age. In Moldova the introduction of such contracts provided school leaders with an opportunity and an instrument to have a conversation with their teachers reaching retirement age about their departure from the system. This idea is currently in the draft of the new Law of Ukraine "On Comprehensive Secondary Education" (which is being prepared by the Ministry of Education and Science). That draft provides for the introduction of fixed-term employment contracts for a term not exceeding one year and possible annual extension of up to 1 year for teachers who have reached retirement age. If implemented into law, such annual contracts could become an important instrument for gradually reducing the number of teaching staff while reducing the number of educational institutions.

iii. Revise the State Sanitary Rules providing details on how close schools need to be located to provide more flexibility and have these norms emphasize the importance of having a quality learning environment provided (as opposed to a nearby school building).[18]

iv. Revise the norms governing the use of the premises of educational institutions to allow for more flexibility in the use of those premises.[19]

[18] The State Sanitary Rules and norms of placement (DSanPiN 5.5.2.008-01, approved by the Decree of the Chief State Sanitary Doctor of Ukraine No. 63 dated August 14, 2001) specify the following: (a) the radius of service from the place of residence to the general educational institution should be no more than 0.5 km of pedestrian accessibility; (b) In rural areas, the placement of schools should not be further than 2 km away and involve no more than 15 minutes of transportation (one way) of students.

[19] Currently, the norms of the Law of Ukraine "On Education" prohibit the use of the premises of educational institutions for non-educational purposes (Part 4 of Article 80).

Policy tool 4: Persuade local actors to take action, and monitor quality and access

Although local authorities will take the lead on the ground[20], MoES' leadership will be critically important to persuade local actors, and to monitor implementation progress, and key outcomes such as drop out rates and learning outcomes. In more details, MOES could:

1) Use school mapping analysis (such as the one done for this note) to suggest to local authorities some potential re-configurations of their network;
2) Ask that local authorities prepare hub school plans – and engage in critical discussions with them if their proposals fall substantially short of what is suggested by the school mapping exercise;
3) Incentivize local authorities to take action early by providing additional resources to early movers;
4) Tweak the per student formula to provide incentives to creating hub schools and closing down affiliated schools.
5) Put in place a task-force to be on stand-by to provide technical assistance to local authorities (e.g. on how to organize community discussions, conduct school mapping exercise, etc.);
6) Put in place a "communication strategy" to explain the need to consolidate (see Kotler and Lee 2018). Such strategy would likely benefit from a study of the political economy considerations that school optimization would need to take into account.
7) Carefully monitor implementation and the potential outcomes that could be affected by consolidation (drop out rates, learning outcomes etc.).

[20] Having local authorities take the lead is essential for at least two reasons: first, they have local knowledge of their schools, roads and communities. Second, by law, (1) they are responsible for ensuring the availability of education, planning and ensuring the development of a network of educational institutions; and (2) they have the right to establish educational institutions, as well as reorganize and liquidate them (Article 66 of the Law "On Education").

B. Introduce teacher certification but with careful thought to implementation and within broader changes to teacher professional development

As part of the MoES' efforts to strengthen quality assurance, it is introducing voluntary certification of teachers (Part 3 of Article 41 of the Law "On Education"), the introduction of which is scheduled for 2019-2020. The Ministry is currently developing a model for certification of teachers in the form of an external assessment of the professional competencies of the teacher through independent testing, self-assessment and study of practical work experience. Voluntary certification is intended to become a tool for material motivation (in accordance with Article 4, Article 61 of the Law "On Education"), and professional growth. Below are some suggestions on how to best introduce teacher certification.

Focus efforts on existing teachers. Teacher licensing (or teacher certification) is designed to raise the quality of those entering -- or already in -- the teaching profession and to maintain, manage, and continually update that quality throughout the teacher's working life. For the moment, while the MOES is concerned with both pre-service and in-service, the latter, the licensing of the stock of existing teachers so that they are in line with the NUS reform, may be more important since few new teachers are needed and relatively few are in fact entering the system.

Introduce certification as part of a systemic strategy to improve teacher performance. Licensing or re-licensing of active teachers is introduced into systems when a quantitative 'jump up' in quality is desired and/or a new status is required. Ideally, this recertification should not be a 'stand-alone' option; a simple training program for upgrading skills. Rather, it should be part of a systemic strategy that takes into consideration the whole continuum of teacher education and development, which includes high quality pre-service teacher education, induction, structured but flexible Continuous Professional Development (CPD), a motivational career path which links CPD to promotion. Teacher licensing underlies all these steps by regulating, controlling, and safeguarding teacher quality. Accreditation of training institutions is also a key component of a modern teacher licensing system. Finally, underlying the entire system should be clear, recognized, and agreed upon standards and competences for all teachers. These take time to develop.

The voluntary approach is a sensible one but it needs to be carefully monitored. Now the approach to licensing that will be adopted in the NUS reform, where teachers volunteer to enter the licensing process, while offering certain advantages (i.e. that only teachers truly interested and motivated to license, do in fact participate), may pose certain challenges that need to be well managed. First, there is the question of harmonizing the status of these newly licensed teachers with their unlicensed colleagues. Secondly, and more importantly, is the question of how to encourage and sustain the licensing movement so that more and more existing teachers do in fact become licensed. For what is needed is a critical mass of these licensed teachers in the system, so that they become the prevailing norm, the new standard of teacher, within the system.

Careful thought need to be put into timing. The question then, really, is one of timing: can the timeline required for the full certification of all teachers be managed? This is where a more comprehensive approach to the NUS reform might be very helpful, one in which relicensing is coordinated with the attrition of teachers and the creation of hub schools. As schools are consolidated into hubs, and as teachers exit the system through attrition, the reform might be able to leverage the fact that the drop in the existing numbers of teachers – coupled with fewer teaching posts and the new possibility of not being hired by school leaders – might nudge existing teachers to want to relicense to give themselves a better chance of being rehired/remaining in the system.

C. Implement the Digital Platform and e-Textbooks in a manner that allows for experimentation, and consider using their roll out to incentivize local authorites to creating hub schools

As part of the implementation of the New Ukrainian School model, the MOES is pursuing a two-pronged approach to digitalization in secondary education: (i) the development of contemporary and interactive electronic textbooks linked to the New Ukrainian School curricula, starting initially with Grade 1; and (ii) the development of a national Digital Platform to foster the ecosystem of educational services and serve as a marketplace and repository for digital content. The MOES has allocated around 100 million UAH in 2018 for the development of both e-textbook prototypes and to begin developing the national Digital Platform.

Given the potentially high costs associated with procuring and distributing ICT equipment and developing digital content, it is important that the MoES test different models, learn from what works and what does not work in different contexts, and adjust the program along the way as it increases in scale. Ukraine's vision for digitalization in the education system is ambitious, and it will take time to pilot, evaluate, and implement. Importantly, the MOES has already initiated a pilot project in 100 schools which will expand over the coming years. It is crucial that this pilot program experiment with different technologies and usage models to determine which combinations are most effective for supporting teaching and learning. This will help to improve cost efficiency of the program while also ensuring that digitalization efforts improve interactions between learners, teachers, and content.

It is also crucial that any efforts at digitalization in schools involve school directors and teachers from the beginning, and be prepared to support capacity building at the school level. E-textbooks and digital learning resources can enhance good teaching in conducive environments with well-trained teachers, but they cannot replace good teaching. Teachers' knowledge and skills with operating ICTs and, more importantly, using ICTs and digital content to enhance and supplement instruction is critical to the success of this initiative, and it should be monitored closely through the pilot.

Consider using expanding access to e-textbooks and particularly the digital platform as an incentive to schools and local authorities to consolidating schools into hubs. Given that this system requires schools to have access to modern ICT equipment (e.g. laptops, tablets) as well as internet connectivity, any investments in such equipment should take place only in those schools that are needed for today's and tomorrow's school population. Furthermore, the envisioned Digital Platform could actually help to facilitate communication and collaboration within the schools in a hub or across hubs, since the Platform would contain tools to support, for example, teacher collaboration on lesson planning and student-teacher dialogue on learning content. In this context, access to ICT equipment, internet connectivity, e-textbooks, and the Digital Platform could be an attractive incentive to stimulate local authorities to take action towards school consolidation.

D. Place more emphasis on measuring learning outcomes and on improving other important education data, and using them to make decisions

Place more emphasis on *measuring* learning to sustain reform momentum and ensure that reform initiatives are being steered by efforts to address gaps in learning. Currently, the only available standardized information about students' achievements of learning outcomes come from the external independent assessment (ZNO, its Ukrainian acronym) carried out by the end of completing general secondary education. The absence of data on learning outcomes for primary and lower secondary means that potential gaps in learning – between students of different backgrounds, or between different types of schools – go unnoticed. The Ministry recognizes this gap and has laid out plans for participating in PISA 2018 and establishing a new system for monitoring the quality of education, starting with a focus on measuring reading and math skills of primary students. Accelerating this work would entail:

a. Ensuring that capacity is built in the newly-formed State Service for Quality of Education, so that it quickly becomes operational;
b. Introducing and rolling out new standardized assessments at the end of primary and lower secondary;
c. Making public such information, and actively using it to guide whether reform initiatives are having an impact. For instance, such assessments could be used to track whether children involved in a school closure do better at their new school than similar children who are left behind in a small school.

To support evidence-based decision making – including providing more reliable estimates of the fiscal impact on some of the proposed changes – better data are needed. Without a database on staff, simulation the fiscal impact of wage increases is subject to too many assumptions (e.g. as Box 4 summarizes). Moreover, with a database of students, it is difficult to accurately monitor drop-out rates, and repetition rates. Such data are especially important to have when having to make big re-organizations of the school network (since many students will be moving from one school to another).

E. Conduct detailed analysis of other sub-sectors of the education sector (incl. pre-school, higher education and technical and vocational sector) with the aim to identify the scope of efficiency savings in those other subsectors

While this policy note focused on the general secondary sector, similar public expenditure work could shed light on the scope for finding efficiency savings in other sub-sectors of the sector, including the preschool, higher education and technical and vocational subsectors. Given that these subsectors suffered from similar underlying weaknesses in the way they have been managed, very likely, savings of similar magnitudes could be found in these sectors.

References

Brandt, Kasper and Thomas Sohnesen. 2018. Estimating teacher redundancies with hierarchical clustering based on travel distances in Ukraine, World Bank mimeo (included as Annex 4)

Coupé, Tom, Anna Olefir, and Juan Diego Alonso. 2016. "Class Size, School Size and the Size of the School Network." Education Economics 24 (3): 329–51. http://dx.doi.org/10.1080/09645292.2015.1015405.

Dal Bo, Ernesto, Frederico Finan, and Martin A. Rossi. 2013: "Strengthening State Capabilities: the role of Financial Incentives in the Call to Public Service", The Quarterly Journal of Economics (2013), 1169–1218. doi:10.1093/qje/qjt008

Del Carpio, Ximena, Olga Kupets, Noël Muller, and Anna Olefir.2017. Skills for a Modern Ukraine • http://dx.doi.org/10.1596/978-1-4648-0890-6

Evans, David. 2018. "What do we learn from increasing teacher salaries in Indonesia? More than the students did", World Bank Blog, https://blogs.worldbank.org/impactevaluations/what-do-we-learn-increasing-teacher-salaries-indonesia-more-students-did

Herczyński, Jan. 2017. "Education Finance in Ukraine: Select Strategic Issues", Institute of Education Research, Warsaw, Poland

Herczyński, Jan. 2018. "Description of Allocation Formula for Education Subvention in 2018", Institute of Education Research, Warsaw, Poland

Institute of Educational Analytics 2018. "Development of hub schools in the context of the general secondary institutions optimization" (presentation)

Lathapipat, Dilaka. 2018. Using School Level Educational Production Function to Determine the Effects of Measured School Characteristics on Average Student Performance for Ukrainian Secondary Schools, See Annex 3

Lee, Nancy R. and Philip Kotler. 2018. Social Marketing: Changing Behaviors for Good, Fifth Edition

Ministry of Education and Science of Ukraine. 2016. The New Ukrainian School. Conceptual Principles of Secondary School Reform

Ministry of Education and Science of Ukraine. 2016. "Hub Schools" Presentation (in Ukrainian).

Ministry of Education and Science of Ukraine. 2017. Action Plan For implementation of state policy towards reforming general secondary education for 2017-2020 – New Ukrainian School

Ministry of Education and Science of Ukraine. 2018. Organizational Framework for the Establishment of Hub Schools, https://mon.gov.ua/ua/tag/oporni-shkoli

Mullis et al. 2012. TIMSS 2011 International Results in Mathematics. International Association for the Evaluation of Educational Achievement (IEA).

OECD. 2017a. OECD Reviews of Integrity in Education: Ukraine 2017

OECD. 2017b. "Education at a Glance 2017: OECD Indicators." Paris: OECD Publishing.

UNESCO. 2015. Education for All 2000-2015: Achievements and Challenges. Paris: UNESCO.

Roseth, Viviana V., Alexandria Valerio, and Marcela Gutiérrez. 2016. Education, Skills and Labor Market Outcomes. Results from large-scale adult skills surveys in urban areas in 12 countries

State Statistics Service of Ukraine. 2017. "Demographic and Social Statistics /Education." http://ukrstat.gov.ua/operativ/menu/menu_e/osv.htm.

Ukrainian Parliament. 2017. Law on Education, unofficial translation

Voytov, Volodymyr. 2003. "Educational Services Financing From Local Budgets" in Ministry of Education and Science of Ukraine. 2003. Reform Strategy for Education in Ukraine: Educational Policy Recommendations, http://www.irf.ua/files/eng/programs_edu_ep_409_en_ref_strategy.pdf

World Bank. 2003. Ukraine - Education reform policy note. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/348821468313765642/Ukraine-Education-reform-policy-note

World Bank. 2008. Ukraine - Improving Intergovernmental Fiscal Relations and Public Health and Education Expenditure Policy : Selected Issues. Washington, DC. World Bank. https://openknowledge.worldbank.org/handle/10986/8006

World Bank. 2011. Ukraine - Equal Access to Quality Education in Ukraine Project (Implementation and Completion Results Report). Washington, DC: World Bank. http://documents.worldbank.org/curated/en/858531468143397357/Ukraine-Equal-Access-to-Quality-Education-in-Ukraine-Project

World Bank. 2017. Ukraine - Public finance review. Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/476521500449393161/Ukraine-Public-finance-review World Bank (2018): Learning to Realize Education's Promise, World Development Report 2018.

Annexes

Annex 1: Summary of the new Law of Ukraine on Education

The new education law lays out four important structural changes

1. Different paths to acquiring skills and competencies.

According to current trends in Europe and worldwide the draft law stipulates three forms of education: formal, non-formal education and informal training. This mechanism provides for recognition of non-formal education and informal training in the system of formal education. This would substantially widen the citizen's opportunities for education. For the first time in Ukraine, the bill guarantees a right of a citizen to choose a form of education (parents of children), first of all, in secondary education, which is in line with the requirements of the European Convention on Human Rights and Fundamental Freedoms (Art. 2 of the First protocol).

2. More professional autonomy to teachers and a push to raise their status

The law lays out a path to improve the working conditions of teachers by reducing bureaucratic controls, widening of academic freedom and creation of space for creativity. Interference of state and local authorities in the educational process will be limited. A teacher will have the freedom to choose the forms to raise his or her qualification. There will also be a dedicated mechanism of voluntary certification of teachers. After its completion teachers would get an appropriate additional reward. The bill provides for a raise of teacher's social status and salary.

3. Putting in place a European structure of education system

The law introduces of a modern classification of formal education that is aligned with the International Standard Classification of Education adopted by UNESCO in 2011 (ISCED-2011). This involves set up of a complete 12-year secondary education. Today the 11-year school in Europe functions only in Ukraine, Russia and Belarus.

4. A more modern division of labor in terms of managing the education sector: the Ministry will set (and, eventually, monitor) learning standards; schools will have more autonomy to pick educational content.

The old paradigm, where, following the post-Soviet tradition, the Ministry defines the education content, will be replaced by modern European model where:

- Government (Ministry) approves standards for competency based education in terms of learning outcomes (what a graduate of a certain level should know, understand and should be able to do);
- Educational institutions develop educational programs, which should guarantee compliance with the state standards; in case the document of education is issued by the state, the educational programs should undergo accreditation;
- Development of curriculum and educational courses is bestowed on the academic autonomy of educational institutions.

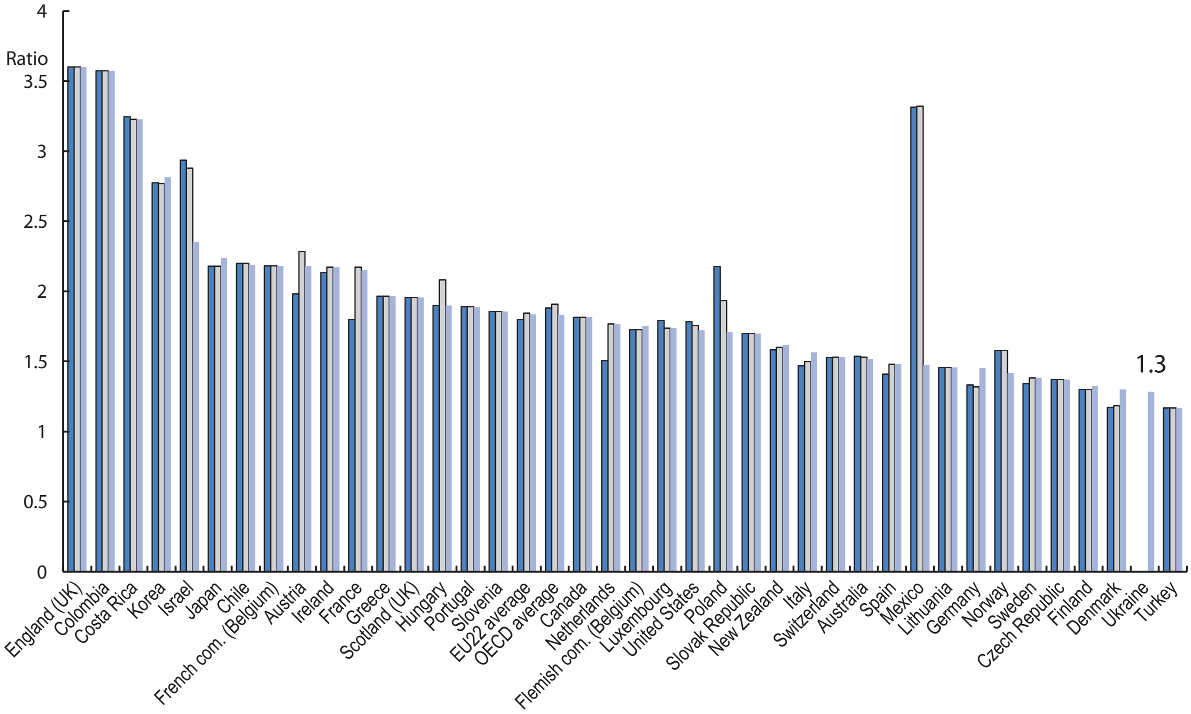
With school getting more autonomy, the Ministry will set up a new quality control agency, modelled on a European model of quality assurance. This new Agency for Education Quality Assurance will be established on the bases of the existing State Inspectorate of Educational Institutions. This agency will deal with the accreditation issues (except of in higher education). Regional divisions of the Agency will provide professional inspection of educational institutions. In order to ensure objective assessment of education quality the bill allows for set up of institutions for external independent evaluation of learning outcomes. A system for education quality monitoring is foreseen.

Annex 2: Tables and figures

Annex table 1: Salary details of teachers of different ranks

Time period	Position salary (18 hours per week)	Markup for years in service (approximately 20 percent)	Markup for occupation prestige (up to 20 percent, up to 30 percent)	1/12 part of the annual remuneration for bona fide work	1/12 part of recovery money during the provision of an annual basic vacation	Other extra pays, markups (type of institution, class supervision, checking of notebooks, etc.) (about 20-22 percent)	Total (average monthly take-home pay)
A teacher without a rank and work experience							
Dec-17	3,152	315	630	262	262	693	5,314
Jan-18	3,735	374	1,121	310	310	822	6,672
Teacher of rank II							
Dec-17	3,392	678	678	282	282	746	6,058
Jan-18	4,000	800	1,200	332	332	800	7,464
Teacher of rank I							
Dec-17	3,632	726	726	301	301	799	6,485
Jan-18	4,264	853	1279	354	354	853	7,957
Teacher of the highest rank							
Dec-17	3,872	774	774	321	321	852	6,914
Jan-18	4,546	909	1,364	377	377	1,000	8,573

Annex figure 1: Salary progression of teachers



Source: OECD AAG 2017. Table D3.6 (web only) and World Bank calculations for Ukraine based on Annex table 1

Annex table 2: Projected Costs and savings of major reform initiatives (bn current price, UAH)

	2018	2019	2020	2021	2022	2023	2024	2025
Spending on New School Reform initiatives (right axis)	**1.81**	**2.04**	**2.28**	**2.51**	**2.75**	**3.00**	**3.27**	**3.56**
Textbooks	0.37	0.42	0.47	0.52	0.57	0.62	0.67	0.73
National e-platform	0.05	0.06	0.07	0.08	0.08	0.09	0.10	0.11
Teachers' professional development	0.39	0.43	0.49	0.54	0.59	0.64	0.70	0.76
Equipment for institutions of general secondary education	1.00	1.12	1.26	1.38	1.52	1.65	1.80	1.96
Spending on expanded hub school program (adding 2600 more hub schools) (right axis)	**1.15**	**1.07**	**2.94**	**5.27**	**6.20**	**1.29**	**1.40**	**1.52**
Busses	0.20	0.21	0.25	0.69	1.19	1.28	1.40	1.52
Ensuring proper education conditions for children with special needs	0.51	0.39	1.31	2.23	2.45	0.00	0.00	0.00
Natural sciences and math study rooms	0.44	0.39	1.31	2.23	2.45	0.00	0.00	0.00
Severance package	0.00	0.06	0.07	0.10	0.11	0.00	0.00	0.00
Public Information Campaign	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.00
Total costs	**1.15**	**1.07**	**2.94**	**5.27**	**6.20**	**1.29**	**1.40**	**1.52**
Total savings from staff reductions	**0.00**	**2.47**	**5.51**	**10.41**	**16.23**	**17.94**	**19.56**	**21.32**
Net savings	**-1.15**	**1.41**	**2.56**	**5.14**	**10.03**	**16.65**	**18.16**	**19.80**

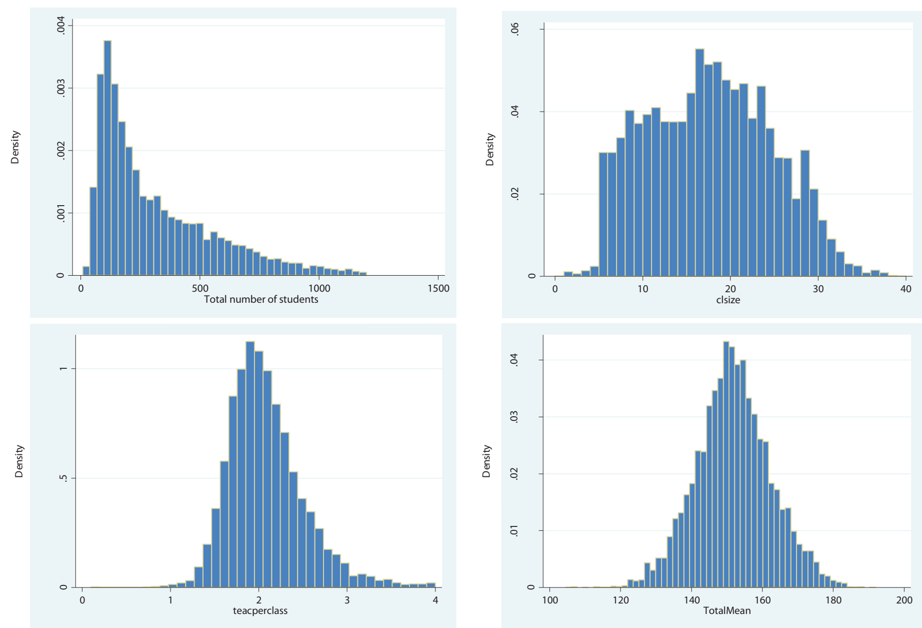
Source: Analysis done for this paper. Excel file with detailed assumptions are available upon request

Annex 3: Using School Level Educational Production Function to Determine the Effects of Measured School Characteristics on Average Student Performance for Ukrainian Secondary Schools[21]

This analysis presented in this annex employs the same school level dataset as Coupe, Olefir, and Diego Alonso (2015) – COD 2015 from here on – to estimate the educational production function for Ukrainian secondary schools. We conceptualize schools as employing some production technology to combine various inputs into producing student learning. In particular, as in COD 2015, we are especially interested in evaluating the effects of school enrolment size and other size-related school characteristics on the average student performance in Ukraine's External Independent Testing (EIT) in 2010. The EIT is the only standardized nation-wide test of student knowledge in Ukraine.

As in COD 2015, we have information on both the 2010 EIT test scores and information on school enrolment and class size variables for 11,683 Ukrainian schools. We drop 3 schools which have no information on the total number of teachers. We further restrict our sample to schools with average class size (in the highest grade level) of 40 students or less (220 schools were dropped), and with average teacher-to-classroom ratio of 4 or less (97 schools were dropped). The final sample for the descriptive statistical analyses presented in Figure 1 and Table 1 contains 11,363 schools.

Figure 1. Distributions of Key Variables for Ukrainian Secondary Schools



The top left chart in Figure 1 shows that Ukrainian secondary school enrolment size is highly right skewed. While the average school size is 328, the median school size is only 233. Around 14 percent of the schools have less than 100 enrolled students.

The average class size is 19.2, but as the top right chart in Figure 1 shows, class size distribution is widely dispersed. Looking at the lower end of the distribution, we can see that about a quarter of all classrooms have less than 15 students.

Another important variable used in the analysis of this study, but not in COD 2015, is the teacher-to-classroom ratio. Regression results obtained in COD 2015, as in most other studies that focus on evaluating the effects of enrolment and class sizes on student learning, could potentially have overlooked this very important factor. Our hypothesis is that the teacher-to-classroom ratio, which is positively associated with school enrolment size, is a key determinant of student

[21] Prepared by Dilaka Lathapipat

performance. This is because the ratio could more closely reflect the level of teacher subject and/or grade level specialization. As will be shown below, larger schools are more able to benefit from scale economies than smaller schools and can allocate teachers in a more effective manner. The bottom left chart in Figure 1 shows that the teacher-to-classroom ratio variable is slightly right skewed with a mean of 2.09 and a median of 2.05.

The performance outcome variable of interest in this study is the school average EIT score in Ukrainian language in 2010. As the bottom right chart in Figure 1 shows, school average test score in 2010 is approximately normally distributed with a mean of 151.9 and a standard deviation of 10.6.

Table 1 presents key school characteristics by school size category. It becomes immediately apparent that average class sizes in small schools are much smaller than in larger schools. The same can be said of the student-teacher ratio. However, the small class sizes and student-teacher ratios do not mean that these schools are offering high quality education to their students. On the contrary, these small schools generally have lower teacher-to-classroom ratio, which could reflect lesser degree of subject and/or grade level specialization of their teachers. If this is indeed the case – an empirical question which we will investigate below – then small schools are offering inferior quality of education at a much higher per-student public cost.

Table 1. Key Characteristics of Ukrainian Secondary Schools by School Size Category

School Size Category	Average class size	Student-teacher ratio	Teacher-classroom ratio	Urban share	Total #Classes	Total #Schools	Average test score
Less than 85	8.05	4.33	1.92	0.07	8,605	1,013	148.2
85 to 109	10.09	5.40	1.82	0.06	10,490	1,021	148.0
110 to 134	11.54	6.30	1.90	0.07	10,555	992	147.9
135 to 159	13.42	7.26	2.01	0.12	9,351	865	148.3
160 to 194	15.09	7.89	2.10	0.18	10,429	939	148.5
195 to 239	17.54	8.83	2.20	0.33	11,376	969	148.0
240 to 299	19.02	9.32	2.20	0.48	10,457	772	148.3
300 to 369	20.20	10.02	2.16	0.66	15,968	1,006	149.3
370 to 469	20.86	10.69	2.12	0.77	19,724	1,017	151.7
470 to 589	22.02	11.35	2.13	0.89	21,455	950	151.7
590 to 759	23.00	11.95	2.13	0.95	25,236	932	154.3
760 or above	24.73	12.85	2.15	0.98	32,267	887	157.5
Overall	19.16	10.61	2.09	0.46	185,913	11,363	151.9

Moreover, we can see from Table 1 that the vast majority of small schools are located in the rural areas of Ukraine. Initial investigation reveals that their average EIT scores are significantly lower than their larger counterparts. A causal investigation of the relationships between these key school characteristics and school performance using regression analysis is the primary purpose of this study.

Ordinary Least Squares Regression Analysis

Table 2 presents the estimation results of educational production function regression models, where we argue that the models are misspecified. In all models, we follow COD 2015 and control for the size of enrolment (in 100 increments), class size, urban indicator, and the share of EIT exams (other than the Ukrainian language exam) taken in Ukrainian.

The estimated regression coefficients of the key variables, namely, the number of students and class size in Model 1 are effectively the same as those reported in COD 2015. An increase in enrolment size of 100 students is associated with an increase of 0.892 points in the average Ukrainian language test score and the coefficient is significant at conventional statistical levels. The size of this effect is very small, however, at around 0.08 standard deviation. Model 2 further controls for school type dummies, but the coefficient estimates of the variables of interest are more or less the same. Class size is found to be insignificant in both specifications and of negligible size.

Table 2: Ordinary Least Squares Regression Results (Misspecified Models)

Variables	Model 1	Model 2	Model 3	Model 4
Number of students ('00)	0.892***	0.877***	1.008***	0.962***
	(0.104)	(0.101)	(0.090)	(0.090)
Class size (highest grade)	-0.188	-0.108	-0.410**	-0.285
	(0.227)	(0.202)	(0.186)	(0.176)
Class size squared	0.006	0.002	0.008*	0.004
	(0.005)	(0.004)	(0.004)	(0.004)
Teacher to classroom ratio			7.256***	5.173***
			(0.570)	(0.513)
Urban	3.487***	1.314**	3.114***	1.443***
	(0.535)	(0.549)	(0.515)	(0.515)
Share Ukrainian exams	9.241***	8.028***	7.450***	6.965***
	(0.619)	(0.536)	(0.535)	(0.504)
Intercept	138.986***	151.565***	128.485***	142.133***
	(2.200)	(1.757)	(2.425)	(2.163)
Control for school type dummies	No	Yes	No	Yes
Dependent variable (Test score in 2010)	Level	Level	Level	Level
Observations	11,361	11,361	11,361	11,361
R-squared	0.209	0.353	0.311	0.400

Robust standard errors in parentheses (clustered at the Rayon level)
*** p<0.01, ** p<0.05, * p<0.1

Rather than controlling for student-teacher ratio as in COD 2015, Models 3 and 4 instead control for the teacher-to-classroom ratio variable. The empirical investigation seems to support our hypothesis that the teacher-to-classroom ratio is an important determinant of student performance. The marginal effect of the variable is statistically significant at conventional levels; a unit increase in the teacher-to-classroom ratio is associated with around 5 to 7 points (0.5 to 0.7 standard deviation) increase in the average test score.

However, we argue that Models 1-4 shown in Table 2, as well as the regression models in COD 2015 are misspecified. The main problem is that we do not observe all potentially important family background characteristics of the student body that capture the quality of early education the students received or the home environments that are conducive to learning. These characteristics are crucial in determining cognitive ability of children. Furthermore, it is conceivable that these relevant family factors are related to both student achievement and the selection of schools and teachers by the parents.[22] In other words, the omitted family background characteristics of the student body are confounding our estimates of the school and teacher effects on performance. The same can be said of the unobserved quality of teachers and other school resources.

[22] For example, highly educated parents are more likely to have prepared their children better since birth to be school ready and to have chosen neighbourhoods with schools that are well-resourced and have high quality teachers.

To mitigate the problem, this study controls for some of these confounding factors by including an urban indicator to capture the socioeconomic status of the student body. However, this is unlikely to be sufficient. A particularly important covariate we have available is average Ukrainian language test score in the 2008 school year. It is a much more convincing argument that prior student achievement can effectively capture the effects of the remaining confounding socioeconomic factors and teacher and school quality. Including the variable as a covariate therefore permits a more valid comparison of student performance across schools.

The estimation results of these more valid models are shown in Table 3. Models 5 and 6 are directly comparable to Models 3 and 4 shown in Table 2. As expected, the estimated marginal effects of enrolment size is greatly diminished, even though the estimates are still statistically significant at the 1 percent level. The effects of class size are also diminished, but to a lesser degree. Furthermore, the quadratic class size variables now turn out to be much more precisely estimated.

Table 3. Ordinary Least Squares Regression Results – Control for Past Student Performance

Variables	Model 5	Model 6	Model 7	Model 8
Number of students ('00)	0.391***	0.412***	0.297***	0.321***
	(0.046)	(0.049)	(0.046)	(0.048)
Class size (highest grade)	-0.295***	-0.264***	-0.296***	-0.257***
	(0.086)	(0.084)	(0.093)	(0.092)
Class size squared	0.006**	0.005**	0.006***	0.005**
	(0.002)	(0.002)	(0.002)	(0.002)
Teacher to classroom ratio	2.736***	2.246***	1.703***	1.171***
	(0.283)	(0.291)	(0.298)	(0.305)
Urban	1.691***	1.154***	1.221***	0.815***
	(0.300)	(0.310)	(0.291)	(0.298)
Share Ukrainian exams	3.230***	3.372***	3.461***	3.589***
	(0.356)	(0.357)	(0.330)	(0.331)
Average test score in 2008	0.674***	0.623***	0.591***	0.550***
	(0.016)	(0.016)	(0.016)	(0.015)
Student-teacher ratio			-0.023	-0.049
			(0.034)	(0.036)
Participation ratio			0.072***	0.071***
			(0.007)	(0.007)
Intercept	41.672***	53.494***	51.241***	61.376***
	(2.432)	(2.546)	(2.372)	(2.303)
Control for school type dummies	No	Yes	No	Yes
Dependent variable (Test score in 2010)	Level	Level	Level	Level
Observations	11,158	11,158	9,527	9,527
R-squared	0.611	0.622	0.602	0.612

Robust standard errors in parentheses (clustered at the Rayon level)
*** p<0.01, ** p<0.05, * p<0.1

Turning now to the teacher-to-classroom ratio variable, it is estimated that a unit increase in the teacher-to-classroom ratio is associated with around 2.2 to 2.7 points (0.21 to 0.26 standard deviation) increase in the average test score. Again, the effects are much smaller in magnitude than those obtained using the misspecified models.

It is important to note that while the EIT test is nation-wide in Ukraine, not all students participate as it is taken only by those who want to continue further studies (higher education). Following COD 2015, we include in Models 7 and 8 a control for EIT test participation ratio.

Furthermore, to test whether the teacher-to-classroom ratio variable is important, rather than the more commonly used student-teacher ratio (as is used in COD 2015), Models 7 and 8 include both of these variables as controls. As expected, once both variables are included, the student-teacher ratio turns out to be small in magnitude and not statistically different from zero.

In the models considered thus far, the key "size of enrolment" and "teacher-to-classroom ratio" variables are entered in level terms. In order to compare the impacts of these variables on school performance, it is informative to look at the effect of a 1 standard deviation increase in each variable. We carry out this exercise and report the results below in Table 4.

Table 4. Ordinary Least Squares Regression Results – Key Variables Specified in Deviation

Variables	Model 9	Model 10	Model 11	Model 12
Number of students (1 Sdev)	1.021***	1.076***	0.759***	0.796***
	(0.121)	(0.129)	(0.121)	(0.126)
Class size (highest grade)	-0.295***	-0.264***	-0.307***	-0.279***
	(0.086)	(0.084)	(0.090)	(0.088)
Class size squared	0.006**	0.005**	0.006***	0.006**
	(0.002)	(0.002)	(0.002)	(0.002)
Teacher to classroom ratio (1 Sdev)	1.152***	0.945***	0.766***	0.599***
	(0.119)	(0.123)	(0.112)	(0.114)
Urban	1.691***	1.154***	1.173***	0.719**
	(0.300)	(0.310)	(0.281)	(0.290)
Share Ukrainian exams	3.230***	3.372***	3.469***	3.604***
	(0.356)	(0.357)	(0.333)	(0.334)
Average test score in 2008	0.674***	0.623***	0.591***	0.550***
	(0.016)	(0.016)	(0.016)	(0.015)
Participation ratio			0.072***	0.072***
			(0.007)	(0.007)
Intercept	41.672***	53.494***	50.959***	60.681***
	(2.432)	(2.546)	(2.338)	(2.270)
Control for school type dummies	No	Yes	No	Yes
Dependent variable (Test score in 2010)	Level	Level	Level	Level
Observations	11,158	11,158	9,527	9,527
R-squared	0.611	0.622	0.602	0.612

Robust standard errors in parentheses (clustered at the Rayon level)
*** p<0.01, ** p<0.05, * p<0.1

It can be seen that once the key variables of interest are specified in deviation, the estimated effects of enrolment size and teacher-to-classroom ratio turn out to be statistically significant and of very similar magnitude across all models. These evidences suggest that enlarging enrolment size, together with staffing classrooms adequately could significantly improve average school performance in Ukraine.

Since we have three years of data, we can carry out a rubustness check of Models 9-12 by using the school average test score in 2009 as the dependent variable instead of the 2010 test score. The results of the exercise are reported below in Table 5.

Table 5: Ordinary Least Squares Regression Results – Key Variables Specified in Deviation (Robustness Check)

Variables	Model 9	Model 10	Model 11	Model 12
Number of students (1 Sdev)	0.729***	0.803***	0.587***	0.641***
	(0.102)	(0.102)	(0.128)	(0.125)
Class size (highest grade)	-0.132*	-0.115	-0.222**	-0.194**
	(0.078)	(0.078)	(0.091)	(0.090)
Class size squared	0.002	0.002	0.004*	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Teacher to classroom ratio (1 Sdev)	1.053***	0.864***	0.908***	0.724***
	(0.106)	(0.111)	(0.107)	(0.111)
Urban	1.320***	0.842***	0.968***	0.538
	(0.293)	(0.303)	(0.326)	(0.332)
Share Ukrainian exams	2.737***	2.857***	2.935***	3.064***
	(0.333)	(0.333)	(0.345)	(0.344)
Average test score in 2008	0.684***	0.637***	0.608***	0.572***
	(0.015)	(0.015)	(0.016)	(0.016)
Participation ratio			0.046***	0.047***
			(0.007)	(0.007)
Intercept	39.956***	50.952***	49.551***	58.705***
	(2.212)	(2.353)	(2.370)	(2.364)
Control for school type dummies	No	Yes	No	Yes
Dependent variable (Test score in 2009)	Level	Level	Level	Level
Observations	11,329	11,329	8,280	8,280
R-squared	0.618	0.628	0.582	0.592

Robust standard errors in parentheses (clustered at the Rayon level)
*** p<0.01, ** p<0.05, * p<0.1

The results from the robustness check once again indicate the significance of enrolment size and teacher-to-classroom ratio as determinants of school performance. However, the importance of teacher-to-classroom ratio is now relatively greater than that of enrolment size as can be inferred from the relative magnitudes of the respective regression coefficients.

Unconditional Quantile Regression Analysis

While OLS regressions in the previous section can be used to estimate the partial effects of the covariates on the performance outcome for an average school, the narrow focus on the mean outcome obscures the effects of the teacher and school characteristics on other important features of the school performance distribution that are of policy relevance. Much richer analysis can be carried out using Unconditional Quantile Regression (UQR), where the estimated marginal effects of the key variables on school performance can be seen for schools ranked throughout the performance quantiles.

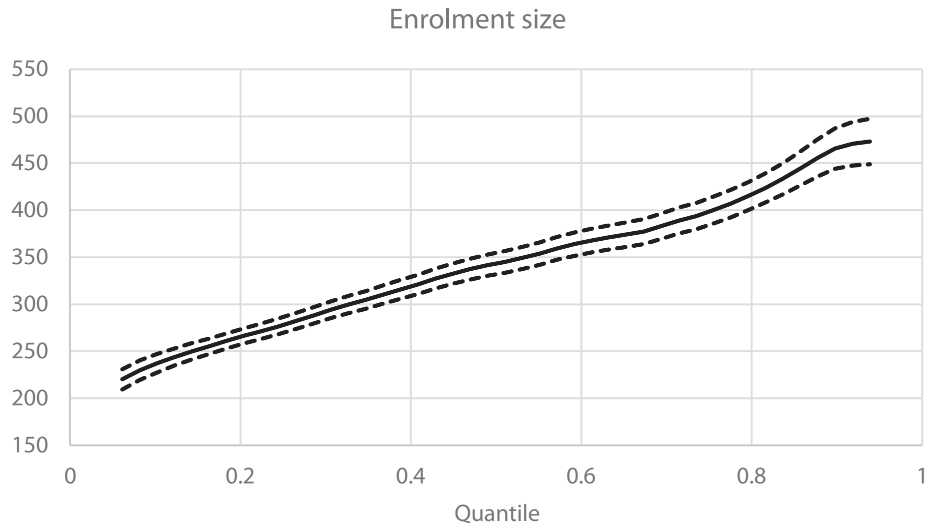
Before we go on to analyze the estimation results from the educational production function model, let us first consider simple relationships between average student performance quantile of the sample schools and enrolment size, class size, and teacher-to-classroom ratio.

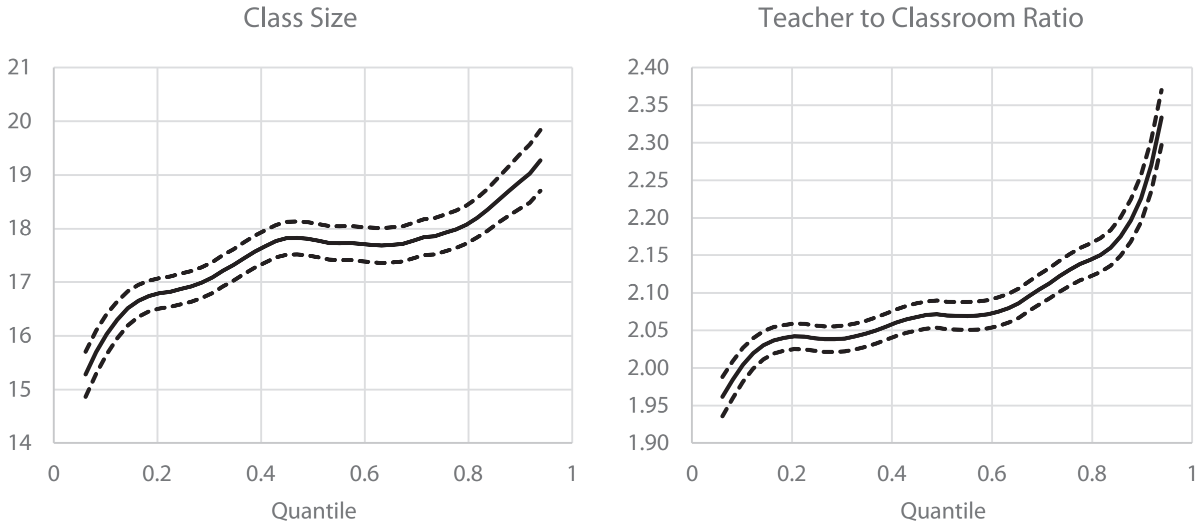
The top graph in Figure 2 presents a line plot of school enrolment size against the quantile ranking of the school student performance in 2010. The average school at the 6th percentile (0.06 quantile) of the performance distribution has an enrolment size of 230, while the average school at the 94th percentile has an enrolment size of 473 and the relationship is approximately linear.

Furthermore, the bottom two graph in the same figure show similar line plots for average class size and the average number of teacher-to-classroom ratio against performance quantile. It is clear that poorer performing schools, on average, have smaller class sizes and lower teacher-to-classroom ratios than better performing schools. The difference in teacher-to-classroom ratios for schools ranked at the 6th and 94th percentiles is about 0.4.

It should be reminded that the charts shown in Figure 2 present simple relationships between the key variables and the school performance quantile ranking and should not be interpreted as having causal relationships. The causal relationships will be estimated under the UQR framework below.

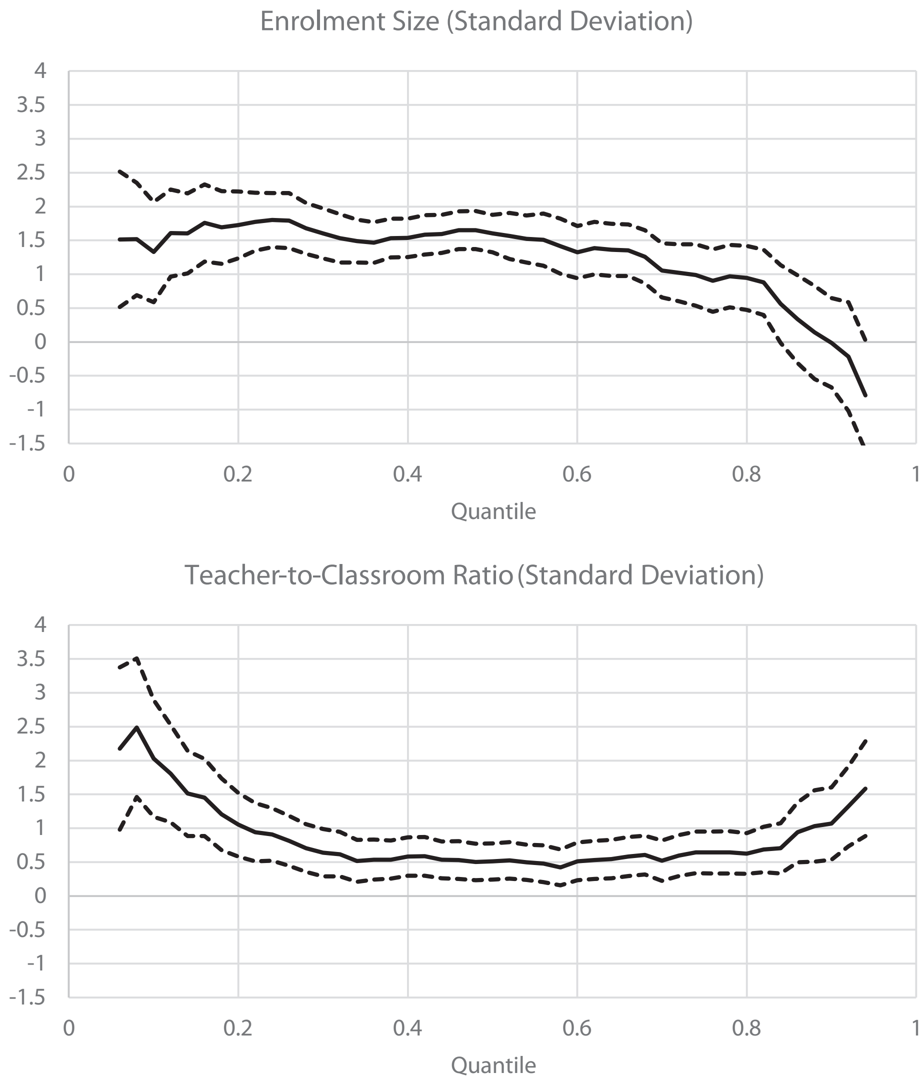
Figure 2. The Relationships between the Key Variables and School Performance Quantile

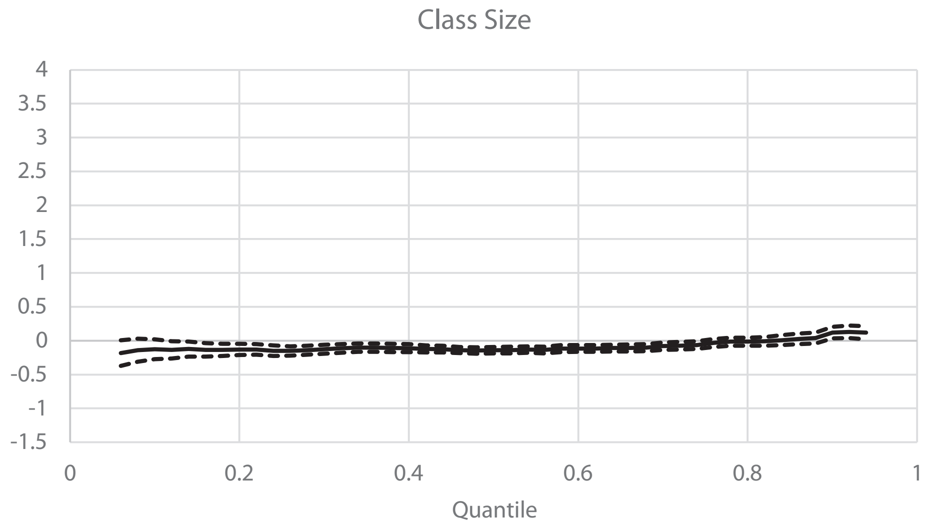




The educational production function specification used to estimate the UQR partial effects or coefficients is the same as that used in Model 10 estimated using OLS above (Table 4), where the enrolment size and teacher-to-classroom ratio variables are expressed in deviation form. The estimated regression coefficients of the three variables of interest are presented graphically in Figure 3 together with the 95 percent confidence band for the estimated effects.

Figure 3. Unconditional Quantile Partial Effects of the Key Variables





Recall that in the models estimated using OLS above, the effects of enrolment size and teacher-to-classroom ratio are similar in magnitudes. However, we can now see from the "Teacher-to-Classroom Ratio" chart in Figure 3 that allocating more teacher to each classroom (by 1 standard deviation or 0.42 teacher) is expected to raise performance for schools ranked at the 8th percentile of the performance distribution by as much as 2.5 points, ceteris paribus. The size of the effect drops down rapidly to within a range of 0.52-0.64 points for schools ranked between the 30th and the 80th percentiles before rising thereafter. The large and positive effects for schools at the bottom end of the performance distribution is not surprising considering the fact that the teacher-to-classroom ratios are much lower for these schools (see Figure 2).

Similarly, the "Enrolment Size" chart indicates that enlarging school enrolment size could benefit lower performing schools more disproportionately. However, it is clear from Figure 3 that the inequality reducing effect of this policy variable is less pronounced than that of the "Teacher-to-Classroom Ratio" variable.

Annex 4: Estimating teacher redundancies with hierarchical clustering based on travel distances in Ukraine[23]

This annex illustrates the potential for findings savings in Ukraine's primary and secondary schools, according to four different options:

1. All primary and secondary schools that have less than 30 minutes of walking between them for the two furthest schools are clustered, and students are allocated optimally at class level across them, with a maximum class size of 30 students for both primary secondary.
2. Secondary schools have up to a 60-minute walk between the two furthest schools. For primary schools, this option is equivalent to option 1. Optimization of classes as in option 1.
3. Secondary schools have up to a 20-minute drive between the two furthest schools. For primary schools, this option is equivalent to option 1. Optimization of classes as in option 1. Driving time is based on Google map estimates of road conditions.
4. Primary and secondary schools have up to 20 minutes driving between the two furthest schools. Optimization of classes as in option 1.

Based on these clustering thresholds, the paper analysis the savings in classes and teachers. Further, the paper analyses how many of the teacher reductions could be achieved through natural retirement.

Leap frogging to the end, skipping the details of data and estimation of savings presented in Section 2 and 3, Annex table 1 shows how the options above could be implemented over time. Each option presents increasingly ambitious reforms, as thresholds for travel distances are slowly increased, allowing time to implement the easiest savings first and build up capacity for implementation, while also allowing an easy adjustment of the numbers of teachers, through retirement.

Annex table 1: Potential phased implementation of reforms

Phase 1 (2019 -21) All walking short distances	Phase 2 (2019 -21) Longer distance walk/drive (secondary) and walking (primary)	Phase 3 (2021 -23) Introduce bus for secondary	Phase 4 (2023 -) Introduce bus for primary
• 4,148 schools join a network where all children can walk • Eliminate classes and let teachers retire at age 62 • Prepare the grounds for further consolidation • Start procuring busses and offering children in "phase 1 isolated schools"	• Further 3,226 schools join a school network • Same as phase 1, except secondary school children allowed to walk for up to 60 minutes.	• Further 4,688 schools join a school network • Introduce busses for secondary students (from phase 1 isolated schools and close down these secondary school classes) • Continue to reduce teachers as they retire at age 62.	• Further 924 schools join a school network • Introduce busses for primary students (from phase 1 isolated schools and close down these schools) • Continue to reduce teachers as they retire at age 62.

1: Data and Methods

Estimating potential savings from school clustering is based on a series of steps and calculations. This section describes the methods applied, the raw data and data generating process. The process has been divided into four subsections: 1) Recovering of school GPS locations, 2) Calculation of travel distances between schools, 3) Clustering of schools, and 4) Calculation of savings potential.

[23] Prepared by Kasper Brandt and Thomas Sohnesen

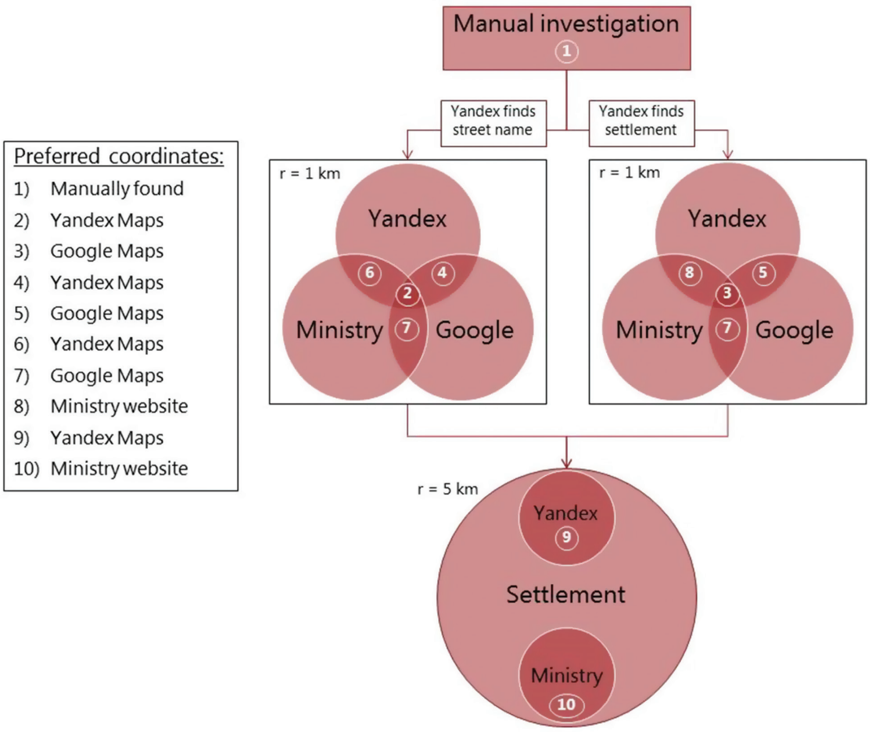
1.1 Recovering school locations

School locations were found on the Ministry of Education's websites. The ministry has a website for each school which include an address and GPS location. To utilize this information in a systematic way a Python code was used to web scape the address and GPS location of all schools. The code is included in appendix 4.1.

A quality check of the web scraped data revealed that the GPS locations on the ministry websites not always provide an exact location. In smaller rural locations the GPS point, at times, refers to a point within the village, and not the exact location of the school. To increase accuracy and confirm location of schools, the addresses were used to look up GPS locations based on the addresses. Addresses were looked up both in Google and Yandex Maps (a dominant search engine in Russian language). Appendix 4.2 shows the Python code used to extract GPS points in Yandex and Goggle Maps. The two mapping services are similar, but the search algorithm and data from school address queries differ. Yandex Maps seem to have a spatial hierarchical search that first finds the oblast, next the rayon, then the settlement, and finally the street name of the school. If Yandex Maps cannot find the street name, it provides the GPS coordinates of the settlement center point. Google Maps on the other hand seem to apply some kind of "smart filter", which means that in cases where the street address cannot be found within the given settlement, it may provide the location of the correct street name in a wrong settlement. Hence, in cases where an exact street name cannot be located, Google Maps may provide a wrong GPS location, while Yandex provides an inaccurate, but close proxy. Unfortunately search queries do not come with data indicating that whether Google Maps or Yandex, located the specific address or not.

Hence, there are three sources of GPS locations (the Ministry of Education, and Yandex and Google based on addresses), which in some cases are not identical. To ensure as high accuracy as possible, school GPS locations where prioritized according to the possible combinations of sources illustrated in Annex figure 1.

Annex figure 1 Determining the preferred GPS coordinates of schools



Leapfrogging to the final analysis, Table 2 shows how many of the estimated class savings that are based on an actual address as supposed to a settlement center point. The table shows that for phase one, 65 percent (7,670/(7,670+4,126)) of savings are based on an exact address only, while accuracy is lower at phase four, where 65 percent (28,729/(28,729+13,524)) of class savings include at least one school based on settlement GPS point. The share of savings based on settlement locations is slightly higher for primary schools, than secondary schools (Annex table 2).

Annex table 2: Class savings by type of GPS location

	Total		Primary		Secondary	
GPS point	**Address**	**Settlement**	**Address**	**Settlement**	**Address**	**Settlement**
Phase 1	7,670	4,126	2,711	1,991	4,959	2,135
Phase 2	9,957	8,413	2,711	1,991	7,246	6,422
Phase 3	11,544	18,347	2,711	1,991	8,833	16,356
Phase 4	13,524	28,729	4,691	12,373	8,833	16,356

1.2 Calculating travel distances

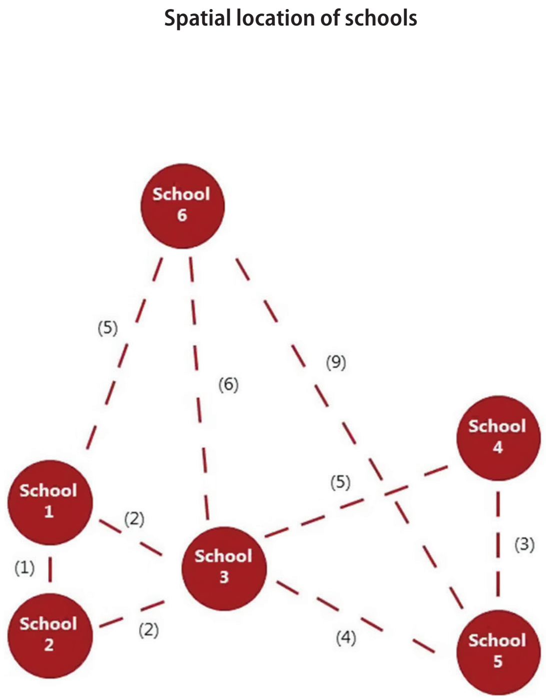
Based on each school's GPS location, the travel distances between schools were extracted from Google Maps. However, most schools are placed so far from each other that there is no reason to calculate the travel distance or consider their clustering potential. Therefore, the more than 14,000 schools were first divided into bins with clustering potential, based on a beeline distance. The analysis of primary and secondary education based on walking uses a beeline distance of five kilometers, while for the analysis of driving options uses a beeline of 12 kilometers. Hence, for each school they are grouped with the schools that are within the beeline based on their GPS points. Appendix 4.4 shows the R-code used for this first stage restriction of clustering according to clustering potential, which also serves to limit the number of travel distances to extract from Google Maps.

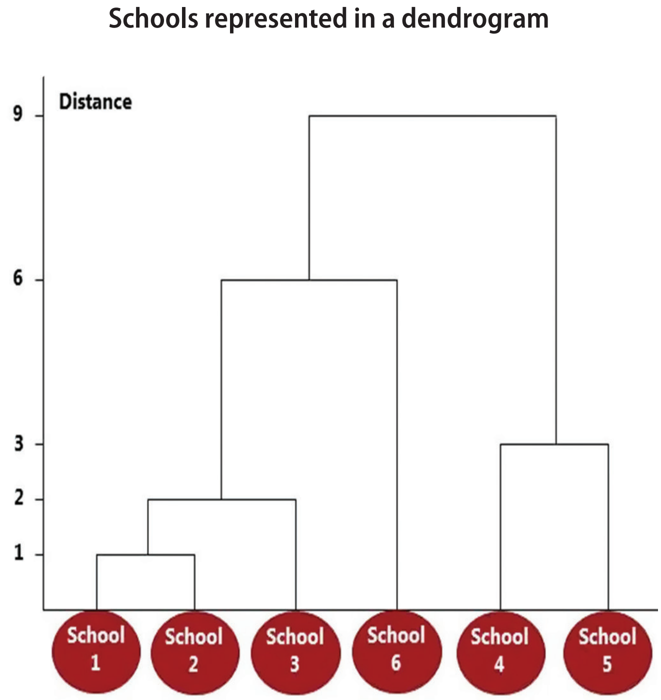
Based on set groups, the walking and driving distances were extracted from Google Maps using a Python code. The code is included in Appendix 4.5. These travel distances are the core input data for the clustering analysis.

1.3 Clustering analysis

Clustering potential was assessed through hierarchical clustering analysis. The method clusters schools such that the furthest distance between two school points is less than the set travel threshold. Annex figure 2 illustrates the method through two diagrams. The diagram on the left shows six schools, with the numbers between representing distances. I.e. there is a distance of five between school one and six, and a distance of two between school two and school three. The same distances and schools are shown in the dendrogram to the right. The dendrogram provides an overview of the clustering potential. For instance, if the maximum travel distance threshold between any school is set to one, only schools one and two would be within distance for clustering. At a threshold of three, schools one, two and three would be clustered, while also school four and five would be clustered. At a threshold of six, school six could be added to schools one through three, while at a threshold of nine all schools would be eligible for the same cluster. In the analysis that follows, the thresholds laid out in the four options in the introduction is used for primary and secondary schools, thereby determining which schools can enter the same cluster. Clustering is done by primary and secondary schooling level, so that classes at primary level is only clustered with other primary classes. Furthermore, the analysis distinguishes between different languages of instructions, meaning that e.g. Russian speaking schools can only be clustered with other Russian speaking schools.

Annex figure 2 Example of clustering – school locations





To get a sense of the potential for clustering Table 3 shows how many schools that can be clustered in the four different phases, and the characteristics of schools that can or cannot be clustered. Even just based on 30-minute walk, 29 percent of schools have clustering potential, while only nine percent are so isolated that if they join a cluster they would be more than 20-minutes drive from the furthest school in that cluster. It's important here to point out that this is the driving distance to the school furthest away in the network. The actual driving distance to nearest or largest school might well be much lower. As such, schools that are not included in a cluster in phase four are on average 6 kilometers away from the nearest school (beeline) (Annex table 3). Google map driving distance estimates generally take into account local conditions.

Annex table 3: Descriptive characteristics for schools with and without clustering potential

	Total	Phase 1		Phase 2		Phase 3		Phase 4	
		Isolated schools	Network schools	Isolated schools	Network schools	Isolated schools	Network schools	Isolated schools	Network schools
Number of schools	14,279	10,131	4,148	6,905	7,374	2,217	12,062	1,293	12,986
Average distance to nearest school (bee-line)	3.2 km	4.2 km	0.7 km	4.9 km	1.6 km	5.0 km	2.9 km	6.0 km	2.9 km
Urban share	22.8%	4.9%	66.6%	2.2%	42.1%	2.8%	26.5%	3.0%	24.8%
Students per school	227	136	452	118	330	88	253	114	239
Students per class	14.9	12.1	21.7	11.6	18.1	11.2	15.6	11.3	15.3
Share with multi-grade class (before networks are created)	31.7%	42.2%	6.0%	46.5%	17.9%	56.6%	27.2%	50.4%	29.9%
Share with multi-grade class (after networks are created)	-	42.2%	1.5%	46.5%	11.6%	56.6%	17.0%	50.4%	10.0%

1.4 Calculating savings potential

Data on number of classes, students, and teachers by grade level is provided by the Ministry of Education and relies on administrative data. The data is from the 2017/18 school year. The analysis focuses on primary and secondary schools, and any specialty schools are excluded. The analysis first calculates the number of classes that could be reduced through clustering. A typical case could be a town as illustrated in annex figure 3. Here, four primary schools are located within 30-minute walk of each other. The average class size in these four schools is 18 and in total they have 60 primary school classes. If students were allocated efficiently so that each class was larger, but still no higher than the maximum of 30 students, a total of only 45 classes is needed. To estimate the number of classes needed, the total number of students in these four schools at each grade level is simply divided by 30 – the maximum class size – and rounded up.

Annex figure 3 Example of clustering in a town

Example 1: School network which can be formed in urban setting

- Town of Dolynska (pop: 20,000).
- Stable elevation of around 180 meters.
- Primary school network.
- Four schools **with primary classes** within walking distance, and two schools (also with primary classes) within driving distance.
- 1,093 primary class students.

TOTAL for all 6 schools	Phase 0 (before)	Phase 1 (walk)	Phase 4 (bus)
Primary school classes	60	45	38
Primary school teachers	80	56 - 67	49 - 59
Primary school class size	18.2	24.3	28.8

Three different teacher/class ratios:

1. Overall average by primary/secondary and urban/rural = 1.278
2. Weighted average teacher/class ratio within network = 1.365
3. Largest school's teacher/class ratio = 1.556

Further information in Annex

23

To estimate potential teacher savings, we need a teacher/class ratio. Three different approaches for calculating teacher/class ratios are considered:

1. Overall average teacher/class ratio by primary/secondary school level and urban/rural network location.
2. Weighted average of network schools' teacher/class ratios. Thus, total number of teachers in network (either primary or secondary school teachers) divided by total number of classes in network (either primary or secondary level).
3. Largest school in network's teacher/class ratio.

It is worth noting here that all three approaches can be seen as conservative estimates, as all three builds on the current teacher/class ratios. If teacher/class ratios are unneededly high due to lack of adjustments to the number of teachers, such inefficiency would still be built into estimated potential savings.

2. Results on savings potential from clustering

To reiterate, results are presented according to four different options:

1. All primary and secondary schools that have less than 30 minutes of walking between them for the two furthest schools are clustered, and students are allocated optimally at class level across them, with a maximum class size of 30 students for both primary and secondary.
2. Secondary schools have up to 60 minutes between the two furthest schools. For primary schools, this option is equivalent to option 1. Optimization of classes as in option 1.
3. Secondary schools have up to 20 minutes driving between the two furthest schools. For primary schools, this option is equivalent to option 1. Optimization of classes as in option 1.
4. Primary and secondary schools have up to 20 minutes driving between the two furthest schools. Optimization of classes as in option 1.

Within each of these four options, the number of redundant classes are estimated based on a maximum of 30 students per class in both primary and secondary schools. This is presented in subsection 1. The number of redundant teachers is based on number of redundant classes, this is shown in subsection 2. Finally, subsection three estimates the number of teacher reductions that could be done through natural old age retirement.

2.1 Class savings

Based on the clustering analysis Annex table 4 shows how many classes that are redundant, if student were allocated more efficiently in clustered schools, according to the four different phases. The table shows that, based on only a maximum of 30 minutes of walking (phase 1) between schools, 29 percent of schools (primary and secondary) have savings potential, and that within these schools, 6.6 percent of classes could be made redundant. However, if schools were allowed to be up to 20 minutes drive apart, 91 percent of schools could be clustered, and 24 percent of classes would be redundant. Appendix 4.8 presents a number of additional tables, breaking down results in Annex table 4 by primary and secondary school and phase. From a spatial point of view, Annex table 5 presents the same result at Oblast level, both as absolute number of classes that can be reduced, and relative to current number of classes.

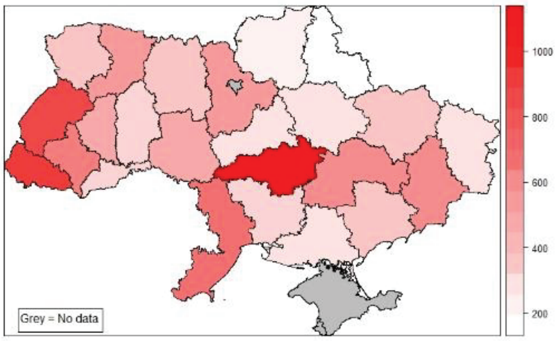
Annex table 4: class savings by phase

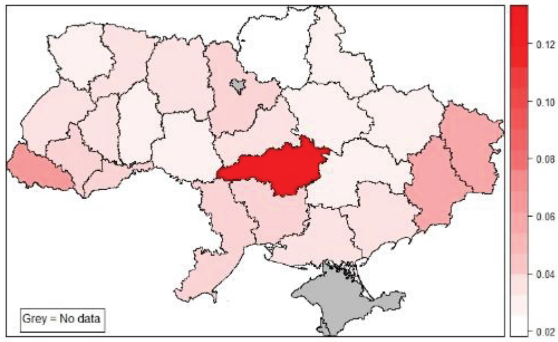
	Phase 1	Phase 2	Phase 3	Phase 4
Classes before networks	177,659	177,659	177,659	177,659
Classes after networks	165,863	159,289	147,768	135,406
Class savings	11,796	18,370	29,891	42,253
Class savings (% of all)	6.6%	10.3%	16.8%	23.8%
Schools affected	29.0%	51.6%	84.5%	90.9%
Students affected	56.3%	66.7%	77.9%	94.3%

Annex table 5 Class savings by Oblast

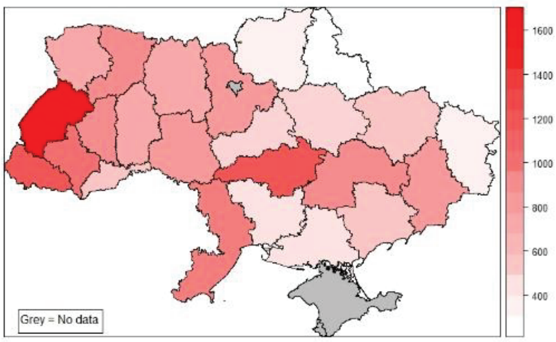
Number of classes (absolute) | **Number of classes (as percentage)**

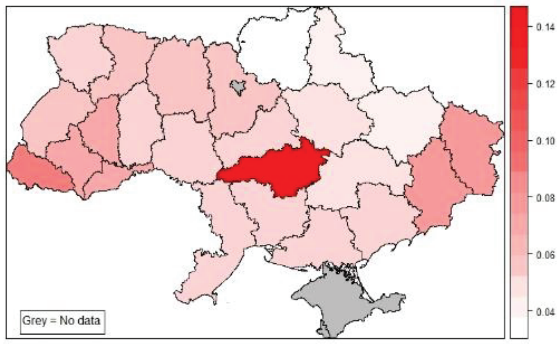
Phase 1



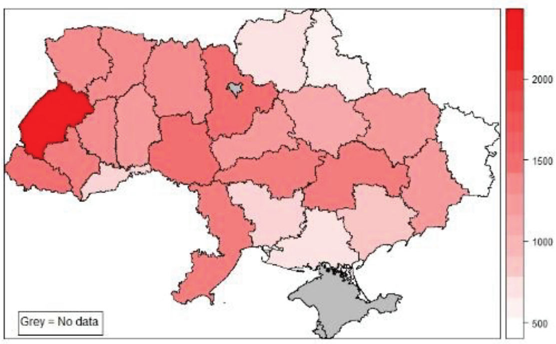


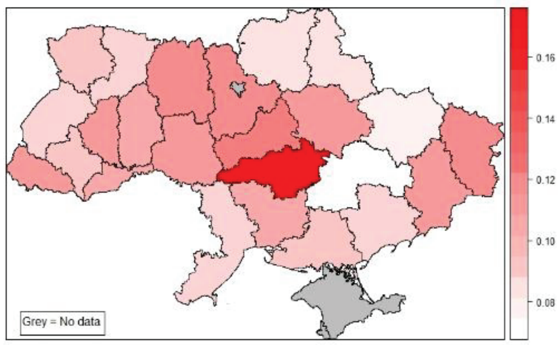
Phase 2



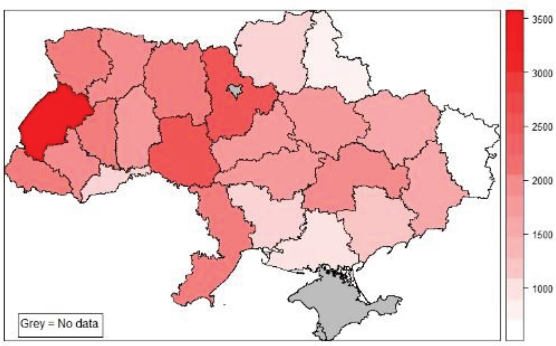


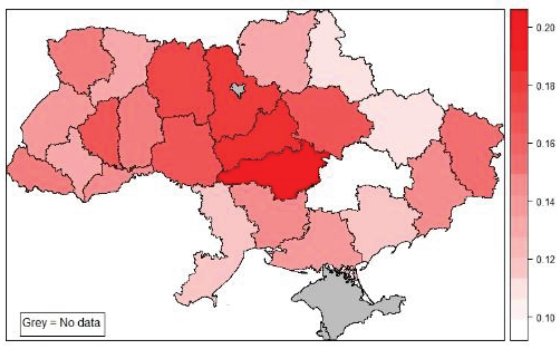
Phase 3





Phase 4





2.1.1 Class savings and walking distance

Option or phase 1, is based on 30 minutes walking, which is arguable an arbitrary cutoff. Annex table 6 provides additional estimates of class savings for phase, based on travel distances of 35 to 45 minutes of walking. The table shows that each additional five minutes of walking roughly bring about 0.5 percent more class savings potential.

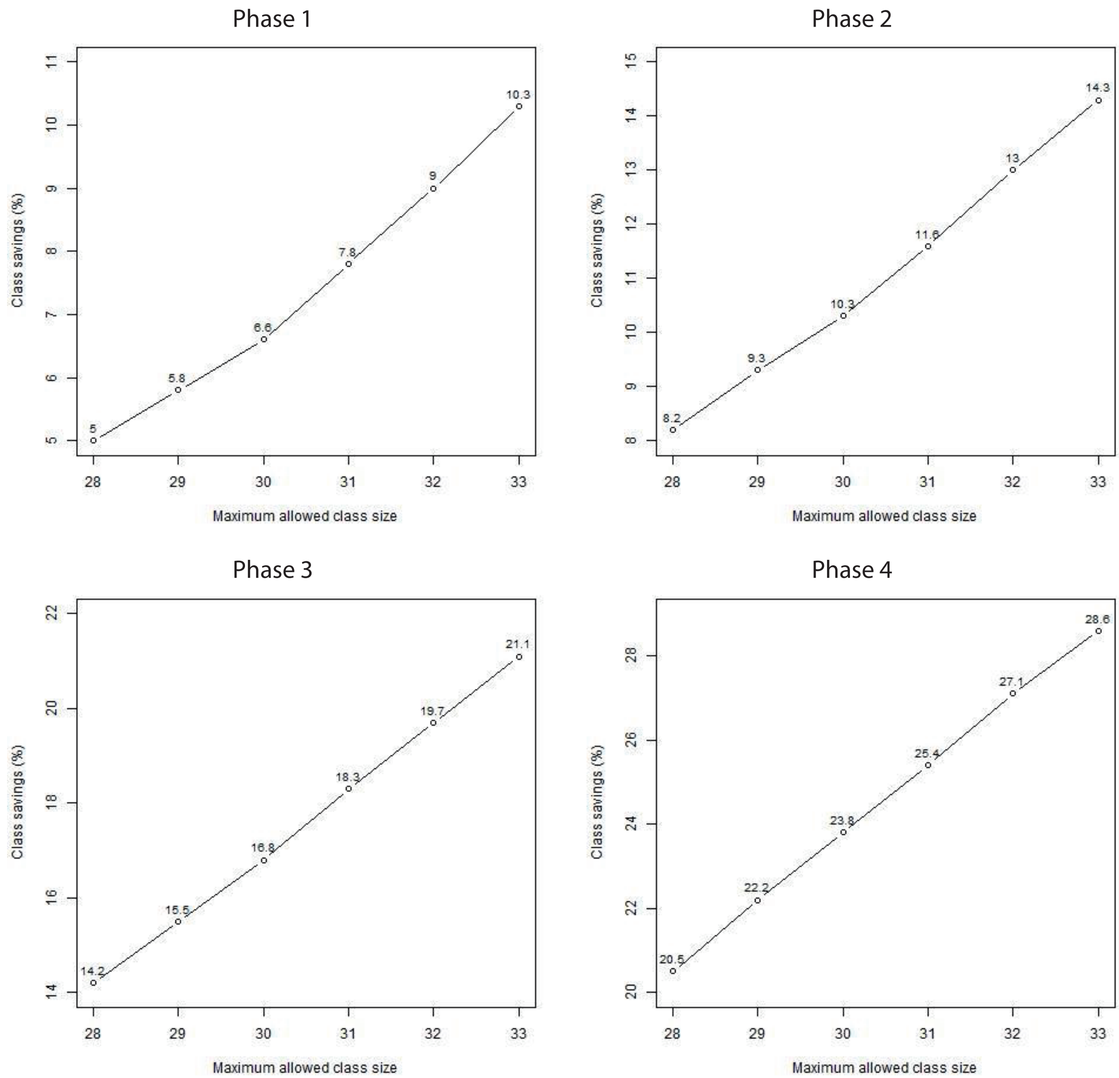
Annex table 6: Travel distance and class savings

	Total savings (30 min limit)	Total savings (35 min limit)	Total savings (40 min limit)	Total savings (45 min limit)
Class savings	11,796	12,580	13,465	14,596
Class savings (% of all)	6.6%	7.1%	7.6%	8.2%

2.2.2 Class savings and class size

The maximum number of students in each class is also a policy variable that will influence saving potential. In Annex table 7 , it is shown that increasing class sizes turns more or less proportionally into larger savings potential. For instance, in Phase 3, each additional larger maximum class size is roughly equivalent to 1.3 percentage points more savings potential.

Annex table 7: Class savings by class size



2.2 Teacher savings

As laid out in the method section, teacher savings are estimated based on three different methods or assumptions regarding the relationship between number of classes and teacher savings. Annex table 8 shows that results are not sensitive to choice of estimation method. It further shows that if schools are allowed to be within a 20-minutes drive between each other, around 24 percent of all teachers would be redundant, while 6.6 percent would be redundant if clustering took place within schools that are within a 30-minutes walk from each other. Appendix 4.9 includes additional tables with result on teacher savings for schools, location and phase.

Annex table 8 Teacher savings by phase

	Phase 1	Phase 2	Phase 3	Phase 4
Class savings	11,796	18,37	29,891	42,253
Teacher savings 1	23,297	35,368	56,385	72,019
Teacher savings 1 (% of all)	7.9%	12.0%	19.2%	24.5%
Teacher savings 2	19,419	31,903	52,665	68,681
Teacher savings 2 (% of all)	6.6%	10.8%	17.9%	23.4%
Teacher savings 3	19,171	31,071	51,606	67,854
Teacher savings 3 (% of all)	6.5%	10.6%	17.6%	23.1%

2.3 Teacher saving that could be obtained through retirement

In Ukraine, 29,278 teachers in primary and secondary schools analyzed in this paper are age 61 or above. Of these, more than 20,000 are employed within the schools that have scope for saving classes and teachers. Hence, one could let these teachers retire, without a replacement and thereby achieve savings, with minimal political costs. Further, another 31,225 teachers are age 55-60, suggesting they will be retiring within the next seven years. Hence, if these school networks are carefully designed and implemented over a 5-7 year horizon, most staff reductions can be made through retirements. Annex table 9 shows that many of the retirements are in fact within the clusters. If further flexibility is assumed, where teachers are moved around to fill gaps within the Oblast or the entire Ukraine, then further reductions in teacher staff could be achieved through retirements.

Annex table 9: Teacher savings and retirement

	Teacher flexibility		
	Move within network	Move within oblast	Move within Ukraine
Retirements with no adjustment period	20,963 – 21,792	27,196 - 29,278	29,278
Retirements with adjustment period of up to 7 years	35,762 – 38,038	48,755 - 55,694	60,503

1.

3. Tables with more details for annex 4 analysis

3.1 Tables with class savings

Annex table 10: Option 1: Walking a maximum of 30 minutes

	Total	Primary (grade 1 to 4)		Secondary (Grade 5 to 12)	
		Urban	Rural	Urban	Rural
Classes before networks	177,659	28,745	45,557	40,758	62,599
Classes after networks	165,863	27,143	42,457	37,119	59,144
Class savings	11,796	1,602	3,100	3,639	3,455
Class savings (% of all)	6.6%	5.6%	6.8%	8.9%	5.5%
Schools affected	29.0%	82.9%	11.4%	84.0%	9.9%
Students affected	56.3%	86.5%	20.1%	87.5%	18.9%

Annex table 11: Option 2: walking up to 60 minutes for secondary

	Total	Primary (grade 1 to 4)		Secondary (Grade 5 to 12)	
		Urban	Rural	Urban	Urban
Classes before networks	177,659	28,745	45,557	41,647	61,710
Classes after networks	159,289	27,143	42,457	36,388	53,301
Class savings	18,370	1,602	3,100	5,259	8,409
Class savings (% of all)	10.3%	5.6%	6.8%	12.6%	13.6%
Schools affected	51.6%	82.9%	11.4%	96.1%	38.2%
Students affected	66.7%	86.5%	20.1%	97.0%	47.0%

Annex table 12: Option 3:

	Total	Primary (grade 1 to 4)		Secondary (Grade 5 to 12)	
		Urban	Rural	Urban	Urban
Classes before networks	177,659	28,745	45,557	43,045	60,312
Classes after networks	147,768	27,143	42,457	36,438	41,730
Class savings	29,891	1,602	3,100	6,607	18,582
Class savings (% of all)	16.8%	5.6%	6.8%	15.3%	30.8%
Schools affected	84.5%	82.9%	11.4%	98.9%	84.4%
Students affected	77.9%	86.5%	20.1%	99.1%	88.4%

Annex table 13: Option 4

	Total	Primary (grade 1 to 4)		Secondary (Grade 5 to 12)	
		Urban	Rural	Urban	Urban
Classes before networks	177,659	30,218	44,084	43,045	60,312
Classes after networks	135,406	27,032	30,206	36,438	41,730
Class savings	42,253	3,186	13,878	6,607	18,582
Class savings (% of all)	23.8%	10.5%	31.5%	15.3%	30.8%
Schools affected	90.9%	98.8%	83.6%	98.9%	84.4%
Students affected	94.3%	99.0%	88.3%	99.1%	88.4%

3.2 Tables with teacher savings

Annex table 14: Teacher savings option 1

	Total	Primary school level (Grade 1 to 4)		Secondary school level (Grade 5 to 12)	
		Urban	Rural	Urban	Rural
Class savings	11,796	1,602	3,100	3,639	3,455
Teacher savings 1	23,297	1,645	5,677	7,436	8,539
Teacher savings 1 (% of all)	7.9%	4.5%	9.5%	9.1%	7.3%
Teacher savings 2	19,419	2,039	4,012	7,136	6,232
Teacher savings 2 (% of all)	6.6%	5.6%	6.7%	8.7%	5.4%
Teacher savings 3	19,171	2,161	4,029	6,770	6,211
Teacher savings 3 (% of all)	6.5%	6.0%	6.8%	8.3%	5.3%

Annex table 15: Teacher savings option 2

	Total	Primary school level (Grade 1 to 4)		Secondary school level (Grade 5 to 12)	
		Urban	Rural	Urban	Rural
Class savings	18,370	1,602	3,100	5,259	8,409
Teacher savings 1	35,368	1,645	5,677	10,461	17,585
Teacher savings 1 (% of all)	12.0%	4.5%	9.5%	12.6%	15.3%
Teacher savings 2	31,903	2,039	4,012	10,271	15,581
Teacher savings 2 (% of all)	10.8%	5.6%	6.7%	12.3%	13.6%
Teacher savings 3	31,071	2,161	4,029	9,255	15,626
Teacher savings 3 (% of all)	10.6%	6.0%	6.8%	11.1%	13.6%

Annex table 16: Teacher savings option 3

	Total	Primary school level (Grade 1 to 4)		Secondary school level (Grade 5 to 12)	
		Urban	Rural	Urban	Rural
Class savings	29,891	1,602	3,100	6,607	18,582
Teacher savings 1	56,385	1,645	5,677	12,448	36,615
Teacher savings 1 (% of all)	19.2%	4.5%	9.5%	14.6%	32.5%
Teacher savings 2	52,665	2,039	4,012	12,744	33,870
Teacher savings 2 (% of all)	17.9%	5.6%	6.7%	14.9%	30.0%
Teacher savings 3	51,606	2,161	4,029	11,352	34,064
Teacher savings 3 (% of all)	17.6%	6.0%	6.8%	13.3%	30.2%

Annex table 17: Teacher savings option 4

	Total	Primary school level (Grade 1 to 4)		Secondary school level (Grade 5 to 12)	
		Urban	Rural	Urban	Rural
Class savings	42,253	3,186	13,878	6,607	18,582
Teacher savings 1	72,019	3,430	19,526	12,448	36,615
Teacher savings 1 (% of all)	24.5%	9.0%	33.7%	14.6%	32.5%
Teacher savings 2	68,681	4,077	17,990	12,744	33,870
Teacher savings 2 (% of all)	23.4%	10.7%	31.0%	14.9%	30.0%
Teacher savings 3	67,854	4,316	18,122	11,352	34,064
Teacher savings 3 (% of all)	23.1%	11.4%	31.2%	13.3%	30.2%

3.3 Further detail on networks

Phase 1

	Primary school level			Secondary school level		
	Total	**Urban**	**Rural**	**Total**	**Urban**	**Rural**
Total schools	13,812	2,982	10,830	13,184	3,104	10,080
Number of school networks that can be formed	1,433	850	583	1,337	867	470
Schools per school network	2.6	3.0	2.1	2.7	3.1	2.1
Students per school network	415	586	165	776	1,014	337
Students per class before networks are created	18.7	20.3	13.5	23.4	24.9	17.6
Students per class after networks are created	20.9	21.6	18.0	26.9	27.7	23.3

Phase 2

	Primary school level			Secondary school level		
	Total	**Urban**	**Rural**	**Total**	**Urban**	**Rural**
Total schools	13,812	2,982	10,830	13,184	3,104	10,080
Number of school networks that can be formed	1,433	850	583	2,507	727	1,780
Schools per school network	2.6	3.0	2.1	2.8	4.4	2.1
Students per school network	415	586	165	548	1,356	218
Students per class before networks are created	18.7	20.3	13.5	20.7	24.5	14.9
Students per class after networks are created	20.9	21.6	18.0	25.6	28.1	21.0

Phase 3

	Primary school level			Secondary school level		
	Total	**Urban**	**Rural**	**Total**	**Urban**	**Rural**
Total schools	13,812	2,982	10,830	13,184	3,104	10,080
Number of school networks that can be formed	1,433	850	583	3,810	685	3,125
Schools per school network	2.6	3.0	2.1	3.1	5.2	2.6
Students per school network	415	586	165	456	1,486	231
Students per class before networks are created	18.7	20.3	13.5	18.3	23.9	13.8
Students per class after networks are created	20.9	21.6	18.0	24.8	28.2	21.2

Phase 4

	Primary school level			Secondary school level		
	Total	**Urban**	**Rural**	**Total**	**Urban**	**Rural**
Total schools	13,812	2,982	10,830	13,184	3,104	10,080
Number of school networks that can be formed	4,022	835	3,187	3,810	685	3,125
Schools per school network	3.0	4.2	2.7	3.1	5.2	2.6
Students per school network	247	690	131	456	1,486	231
Students per class before networks are created	14.6	19.3	11.0	18.3	23.9	13.8
Students per class after networks are created	19.4	21.6	17.1	24.8	28.2	21.2

3.3 Additional school data on Example 1

	School ID					
	13908	13909	13910	13911	13914 (isolated pri.)	13917 (isolated pri./sec.)
Secondary school classes	9	11	30	32	7	6
Secondary school students	180	210	493	467	73	56
Secondary school teachers	16	18	48	47	10	9
Teachers aged 55+	2	3	9	12	1	6
Primary school classes	6	8	18	20	4	4
Primary school students	122	170	369	367	35	30
Primary school teachers	8	9	28	26	5	4
Empty classrooms	18	14	54	38	17	4

4. Computer code (R-code and Python code) used for analysis

Code for web scraping

```
#################################################
### Scrape the address of each school from the ministry ###
#################################################
input1 = codecs.open("Your_location /website.txt", 'r', 'utf-8')
reader = csv.reader(input1, delimiter='+')
website = [row[0] for row in reader]

full_address = []

print(datetime.datetime.now())
for i in range(0,len(website)):
    time.sleep(1)
    try:
        page = requests.get(website[i])
        table = html.fromstring(page.text)
        if table.xpath('//table/tr/td')[9].text[:7] == "Україна":
            full_address.append(str(table.xpath('//table/tr/td')[9].text))
        else:
            if table.xpath('//table/tr/td')[8].text[:7] == "Україна":
                full_address.append(str(table.xpath('//table/tr/td')[8].text))
            else:
                if table.xpath('//table/tr/td')[10].text[:7] == "Україна":
                    full_address.append(str(table.xpath('//table/tr/td')[10].text))
                else:
                    if table.xpath('//table/tr/td')[11].text[:7] == "Україна":
                        full_address.append(str(table.xpath('//table/tr/td')[11].text))
                    else:
                        if table.xpath('//table/tr/td')[7].text[:7] == "Україна":
                            full_address.append(str(table.xpath('//table/tr/td')[7].text))
                        else:
                            print("Wrong row input " + str(i))
                            full_address.append('wrong_row_input_' + str(i))
    except:
        print('Error input ' + str(i))
        full_address.append('error_input_' + str(i))
        continue
print(datetime.datetime.now())
```

Python code for looking up addresses in yandex and google maps

```
######
### Settlements and full addresses to be looked up in Google Maps
######
input2 = codecs.open("your location/settlement_try1.txt", 'r', 'utf-8')
input3 = codecs.open("your location /settlement_try2.txt", 'r', 'utf-8')
input4 = codecs.open("your location /full_address1.txt", 'r', 'utf-8')
input5 = codecs.open("your location /full_address2.txt", 'r', 'utf-8')
reader = csv.reader(input2, delimiter='+')
settlement1 = [row[0] for row in reader]
reader = csv.reader(input3, delimiter='+')
settlement2 = [row[0] for row in reader]
reader = csv.reader(input4, delimiter='+')
address1 = [row[0] for row in reader]
reader = csv.reader(input5, delimiter='+')
address2 = [row[0] for row in reader]

google_maps = GoogleMaps(api_key='YOUR_API_KEY')
gmaps = googlemaps.Client(key='YOUR_API_KEY')
geolocator = Nominatim()
settlement_lat = []
settlement_lon = []
address_lat = []
address_lon = []
address_error = []
distance_settlement_address = []
distance_settlement_website = []
gps_lat = []
gps_lon = []
gps_final = []
gps_source = []

############################################################
### Get the GPS coordinates for each school from Yandex ###
############################################################
# Load and save yandex lists with latitude and longitude (remember to save twice for some reason)
input1 = codecs.open("your location/yandex_lat.txt", 'r', 'utf-8')
input2 = codecs.open("your location /yandex_lon.txt", 'r', 'utf-8')
reader = csv.reader(input1, delimiter='+')
yandex_lat = [row[0] for row in reader]
reader = csv.reader(input2, delimiter='+')
yandex_lon = [row[0] for row in reader]

#yandex_lat_file = open('Z:/wb_school_clustering_ukraine/school_lists/yandex_lat.txt', 'w') # save
what has been fetched today
#yandex_lon_file = open('Z:/wb_school_clustering_ukraine/school_lists/yandex_lon.txt', 'w') # save
what has been fetched today
#for item in yandex_lat:
#    yandex_lat_file.write("%s\n" % item)
#for item in yandex_lon:
#    yandex_lon_file.write("%s\n" % item)

# URL inputs for fetching location from Yandex
input1 = codecs.open("Z:/wb_school_clustering_ukraine/school_lists/yandex_url.txt", 'r', 'utf-8')
input2 = codecs.open("Z:/wb_school_clustering_ukraine/school_lists/yandex_id.txt", 'r', 'utf-8')
reader = csv.reader(input1, delimiter='+')
yandex_url = [row[0] for row in reader]
reader = csv.reader(input2, delimiter='+')
yandex_id = [row[0] for row in reader]

yandex_lat = []
yandex_lon = []

# problem in 4444 - should be lat: 48.650149, lon: 29.199101
# problem in 8721 - should be lat: 49.921896, lon: 35.873619
print(datetime.datetime.now())
for i in range(0,len(yandex_url)):
    time.sleep(0.5)
    page = requests.get(yandex_url[i])
    table = html.fromstring(page.text)
    soup = BeautifulSoup(page.text, 'html.parser')
```

```
    test1 = str(soup.find_all(text=re.compile('center')))
    test2 = test1.split('"center":[')
    test3 = test2[1].split('],"bbox"')
    test4 = test3[0].split(',')
    yandex_lat.append(float(test4[1]))
    yandex_lon.append(float(test4[0]))
print(datetime.datetime.now())

#yandex_lat.append(48.650149) # for line 4444 (id = 7664)
#yandex_lon.append(29.199101) # for line 4444 (id = 7664)
#yandex_lat.append(49.921896) # for line 8721 (id = 13638)
#yandex_lon.append(35.873619) # for line 8721 (id = 13638)
```

Python code for selection of GPS locations

```
# Find preferred latitude and longitude
preferred_lat = []
preferred_lon = []
GPS_source = []

for i in range(0,len(yandex_url)):
    if distance_yandex_google[i]<1 and distance_yandex_website[i]<1:
        preferred_lat.append(yandex_lat[i])
        preferred_lon.append(yandex_lon[i])
        GPS_source.append('found_yandex_google_ministry')
    elif distance_yandex_google[i]<1:
        preferred_lat.append(yandex_lat[i])
        preferred_lon.append(yandex_lon[i])
        GPS_source.append('found_yandex_google')
    elif distance_yandex_website[i]<1:
        preferred_lat.append(yandex_lat[i])
        preferred_lon.append(yandex_lon[i])
        GPS_source.append('found_yandex_ministry')
    elif distance_yandex_settlement[i]<2:
        preferred_lat.append(yandex_lat[i])
        preferred_lon.append(yandex_lon[i])
        GPS_source.append('found_yandex')
    else:
        preferred_lat.append(yandex_lat[i])
        preferred_lon.append(yandex_lon[i])
        GPS_source.append('lookup')


preferred_lat = []
preferred_lon = []
GPS_source = []

input1 = codecs.open("Z:/wb_school_clustering_ukraine/school_lists/urban_rural.txt", 'r', 'utf-8')
reader = csv.reader(input1, delimiter='+')
urban_rural = [row[0] for row in reader]

for i in range(0,len(yandex_url)):
    if distance_yandex_google[i]<1 and distance_yandex_website[i]<1:
        preferred_lat.append(yandex_lat[i])
        preferred_lon.append(yandex_lon[i])
        GPS_source.append('found_yandex_google_ministry')
    elif distance_yandex_google[i]<1:
        preferred_lat.append(yandex_lat[i])
        preferred_lon.append(yandex_lon[i])
        GPS_source.append('found_yandex_google')
    elif urban_rural[i]=='Місто':
        if distance_yandex_website[i]<1:
            preferred_lat.append(yandex_lat[i])
            preferred_lon.append(yandex_lon[i])
            GPS_source.append('found_yandex_ministry')
        elif distance_google_website[i]<1:
            preferred_lat.append(google_lat[i])
            preferred_lon.append(google_lon[i])
            GPS_source.append('found_google_ministry')
        elif distance_yandex_settlement[i]<2:
            preferred_lat.append(yandex_lat[i])
            preferred_lon.append(yandex_lon[i])
```

```
            GPS_source.append('found_yandex')
        else:
            preferred_lat.append(yandex_lat[i])
            preferred_lon.append(yandex_lon[i])
            GPS_source.append('lookup')
    elif urban_rural[i]!='Місто':
        if distance_google_website[i]<1:
            preferred_lat.append(google_lat[i])
            preferred_lon.append(google_lon[i])
            GPS_source.append('found_google_ministry')
        elif distance_yandex_website[i]<1:
            preferred_lat.append(yandex_lat[i])
            preferred_lon.append(yandex_lon[i])
            GPS_source.append('found_yandex_ministry')
        elif distance_yandex_settlement[i]<2:
            preferred_lat.append(yandex_lat[i])
            preferred_lon.append(yandex_lon[i])
            GPS_source.append('found_yandex')
        else:
            preferred_lat.append(yandex_lat[i])
            preferred_lon.append(yandex_lon[i])
            GPS_source.append('lookup')
```

R-code for limiting combinations to be used in google distances calculations.

```
###################################################################################################
###        Generate list of school pairs with less than 5 km between them (bee-line distance)
              ###
###################################################################################################
# Load input for R
input <- read.table("R_input/school_data.txt", header=TRUE)

# Some schools need to be excluded as key information is missing
input <- input[input$school_drop==0,]

# Set maximum allowed walking distance (bee-line) between schools
max_distance_w = 5

# Generate dummies for whether a school offers primary school classes and secondary school classes
for (i in 1:nrow(input)) {
  if (input$classes_psei[i]>0 | input$classes_gei[i]>0 | input$classes_grade2[i]>0 | input$classes_
grade3[i]>0 | input$classes_grade4[i]>0) {
    input$classes_pri[i] <- 1
  } else {input$classes_pri[i] <- 0}

  if (sum(input$classes_grade5[i],input$classes_grade6[i],input$classes_grade7[i],input$classes_
grade8[i],input$classes_grade9[i],
          input$classes_grade10[i],input$classes_grade11[i],input$classes_grade12[i]) > 0) {
    input$classes_sec[i] <- 1
  } else {input$classes_sec[i] <- 0}
}

# Load coordinates and plot to see them
coordinates <- cbind(input$preferred_lon, input$preferred_lat, input$school_id)
plot(coordinates[,1:2])

# Find bee-line distances in kilometers between all schools
distance_matrix <- distm(coordinates[,1:2], fun = distHaversine)/1000

# Generate empty matrix to be filled out with ones for school pairs with distance below maximum
allowed
matrix_temp <- matrix(0, nrow=nrow(distance_matrix), ncol=ncol(distance_matrix))

# Create empty list to be filled out with school pairs where distance is below maximum allowed
lookup_google <- setNames(data.frame(matrix(ncol = 3, nrow = 0)), c("school1", "school2", "bee_
distance"))

temp_row <- setNames(data.frame(matrix(ncol = 3, nrow = 1)), c("school1", "school2", "bee_
distance"))

for (i in 1:nrow(distance_matrix)) {
  for (j in 1:ncol(distance_matrix)) {
    if (distance_matrix[i,j] < max_distance_w & i>j) {
```

```
      matrix_temp[i,j] = 1
      temp_row$school1[1] <- i
      temp_row$school2[1] <- j
      temp_row$bee_distance[1] <- distance_matrix[i,j]
      lookup_google <- rbind(lookup_google, temp_row)
    }
  }
}

lookup_google$pair_id <- list(0) # school pair id
for (i in 1:nrow(lookup_google)) {
  lookup_google$pair_id[i] <- i}
lookup_google$pair_id <- as.numeric(lookup_google$pair_id)

# Load school id to be merged with lookup_google
school_id <- data.frame(input$school_id) # school id
colnames(school_id) <- 'school_id1'

school_id$school1 <- list(nrow(school_id)) # list from 1 to number of schools (current school id in
lookup_google)
for (i in 1:nrow(school_id)) {
  school_id$school1[i] <- i}

lookup_google <- merge(lookup_google, school_id, by="school1") # merge school id of school1 in
lookup_google

colnames(school_id) <- c('school_id2', 'school2')
lookup_google <- merge(lookup_google, school_id, by="school2") # merge school id of school2 in
lookup_google

# Merge coordinates of the two schools to be looked up
lookup_google <- merge(lookup_google, coordinates, by.x="school_id1", by.y="V3")
colnames(lookup_google)[7:8] <- c('longitude1', 'latitude1')
lookup_google <- merge(lookup_google, coordinates, by.x="school_id2", by.y="V3")
colnames(lookup_google)[9:10] <- c('longitude2', 'latitude2')

# Order to get the same order as original lookup_google
lookup_google <- lookup_google[order(lookup_google$pair_id),]
lookup_google <- lookup_google[, c(2,1,4,3,5,6,7,8,9,10)] # change column order

# Exporting school pairs to be looked up for travel distance (using Python)
write.table(lookup_google, "R_output/school_pairs_lookup.txt", sep=",", row.names=FALSE)
Code for look-up travel distances in google map
Code for clustering analysis
Code for savings analysis
```

Python code for travel distance between school from google maps

```
##################
### Travel distances and duration between schools (walking 5 km bee-line max)
##################
school_pairs = np.loadtxt(Your_location/school_pairs_lookup.txt', \
                          skiprows=1, delimiter=',')

distance_w = []
duration_w = []

# Lookup strategy: 0 to 7979, manual, 7980 to 10202, manual, 10203 to 19705,
#                  manual, 19706 to 31069, manual, manual, 31071 to 31112,
#                  manual, 31113 to 35832, manual, manual, manual, manual,
#                  manual, 35837 to 36630, manual, 36631 to 44680, manual,
#                  44681 to end.
print(datetime.datetime.now())
for i in range(44681,len(school_pairs[:,1])):
    page = requests.get('https://maps.googleapis.com/maps/api/distancematrix/json?origins=' + \
      str(school_pairs[i,7]) + ',' + str(school_pairs[i,6]) + '&destinations=' + \
      str(school_pairs[i,9]) + ',' + str(school_pairs[i,8]) + \
      '&key=YOUR_API_KEY' + '&mode=walking')

    page_data = json.loads(page.text)

    distance_w.append(page_data['rows'][0]['elements'][0]['distance']['text'])
```

```
    duration_w.append(page_data['rows'][0]['elements'][0]['duration']['text'])
print(datetime.datetime.now())

distance_w.append('7.7 km') # i = 7979
duration_w.append('1 hour 36 mins') # i = 7979

distance_w.append('5.8 km') # i = 10202
duration_w.append('1 hour 13 mins') # i = 10202

distance_w.append('5.5 km') # i = 19705
duration_w.append('1 hour 9 mins') # i = 19705

distance_w.append('5.4 km') # i = 31069
duration_w.append('1 hour 4 mins') # i = 31069

distance_w.append('6.1 km') # i = 31070
duration_w.append('1 hour 16 mins') # i = 31070

distance_w.append('5.5 km') # i = 31112
duration_w.append('1 hour 9 mins') # i = 31112

distance_w.append('50 km') # i = 35832
duration_w.append('10 hours 25 mins') # i = 35832

distance_w.append('50 km') # i = 35833
duration_w.append('10 hours 25 mins') # i = 35833

distance_w.append('50 km') # i = 35834
duration_w.append('10 hours 25 mins') # i = 35834

distance_w.append('50 km') # i = 35835
duration_w.append('10 hours 25 mins') # i = 35835

distance_w.append('50 km') # i = 35836
duration_w.append('10 hours 25 mins') # i = 35836

distance_w.append('5.4 km') # i = 36630
duration_w.append('1 hour 4 mins') # i = 36630

distance_w.append('4.4 km') # i = 44680
duration_w.append('56 mins') # i = 44680


distance_w2 = []
duration_w2 = []
for i in range(0,len(distance_w)):
    temp1 = distance_w[i].split()
    if temp1[1] == 'm':
        distance_w2.append(float(temp1[0])/1000)
    elif temp1[1] == 'km':
        distance_w2.append(float(temp1[0]))
    else:
        distance_w2.append('Error')

    temp2 = duration_w[i].split()
    if temp2[1][0:3] == 'hou':
        duration_w2.append(float(temp2[0]) * 60 + float(temp2[2]))
    elif temp2[1][0:3] == 'min':
        duration_w2.append(float(temp2[0]))
    else:
        duration_w2.append('Error')

# Save travel distances and duration (walking 5 km bee-line max)
travel_distance_file = open('Your_location/travel_distance_5w.txt', 'w')
travel_duration_file = open(Your_location /travel_duration_5w.txt', 'w')
for item in distance_w2:
    travel_distance_file.write("%s\n" % item)
for item in duration_w2:
    travel_duration_file.write("%s\n" % item)
```

R-code for clustering of schools

```
####################################################################################################
#
### STEP 4  Cluster schools for primary school level (set maximum allowed walking duration first) ###
###         These networks are used for phase 1, phase 2, and phase 3
###
####################################################################################################
#
# Set maximum allowed walking duration between schools for primary school students
max_duration_w = 30

# Drop school pairs without primary school classes, with walking duration above maximum allowed, and
different languages
school_pairs1 <- school_pairs[school_pairs$duration_w<=max_duration_w,]
school_pairs1 <- school_pairs1[school_pairs1$classes_pri1+school_pairs1$classes_pri2==2,]
school_pairs1 <- school_pairs1[school_pairs1$language1==school_pairs1$language2,]

# Generate empty matrix to be filled with walking durations less than 30 minutes
duration_matrix_30w <- matrix(100, nrow=nrow(input), ncol=nrow(input))

for (row in 1:nrow(school_pairs1)) {
  school1_temp <- school_pairs1[row, "school1"]
  school2_temp  <- school_pairs1[row, "school2"]
  duration_temp <- school_pairs1[row, "duration_w"]
  duration_matrix_30w[school1_temp,school2_temp] = duration_temp
}

# Change duration matrix between schools to Class 'dist' atomic
duration_list_30w <- as.dist(duration_matrix_30w)

# Generate a dendogram used for the identification of clusters
dendogram <- hclust(duration_list_30w)

# Identify clusters based on the dendogram and maximum allowed walking duration between schools
clusters_30w <- cutree(dendogram, h = max_duration_w)

# Change to data frame and merge with main data
clusters_30w_df <- as.data.frame(clusters_30w)

clusters_pri <- cbind(clusters_30w_df, input)
colnames(clusters_pri)[1] <- "cluster_pri_id"

# Exporting clusters for primary school classes (phase 1 networks are the same as phase 2 and 3
networks)
write.table(clusters_pri, "R_output/clusters_primary_classes_phase1.txt", sep=";", row.names=FALSE)
write.table(clusters_pri, "R_output/clusters_primary_classes_phase2.txt", sep=";", row.names=FALSE)
write.table(clusters_pri, "R_output/clusters_primary_classes_phase3.txt", sep=";", row.names=FALSE)


################################################################################################
### STEP 5: Cluster schools for secondary school level based on two different approaches:     ###
###         1) walking duration below 30 minutes (used for phase 1)                             ###
###         2) biking duration below 30 minutes, assuming speed of 10 km/h (used for phase 2) ###
################################################################################################
# Set maximum allowed walking distance/duration between schools for secondary school students
max_duration_w = 30
max_distance_w = 5

# Drop school pairs without secondary school classes, with walking duration above 30 min, and
different languages
school_pairs2 <- school_pairs[school_pairs$duration_w<=max_duration_w,]
school_pairs2 <- school_pairs2[school_pairs2$classes_sec1+school_pairs2$classes_sec2==2,]
school_pairs2 <- school_pairs2[school_pairs2$language1==school_pairs2$language2,]

# Drop school pairs without secondary school classes, with walking distance above 5 km, and different
languages
school_pairs3 <- school_pairs[school_pairs$distance_w<=max_distance_w,]
school_pairs3 <- school_pairs3[school_pairs3$classes_sec1+school_pairs3$classes_sec2==2,]
school_pairs3 <- school_pairs3[school_pairs3$language1==school_pairs3$language2,]

# Generate empty matrix to be filled with walking durations less than 30 minutes
duration_matrix_30w_sec <- matrix(100, nrow=nrow(input), ncol=nrow(input))

for (row in 1:nrow(school_pairs2)) {
```

```
  school1_temp <- school_pairs2[row, "school1"]
  school2_temp  <- school_pairs2[row, "school2"]
  distance_temp <- school_pairs2[row, "distance_w"]
  duration_matrix_30w_sec[school1_temp,school2_temp] = distance_temp
}

# Generate empty matrix to be filled with walking distances less than 5 km
distance_matrix_5w <- matrix(100, nrow=nrow(input), ncol=nrow(input))

for (row in 1:nrow(school_pairs3)) {
  school1_temp <- school_pairs3[row, "school1"]
  school2_temp  <- school_pairs3[row, "school2"]
  distance_temp <- school_pairs3[row, "distance_w"]
  distance_matrix_5w[school1_temp,school2_temp] = distance_temp
}

# Change duration matrix between schools to Class 'dist' atomic
duration_list_30w_sec <- as.dist(duration_matrix_30w_sec)
distance_list_5w <- as.dist(distance_matrix_5w)

# Generate a dendogram used for the identification of clusters
dendogram1 <- hclust(duration_list_30w_sec)
dendogram2 <- hclust(distance_list_5w)

# Identify clusters based on the dendogram and maximum allowed walking duration/distance between
schools
clusters_30w <- cutree(dendogram1, h = max_duration_w)
clusters_5w <- cutree(dendogram2, h = max_distance_w)

# Change to data frame and merge with main data
clusters_30w_df <- as.data.frame(clusters_30w)
clusters_5w_df <- as.data.frame(clusters_5w)

clusters_sec1 <- cbind(clusters_30w_df, input)
clusters_sec2 <- cbind(clusters_5w_df, input)
colnames(clusters_sec1)[1] <- "cluster_sec_id"
colnames(clusters_sec2)[1] <- "cluster_sec_id"

# Exporting clusters for secondary school classes
write.table(clusters_sec1, "R_output/clusters_secondary_classes_phase1.txt", sep=";", row.
names=FALSE)
write.table(clusters_sec2, "R_output/clusters_secondary_classes_phase2.txt", sep=";", row.
names=FALSE)


####################################################################################################
###########
### STEP 6_1: Cluster schools for primary and secondary school level based on driving duration below
20 min                  ###
###           (used for phase 3 and phase 4)
                              ###
###           Procedure: if a primary school class is isolated in phase 1, find out if there is a
school                                ###
###                      or a school network within 20 minutes of driving and let the isolated class
join                          ###
###                      the closest phase 1 school or school network
                                      ###
###                      if a secondary school class is isolated in phase 2, do the same as for
isolated                               ###
###                      primary classes.
                              ###
####################################################################################################
###########
# Set maximum allowed distance (bee-line) between schools to 12 km and maximum allowed driving time
to 20 minutes
max_distance_d = 12
max_duration_d = 20
remove_mountain = 0 # set to 1 if we want to exclude mountain settlements


# Load data
input <- read.table("R_input/school_data.txt", header=TRUE)
input <- input[input$school_drop==0,]
```

```
clusters_pri <- read.table("R_output/clusters_primary_classes_phase1.txt", header=TRUE, sep=";")
clusters_sec <- read.table("R_output/clusters_secondary_classes_phase2.txt", header=TRUE, sep=";")

for (i in 1:nrow(input)) {
  if (input$classes_psei[i]>0 | input$classes_gei[i]>0 | input$classes_grade2[i]>0 | input$classes_
grade3[i]>0 | input$classes_grade4[i]>0) {
    input$classes_pri[i] <- 1
  } else {input$classes_pri[i] <- 0}

  if (sum(input$classes_grade5[i],input$classes_grade6[i],input$classes_grade7[i],input$classes_
grade8[i],input$classes_grade9[i],
          input$classes_grade10[i],input$classes_grade11[i],input$classes_grade12[i]) > 0) {
    input$classes_sec[i] <- 1
  } else {input$classes_sec[i] <- 0}
}


# Find center point of clusters with more than one school
count_temp <- as.data.frame(table(clusters_pri$cluster_pri_id))
colnames(count_temp) <- c("cluster_pri_id", "clusters_count")
clusters_pri <- merge(clusters_pri, count_temp, by.x="cluster_pri_id", by.y="cluster_pri_id")

temp_gps <- by(clusters_pri, clusters_pri$cluster_pri_id, function(x){
  clusters_temp <- x[,c('cluster_pri_id', 'preferred_lat', 'preferred_lon')]
  cluster_gps <- colMeans(clusters_temp[,1:3])
})
temp_gps <- as.data.frame(t(sapply(temp_gps, I))) # change to data frame
colnames(temp_gps) <- c('cluster_pri_id', 'cluster_lat', 'cluster_lon')

clusters_pri <- merge(clusters_pri, temp_gps, by='cluster_pri_id')


count_temp <- as.data.frame(table(clusters_sec$cluster_sec_id))
colnames(count_temp) <- c("cluster_sec_id", "clusters_count")
clusters_sec <- merge(clusters_sec, count_temp, by.x="cluster_sec_id", by.y="cluster_sec_id")

temp_gps <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){
  clusters_temp <- x[,c('cluster_sec_id', 'preferred_lat', 'preferred_lon')]
  cluster_gps <- colMeans(clusters_temp[,1:3])
})
temp_gps <- as.data.frame(t(sapply(temp_gps, I))) # change to data frame
colnames(temp_gps) <- c('cluster_sec_id', 'cluster_lat', 'cluster_lon')

clusters_sec <- merge(clusters_sec, temp_gps, by='cluster_sec_id')


# Only keep one school for each cluster
clusters_pri_d <- clusters_pri[!duplicated(clusters_pri$cluster_pri_id),]
clusters_sec_d <- clusters_sec[!duplicated(clusters_sec$cluster_sec_id),]


# Find cluster pairs with bee-line distance below 12 km (clusters with more than one school cannot
be paired)
# This is similar to what is done in STEP 1, except in this step we examine clusters rather than
schools.
coordinates_pri <- cbind(clusters_pri_d$cluster_lon, clusters_pri_d$cluster_lat, clusters_pri_
d$cluster_pri_id) # "distm" sees first column as longitude and the second as latitude
coordinates_sec <- cbind(clusters_sec_d$cluster_lon, clusters_sec_d$cluster_lat, clusters_sec_
d$cluster_sec_id) # "distm" sees first column as longitude and the second as latitude

distance_matrix_pri <- distm(coordinates_pri[,1:2], fun = distHaversine)/1000 # find distances in
kilometers between all clusters
distance_list_pri <- as.dist(distance_matrix_pri) # change to data frame
distance_matrix_sec <- distm(coordinates_sec[,1:2], fun = distHaversine)/1000 # find distances in
kilometers between all clusters
distance_list_sec <- as.dist(distance_matrix_sec) # change to data frame

matrix_temp_pri <- matrix(0, nrow=nrow(distance_matrix_pri), ncol=ncol(distance_matrix_pri)) # empty
matrix to be filled with ones for cluster pairs
matrix_temp_sec <- matrix(0, nrow=nrow(distance_matrix_sec), ncol=ncol(distance_matrix_sec)) # empty
matrix to be filled with ones for cluster pairs
```

```
lookup_google_pri <- setNames(data.frame(matrix(ncol = 3, nrow = 0)), c("cluster1", "cluster2",
"bee_distance")) # empty list to be filled out with school pairs
lookup_google_sec <- setNames(data.frame(matrix(ncol = 3, nrow = 0)), c("cluster1", "cluster2",
"bee_distance")) # empty list to be filled out with school pairs

temp_row_pri <- setNames(data.frame(matrix(ncol = 3, nrow = 1)), c("cluster1", "cluster2", "bee_
distance"))
temp_row_sec <- setNames(data.frame(matrix(ncol = 3, nrow = 1)), c("cluster1", "cluster2", "bee_
distance"))

for (i in 1:nrow(distance_matrix_pri)) {
  for (j in 1:ncol(distance_matrix_pri)) {
    if (i<=j) next
    if (distance_matrix_pri[i,j] >= max_distance_d) next
    if (distance_matrix_pri[i,j] < max_distance_d & i>j & (clusters_pri_d$clusters_count[i]==1 |
clusters_pri_d$clusters_count[j]==1) & clusters_pri_d$classes_pri[i]==1 & clusters_pri_d$classes_
pri[j]==1) {
      matrix_temp_pri[i,j] = 1
      temp_row_pri$cluster1[1] <- clusters_pri_d$cluster_pri_id[i]
      temp_row_pri$cluster2[1] <- clusters_pri_d$cluster_pri_id[j]
      temp_row_pri$bee_distance[1] <- distance_matrix_pri[i,j]
      lookup_google_pri <- rbind(lookup_google_pri, temp_row_pri)}}}
for (i in 1:nrow(distance_matrix_sec)) {
  for (j in 1:ncol(distance_matrix_sec)) {
    if (i<=j) next
    if (distance_matrix_sec[i,j] >= max_distance_d) next
    if (distance_matrix_sec[i,j] < max_distance_d & i>j & (clusters_sec_d$clusters_count[i]==1 |
clusters_sec_d$clusters_count[j]==1) & clusters_sec_d$classes_sec[i]==1 & clusters_sec_d$classes_
sec[j]==1) {
      matrix_temp_sec[i,j] = 1
      temp_row_sec$cluster1[1] <- clusters_sec_d$cluster_sec_id[i]
      temp_row_sec$cluster2[1] <- clusters_sec_d$cluster_sec_id[j]
      temp_row_sec$bee_distance[1] <- distance_matrix_sec[i,j]
      lookup_google_sec <- rbind(lookup_google_sec, temp_row_sec)}}}

lookup_google_pri$pair_id <- list(0) # cluster pair id (primary)
for (i in 1:nrow(lookup_google_pri)) {
  lookup_google_pri$pair_id[i] <- i}
lookup_google_pri$pair_id <- as.numeric(lookup_google_pri$pair_id)
lookup_google_sec$pair_id <- list(0) # cluster pair id (secondary)
for (i in 1:nrow(lookup_google_sec)) {
  lookup_google_sec$pair_id[i] <- i}
lookup_google_sec$pair_id <- as.numeric(lookup_google_sec$pair_id)

lookup_google_pri <- merge(lookup_google_pri, coordinates_pri, by.x="cluster1", by.y="V3") # merge
coordinates of the two clusters to be looked up
colnames(lookup_google_pri)[5:6] <- c('longitude1', 'latitude1')
lookup_google_pri <- merge(lookup_google_pri, coordinates_pri, by.x="cluster2", by.y="V3")
colnames(lookup_google_pri)[7:8] <- c('longitude2', 'latitude2')
lookup_google_sec <- merge(lookup_google_sec, coordinates_sec, by.x="cluster1", by.y="V3") # merge
coordinates of the two clusters to be looked up
colnames(lookup_google_sec)[5:6] <- c('longitude1', 'latitude1')
lookup_google_sec <- merge(lookup_google_sec, coordinates_sec, by.x="cluster2", by.y="V3")
colnames(lookup_google_sec)[7:8] <- c('longitude2', 'latitude2')

lookup_google_pri <- lookup_google_pri[order(lookup_google_pri$pair_id),] # order to get the same
order as original lookup_google_pri
lookup_google_pri <- lookup_google_pri[, c(2,1,3,4,5,6,7,8)] # change column order
lookup_google_sec <- lookup_google_sec[order(lookup_google_sec$pair_id),] # order to get the same
order as original lookup_google_sec
lookup_google_sec <- lookup_google_sec[, c(2,1,3,4,5,6,7,8)] # change column order


# Exporting cluster pairs to be looked up for driving travel distance (using Python)
write.table(lookup_google_pri, "R_output/cluster_pairs_lookup_pri.txt", sep=",", row.names=FALSE)
write.table(lookup_google_sec, "R_output/cluster_pairs_lookup_sec.txt", sep=",", row.names=FALSE)
```

R-code for savings analysis

```
###################################################
### STEP 7: Class and teacher savings           ###
###         (manually set phase of interest)    ###
```

```
###         (can be run without steps 1-6)      ###
##################################################
# Set the phase of interest and maximum number of students allowed per class
phase = 4
max_students = 30
remove_mountain = 0


# Load data if you don't want to run steps 1-6
input <- read.table("R_input/school_data.txt", header=TRUE)
input <- input[input$school_drop==0,]
input <- input[input$oblast_id!=13,]

if(phase==1) {
  clusters_pri <- read.table("R_output/clusters_primary_classes_phase1.txt", header=TRUE, sep=";")
  clusters_sec <- read.table("R_output/clusters_secondary_classes_phase1.txt", header=TRUE, sep=";")
} else if(phase==2) {
    clusters_pri <- read.table("R_output/clusters_primary_classes_phase2.txt", header=TRUE, sep=";")
    clusters_sec <- read.table("R_output/clusters_secondary_classes_phase2.txt", header=TRUE,
sep=";")
  } else if(phase==3) {
      clusters_pri <- read.table("R_output/clusters_primary_classes_phase3.txt", header=TRUE,
sep=";")
      clusters_sec <- read.table("R_output/clusters_secondary_classes_phase3.txt", header=TRUE,
sep=";")
    } else if(phase==4) {
        clusters_pri <- read.table("R_output/clusters_primary_classes_phase4.txt", header=TRUE,
sep=";")
        clusters_sec <- read.table("R_output/clusters_secondary_classes_phase4.txt", header=TRUE,
sep=";")
    }

if(remove_mountain==1){
  clusters_pri <- read.table("R_output/clusters_primary_classes_phase4_no_mountain.txt",
header=TRUE, sep=";")
  clusters_sec <- read.table("R_output/clusters_secondary_classes_phase4_no_mountain.txt",
header=TRUE, sep=";")
}

clusters_pri <- clusters_pri[clusters_pri$oblast_id!=13,]
clusters_sec <- clusters_sec[clusters_sec$oblast_id!=13,]

for (i in 1:nrow(input)) {
  if (input$classes_psei[i]>0 | input$classes_gei[i]>0 | input$classes_grade2[i]>0 | input$classes_
grade3[i]>0 | input$classes_grade4[i]>0) {
    input$classes_pri[i] <- 1
  } else {input$classes_pri[i] <- 0}

  if (sum(input$classes_grade5[i],input$classes_grade6[i],input$classes_grade7[i],input$classes_
grade8[i],input$classes_grade9[i],
          input$classes_grade10[i],input$classes_grade11[i],input$classes_grade12[i]) > 0) {
    input$classes_sec[i] <- 1
  } else {input$classes_sec[i] <- 0}
}


# Number of schools in cluster
count_pri <- as.data.frame(table(clusters_pri$cluster_pri_id))
colnames(count_pri) <- c("cluster_pri_id", "schools_cluster_pri")
count_pri <- merge(clusters_pri[,c("cluster_pri_id","school_id")], count_pri, by="cluster_pri_id")
clusters_pri <- merge(count_pri[,c("school_id", "schools_cluster_pri")], clusters_pri, by="school_
id")

count_sec <- as.data.frame(table(clusters_sec$cluster_sec_id))
colnames(count_sec) <- c("cluster_sec_id", "schools_cluster_sec")
count_sec <- merge(clusters_sec[,c("cluster_sec_id","school_id")], count_sec, by="cluster_sec_id")
clusters_sec <- merge(count_sec[,c("school_id", "schools_cluster_sec")], clusters_sec, by="school_
id")


# Classes for each grade before clustering (at cluster level) (first for primary clusters and next
for secondary clusters)
clusters_pri$classes_pri_number <- 0
```

```
for (i in 1:nrow(clusters_pri)) {
  clusters_pri$classes_pri_number[i] = sum(clusters_pri[i,c("classes_preparatory","classes_
psei","classes_gei",
                                                            "classes_grade2","classes_
grade3","classes_grade4")])}
classes_before_pri <- by(clusters_pri, clusters_pri$cluster_pri_id, function(x){ # classes before
clustering within each primary cluster
  id_temp <- mean(x$cluster_pri_id)
  sum <- colSums(x[,c("classes_preparatory","classes_psei","classes_gei","classes_grade2","classes_
grade3","classes_grade4")])
  final <- cbind(id_temp, sum)})

classes_before_pri <- as.data.frame(t(sapply(classes_before_pri, I))) # change to data frame
classes_before_pri <- classes_before_pri[,(ncol(classes_before_pri)/2):ncol(classes_before_pri)] #
drop duplicates of cluster id
colnames(classes_before_pri) <- c("cluster_pri_id","classes_preparatory_pri","classes_psei_
pri","classes_gei_pri",
                                  "classes_grade2_pri","classes_grade3_pri","classes_grade4_pri")


clusters_sec$classes_sec_number <- 0
for (i in 1:nrow(clusters_sec)) {
  clusters_sec$classes_sec_number[i] = sum(clusters_sec[i,c("classes_grade5","classes_
grade6","classes_grade7",
                                                            "classes_grade8","classes_
grade9","classes_grade10",
                                                            "classes_grade11","classes_grade12")])}
classes_before_sec <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){ # classes before
clustering within each secondary cluster
  id_temp <- mean(x$cluster_sec_id)
  sum <- colSums(x[,c("classes_grade5","classes_grade6","classes_grade7","classes_grade8","classes_
grade9","classes_grade10",
                      "classes_grade11","classes_grade12")])
  final <- cbind(id_temp, sum)})

classes_before_sec <- as.data.frame(t(sapply(classes_before_sec, I))) # change to data frame
classes_before_sec <- classes_before_sec[,(ncol(classes_before_sec)/2):ncol(classes_before_sec)] #
drop duplicates of cluster id
colnames(classes_before_sec) <- c("cluster_sec_id","classes_grade5_sec","classes_grade6_
sec","classes_grade7_sec",
                                  "classes_grade8_sec","classes_grade9_sec","classes_grade10_sec",
                                  "classes_grade11_sec","classes_grade12_sec")


# Students for each grade (at cluster level) (first for primary clusters and next for secondary
clusters)
clusters_pri$students_pri_number <- 0
for (i in 1:nrow(clusters_pri)) {
  clusters_pri$students_pri_number[i] = sum(clusters_pri[i,c("students_preparatory","students_
psei","students_gei",
                                                            "students_grade2","students_
grade3","students_grade4")])}
students_pri <- by(clusters_pri, clusters_pri$cluster_pri_id, function(x){ # students within each
primary cluster
  id_temp <- mean(x$cluster_pri_id)
  sum <- colSums(x[,c("students_preparatory","students_psei","students_gei","students_
grade2","students_grade3","students_grade4")])
  final <- cbind(id_temp, sum)})

students_pri <- as.data.frame(t(sapply(students_pri, I))) # change to data frame
students_pri <- students_pri[,(ncol(students_pri)/2):ncol(students_pri)] # drop duplicates of
cluster id
colnames(students_pri) <- c("cluster_pri_id","students_preparatory_pri","students_psei_
pri","students_gei_pri",
                            "students_grade2_pri","students_grade3_pri","students_grade4_pri")


clusters_sec$students_sec_number <- 0
for (i in 1:nrow(clusters_sec)) {
  clusters_sec$students_sec_number[i] = sum(clusters_sec[i,c("students_grade5","students_
grade6","students_grade7",
                                                            "students_grade8","students_
grade9","students_grade10",
```

```
                                                              "students_grade11","students_
grade12")])}
students_sec <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){ # students within each
secondary cluster
  id_temp <- mean(x$cluster_sec_id)
  sum <- colSums(x[,c("students_grade5","students_grade6","students_grade7","students_
grade8","students_grade9","students_grade10",
                      "students_grade11","students_grade12")])
  final <- cbind(id_temp, sum)})

students_sec <- as.data.frame(t(sapply(students_sec, I))) # change to data frame
students_sec <- students_sec[,(ncol(students_sec)/2):ncol(students_sec)] # drop duplicates of
cluster id
colnames(students_sec) <- c("cluster_sec_id","students_grade5_sec","students_grade6_sec","students_
grade7_sec",
                            "students_grade8_sec","students_grade9_sec","students_grade10_sec",
                            "students_grade11_sec","students_grade12_sec")


# Classes for each grade after clustering (at cluster level) (first for primary clusters and next for
secondary clusters)
# If schools have been allowed to have multi-grade classes or allowed to exceed the limit, we assume
they will continue to do so
classes_after_pri <- classes_before_pri
for (i in 1:nrow(classes_after_pri)) {
  classes_after_pri[i,2:ncol(classes_after_pri)] <- round(students_pri[i,2:ncol(classes_after_pri)]/
max_students + 0.4999)
  for (j in 2:ncol(classes_after_pri)) {
    if (classes_after_pri[i,j] > classes_before_pri[i,j]) {
      classes_after_pri[i,j] <- classes_before_pri[i,j]
    }
  }
}

classes_after_sec <- classes_before_sec
for (i in 1:nrow(classes_after_sec)) {
  classes_after_sec[i,2:ncol(classes_after_sec)] <- round(students_sec[i,2:ncol(classes_after_sec)]/
max_students + 0.4999)
  for (j in 2:ncol(classes_after_sec)) {
    if (classes_after_sec[i,j] > classes_before_sec[i,j]) {
      classes_after_sec[i,j] <- classes_before_sec[i,j]
    }
  }
}


### TABLE: MAIN RESULTS SAVINGS OF CLASSES ###
# Total savings of classes
table1 <- matrix(0, nrow=6, ncol=5)
table1[1,1] <- sum(classes_before_pri[2:7]) + sum(classes_before_sec[2:9])
table1[2,1] <- sum(classes_after_pri[2:7]) + sum(classes_after_sec[2:9])
table1[3,1] <- table1[1,1] - table1[2,1]
table1[4,1] <- table1[3,1] / table1[1,1]

# Savings of classes divided by urban or rural clusters
urban_cluster_pri <- by(clusters_pri, clusters_pri$cluster_pri_id, function(x){ # urban or rural
cluster
  mean_urban <- mean(x$urban_dummy)
  if (mean_urban >= 0.5) {
    urban_cluster <- 1}
  if (mean_urban < 0.5) {
    urban_cluster <- 0}
  final <- urban_cluster
})
urban_cluster_pri <- as.data.frame(sapply(urban_cluster_pri, I)) # change to data frame
colnames(urban_cluster_pri)[1] <- "urban_cluster_pri"
savings_classes_pri <- cbind(as.data.frame(rowSums((classes_before_pri - classes_after_pri),2:7)),
urban_cluster_pri)
colnames(savings_classes_pri) <- c("class_savings_pri", "urban_cluster_pri")
classes_before_pri <- cbind(classes_before_pri,urban_cluster_pri)
classes_after_pri <- cbind(classes_after_pri,urban_cluster_pri)

urban_cluster_sec <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){ # urban or rural
```

```
cluster
  mean_urban <- mean(x$urban_dummy)
  if (mean_urban >= 0.5) {
    urban_cluster <- 1}
  if (mean_urban < 0.5) {
    urban_cluster <- 0}
  final <- urban_cluster
})
urban_cluster_sec <- as.data.frame(sapply(urban_cluster_sec, I)) # change to data frame
colnames(urban_cluster_sec)[1] <- "urban_cluster_sec"
savings_classes_sec <- cbind(as.data.frame(rowSums((classes_before_sec - classes_after_sec),2:7)),
urban_cluster_sec)
colnames(savings_classes_sec) <- c("class_savings_sec", "urban_cluster_sec")
classes_before_sec <- cbind(classes_before_sec,urban_cluster_sec)
classes_after_sec <- cbind(classes_after_sec,urban_cluster_sec)

table1[1,2] <- sum(classes_before_pri[classes_before_pri$urban_cluster_pri==1,2:7])
table1[2,2] <- sum(classes_after_pri[classes_after_pri$urban_cluster_pri==1,2:7])
table1[3,2] <- table1[1,2] - table1[2,2]
table1[4,2] <- table1[3,2] / table1[1,2]
table1[1,3] <- sum(classes_before_pri[classes_before_pri$urban_cluster_pri==0,2:7])
table1[2,3] <- sum(classes_after_pri[classes_after_pri$urban_cluster_pri==0,2:7])
table1[3,3] <- table1[1,3] - table1[2,3]
table1[4,3] <- table1[3,3] / table1[1,3]
table1[1,4] <- sum(classes_before_sec[classes_before_sec$urban_cluster_sec==1,2:9])
table1[2,4] <- sum(classes_after_sec[classes_after_sec$urban_cluster_sec==1,2:9])
table1[3,4] <- table1[1,4] - table1[2,4]
table1[4,4] <- table1[3,4] / table1[1,4]
table1[1,5] <- sum(classes_before_sec[classes_before_sec$urban_cluster_sec==0,2:9])
table1[2,5] <- sum(classes_after_sec[classes_after_sec$urban_cluster_sec==0,2:9])
table1[3,5] <- table1[1,5] - table1[2,5]
table1[4,5] <- table1[3,5] / table1[1,5]


# Schools affected
clusters_pri <- merge(clusters_pri, urban_cluster_pri, by.x="cluster_pri_id", by.y=0)
clusters_sec <- merge(clusters_sec, urban_cluster_sec, by.x="cluster_sec_id", by.y=0)

students_pri$students_total_pri <- rowSums(students_pri[,2:7])
students_sec$students_total_sec <- rowSums(students_sec[,2:9])
schools_students_affected <- merge(clusters_pri[,c("cluster_pri_id","school_id","schools_cluster_
pri","urban_cluster_pri")],
                                   students_pri[,c("cluster_pri_id","students_total_pri")],
by="cluster_pri_id")
schools_students_affected <- merge(schools_students_affected,
                                   clusters_sec[,c("cluster_sec_id","school_id","schools_cluster_
sec","urban_cluster_sec")],
                                   by="school_id")
schools_students_affected <- merge(schools_students_affected,students_sec[,c("cluster_sec_
id","students_total_sec")],
                                   by="cluster_sec_id")

table1[5,1] <- nrow(schools_students_affected[schools_students_affected$schools_cluster_pri>1 |
                                              schools_students_affected$schools_cluster_sec>1,]) /
  nrow(schools_students_affected)
table1[5,2] <- nrow(schools_students_affected[schools_students_affected$schools_cluster_pri>1 &
                                              schools_students_affected$urban_cluster_pri==1 &
schools_students_affected$students_total_pri>0,]) /
  nrow(schools_students_affected[schools_students_affected$urban_cluster_pri==1 & schools_students_
affected$students_total_pri>0,])
table1[5,3] <- nrow(schools_students_affected[schools_students_affected$schools_cluster_pri>1 &
                                              schools_students_affected$urban_cluster_pri==0 &
schools_students_affected$students_total_pri>0,]) /
  nrow(schools_students_affected[schools_students_affected$urban_cluster_pri==0 & schools_students_
affected$students_total_pri>0,])
table1[5,4] <- nrow(schools_students_affected[schools_students_affected$schools_cluster_sec>1 &
                                              schools_students_affected$urban_cluster_sec==1 &
schools_students_affected$students_total_sec>0,]) /
  nrow(schools_students_affected[schools_students_affected$urban_cluster_sec==1 & schools_students_
affected$students_total_sec>0,])
table1[5,5] <- nrow(schools_students_affected[schools_students_affected$schools_cluster_sec>1 &
                                              schools_students_affected$urban_cluster_sec==0 &
schools_students_affected$students_total_sec>0,]) /
```

```
  nrow(schools_students_affected[schools_students_affected$urban_cluster_sec==0 & schools_students_
affected$students_total_sec>0,])

# Students affected
for (i in 1:nrow(schools_students_affected)) {
  schools_students_affected$students_total_pri[i] <- schools_students_affected$students_total_pri[i]/
schools_students_affected$schools_cluster_pri[i]
  schools_students_affected$students_total_sec[i] <- schools_students_affected$students_total_sec[i]/
schools_students_affected$schools_cluster_sec[i]}

table1[6,1] <- (sum(schools_students_affected$students_total_pri[schools_students_affected$schools_
cluster_pri>1]) +
                 sum(schools_students_affected$students_total_sec[schools_students_affected$schools_
cluster_sec>1])) /
  (sum(schools_students_affected$students_total_pri) + sum(schools_students_affected$students_total_
sec))
table1[6,2] <- sum(schools_students_affected$students_total_pri[schools_students_affected$schools_
cluster_pri>1 &
                                                                 schools_students_affected$urban_
cluster_pri==1]) /
  sum(schools_students_affected$students_total_pri[schools_students_affected$urban_cluster_pri==1])
table1[6,3] <- sum(schools_students_affected$students_total_pri[schools_students_affected$schools_
cluster_pri>1 &
                                                                 schools_students_affected$urban_
cluster_pri==0]) /
  sum(schools_students_affected$students_total_pri[schools_students_affected$urban_cluster_pri==0])
table1[6,4] <- sum(schools_students_affected$students_total_sec[schools_students_affected$schools_
cluster_sec>1 &
                                                                 schools_students_affected$urban_
cluster_sec==1]) /
  sum(schools_students_affected$students_total_sec[schools_students_affected$urban_cluster_sec==1])
table1[6,5] <- sum(schools_students_affected$students_total_sec[schools_students_affected$schools_
cluster_sec>1 &
                                                                 schools_students_affected$urban_
cluster_sec==0]) /
  sum(schools_students_affected$students_total_sec[schools_students_affected$urban_cluster_sec==0])


### TABLE: SAVINGS OF TEACHERS ###
# Teacher/class ratios for each school (assume share of music etc teachers are split between primary
and secondary classes)
teacher_class_sec <- clusters_sec[clusters_sec$classes_sec_number>0 & clusters_sec$students_sec_
number>0
                                  ,c("cluster_sec_id","school_id","urban_dummy","teachers_
pri","teachers_sec",
                                     "teachers_other","classes_sec_number","students_sec_number")]

teacher_class_sec$teachers_sec2 <- teacher_class_sec$teachers_sec + teacher_class_sec$teachers_
other*
  teacher_class_sec$teachers_sec/(teacher_class_sec$teachers_pri + teacher_class_sec$teachers_sec)

teacher_class_sec$ratio_sec <- teacher_class_sec$teachers_sec2/teacher_class_sec$classes_sec_number

teacher_class_pri <- clusters_pri[clusters_pri$classes_pri_number>0 & clusters_pri$students_pri_
number>0
                                  ,c("cluster_pri_id","school_id","urban_dummy","teachers_
pri","teachers_sec",
                                     "teachers_other","classes_pri_number","students_pri_number")]

teacher_class_pri$teachers_pri2 <- teacher_class_pri$teachers_pri + teacher_class_pri$teachers_
other*
  teacher_class_pri$teachers_pri/(teacher_class_pri$teachers_pri + teacher_class_pri$teachers_sec)

teacher_class_pri$ratio_pri <- teacher_class_pri$teachers_pri2/teacher_class_pri$classes_pri_number

input$teachers_pri2 <- input$teachers_pri + input$teachers_other*input$teachers_pri/(input$teachers_
pri + input$teachers_sec)
input$teachers_sec2 <- input$teachers_sec + input$teachers_other*input$teachers_sec/(input$teachers_
pri + input$teachers_sec)

input$ratio_pri <- 0
input$ratio_sec <- 0
for (i in 1:nrow(input)) {
```

```
  input$ratio_pri[i] <- input$teachers_pri2[i]/sum(input[i,c("classes_preparatory","classes_
psei","classes_gei","classes_grade2",
                                                             "classes_grade3","classes_grade4")])
  input$ratio_sec[i] <- input$teachers_sec2[i]/sum(input[i,c("classes_grade5","classes_
grade6","classes_grade7","classes_grade8",
                                                             "classes_grade9","classes_
grade10","classes_grade11","classes_grade12")])}

# Teacher savings using overall average ratios by primary/secondary and urban/rural
ratio_pri_urb <- mean(input[input$urban_dummy==1 & input$students_total>50 & is.finite(input$ratio_
pri),ncol(input)-1], na.rm=T)
ratio_pri_rur <- mean(input[input$urban_dummy==0 & input$students_total>50 & is.finite(input$ratio_
pri),ncol(input)-1], na.rm=T)
ratio_sec_urb <- mean(input[input$urban_dummy==1 & input$students_total>50 & is.finite(input$ratio_
sec),ncol(input)], na.rm=T)
ratio_sec_rur <- mean(input[input$urban_dummy==0 & input$students_total>50 & is.finite(input$ratio_
sec),ncol(input)], na.rm=T)

input <- merge(input, clusters_pri[,c("school_id","cluster_pri_id")], by="school_id")
teachers_before_pri <- by(input, input$cluster_pri_id, function(x){
  final <- round(sum(x$teachers_pri2, na.rm=TRUE))})
teachers_before_pri <- as.data.frame(sapply(teachers_before_pri, I)) # change to data frame
colnames(teachers_before_pri)[1] <- "teachers"
classes_before_pri <- cbind(classes_before_pri,teachers_before_pri)

input <- merge(input, clusters_sec[,c("school_id","cluster_sec_id")], by="school_id")
teachers_before_sec <- by(input, input$cluster_sec_id, function(x){
  final <- round(sum(x$teachers_sec2, na.rm=TRUE))})
teachers_before_sec <- as.data.frame(sapply(teachers_before_sec, I)) # change to data frame
colnames(teachers_before_sec)[1] <- "teachers"
classes_before_sec <- cbind(classes_before_sec,teachers_before_sec)

savings_teachers_pri1 <- 0
for (i in 1:nrow(classes_before_pri)) {
  if (classes_before_pri$urban_cluster_pri[i]==1) {
    savings_teachers_pri1[i] <- classes_before_pri[i,9] - round(rowSums(classes_after_
pri[i,2:7])*ratio_pri_urb)}
  else { savings_teachers_pri1[i] <- classes_before_pri[i,9] - round(rowSums(classes_after_
pri[i,2:7])*ratio_pri_rur)}
}
savings_teachers_pri1 <- as.data.frame(savings_teachers_pri1)

savings_teachers_sec1 <- 0
for (i in 1:nrow(classes_before_sec)) {
  if (classes_before_sec$urban_cluster_sec[i]==1) {
    savings_teachers_sec1[i] <- classes_before_sec[i,11] - round(rowSums(classes_after_
sec[i,2:9])*ratio_sec_urb)}
  else { savings_teachers_sec1[i] <- classes_before_sec[i,11] - round(rowSums(classes_after_
sec[i,2:9])*ratio_sec_rur)}
}
savings_teachers_sec1 <- as.data.frame(savings_teachers_sec1)

table2 <- matrix(0, nrow=7, ncol=5)
table2[2,2] <- sum(classes_before_pri[classes_before_pri$urban_cluster_pri==1,9]) -
sum(round(rowSums(classes_after_pri[classes_after_pri$urban_cluster_pri==1,2:7])*ratio_pri_urb))
table2[2,3] <- sum(classes_before_pri[classes_before_pri$urban_cluster_pri==0,9]) -
sum(round(rowSums(classes_after_pri[classes_after_pri$urban_cluster_pri==0,2:7])*ratio_pri_rur))
table2[2,4] <- sum(classes_before_sec[classes_before_sec$urban_cluster_sec==1,11]) -
sum(round(rowSums(classes_after_sec[classes_after_sec$urban_cluster_sec==1,2:9])*ratio_sec_urb))
table2[2,5] <- sum(classes_before_sec[classes_before_sec$urban_cluster_sec==0,11]) -
sum(round(rowSums(classes_after_sec[classes_after_sec$urban_cluster_sec==0,2:9])*ratio_sec_rur))
table2[2,1] <- table2[2,2] + table2[2,3] + table2[2,4] + table2[2,5]

table2[3,1] <- table2[2,1] / (sum(teachers_before_pri) + sum(teachers_before_sec))
table2[3,2] <- table2[2,2] / sum(classes_before_pri[classes_before_pri$urban_cluster_pri==1,9])
table2[3,3] <- table2[2,3] / sum(classes_before_pri[classes_before_pri$urban_cluster_pri==0,9])
table2[3,4] <- table2[2,4] / sum(classes_before_sec[classes_before_sec$urban_cluster_sec==1,11])
table2[3,5] <- table2[2,5] / sum(classes_before_sec[classes_before_sec$urban_cluster_sec==0,11])


# Teacher savings using weighted average of ratios within a network
ratio_pri_cluster <- 0
ratio_sec_cluster <- 0
```

```
for (i in 1:nrow(classes_before_pri)) {
  ratio_pri_cluster[i] <- teachers_before_pri[i,1] / sum(classes_before_pri[i,2:7])}
for (i in 1:nrow(classes_before_sec)) {
  ratio_sec_cluster[i] <- teachers_before_sec[i,1] / sum(classes_before_sec[i,2:9])}
ratio_pri_cluster1 <- as.data.frame(sapply(ratio_pri_cluster,I))
ratio_sec_cluster1 <- as.data.frame(sapply(ratio_sec_cluster,I))

savings_classes_pri <- 0
savings_teachers_pri2 <- 0
teachers_after_pri <- 0
for (i in 1:nrow(classes_before_pri)) { # find savings for each primary cluster
  savings_classes_pri[i] <- sum(classes_before_pri[i,2:(ncol(classes_before_pri)-2)])-sum(classes_
after_pri[i,2:(ncol(classes_after_pri)-1)])
  teachers_after_pri[i] <- round((sum(classes_before_pri[i,2:(ncol(classes_before_pri)-2)])-savings_
classes_pri[i])*ratio_pri_cluster1[i,])
  savings_teachers_pri2[i] <- teachers_before_pri[i,1] - teachers_after_pri[i]}

savings_teachers_pri2 <- cbind(as.data.frame(sapply(savings_teachers_pri2,I)),urban_cluster_pri) #
change to data frame

savings_classes_sec <- 0
savings_teachers_sec2 <- 0
teachers_after_sec <- 0
for (i in 1:nrow(classes_before_sec)) { # find savings for each secondary cluster
  savings_classes_sec[i] <- sum(classes_before_sec[i,2:(ncol(classes_after_sec)-1)])-sum(classes_
after_sec[i,2:(ncol(classes_after_sec)-1)])
  teachers_after_sec[i] <- round((sum(classes_before_sec[i,2:(ncol(classes_before_sec)-2)])-savings_
classes_sec[i])*ratio_sec_cluster1[i,])
  savings_teachers_sec2[i] <- teachers_before_sec[i,1] - teachers_after_sec[i]}

savings_teachers_sec2 <- cbind(as.data.frame(sapply(savings_teachers_sec2,I)),urban_cluster_sec) #
change to data frame

table2[4,1] <- sum(savings_teachers_pri2[,1], na.rm=TRUE)+sum(savings_teachers_sec2[,1], na.rm=TRUE)
table2[4,2] <- sum(savings_teachers_pri2[savings_teachers_pri2$urban_cluster==1,1], na.rm=TRUE)
table2[4,3] <- sum(savings_teachers_pri2[savings_teachers_pri2$urban_cluster==0,1], na.rm=TRUE)
table2[4,4] <- sum(savings_teachers_sec2[savings_teachers_sec2$urban_cluster==1,1], na.rm=TRUE)
table2[4,5] <- sum(savings_teachers_sec2[savings_teachers_sec2$urban_cluster==0,1], na.rm=TRUE)

table2[5,1] <- table2[4,1] / (sum(teachers_before_pri) + sum(teachers_before_sec))
table2[5,2] <- table2[4,2] / sum(classes_before_pri[classes_before_pri$urban_cluster_pri==1,9])
table2[5,3] <- table2[4,3] / sum(classes_before_pri[classes_before_pri$urban_cluster_pri==0,9])
table2[5,4] <- table2[4,4] / sum(classes_before_sec[classes_before_sec$urban_cluster_sec==1,11])
table2[5,5] <- table2[4,5] / sum(classes_before_sec[classes_before_sec$urban_cluster_sec==0,11])


# Teacher savings using largest school in the network's teacher/class ratio
input$max_size_cluster_pri <- rowSums(input[,29:34])
input$max_size_cluster_sec <- rowSums(input[,35:42])
students_max_cluster_pri <- aggregate(input$max_size_cluster_pri ~ input$cluster_pri_id, input, max)
students_max_cluster_sec <- aggregate(input$max_size_cluster_sec ~ input$cluster_sec_id, input, max)
colnames(students_max_cluster_pri) <- c("cluster_pri_id", "max_size_cluster_pri")
colnames(students_max_cluster_sec) <- c("cluster_sec_id", "max_size_cluster_sec")

students_max_cluster_pri <- merge(students_max_cluster_pri, input[,c("cluster_pri_id","max_size_
cluster_pri","ratio_pri")], by=c("cluster_pri_id","max_size_cluster_pri"))
students_max_cluster_sec <- merge(students_max_cluster_sec, input[,c("cluster_sec_id","max_size_
cluster_sec","ratio_sec")], by=c("cluster_sec_id","max_size_cluster_sec"))

students_max_cluster_pri <-aggregate(students_max_cluster_pri, by=list(students_max_cluster_
pri$cluster_pri_id), FUN=mean, na.rm=TRUE)
students_max_cluster_sec <-aggregate(students_max_cluster_sec, by=list(students_max_cluster_
sec$cluster_sec_id), FUN=mean, na.rm=TRUE)

ratio_pri_cluster2 <- as.data.frame(students_max_cluster_pri[,4])
ratio_sec_cluster2 <- as.data.frame(students_max_cluster_sec[,4])

savings_classes_pri <- 0
savings_teachers_pri3 <- 0
teachers_after_pri <- 0
for (i in 1:nrow(classes_before_pri)) { # find savings for each primary cluster
  savings_classes_pri[i] <- sum(classes_before_pri[i,2:(ncol(classes_before_pri)-2)])-sum(classes_
after_pri[i,2:(ncol(classes_after_pri)-1)])
```

```
  teachers_after_pri[i] <- round((sum(classes_before_pri[i,2:(ncol(classes_before_pri)-2)])-savings_
classes_pri[i])*ratio_pri_cluster2[i,])
  savings_teachers_pri3[i] <- teachers_before_pri[i,1] - teachers_after_pri[i]}

savings_teachers_pri3 <- cbind(as.data.frame(sapply(savings_teachers_pri3,I)),urban_cluster_pri) #
change to data frame

savings_classes_sec <- 0
savings_teachers_sec3 <- 0
teachers_after_sec <- 0
for (i in 1:nrow(classes_before_sec)) { # find savings for each secondary cluster
  savings_classes_sec[i] <- sum(classes_before_sec[i,2:(ncol(classes_after_sec)-1)])-sum(classes_
after_sec[i,2:(ncol(classes_after_sec)-1)])
  teachers_after_sec[i] <- round((sum(classes_before_sec[i,2:(ncol(classes_before_sec)-2)])-savings_
classes_sec[i])*ratio_sec_cluster2[i,])
  savings_teachers_sec3[i] <- teachers_before_sec[i,1] - teachers_after_sec[i]}

savings_teachers_sec3 <- cbind(as.data.frame(sapply(savings_teachers_sec3,I)),urban_cluster_sec) #
change to data frame

table2[6,1] <- sum(savings_teachers_pri3[,1], na.rm=TRUE)+sum(savings_teachers_sec3[,1], na.rm=TRUE)
table2[6,2] <- sum(savings_teachers_pri3[savings_teachers_pri3$urban_cluster==1,1], na.rm=TRUE)
table2[6,3] <- sum(savings_teachers_pri3[savings_teachers_pri3$urban_cluster==0,1], na.rm=TRUE)
table2[6,4] <- sum(savings_teachers_sec3[savings_teachers_sec3$urban_cluster==1,1], na.rm=TRUE)
table2[6,5] <- sum(savings_teachers_sec3[savings_teachers_sec3$urban_cluster==0,1], na.rm=TRUE)

table2[7,1] <- table2[6,1] / (sum(teachers_before_pri) + sum(teachers_before_sec))
table2[7,2] <- table2[6,2] / sum(classes_before_pri[classes_before_pri$urban_cluster_pri==1,9])
table2[7,3] <- table2[6,3] / sum(classes_before_pri[classes_before_pri$urban_cluster_pri==0,9])
table2[7,4] <- table2[6,4] / sum(classes_before_sec[classes_before_sec$urban_cluster_sec==1,11])
table2[7,5] <- table2[6,5] / sum(classes_before_sec[classes_before_sec$urban_cluster_sec==0,11])


# Savings from "safe" (schools based on exact address found) and "uncertain" (schools based on
settlement found) clusters
clusters_pri$safe_school <- 1
clusters_sec$safe_school <- 1

for (i in 1:nrow(clusters_pri)) {
  if (clusters_pri$gps_source[i]=="ministry" | clusters_pri$gps_source[i]=="yandex" & clusters_
pri$yandex_type[i]=="non_address") {
    clusters_pri$safe_school[i] <- 0}}

for (i in 1:nrow(clusters_sec)) {
  if (clusters_sec$gps_source[i]=="ministry" | clusters_sec$gps_source[i]=="yandex" & clusters_
sec$yandex_type[i]=="non_address") {
    clusters_sec$safe_school[i] <- 0}}

safe_cluster_pri <- by(clusters_pri, clusters_pri$cluster_pri_id, function(x){ # safe or uncertain
cluster (primary)
  mean_safe <- mean(x$safe_school)
  if (mean_safe >= 0.75) { # at least three out of four schools are based on addresses
    safe_cluster <- 1}
  if (mean_safe < 0.75) {
    safe_cluster <- 0}
  final <- safe_cluster
})
safe_cluster_pri <- as.data.frame(sapply(safe_cluster_pri, I)) # change to data frame
colnames(safe_cluster_pri)[1] <- "safe_cluster_pri"

safe_cluster_sec <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){ # safe or uncertain
cluster (secondary)
  mean_safe <- mean(x$safe_school)
  if (mean_safe >= 0.75) { # at least three out of four schools are based on addresses
    safe_cluster <- 1}
  if (mean_safe < 0.75) {
    safe_cluster <- 0}
  final <- safe_cluster
})
safe_cluster_sec <- as.data.frame(sapply(safe_cluster_sec, I)) # change to data frame
colnames(safe_cluster_sec)[1] <- "safe_cluster_sec"

savings_classes_pri <- as.data.frame(savings_classes_pri)
```

```
savings_classes_sec <- as.data.frame(savings_classes_sec)

savings_classes_pri <- cbind(savings_classes_pri, safe_cluster_pri)
savings_classes_sec <- cbind(savings_classes_sec, safe_cluster_sec)

table3 <- matrix(0, nrow=1, ncol=6)
table3[1,3] <- sum(savings_classes_pri[savings_classes_pri$safe_cluster_pri==1,1]) # safe primary
class savings
table3[1,4] <- sum(savings_classes_pri[savings_classes_pri$safe_cluster_pri==0,1]) # uncertain
primary class savings
table3[1,5] <- sum(savings_classes_sec[savings_classes_sec$safe_cluster_sec==1,1]) # safe secondary
class savings
table3[1,6] <- sum(savings_classes_sec[savings_classes_sec$safe_cluster_sec==0,1]) # uncertain
secondary class savings
table3[1,1] <- table3[1,3]+table3[1,5]
table3[1,2] <- table3[1,4]+table3[1,6]


#######################################
### Savings through retirements     ###
### (first run STEP 7 with phase 4) ###
#######################################
# First at cluster level
clusters_sec$teachers_senior1 <- clusters_sec$teachers_61
clusters_sec$teachers_senior2 <- clusters_sec$teachers_55 + clusters_sec$teachers_61

teachers_senior_cluster <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){
  sum1 <- sum(x$teachers_senior1)
  sum2 <- sum(x$teachers_senior2)
  sum <- cbind(sum1,sum2)})
teachers_senior_cluster <- as.data.frame(t(sapply(teachers_senior_cluster,I))) # change to data
frame
colnames(teachers_senior_cluster) <- c("teachers_senior1","teachers_senior2")

savings_teachers_pri <- cbind(savings_teachers_pri1,savings_teachers_pri2[,1],savings_teachers_
pri3[,1])
colnames(savings_teachers_pri) <- c("savings1", "savings2", "savings3")

savings_teachers_sec <- cbind(savings_teachers_sec1,savings_teachers_sec2[,1],savings_teachers_
sec3[,1])
colnames(savings_teachers_sec) <- c("savings1", "savings2", "savings3")

teachers_senior_cluster <- cbind(teachers_senior_cluster, savings_teachers_sec)

for (i in 1:nrow(teachers_senior_cluster)) {
  if (teachers_senior_cluster[i,1] > teachers_senior_cluster[i,3] & !is.na(teachers_senior_
cluster[i,3])) {
    teachers_senior_cluster[i,6] <- teachers_senior_cluster[i,3]}
  else teachers_senior_cluster[i,6] <- teachers_senior_cluster[i,1]

  if (teachers_senior_cluster[i,1] > teachers_senior_cluster[i,4] & !is.na(teachers_senior_
cluster[i,4])) {
    teachers_senior_cluster[i,7] <- teachers_senior_cluster[i,4]}
  else teachers_senior_cluster[i,7] <- teachers_senior_cluster[i,1]

  if (teachers_senior_cluster[i,1] > teachers_senior_cluster[i,5] & !is.na(teachers_senior_
cluster[i,5])) {
    teachers_senior_cluster[i,8] <- teachers_senior_cluster[i,5]}
  else teachers_senior_cluster[i,8] <- teachers_senior_cluster[i,1]

  if (teachers_senior_cluster[i,2] > teachers_senior_cluster[i,3] & !is.na(teachers_senior_
cluster[i,3])) {
    teachers_senior_cluster[i,9] <- teachers_senior_cluster[i,3]}
  else teachers_senior_cluster[i,9] <- teachers_senior_cluster[i,2]

  if (teachers_senior_cluster[i,2] > teachers_senior_cluster[i,4] & !is.na(teachers_senior_
cluster[i,4])) {
    teachers_senior_cluster[i,10] <- teachers_senior_cluster[i,4]}
  else teachers_senior_cluster[i,10] <- teachers_senior_cluster[i,2]

  if (teachers_senior_cluster[i,2] > teachers_senior_cluster[i,5] & !is.na(teachers_senior_
cluster[i,5])) {
    teachers_senior_cluster[i,11] <- teachers_senior_cluster[i,5]}
```

```
  else teachers_senior_cluster[i,11] <- teachers_senior_cluster[i,2]

  if (teachers_senior_cluster[i,3] < 0) {
    teachers_senior_cluster[i,6] <- 0
    teachers_senior_cluster[i,9] <- 0}

  if (teachers_senior_cluster[i,4] < 0 & !is.na(teachers_senior_cluster[i,5])) {
    teachers_senior_cluster[i,7] <- 0
    teachers_senior_cluster[i,10] <- 0}

  if (teachers_senior_cluster[i,5] < 0 & !is.na(teachers_senior_cluster[i,5])) {
    teachers_senior_cluster[i,8] <- 0
    teachers_senior_cluster[i,11] <- 0}}

min(table2[2,1],table2[4,1],table2[6,1]) # minimum teacher savings
min(sum(teachers_senior_cluster[,6]),sum(teachers_senior_cluster[,7]),sum(teachers_senior_
cluster[,8])) # retirements with no adjustment period
max(sum(teachers_senior_cluster[,6]),sum(teachers_senior_cluster[,7]),sum(teachers_senior_
cluster[,8])) # retirements with no adjustment period
min(sum(teachers_senior_cluster[,9]),sum(teachers_senior_cluster[,10]),sum(teachers_senior_
cluster[,11])) # retirements with 5-7 year adjustment period
max(sum(teachers_senior_cluster[,9]),sum(teachers_senior_cluster[,10]),sum(teachers_senior_
cluster[,11])) # retirements with 5-7 year adjustment period

# Secondly at oblast level
getmode <- function(v) { # Create mode function
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]}

oblast_cluster_pri <- by(clusters_pri, clusters_pri$cluster_pri_id, function(x){ # Oblast id for
each cluster
  oblast_cluster_pri <- getmode(x$oblast_id)})
oblast_cluster_pri <- as.data.frame(sapply(oblast_cluster_pri,I)) # change to data frame
colnames(oblast_cluster_pri)[1] <- "oblast_id"

oblast_cluster_sec <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){
  oblast_cluster_sec<- getmode(x$oblast_id)})
oblast_cluster_sec <- as.data.frame(sapply(oblast_cluster_sec,I)) # change to data frame
colnames(oblast_cluster_sec)[1] <- "oblast_id"

oblast_cluster_pri <- cbind(oblast_cluster_pri, savings_teachers_pri) # Combine savings of teachers
with oblast id for each cluster
oblast_cluster_sec <- cbind(oblast_cluster_sec, savings_teachers_sec)

teachers_oblast_savings_pri1 <- by(oblast_cluster_pri, oblast_cluster_pri$oblast_id, function(x){
  teachers_oblast_savings_pri1 <- sum(x[,2],na.rm=T)})
teachers_oblast_savings_pri1 <- as.data.frame(sapply(teachers_oblast_savings_pri1,I)) # change to
data frame
colnames(teachers_oblast_savings_pri1)[1] <- "teacher_savings_pri1"
teachers_oblast_savings_pri2 <- by(oblast_cluster_pri, oblast_cluster_pri$oblast_id, function(x){
  teachers_oblast_savings_pri2 <- sum(x[,3],na.rm=T)})
teachers_oblast_savings_pri2 <- as.data.frame(sapply(teachers_oblast_savings_pri2,I)) # change to
data frame
colnames(teachers_oblast_savings_pri2)[1] <- "teacher_savings_pri2"
teachers_oblast_savings_pri3 <- by(oblast_cluster_pri, oblast_cluster_pri$oblast_id, function(x){
  teachers_oblast_savings_pri3 <- sum(x[,4],na.rm=T)})
teachers_oblast_savings_pri3 <- as.data.frame(sapply(teachers_oblast_savings_pri3,I)) # change to
data frame
colnames(teachers_oblast_savings_pri3)[1] <- "teacher_savings_pri3"
teachers_oblast_savings_pri <- cbind(teachers_oblast_savings_pri1,teachers_oblast_savings_
pri2,teachers_oblast_savings_pri3)

teachers_oblast_savings_sec1 <- by(oblast_cluster_sec, oblast_cluster_sec$oblast_id, function(x){
  teachers_oblast_savings_sec1 <- sum(x[,2],na.rm=T)})
teachers_oblast_savings_sec1 <- as.data.frame(sapply(teachers_oblast_savings_sec1,I)) # change to
data frame
colnames(teachers_oblast_savings_sec1)[1] <- "teacher_savings_sec1"
teachers_oblast_savings_sec2 <- by(oblast_cluster_sec, oblast_cluster_sec$oblast_id, function(x){
  teachers_oblast_savings_sec2 <- sum(x[,3],na.rm=T)})
teachers_oblast_savings_sec2 <- as.data.frame(sapply(teachers_oblast_savings_sec2,I)) # change to
data frame
colnames(teachers_oblast_savings_sec2)[1] <- "teacher_savings_sec2"
teachers_oblast_savings_sec3 <- by(oblast_cluster_sec, oblast_cluster_sec$oblast_id, function(x){
```

```
  teachers_oblast_savings_sec3 <- sum(x[,4],na.rm=T)})
teachers_oblast_savings_sec3 <- as.data.frame(sapply(teachers_oblast_savings_sec3,I)) # change to
data frame
colnames(teachers_oblast_savings_sec3)[1] <- "teacher_savings_sec3"
teachers_oblast_savings_sec <- cbind(teachers_oblast_savings_sec1,teachers_oblast_savings_
sec2,teachers_oblast_savings_sec3)

teachers_oblast_savings_all <- teachers_oblast_savings_pri + teachers_oblast_savings_sec # Total
oblast teacher savings

teachers_senior_oblast <- by(clusters_sec, clusters_sec$oblast_id, function(x){
  sum1 <- sum(x$teachers_61)
  sum2 <- sum(x$teachers_61 + x$teachers_55)
  sum <- cbind(sum1,sum2)})
teachers_senior_oblast <- as.data.frame(t(sapply(teachers_senior_oblast,I))) # change to data frame
colnames(teachers_senior_oblast) <- c("teachers_senior1","teachers_senior2")

teachers_senior_oblast <- cbind(teachers_senior_oblast, teachers_oblast_savings_all)

for (i in 1:nrow(teachers_senior_oblast)) {
  if (teachers_senior_oblast[i,1] > teachers_senior_oblast[i,3]) {
    teachers_senior_oblast[i,6] <- teachers_senior_oblast[i,3]}
  else teachers_senior_oblast[i,6] <- teachers_senior_oblast[i,1]

  if (teachers_senior_oblast[i,1] > teachers_senior_oblast[i,4]) {
    teachers_senior_oblast[i,7] <- teachers_senior_oblast[i,4]}
  else teachers_senior_oblast[i,7] <- teachers_senior_oblast[i,1]

  if (teachers_senior_oblast[i,1] > teachers_senior_oblast[i,5]) {
    teachers_senior_oblast[i,8] <- teachers_senior_oblast[i,5]}
  else teachers_senior_oblast[i,8] <- teachers_senior_oblast[i,1]

  if (teachers_senior_oblast[i,2] > teachers_senior_oblast[i,3]) {
    teachers_senior_oblast[i,9] <- teachers_senior_oblast[i,3]}
  else teachers_senior_oblast[i,9] <- teachers_senior_oblast[i,2]

  if (teachers_senior_oblast[i,2] > teachers_senior_oblast[i,4]) {
    teachers_senior_oblast[i,10] <- teachers_senior_oblast[i,4]}
  else teachers_senior_oblast[i,10] <- teachers_senior_oblast[i,2]

  if (teachers_senior_oblast[i,2] > teachers_senior_oblast[i,5]) {
    teachers_senior_oblast[i,11] <- teachers_senior_oblast[i,5]}
  else teachers_senior_oblast[i,11] <- teachers_senior_oblast[i,2]}

min(table2[2,1],table2[4,1],table2[6,1]) # minimum teacher savings
min(sum(teachers_senior_oblast[,6]),sum(teachers_senior_oblast[,7]),sum(teachers_senior_oblast[,8]))
# retirements at oblast level with no adjustment period
max(sum(teachers_senior_oblast[,6]),sum(teachers_senior_oblast[,7]),sum(teachers_senior_oblast[,8]))
# retirements at oblast level with no adjustment period
min(sum(teachers_senior_oblast[,9]),sum(teachers_senior_oblast[,10]),sum(teachers_senior_
oblast[,11])) # retirements at oblast level with 5-7 year adjustment period
max(sum(teachers_senior_oblast[,9]),sum(teachers_senior_oblast[,10]),sum(teachers_senior_
oblast[,11])) # retirements at oblast level with 5-7 year adjustment period


#################################
### Teacher savings by oblast ###
#################################
# Sum of teachers for each oblast before clustering (Kiev has no information on current teachers)
teachers_oblast_before <- by(input, input$oblast_id, function(x){
  sum <- sum(x$teachers_total)})
teachers_oblast_before <- as.data.frame(sapply(teachers_oblast_before,I)) # change to data frame
colnames(teachers_oblast_before)[1] <- "teachers"
teachers_oblast_before$oblast_id <- c(1:12,14:25) # change to "c(1:25)" if Kiev is included

# Create mode function
getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}

# Oblast id for each cluster
oblast_cluster_pri <- by(clusters_pri, clusters_pri$cluster_pri_id, function(x){
```

```
  oblast_cluster_pri <- getmode(x$oblast_id)})
oblast_cluster_pri <- as.data.frame(sapply(oblast_cluster_pri,I)) # change to data frame
colnames(oblast_cluster_pri)[1] <- "oblast_id"

oblast_cluster_sec <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){
   oblast_cluster_sec<- getmode(x$oblast_id)})
oblast_cluster_sec <- as.data.frame(sapply(oblast_cluster_sec,I)) # change to data frame
colnames(oblast_cluster_sec)[1] <- "oblast_id"

# Combine savings of teachers with oblast id for each cluster
oblast_cluster_pri <- cbind(oblast_cluster_pri, savings_classes_pri)
oblast_cluster_sec <- cbind(oblast_cluster_sec, savings_classes_sec)

# Find absolute and relative teacher savings for each oblast
teachers_oblast_savings_pri <- by(oblast_cluster_pri, oblast_cluster_pri$oblast_id, function(x){
  teachers_oblast_savings_pri <- sum(x[,2], na.rm=T)})
teachers_oblast_savings_pri <- as.data.frame(sapply(teachers_oblast_savings_pri,I)) # change to data
frame
colnames(teachers_oblast_savings_pri)[1] <- "teacher_savings_pri"

teachers_oblast_savings_sec <- by(oblast_cluster_sec, oblast_cluster_sec$oblast_id, function(x){
  teachers_oblast_savings_sec <- sum(x[,2], na.rm=T)})
teachers_oblast_savings_sec <- as.data.frame(sapply(teachers_oblast_savings_sec,I)) # change to data
frame
colnames(teachers_oblast_savings_sec)[1] <- "teacher_savings_sec"

teachers_oblast_savings_all <- teachers_oblast_savings_pri + teachers_oblast_savings_sec

teachers_oblast_savings_all$savings_percent <- teachers_oblast_savings_all$teacher_savings/teachers_
oblast_before$teachers
#teachers_oblast_savings_all[13,1:2] <- NaN # uncomment if Kiev is included

colnames(teachers_oblast_savings_all) <- c("savings_absolute","savings_percent")
teachers_oblast_savings_all$oblast_id <- c(1:12,14:25)

# Load map of oblasts and give id corresponding to id in dataset
ukraine_adm1 <- readOGR(dsn = "R_input", layer = "ukr_admbnda_adm1_q2_sspe_20171221")

ukraine_adm1@data$oblast_id <- c(26,27,13,25,24,23,22,21,20,18,16,15,11,10,9,7,5,4,3,1,19,17,2,8,14,
6,12)

# Merge savings on to map data and plot
ukraine_adm1 <- merge(ukraine_adm1, teachers_oblast_savings_all, by="oblast_id")
ukraine_adm1@data$savings_percent[1:3] <- NaN
ukraine_adm1@data$savings_absolute[1:3] <- NaN

missing_oblasts <- ukraine_adm1
missing_oblasts <- missing_oblasts[1:3,]
missing_oblasts@data[,26:27] <- 0

pal2 <- colorRampPalette(c("white", "red"))
spplot(ukraine_adm1, "savings_percent", col="black", col.regions=pal2(19),
       sp.layout = list("sp.polygons", ukraine_adm1[1:3,], pch = 16, cex = 2, col = "black",
fill="grey"))
grid.rect(x = unit(4.75,'lines'), y = unit(3,'lines'),
          width = unit(0.15, "npc"), height = unit(0.06, "npc"))
grid.text("Grey = No data", x=unit(0.12, "npc"), y=unit(0.12, "npc"))

pal2 <- colorRampPalette(c("white", "red"))
spplot(ukraine_adm1, "savings_absolute", col="black", col.regions=pal2(19),
       sp.layout = list("sp.polygons", ukraine_adm1[1:3,], pch = 16, cex = 2, col = "black",
fill="grey"))
grid.rect(x = unit(4.75,'lines'), y = unit(3,'lines'),
          width = unit(0.15, "npc"), height = unit(0.06, "npc"))
grid.text("Grey = No data", x=unit(0.12, "npc"), y=unit(0.12, "npc"))


###################################################################
### Figure with savings dependent on maximum allowed class size ###
###################################################################
# Set the phase of interest
phase = 1
```

```
# Used for sensitivity analysis allowing for different maximum class sizes
savings_max_size <- matrix(0, nrow=6, ncol=3)

for (k in 28:33) {
  max_students = k

  # Load data if you don’t want to run step 1, 2, 3, and 4.
  input <- read.table(“R_input/school_data.txt”, header=TRUE)
  input <- input[input$school_drop==0,]
  input <- input[input$oblast_id!=13,]

  if(phase==1) {
    clusters_pri <- read.table(“R_output/clusters_primary_classes_phase1.txt”, header=TRUE, sep=”;”)
    clusters_sec <- read.table(“R_output/clusters_secondary_classes_phase1.txt”, header=TRUE,
sep=”;”)
  } else if(phase==2) {
    clusters_pri <- read.table(“R_output/clusters_primary_classes_phase2.txt”, header=TRUE, sep=”;”)
    clusters_sec <- read.table(“R_output/clusters_secondary_classes_phase2.txt”, header=TRUE,
sep=”;”)
  } else if(phase==3) {
    clusters_pri <- read.table(“R_output/clusters_primary_classes_phase3.txt”, header=TRUE, sep=”;”)
    clusters_sec <- read.table(“R_output/clusters_secondary_classes_phase3.txt”, header=TRUE,
sep=”;”)
  } else if(phase==4) {
    clusters_pri <- read.table(“R_output/clusters_primary_classes_phase4.txt”, header=TRUE, sep=”;”)
    clusters_sec <- read.table(“R_output/clusters_secondary_classes_phase4.txt”, header=TRUE,
sep=”;”)
  }

  clusters_pri <- clusters_pri[clusters_pri$oblast_id!=13,]
  clusters_sec <- clusters_sec[clusters_sec$oblast_id!=13,]

  for (i in 1:nrow(input)) {
    if (input$classes_psei[i]>0 | input$classes_gei[i]>0 | input$classes_grade2[i]>0 |
input$classes_grade3[i]>0 | input$classes_grade4[i]>0) {
      input$classes_pri[i] <- 1
    } else {input$classes_pri[i] <- 0}

    if (sum(input$classes_grade5[i],input$classes_grade6[i],input$classes_grade7[i],input$classes_
grade8[i],input$classes_grade9[i],
            input$classes_grade10[i],input$classes_grade11[i],input$classes_grade12[i]) > 0) {
      input$classes_sec[i] <- 1
    } else {input$classes_sec[i] <- 0}
  }


  # Number of schools in cluster
  count_pri <- as.data.frame(table(clusters_pri$cluster_pri_id))
  colnames(count_pri) <- c(“cluster_pri_id”, “schools_cluster_pri”)
  count_pri <- merge(clusters_pri[,c(“cluster_pri_id”,”school_id”)], count_pri, by=”cluster_pri_id”)
  clusters_pri <- merge(count_pri[,c(“school_id”, “schools_cluster_pri”)], clusters_pri, by=”school_
id”)

  count_sec <- as.data.frame(table(clusters_sec$cluster_sec_id))
  colnames(count_sec) <- c(“cluster_sec_id”, “schools_cluster_sec”)
  count_sec <- merge(clusters_sec[,c(“cluster_sec_id”,”school_id”)], count_sec, by=”cluster_sec_id”)
  clusters_sec <- merge(count_sec[,c(“school_id”, “schools_cluster_sec”)], clusters_sec, by=”school_
id”)


  # Classes for each grade before clustering (at cluster level) (first for primary clusters and next
for secondary clusters)
  clusters_pri$classes_pri_number <- 0
  for (i in 1:nrow(clusters_pri)) {
    clusters_pri$classes_pri_number[i] = sum(clusters_pri[i,c(“classes_preparatory”,”classes_
psei”,”classes_gei”,
                                                              “classes_grade2”,”classes_
grade3”,”classes_grade4”)])}
  classes_before_pri <- by(clusters_pri, clusters_pri$cluster_pri_id, function(x){ # classes before
clustering within each primary cluster
    id_temp <- mean(x$cluster_pri_id)
    sum <- colSums(x[,c(“classes_preparatory”,”classes_psei”,”classes_gei”,”classes_
```

```
grade2","classes_grade3","classes_grade4")])
    final <- cbind(id_temp, sum)})

  classes_before_pri <- as.data.frame(t(sapply(classes_before_pri, I))) # change to data frame
  classes_before_pri <- classes_before_pri[,(ncol(classes_before_pri)/2):ncol(classes_before_pri)] #
drop duplicates of cluster id
  colnames(classes_before_pri) <- c("cluster_pri_id","classes_preparatory_pri","classes_psei_
pri","classes_gei_pri",
                                    "classes_grade2_pri","classes_grade3_pri","classes_grade4_pri")


  clusters_sec$classes_sec_number <- 0
  for (i in 1:nrow(clusters_sec)) {
    clusters_sec$classes_sec_number[i] = sum(clusters_sec[i,c("classes_grade5","classes_
grade6","classes_grade7",
                                                              "classes_grade8","classes_
grade9","classes_grade10",
                                                              "classes_grade11","classes_
grade12")])}
  classes_before_sec <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){ # classes before
clustering within each secondary cluster
    id_temp <- mean(x$cluster_sec_id)
    sum <- colSums(x[,c("classes_grade5","classes_grade6","classes_grade7","classes_
grade8","classes_grade9","classes_grade10",
                        "classes_grade11","classes_grade12")])
    final <- cbind(id_temp, sum)})

  classes_before_sec <- as.data.frame(t(sapply(classes_before_sec, I))) # change to data frame
  classes_before_sec <- classes_before_sec[,(ncol(classes_before_sec)/2):ncol(classes_before_sec)] #
drop duplicates of cluster id
  colnames(classes_before_sec) <- c("cluster_sec_id","classes_grade5_sec","classes_grade6_
sec","classes_grade7_sec",
                                    "classes_grade8_sec","classes_grade9_sec","classes_grade10_sec",
                                    "classes_grade11_sec","classes_grade12_sec")


  # Students for each grade (at cluster level) (first for primary clusters and next for secondary
clusters)
  clusters_pri$students_pri_number <- 0
  for (i in 1:nrow(clusters_pri)) {
    clusters_pri$students_pri_number[i] = sum(clusters_pri[i,c("students_preparatory","students_
psei","students_gei",
                                                               "students_grade2","students_
grade3","students_grade4")])}
  students_pri <- by(clusters_pri, clusters_pri$cluster_pri_id, function(x){ # students within each
primary cluster
    id_temp <- mean(x$cluster_pri_id)
    sum <- colSums(x[,c("students_preparatory","students_psei","students_gei","students_
grade2","students_grade3","students_grade4")])
    final <- cbind(id_temp, sum)})

  students_pri <- as.data.frame(t(sapply(students_pri, I))) # change to data frame
  students_pri <- students_pri[,(ncol(students_pri)/2):ncol(students_pri)] # drop duplicates of
cluster id
  colnames(students_pri) <- c("cluster_pri_id","students_preparatory_pri","students_psei_
pri","students_gei_pri",
                              "students_grade2_pri","students_grade3_pri","students_grade4_pri")


  clusters_sec$students_sec_number <- 0
  for (i in 1:nrow(clusters_sec)) {
    clusters_sec$students_sec_number[i] = sum(clusters_sec[i,c("students_grade5","students_
grade6","students_grade7",
                                                               "students_grade8","students_
grade9","students_grade10",
                                                               "students_grade11","students_
grade12")])}
  students_sec <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){ # students within each
secondary cluster
    id_temp <- mean(x$cluster_sec_id)
    sum <- colSums(x[,c("students_grade5","students_grade6","students_grade7","students_
grade8","students_grade9","students_grade10",
                        "students_grade11","students_grade12")])
```

```
    final <- cbind(id_temp, sum)})

  students_sec <- as.data.frame(t(sapply(students_sec, I))) # change to data frame
  students_sec <- students_sec[,(ncol(students_sec)/2):ncol(students_sec)] # drop duplicates of
cluster id
  colnames(students_sec) <- c("cluster_sec_id","students_grade5_sec","students_grade6_
sec","students_grade7_sec",
                              "students_grade8_sec","students_grade9_sec","students_grade10_sec",
                              "students_grade11_sec","students_grade12_sec")


  # Classes for each grade after clustering (at cluster level) (first for primary clusters and next
for secondary clusters)
  # If schools have been allowed to have multi-grade classes or allowed to exceed the limit, we
assume they will continue to do so
  classes_after_pri <- classes_before_pri
  for (i in 1:nrow(classes_after_pri)) {
    classes_after_pri[i,2:ncol(classes_after_pri)] <- round(students_pri[i,2:ncol(classes_after_
pri)]/max_students + 0.4999)
    for (j in 2:ncol(classes_after_pri)) {
      if (classes_after_pri[i,j] > classes_before_pri[i,j]) {
        classes_after_pri[i,j] <- classes_before_pri[i,j]
      }
    }
  }

  classes_after_sec <- classes_before_sec
  for (i in 1:nrow(classes_after_sec)) {
    classes_after_sec[i,2:ncol(classes_after_sec)] <- round(students_sec[i,2:ncol(classes_after_
sec)]/max_students + 0.4999)
    for (j in 2:ncol(classes_after_sec)) {
      if (classes_after_sec[i,j] > classes_before_sec[i,j]) {
        classes_after_sec[i,j] <- classes_before_sec[i,j]
      }
    }
  }


  ### TABLE: MAIN RESULTS SAVINGS OF CLASSES ###
  # Total savings of classes
  table1 <- matrix(0, nrow=6, ncol=5)
  table1[1,1] <- sum(classes_before_pri[2:7]) + sum(classes_before_sec[2:9])
  table1[2,1] <- sum(classes_after_pri[2:7]) + sum(classes_after_sec[2:9])
  table1[3,1] <- table1[1,1] - table1[2,1]
  table1[4,1] <- table1[3,1] / table1[1,1]

  # Savings of classes divided by urban or rural clusters
  urban_cluster_pri <- by(clusters_pri, clusters_pri$cluster_pri_id, function(x){ # urban or rural
cluster
    mean_urban <- mean(x$urban_dummy)
    if (mean_urban >= 0.5) {
      urban_cluster <- 1}
    if (mean_urban < 0.5) {
      urban_cluster <- 0}
    final <- urban_cluster
  })
  urban_cluster_pri <- as.data.frame(sapply(urban_cluster_pri, I)) # change to data frame
  colnames(urban_cluster_pri)[1] <- "urban_cluster_pri"
  savings_classes_pri <- cbind(as.data.frame(rowSums((classes_before_pri - classes_after_pri),2:7)),
urban_cluster_pri)
  colnames(savings_classes_pri) <- c("class_savings_pri", "urban_cluster_pri")
  classes_before_pri <- cbind(classes_before_pri,urban_cluster_pri)
  classes_after_pri <- cbind(classes_after_pri,urban_cluster_pri)

  urban_cluster_sec <- by(clusters_sec, clusters_sec$cluster_sec_id, function(x){ # urban or rural
cluster
    mean_urban <- mean(x$urban_dummy)
    if (mean_urban >= 0.5) {
      urban_cluster <- 1}
    if (mean_urban < 0.5) {
      urban_cluster <- 0}
    final <- urban_cluster
  })
```

```
  urban_cluster_sec <- as.data.frame(sapply(urban_cluster_sec, I)) # change to data frame
  colnames(urban_cluster_sec)[1] <- "urban_cluster_sec"
  savings_classes_sec <- cbind(as.data.frame(rowSums((classes_before_sec - classes_after_sec),2:7)), 
urban_cluster_sec)
  colnames(savings_classes_sec) <- c("class_savings_sec", "urban_cluster_sec")
  classes_before_sec <- cbind(classes_before_sec,urban_cluster_sec)
  classes_after_sec <- cbind(classes_after_sec,urban_cluster_sec)

  table1[1,2] <- sum(classes_before_pri[classes_before_pri$urban_cluster_pri==1,2:7])
  table1[2,2] <- sum(classes_after_pri[classes_after_pri$urban_cluster_pri==1,2:7])
  table1[3,2] <- table1[1,2] - table1[2,2]
  table1[4,2] <- table1[3,2] / table1[1,2]
  table1[1,3] <- sum(classes_before_pri[classes_before_pri$urban_cluster_pri==0,2:7])
  table1[2,3] <- sum(classes_after_pri[classes_after_pri$urban_cluster_pri==0,2:7])
  table1[3,3] <- table1[1,3] - table1[2,3]
  table1[4,3] <- table1[3,3] / table1[1,3]
  table1[1,4] <- sum(classes_before_sec[classes_before_sec$urban_cluster_sec==1,2:9])
  table1[2,4] <- sum(classes_after_sec[classes_after_sec$urban_cluster_sec==1,2:9])
  table1[3,4] <- table1[1,4] - table1[2,4]
  table1[4,4] <- table1[3,4] / table1[1,4]
  table1[1,5] <- sum(classes_before_sec[classes_before_sec$urban_cluster_sec==0,2:9])
  table1[2,5] <- sum(classes_after_sec[classes_after_sec$urban_cluster_sec==0,2:9])
  table1[3,5] <- table1[1,5] - table1[2,5]
  table1[4,5] <- table1[3,5] / table1[1,5]


  # Schools affected
  clusters_pri <- merge(clusters_pri, urban_cluster_pri, by.x="cluster_pri_id", by.y=0)
  clusters_sec <- merge(clusters_sec, urban_cluster_sec, by.x="cluster_sec_id", by.y=0)

  students_pri$students_total_pri <- rowSums(students_pri[,2:7])
  students_sec$students_total_sec <- rowSums(students_sec[,2:9])
  schools_students_affected <- merge(clusters_pri[,c("cluster_pri_id","school_id","schools_cluster_
pri","urban_cluster_pri")],
                                   students_pri[,c("cluster_pri_id","students_total_pri")], 
by="cluster_pri_id")
  schools_students_affected <- merge(schools_students_affected,
                                   clusters_sec[,c("cluster_sec_id","school_id","schools_cluster_
sec","urban_cluster_sec")],
                                   by="school_id")
  schools_students_affected <- merge(schools_students_affected,students_sec[,c("cluster_sec_
id","students_total_sec")],
                                   by="cluster_sec_id")

  table1[5,1] <- nrow(schools_students_affected[schools_students_affected$schools_cluster_pri>1 |
                                                 schools_students_affected$schools_cluster_sec>1,]) 
/
    nrow(schools_students_affected)
  table1[5,2] <- nrow(schools_students_affected[schools_students_affected$schools_cluster_pri>1 &
                                                 schools_students_affected$urban_cluster_pri==1 & 
schools_students_affected$students_total_pri>0,]) /
    nrow(schools_students_affected[schools_students_affected$urban_cluster_pri==1 & schools_students_
affected$students_total_pri>0,])
  table1[5,3] <- nrow(schools_students_affected[schools_students_affected$schools_cluster_pri>1 &
                                                 schools_students_affected$urban_cluster_pri==0 & 
schools_students_affected$students_total_pri>0,]) /
    nrow(schools_students_affected[schools_students_affected$urban_cluster_pri==0 & schools_students_
affected$students_total_pri>0,])
  table1[5,4] <- nrow(schools_students_affected[schools_students_affected$schools_cluster_sec>1 &
                                                 schools_students_affected$urban_cluster_sec==1 & 
schools_students_affected$students_total_sec>0,]) /
    nrow(schools_students_affected[schools_students_affected$urban_cluster_sec==1 & schools_students_
affected$students_total_sec>0,])
  table1[5,5] <- nrow(schools_students_affected[schools_students_affected$schools_cluster_sec>1 &
                                                 schools_students_affected$urban_cluster_sec==0 & 
schools_students_affected$students_total_sec>0,]) /
    nrow(schools_students_affected[schools_students_affected$urban_cluster_sec==0 & schools_students_
affected$students_total_sec>0,])

  # Students affected
  for (i in 1:nrow(schools_students_affected)) {
    schools_students_affected$students_total_pri[i] <- schools_students_affected$students_total_
pri[i]/schools_students_affected$schools_cluster_pri[i]
```

```
    schools_students_affected$students_total_sec[i] <- schools_students_affected$students_total_
sec[i]/schools_students_affected$schools_cluster_sec[i]}

  table1[6,1] <- (sum(schools_students_affected$students_total_pri[schools_students_affected$schools_
cluster_pri>1]) +
                    sum(schools_students_affected$students_total_sec[schools_students_
affected$schools_cluster_sec>1])) /
    (sum(schools_students_affected$students_total_pri) + sum(schools_students_affected$students_total_
sec))
  table1[6,2] <- sum(schools_students_affected$students_total_pri[schools_students_affected$schools_
cluster_pri>1 &
                                                                    schools_students_affected$urban_
cluster_pri==1]) /
    sum(schools_students_affected$students_total_pri[schools_students_affected$urban_cluster_pri==1])
  table1[6,3] <- sum(schools_students_affected$students_total_pri[schools_students_affected$schools_
cluster_pri>1 &
                                                                    schools_students_affected$urban_
cluster_pri==0]) /
    sum(schools_students_affected$students_total_pri[schools_students_affected$urban_cluster_pri==0])
  table1[6,4] <- sum(schools_students_affected$students_total_sec[schools_students_affected$schools_
cluster_sec>1 &
                                                                    schools_students_affected$urban_
cluster_sec==1]) /
    sum(schools_students_affected$students_total_sec[schools_students_affected$urban_cluster_sec==1])
  table1[6,5] <- sum(schools_students_affected$students_total_sec[schools_students_affected$schools_
cluster_sec>1 &
                                                                    schools_students_affected$urban_
cluster_sec==0]) /
    sum(schools_students_affected$students_total_sec[schools_students_affected$urban_cluster_sec==0])


  # Used for sensitivity analysis allowing for different maximum class sizes ("tab" to indent and
"shift"+"tab" to remove indent)
  savings_max_size[k-27,1] <- k
  savings_max_size[k-27,2] <- table1[3,1]
  savings_max_size[k-27,3] <- table1[4,1] * 100
}

savings_max_size_all_decimals <- as.data.frame(savings_max_size)
savings_max_size <- as.data.frame(savings_max_size)
savings_max_size[,3] <- round(savings_max_size[,3], digits=1)
colnames(savings_max_size) <- c("max_size","class_savings_abs","class_savings_rel")

if (phase==1) {
  jpeg('R_output/savings_max_size1.jpg') # tells where to save the graph
  plot(range(28:33), range(5:11), type="n", xlab="Maximum allowed class size",ylab="Class savings
(%)")
  lines(savings_max_size$max_size, savings_max_size$class_savings_rel, type="b")
  text(savings_max_size$max_size, savings_max_size$class_savings_rel, labels=savings_max_size$class_
savings_rel, cex=0.8, pos=3)

  dev.off() # to shut off the saving of the plot
  } else if(phase==2) {
      jpeg('R_output/savings_max_size2.jpg') # tells where to save the graph
      plot(range(28:33), range(14:22), type="n", xlab="Maximum allowed class size",ylab="Class
savings (%)")
      lines(savings_max_size$max_size, savings_max_size$class_savings_rel, type="b")
      text(savings_max_size$max_size, savings_max_size$class_savings_rel, labels=savings_max_
size$class_savings_rel, cex=0.8, pos=3)

      dev.off() # to shut off the saving of the plot
  } else if(phase==3) {
      jpeg('R_output/savings_max_size3.jpg') # tells where to save the graph
      plot(range(28:33), range(20:29), type="n", xlab="Maximum allowed class size",ylab="Class
savings (%)")
      lines(savings_max_size$max_size, savings_max_size$class_savings_rel, type="b")
      text(savings_max_size$max_size, savings_max_size$class_savings_rel, labels=savings_max_
size$class_savings_rel, cex=0.8, pos=3)

      dev.off() # to shut off the saving of the plot
  } else if(phase==4) {
      jpeg('R_output/savings_max_size3.jpg') # tells where to save the graph
      plot(range(28:33), range(20:29), type="n", xlab="Maximum allowed class size",ylab="Class
```

```
savings (%)”)
      lines(savings_max_size$max_size, savings_max_size$class_savings_rel, type=”b”)
      text(savings_max_size$max_size, savings_max_size$class_savings_rel, labels=savings_max_
size$class_savings_rel, cex=0.8, pos=3)

      dev.off() # to shut off the saving of the plot
  }


###########################################################
### Descriptives on non-clustered and clustered schools ###
###########################################################
# Identify which schools have not been clustered
count_pri <- as.data.frame(table(clusters_pri$cluster_pri_id))
colnames(count_pri) <- c(“cluster_pri_id”, “schools_cluster_pri”)
count_pri <- merge(clusters_pri[,c(“cluster_pri_id”,”school_id”)], count_pri, by=”cluster_pri_id”)
input2 <- merge(count_pri, input, by=”school_id”)

count_sec <- as.data.frame(table(clusters_sec$cluster_sec_id))
colnames(count_sec) <- c(“cluster_sec_id”, “schools_cluster_sec”)
count_sec <- merge(clusters_sec[,c(“cluster_sec_id”,”school_id”)], count_sec, by=”cluster_sec_id”)
input2 <- merge(count_sec, input2, by=”school_id”)

input2$remote <- 1
for (i in 1:nrow(input2)) {
  if (input2$schools_cluster_pri[i]>1 | input2$schools_cluster_sec[i]>1) {
    input2$remote[i] <- 0}}

# Number of schools non-clustered and clustered
table_charac1 <- matrix(0, nrow=7, ncol=2)
table_charac1[1,1] <- sum(input2$remote)
table_charac1[1,2] <- sum(input2$remote==0)

# Distance to nearest school
coordinates <- cbind(input$preferred_lon, input$preferred_lat, input$school_id)
distance_matrix2 <- distm(coordinates[,1:2], fun = distHaversine)/1000
diag(distance_matrix2) <- 1000 # set diagonal to 1000 to make sure the school itself is not the
closest school

distance_minimum <- sapply(seq(nrow(distance_matrix2)), function(i) { # minimum distance of all
schools
  j <- which.min(distance_matrix2[i,])
  distance_matrix2[i,j]})

input2$distance_minimum <- as.data.frame(distance_minimum) # convert to data frame
colnames(input2)[ncol(input2)] <- “distance_minimum”

mean(input2$distance_minimum[,]) # 3.2 km for all schools
table_charac1[2,1] <- mean(input2$distance_minimum[input2$remote==1,])
table_charac1[2,2] <- mean(input2$distance_minimum[input2$remote==0,])

# Urban share
mean(input2$urban_dummy) # 22.8% for all schools
table_charac1[3,1] <- mean(input2$urban_dummy[input2$remote==1])
table_charac1[3,2] <- mean(input2$urban_dummy[input2$remote==0])

# Number of students per school
mean(input2$students_total) # 227 for all schools
table_charac1[4,1] <- mean(input2$students_total[input2$remote==1])
table_charac1[4,2] <- mean(input2$students_total[input2$remote==0])

# Number of students per class
input2$students_per_class <- input2$students_total/input2$classes_total
mean(input2$students_per_class) # 14.9 for all schools
table_charac1[5,1] <- mean(input2$students_per_class[input2$remote==1])
table_charac1[5,2] <- mean(input2$students_per_class[input2$remote==0])

# Share with multi-grade classes
input2$multi_grade <- 0
for (i in 1:nrow(input2)) {
  if (input2$students_preparatory[i]>0 & input2$classes_preparatory[i]==0 | input2$students_
psei[i]>0 & input2$classes_psei[i]==0 |
      input2$students_gei[i]>0 & input2$classes_gei[i]==0 | input2$students_grade2[i]>0 &
```

```
input2$classes_grade2[i]==0 |
      input2$students_grade3[i]>0 & input2$classes_grade3[i]==0 | input2$students_grade4[i]>0 &
input2$classes_grade4[i]==0 |
      input2$students_grade5[i]>0 & input2$classes_grade5[i]==0 | input2$students_grade6[i]>0 &
input2$classes_grade6[i]==0 |
      input2$students_grade7[i]>0 & input2$classes_grade7[i]==0 | input2$students_grade8[i]>0 &
input2$classes_grade8[i]==0 |
      input2$students_grade9[i]>0 & input2$classes_grade9[i]==0 | input2$students_grade10[i]>0 &
input2$classes_grade10[i]==0 |
      input2$students_grade11[i]>0 & input2$classes_grade11[i]==0 | input2$students_grade12[i]>0 &
input2$classes_grade12[i]==0) {
    input2$multi_grade[i] <- 1}}

mean(input2$multi_grade) # 31.7% for non-clustered schools
table_charac1[6,1] <- mean(input2$multi_grade[input2$remote==1])
table_charac1[6,2] <- mean(input2$multi_grade[input2$remote==0])

# Schools entering cluster and avoiding multi-grade classes
multi_grade_cluster_pri <- setNames(data.frame(matrix(ncol = 2, nrow = nrow(classes_after_pri))),
c("multi_grade_pri", "cluster_pri_id"))
for (i in 1:nrow(classes_after_pri)) {
  if (classes_after_pri[i,2]==0 & students_pri[i,2]>0 | classes_after_pri[i,3]==0 & students_
pri[i,3]>0 |
      classes_after_pri[i,4]==0 & students_pri[i,4]>0 | classes_after_pri[i,5]==0 & students_
pri[i,5]>0 |
      classes_after_pri[i,6]==0 & students_pri[i,6]>0 | classes_after_pri[i,7]==0 & students_
pri[i,7]>0) {
    multi_grade_cluster_pri$multi_grade_pri[i] <- 1}
  else {multi_grade_cluster_pri$multi_grade_pri[i] <- 0}
  multi_grade_cluster_pri$cluster_pri_id[i] <- classes_after_pri[i,1]}

multi_grade_cluster_sec <- setNames(data.frame(matrix(ncol = 2, nrow = nrow(classes_after_sec))),
c("multi_grade_sec", "cluster_sec_id"))
for (i in 1:nrow(classes_after_sec)) {
  if (classes_after_sec[i,2]==0 & students_sec[i,2]>0 | classes_after_sec[i,3]==0 & students_
sec[i,3]>0 |
      classes_after_sec[i,4]==0 & students_sec[i,4]>0 | classes_after_sec[i,5]==0 & students_
sec[i,5]>0 |
      classes_after_sec[i,6]==0 & students_sec[i,6]>0 | classes_after_sec[i,7]==0 & students_
sec[i,7]>0 |
      classes_after_sec[i,8]==0 & students_sec[i,8]>0 | classes_after_sec[i,9]==0 & students_
sec[i,9]>0) {
    multi_grade_cluster_sec$multi_grade_sec[i] <- 1}
  else {multi_grade_cluster_sec$multi_grade_sec[i] <- 0}
  multi_grade_cluster_sec$cluster_sec_id[i] <- classes_after_sec[i,1]}

input2 <- merge(input2, multi_grade_cluster_pri, by.x="cluster_pri_id.y", by.y="cluster_pri_id")
input2 <- merge(input2, multi_grade_cluster_sec, by.x="cluster_sec_id.y", by.y="cluster_sec_id")
input2$multi_grade_after <- (input2$multi_grade==1 & (input2$multi_grade_pri==1 | input2$multi_
grade_sec==1))

mean(input2$multi_grade_after)
table_charac1[7,1] <- mean(input2$multi_grade_after[input2$remote==1])
table_charac1[7,2] <- mean(input2$multi_grade_after[input2$remote==0])


###########################################################################
### Descriptives on clusters with more than one school before and after ###
###########################################################################
# Number of clusters
urban_cluster_pri$cluster_pri_id <- classes_after_pri$cluster_pri_id
urban_cluster_sec$cluster_sec_id <- classes_after_sec$cluster_sec_id

input2 <- merge(input2, urban_cluster_pri, by.x="cluster_pri_id.y", by.y="cluster_pri_id")
input2 <- merge(input2, urban_cluster_sec, by.x="cluster_sec_id.y", by.y="cluster_sec_id")

nrow(input2[input2$classes_pri==1,]) # 13,812 primary schools
nrow(input2[input2$urban_dummy==1 & input2$classes_pri==1,]) #
nrow(input2[input2$urban_dummy==0 & input2$classes_pri==1,]) #
nrow(input2[input2$classes_sec==1,]) # 13,184 secondary schools
nrow(input2[input2$urban_dummy==1 & input2$classes_sec==1,]) #
nrow(input2[input2$urban_dummy==0 & input2$classes_sec==1,]) #
```

```
table_charac2 <- matrix(0, nrow=5, ncol=6)
table_charac2[1,1] <- length(unique(input2$cluster_pri_id.y[input2$schools_cluster_pri>1]))
table_charac2[1,2] <- length(unique(input2$cluster_pri_id.y[input2$schools_cluster_pri>1 &
input2$urban_cluster_pri==1])) # primary school urban clusters
table_charac2[1,3] <- length(unique(input2$cluster_pri_id.y[input2$schools_cluster_pri>1 &
input2$urban_cluster_pri==0])) # primary school rural clusters
table_charac2[1,4] <- length(unique(input2$cluster_sec_id.y[input2$schools_cluster_sec>1]))
table_charac2[1,5] <- length(unique(input2$cluster_sec_id.y[input2$schools_cluster_sec>1 &
input2$urban_cluster_sec==1])) # secondary school urban clusters
table_charac2[1,6] <- length(unique(input2$cluster_sec_id.y[input2$schools_cluster_sec>1 &
input2$urban_cluster_sec==0])) # secondary school rural clusters

# Schools per cluster
table_charac2[2,1] <- mean(input2[!duplicated(input2$cluster_pri_id.y) & input2$schools_cluster_
pri>1,"schools_cluster_pri"])
table_charac2[2,2] <- mean(input2[!duplicated(input2$cluster_pri_id.y) & input2$schools_cluster_
pri>1 & input2$urban_cluster_pri==1,"schools_cluster_pri"])
table_charac2[2,3] <- mean(input2[!duplicated(input2$cluster_pri_id.y) & input2$schools_cluster_
pri>1 & input2$urban_cluster_pri==0,"schools_cluster_pri"])
table_charac2[2,4] <- mean(input2[!duplicated(input2$cluster_sec_id.y) & input2$schools_cluster_
sec>1,"schools_cluster_sec"])
table_charac2[2,5] <- mean(input2[!duplicated(input2$cluster_sec_id.y) & input2$schools_cluster_
sec>1 & input2$urban_cluster_sec==1,"schools_cluster_sec"])
table_charac2[2,6] <- mean(input2[!duplicated(input2$cluster_sec_id.y) & input2$schools_cluster_
sec>1 & input2$urban_cluster_sec==0,"schools_cluster_sec"])

# Students per cluster
table_charac2[3,1] <- sum(input2[input2$schools_cluster_pri>1,c("students_preparatory","students_
psei","students_grade2",
                                                        "students_grade3","students_grade4")])/
table_charac2[1,1]
table_charac2[3,2] <- sum(input2[input2$schools_cluster_pri>1 & input2$urban_cluster_
pri==1,c("students_preparatory","students_psei","students_grade2",
                                                                                     "students_
grade3","students_grade4")])/table_charac2[1,2]
table_charac2[3,3] <- sum(input2[input2$schools_cluster_pri>1 & input2$urban_cluster_
pri==0,c("students_preparatory","students_psei","students_grade2",
                                                                                     "students_
grade3","students_grade4")])/table_charac2[1,3]
table_charac2[3,4] <- sum(input2[input2$schools_cluster_sec>1,c("students_grade5","students_
grade6","students_grade7",
                                                        "students_grade8","students_
grade9","students_grade10",
                                                        "students_grade11","students_grade12")])/
table_charac2[1,4]
table_charac2[3,5] <- sum(input2[input2$schools_cluster_sec>1 & input2$urban_cluster_
sec==1,c("students_grade5","students_grade6","students_grade7",
                                                                                     "students_
grade8","students_grade9","students_grade10",
                                                                                     "students_
grade11","students_grade12")])/table_charac2[1,5]
table_charac2[3,6] <- sum(input2[input2$schools_cluster_sec>1 & input2$urban_cluster_
sec==0,c("students_grade5","students_grade6","students_grade7",
                                                                                     "students_
grade8","students_grade9","students_grade10",
                                                                                     "students_
grade11","students_grade12")])/table_charac2[1,6]

# Students per class before clustering (maybe change "classes_before_pri.y" to "classes_before_pri")
classes_before_pri$classes_before_pri <- rowSums(classes_before_pri[,2:7])
classes_before_sec$classes_before_sec <- rowSums(classes_before_sec[,2:9])
input2 <- merge(input2, classes_before_pri[,c("cluster_pri_id","classes_before_pri")],
by.x="cluster_pri_id.y", by.y="cluster_pri_id")
input2 <- merge(input2, classes_before_sec[,c("cluster_sec_id","classes_before_sec")],
by.x="cluster_sec_id.y", by.y="cluster_sec_id")

table_charac2[4,1] <- table_charac2[1,1]*table_charac2[3,1]/sum(input2[!duplicated(input2$cluster_
pri_id.y) & input2$schools_cluster_pri>1,"classes_before_pri"])
table_charac2[4,2] <- table_charac2[1,2]*table_charac2[3,2]/sum(input2[!duplicated(input2$cluster_
pri_id.y) & input2$schools_cluster_pri>1 & input2$urban_cluster_pri==1,"classes_before_pri"])
table_charac2[4,3] <- table_charac2[1,3]*table_charac2[3,3]/sum(input2[!duplicated(input2$cluster_
pri_id.y) & input2$schools_cluster_pri>1 & input2$urban_cluster_pri==0,"classes_before_pri"])
table_charac2[4,4] <- table_charac2[1,4]*table_charac2[3,4]/sum(input2[!duplicated(input2$cluster_
```

```
sec_id.y) & input2$schools_cluster_sec>1,"classes_before_sec"])
table_charac2[4,5] <- table_charac2[1,5]*table_charac2[3,5]/sum(input2[!duplicated(input2$cluster_
sec_id.y) & input2$schools_cluster_sec>1 & input2$urban_cluster_sec==1,"classes_before_sec"])
table_charac2[4,6] <- table_charac2[1,6]*table_charac2[3,6]/sum(input2[!duplicated(input2$cluster_
sec_id.y) & input2$schools_cluster_sec>1 & input2$urban_cluster_sec==0,"classes_before_sec"])

# Students per class after clustering (maybe change "classes_after_pri.y" to "classes_after_pri")
classes_after_pri$classes_after_pri <- rowSums(classes_after_pri[,2:7])
classes_after_sec$classes_after_sec <- rowSums(classes_after_sec[,2:9])
input2 <- merge(input2, classes_after_pri[,c("cluster_pri_id","classes_after_pri")], by.x="cluster_
pri_id.y", by.y="cluster_pri_id")
input2 <- merge(input2, classes_after_sec[,c("cluster_sec_id","classes_after_sec")], by.x="cluster_
sec_id.y", by.y="cluster_sec_id")

table_charac2[5,1] <- table_charac2[1,1]*table_charac2[3,1]/sum(input2[!duplicated(input2$cluster_
pri_id.y) & input2$schools_cluster_pri>1,"classes_after_pri"])
table_charac2[5,2] <- table_charac2[1,2]*table_charac2[3,2]/sum(input2[!duplicated(input2$cluster_
pri_id.y) & input2$schools_cluster_pri>1 & input2$urban_cluster_pri==1,"classes_after_pri"])
table_charac2[5,3] <- table_charac2[1,3]*table_charac2[3,3]/sum(input2[!duplicated(input2$cluster_
pri_id.y) & input2$schools_cluster_pri>1 & input2$urban_cluster_pri==0,"classes_after_pri"])
table_charac2[5,4] <- table_charac2[1,4]*table_charac2[3,4]/sum(input2[!duplicated(input2$cluster_
sec_id.y) & input2$schools_cluster_sec>1,"classes_after_sec"])
table_charac2[5,5] <- table_charac2[1,5]*table_charac2[3,5]/sum(input2[!duplicated(input2$cluster_
sec_id.y) & input2$schools_cluster_sec>1 & input2$urban_cluster_sec==1,"classes_after_sec"])
table_charac2[5,6] <- table_charac2[1,6]*table_charac2[3,6]/sum(input2[!duplicated(input2$cluster_
sec_id.y) & input2$schools_cluster_sec>1 & input2$urban_cluster_sec==0,"classes_after_sec"])


##################################################################################
### Do mountain settlements have different driving speed than other settlements? ###
### Find clusters_sec_d and cluster_pairs_sec and merge on cluster1 and cluster2 ###
##################################################################################
temp_data <- merge(cluster_pairs_sec, clusters_sec_d[,c("cluster_sec_id","mountain")],
by.x="cluster1", by.y="cluster_sec_id")
colnames(temp_data)[ncol(temp_data)] <- "mountain1"
temp_data <- merge(temp_data, clusters_sec_d[,c("cluster_sec_id","mountain")], by.x="cluster2",
by.y="cluster_sec_id")
colnames(temp_data)[ncol(temp_data)] <- "mountain2"

write.table(temp_data, "R_output/driving_mountain.txt", sep=",", row.names=FALSE)
```