

**Children of Immigrants Longitudinal Survey
in Four European Countries
CILS4EU**

Technical Report

Wave 1 – 2010/2011

Version: 1.2.0

Year: 2016

Citation: CILS4EU. 2016. Children of Immigrants Longitudinal Survey in Four European Countries. Technical Report. Wave 1 – 2010/2011, v1.2.0. Mannheim: Mannheim University.

Content

1	Aim of the study	1
2	Sampling design	2
2.1	Target Population	2
2.2	Coverage	3
2.3	Sample Selection	6
2.3.1	School Sample Selection	7
2.3.2	Class Sample Selection	20
2.3.3	Sample Selection within classes	23
2.4	Response rates and sample sizes	24
2.4.1	Response rates	24
2.4.2	Sample sizes	29
2.4.3	Composition of the sample	31
3	Weighting	33
3.1	Design weights	34
3.2	Adjustment weights	36
3.3	Final Student Weight	37
3.4	House Weight	37
3.5	Senate Weight	37
4	Development of the instruments	38
4.1	Instruments for the student survey	38
4.1.1	Student main questionnaire	38
4.1.2	Name generator – “Five best friends”	39
4.1.3	Achievement test	39
4.1.4	Sociometric instrument	42

4.2	Parental questionnaire	43
4.3	Teachers' questionnaire	44
4.4	Adaption and translation of the instruments	44
4.4.1	Adaptions and country specific questions	45
4.4.2	Translation of the instruments	46
4.4.3	Minority languages	47
5	Fieldwork	49
5.1	Students' survey	49
5.2	Parental survey	51
5.3	Teacher survey	52
	References	53

1 Aim of the study

The project focuses on the intergenerational integration of children of immigrants in four selected European countries: Germany, the Netherlands, Sweden, and the United Kingdom. Funded within the NORFACE¹ programme, it is the first comprehensive and fully-standardized panel study on this topic in Europe.

Children of immigrants and their ethnic majority peers at age 14 are interviewed in the school year 2010/2011, as well as their parents and teachers, with a subsequent follow up of the adolescents over the next two years, thus covering a crucial, formative period of their lives. Based on these data, it is possible to investigate the complex causal interplay between the processes of structural, social, and cultural integration. The project starts from the assumption that only thereby can one account for the important differences between countries, ethnic groups, and domains of life. There is a substantial body of prior research on the integration of the second generation in Europe that has already taken steps to reveal these important differences. However, CILS4EU is the first project to collect data on immigrants' children in Europe that satisfies different important needs sufficiently: large-scale, strictly comparative, theory-guided, multilevel and longitudinal. All data will be made available to the international research community for public use. Thus, in addition to the project team's own substantive research contributions, the aim is to build an enduring infrastructure for continuing research on the intergenerational integration of immigrants in Europe.

This technical report provides an outline of the sampling of the students and their families, the fieldwork procedures and the development of the different instruments.

¹ NORFACE stands for New Opportunities for Research Funding Agency Co-operation in Europe (<http://www.norface.org>). NORFACE is a partnership between fifteen research councils to increase co-operation in research and research policy in Europe.

2 Sampling design

Given the comparative scope of the study, a standardized sample selection in all countries is fundamental. CILS4EU used a stratified three-stage sample design, with first stage sampling units being schools that were selected with probability proportional to size. The second stage sampling units are classes within these schools, whereas the third stage sampling units are the students enrolled in these classes. This chapter describes the target population (2.1), provides an overview of the coverage (2.2), the general sampling design (2.3) as well as the response rates and achieved sample sizes on school, student, parent, and teacher level (2.4).

2.1 Target Population

The general idea behind CILS4EU was to survey 14-year-old children with and without an immigrant background. One can approach 14-year-old children either at home or at school. For two major reasons we decided using a school-based sampling approach. Firstly, register-based sampling is not possible in the United Kingdom, and interviewing at home would require comprehensive screening or focused enumeration, both of which are prohibitively expensive. Secondly, and even more important, school sampling provides context information, including information from co-ethnic and interethnic peers, which is missing in most studies but extremely relevant from a theoretical point of view.

Considering these advantages of a school-based sampling approach, the target population of CILS4EU consists of students attending the school grade in Dutch, English, German and Swedish schools in which most of the students already are (or will become) 14-years old. These are the 3rd grades of secondary schools in the Netherlands, the 8th grades in Sweden, the 9th grades in Germany, and the 10th grades in the United Kingdom. Whenever in the following the term “students” is used, students falling in this definition of the CILS4EU-target population are meant.

To supplement the measures obtained through the students’ survey, we aimed to survey also the parents as well as the teachers of the participating students. Regarding the parental survey, the parents of each student in the sampled target grade were seen as the target population, whereas the form teacher of the sampled target grades represented the target

population for the teacher survey. In contrast to the students' survey, the parents as well as the teachers were interviewed in a cross-sectional manner i.e. parallel to the first wave of the students' survey end of 2010/beginning of 2011.

2.2 Coverage

To provide valid measures of children's structural, social, and cultural integration in the participating countries, the sample of students had to be selected in a way that ensures representation of the full target population of each respective country. Therefore, the aim was to follow a full-coverage strategy in all countries, desisting from the exclusion of areas with low immigrant proportion or of small and remote geographical regions due to cost reasons. Nevertheless, several deviations from this strategy have to be reported. Firstly, and despite the aimed full coverage strategy, some countries had to remove larger groups of schools from specific regions (*reduced population coverage*). Secondly, it was sometimes necessary to exclude complete schools from the sampling frame (*exclusion of schools*). And finally, students had to be excluded within participating schools (*exclusion of students within participating schools*).

Before going into detail of reduced coverage, an important prior decision regarding the fieldwork in the United Kingdom has to be reported. For the United Kingdom, we decided to restrain the geographical area to England rather to include Scotland, Wales and Northern Ireland as well. The main reason for this restriction was the sometimes large institutional differences between the educational systems in the different countries. These differences would have made separate studies necessary but the distribution of the overall expected number of cases on student level on the four different countries with resulting smaller number of cases within the countries (especially with a further differentiation with respect to ethnic origin) would have meant almost no possibility to analyse the consequences of such institutional differences effectively. Furthermore, Welsh students are often taught in Welsh, which would have led to further problems. Therefore, the restriction to survey only schools in England rather in the United Kingdom seems to be meaningful in several ways. In the following, the term "England" will be used instead of "United Kingdom", and the coverage of England will be reported subsequently.

Reduced population coverage

For Germany, we intended to cover all Federal States in the sample. However, Bavaria refused to participate in the study. As a consequence, only 15 Federal States are included in the German sample.

Exclusions on school level

One deviation from the full-coverage-strategy that applies to all countries is the handling of small schools. Although using the technique *sampling with probability proportional to size* (PPS) that ensures a small inclusion probability for small schools, we decided to exclude very small schools from the sampling frame. Very small schools are defined as schools with number of students in the target class level smaller than one quarter of the average class size, calculated over all schools with two or more classes in the relevant grade. However, these very small schools are only excluded if the total number of students in these schools is less than 2% of the total target population. Exclusions of small schools is mainly due to cost reasons.

Furthermore, special schools for mentally and physically disabled children were excluded from the school population prior to sampling. In Germany, however, schools for students with learning disabilities were included in the school population (“Förderschulen mit dem Schwerpunkt Lernen”), as this is a quite relevant school type for immigrant children (Kornmann and Kornmann 2003). In addition, private and boarding schools were not registered in the sampling frame initially. Therefore, the exclusion rate cannot be calculated for these school types.

In contrast, the Dutch team excluded schools for students with learning disabilities from the sampling frame (“praktijkonderwijs”).

England also excluded several schools from the sampling frame of the main survey. Most of these schools were special schools or children’s homes. Some minor exclusions were

necessary as some schools were about to close during the fieldwork time. Lastly, very few schools were excluded as no school data was available for these schools.²

In Sweden, resource schools, hospital schools, special day schools and other special units as well as non-active schools were deleted from the population before the school sample was drawn.

The exclusion rates displayed in Table 1 refer to the exclusion rate of schools, not students. As mostly small schools were excluded (“very small schools” as an exclusion criterion as well as the usually smaller class/school-sizes in special schools), the exclusion rate on students’ level is significantly smaller.

Exclusions within schools

In addition to these exclusions on school level, it was also possible to exclude single students within an eligible and sampled school. This is the case for intellectually and functionally disabled students who are disabled in a way that they cannot perform the CILS4EU questionnaire and achievement test. The same holds true for children with inadequate language skills in the majority language that do not allow filling out the questionnaire. In all countries, exclusions within schools are negligible.

Table 1 shows the population coverage and the exclusion rates on school level and within schools. The comparable high exclusion rates in England on school level are due to the exclusion of Special Schools (for both independent and state schools) and of Children’s homes (independent schools).

² In addition to these exclusions, several schools were excluded due to their participation in other surveys conducted by NatCen, the institute being responsible for the fieldwork in England. The main argument for excluding these schools was that convincing them to take part in two studies in rapid succession would have been an extremely difficult task due to time constraints schools usually have to face. That was also the reason to exclude six schools being issued by NatCen for the pilot of CILS4EU. As all of these schools were sampled in a similar or exactly the same way as the CILS4EU schools, the assumption is plausible that the exclusion of these schools doesn’t unduly bias the remaining sample. Therefore, these schools are not listed here as school level exclusions.

Table 1: Population coverage and exclusions

	Population coverage (%)	School level exclusions (%)	Within-school exclusions (%)
England*	100/100	22.0/33.6	0/0
Germany [†]	84.9	2.4	0.1
Netherlands [‡]	100	6.8	0
Sweden	100	11.5	0

Note: * First figure in each column represents the coverage for state/maintained schools, the second figure the one for independent schools.

[†] The reduced population coverage is due to the exclusion of the Federal State of Bavaria. However, the 84.9% refer to the coverage *after* school level exclusions were made. Therefore, the population coverage refers to the coverage of all eligible schools in the sampling frame.

[‡] Special schools were excluded, but there is no information available how many schools are affected by this. Therefore, exclusion rate applies only to Dutch schools *without* special schools.

2.3 Sample Selection³

In the school year 2010/2011 we intended to interview 4,000 students at minimum in each country, comprising at least 1,500 students with an immigrant background. To achieve this, we aimed to sample at least 100 schools in each country. This is an adequate balance between having sufficient schools (thus reducing sampling error on that level and, as a consequence, allowing meaningful comparisons between countries) and having sufficient students within each school (thus allowing us to compute reliable context measures for multi-level analyses).

In all countries, the sampling design used for CILS4EU was a three-stage stratified sample design. The *first-stage sampling units* consisted of individual schools enrolling our relevant target grades, i.e. grades with mainly 14-year-old students. After the exclusions on school level described in the previous section, schools were sampled from a comprehensive national list of all eligible schools (school sampling frame) with probabilities that were proportional to the size of the school (PPS). Prior to sampling, schools in the sampling frame were assigned to mutually exclusive groups (explicit strata) according to the proportion of students with an immigrant background in the schools. The aim of this stratification-strategy was to oversample schools with high proportions of students with an immigrant background to achieve the desired case numbers of children with an immigrant background.

³ The Sampling Manual used in all countries to select the first-, second-, and third-stage sampling units (“CILS4EU_Sampling Manual.pdf”) as well as parts of this section are based on close collaboration with the IEA Data Processing Center DPC (<http://www.iea-dpc.de>; i.e. Heiko Sibbers, Carolin Vandenplas and Olaf Zuehlke). Cf. also Zuehlke (2011): “Sampling design and implementation” in the ICCS Technical Report, pp 59–68.

The *second-stage sampling units* were classes in our target grades within sampled schools. To reduce cluster effects (e.g. by including very large schools), we selected two school classes at random wherever more than two classes were available. In cases where the school enrolled only one or two classes in the relevant age group, this/these class/classes was/were selected.

The main *third-stage sampling units* were students within sampled classes. At this last stage, all students within a sampled grade were included in the sample, with the exception of those who were excluded due to the rules in section 2.2, *Exclusions within schools*. Furthermore, we also sampled at this third stage the parents and the teachers for the parental and the teachers' survey. A detailed description of the selection of the three sampling units as well as minor variations from this standard procedure in different countries is discussed in the following subsections (2.3.1, 2.3.2 and 2.3.3).

2.3.1 *School Sample Selection*

In the following, we will go into detail about the source and the design of the school sampling frame (2.3.1.1), the explicit and implicit stratification criteria (2.3.1.2), the selection (2.3.1.3) and finally the recruitment of the schools (2.3.1.4).

2.3.1.1 *Sources and design of the sampling frame*

The schools for the school sample were selected out of all schools enrolling our target grades in each country. The basis of selection was a comprehensive school list comprising all schools enrolling students in the relevant age group, deducting the schools being excluded as described in section 2.2. In addition to the name of the school, the school list also comprised information about the address of the school; a measure of the size of the school, which is the number of classes or the number of students in the relevant grade level; the proportion of foreign national or immigrant background students in the school; and information about school type (where applicable, e.g. in England, Germany and the Netherlands). The source of the sampling frame in each country as well as the information contained is listed in Table 2.

Table 2: Source of and information in the sampling frame in the different countries

Source of the sampling frame	Information included
England <ul style="list-style-type: none"> • National Pupil Database (NPD) – for state maintained schools • Department of Education – for independent schools 	<ul style="list-style-type: none"> • Name of the school • Address of the school • Region • School type • Number of students in grade 10 • Proportion of non-White British students in grade 10 (for state schools only)
Germany <ul style="list-style-type: none"> • The school list for each German federal state was delivered by the local authorities and statistical offices. • The resulting school sampling frame for Germany is a unique source of information that is also used for the National Educational Panel Study (NEPS) 	<ul style="list-style-type: none"> • Name of the school • Address of the school • Possible contact person • Region • School type • Number of students in grade 9 • Proportion of foreign nationals in grade 9
The Netherlands <ul style="list-style-type: none"> • The school list was delivered by Dienst Uitvoering Onderwijs (DUO/CFI), a Dutch governmental organization responsible for – among other things – documenting numbers and characteristics of all Dutch schools and the students attending these schools. • The list provides information about each secondary school (except for the special schools) in the Netherlands measured at October 1st, 2008. 	<ul style="list-style-type: none"> • Name of the school • Address of the school • Denomination of the school • Educational level/school track • Number of students with Iwo-indication (Iwo= leerweg ondersteuning; students needing extra coaching) • Number of students in grade 3 • Sex of students • Ethnicity of students (ethnicity of students is based on country of birth of student, father and mother according to CBS definition, grouped by natives, Western immigrants, Turkish, Moroccan, Surinamese, Antillean, other non-Western immigrants, unknown)
Sweden <ul style="list-style-type: none"> • The Pupil Register of the Swedish National Agency for Education (“Skolverkets elevregister”) • The School Register 	<ul style="list-style-type: none"> • Name of school • Address of the school • Total number of students in grade 7 of the previous school year • Proportion of students with foreign background in grade 7 of the previous school year (students with at least one biological parent born outside the Western hemisphere)

A “school” in the school list was defined as an administrative unit rather than a building. Especially in the Dutch case, different school types are offered/taught within the same building. Each school type represents an administrative unit, resulting in different administrative units being located within one building. All students within these administrative units were listed on mutually exclusive school lists, so that one student had only the chance to get selected once.

2.3.1.2 Stratification and Ordering within the Sampling frame

Prior to the selection of schools out of the sampling frame, schools were explicitly and implicitly stratified. The stratification procedure as well as the sorting of the sampling units following from the implicit stratification is described in the following sections.

2.3.1.2.1 Explicit Stratification

To ensure the desired number of immigrant background children in the sample and to guarantee enough variance on the theoretical relevant variable “immigrant proportion in school”, schools were stratified within the sampling frame according to their immigrant proportion in the relevant grade. In doing so, we were able to oversample schools with high immigrant proportions in the respective target grade. As boundaries for the four different strata, we used the following immigrant proportions (ip)⁴:

Stratum 1: $0 \leq ip < 0.1$

Stratum 2: $0.1 \leq ip < 0.3$

Stratum 3: $0.3 \leq ip < 0.6$

Stratum 4: $0.6 \leq ip \leq 1$

These different strata are then treated as separate sampling frames when selecting the schools.

The information that could be used to approximate the criterion “immigrant proportion” in grade level differed considerably between the countries. Therefore, and in order to follow the stratification approach as comparable as possible in all countries, we had to use other criteria that are closely connected to the “immigrant proportion”, although they do not perfectly substitute our initial explicit stratification characteristic. These criteria are described in the following. Furthermore, some countries tried to overcome the inaccuracies resulting from the use of the available information compared to the criterion “immigrant background” and applied some adaptations to the supplementary criteria, which are also described in the next paragraphs.

⁴ In the Netherlands, slightly different boundaries were used: Stratum 1: $0 \leq ip \leq 0.1$ / Stratum 2: $0.1 < ip \leq 0.3$ / Stratum 3: $0.3 < ip \leq 0.6$ / Stratum 4: $0.6 < ip \leq 1$

England

In England, the explicit stratification variable was the proportion of non-White British students, as only this information and not the heritage of the students is recorded in the National Students Database (NPD). However, “Understanding Society” as a nationally representative longitudinal survey contains both kind of information; therefore, the data was used to estimate the proportion of students with an immigrant background, which is 79%. This means that the “non-White British student”-category is a broader category than the “immigrant”-category, meaning that the proportion of non-White British students in a school was multiplied by 0.79 in order to get the anticipated proportion of students with an immigrant heritage.

In contrast to the maintained or state schools, independent schools provide no information about the proportion of non-White British students or similar measures. Therefore, independent schools were located in a different explicit stratum and were sampled separately.

Germany

In the German school sampling frame, no information about the immigrant background was available for some Federal States either. Instead, information about the nationality of the child was used as the stratification characteristic. By using nationality instead of immigrant background, however, one underestimates the proportion of children with an immigrant background in schools, as all naturalised children with an immigrant background are seen as native Germans, not as immigrants/children of immigrants. In contrast to the situation in England, the available information provides a narrower category compared to the categorisation of students we are aiming for.

To correct for this, information from prior school surveys conducted at the University of Mannheim where both information was accessible was used to calculate a correction factor to adjust the proportion of children with foreign nationality to the proportion of children with immigrant background. The correction factor was calculated by dividing the proportion of actual immigrants through the proportion of foreign nationals in the respective

data sets. For example, having 10% foreign nationals in one of the surveys used and 12% actual immigrants, the correction factor would be 1.2. This means that when facing a certain proportion of foreign nationals in the data set, one can expect that the actual immigrant proportion is 1.2 times higher than the reported proportion of foreign nationals. As it turned out that the factor differs between different strata (i.e. higher rates of naturalisation in schools with low immigrant proportions), four different correction factors were calculated.

Table 3 shows the correction factors for the different strata. These factors mean the following: In schools within the stratum 1 ($0 \leq ip < 0.1$) with an exemplary proportion of foreign nationals of 5%, actually 10% of students with an immigrant background are expected

Table 3: Correction Factor for immigrant proportion in Germany

Stratum	Correction factor
1	2.0
2	1.7
3	1.5
4	1.2

to be found. Before the actual stratification of the schools, the share of foreign students in the school was multiplied with the correction factor, resulting in an approximation of the proportion of immigrant students in the school. The schools were then ordered by their (newly calculated) proxy information for “immigrant proportion” and grouped into the different strata in the sampling frame with the boundaries described above.

Netherlands

In the Netherlands, immigration background is defined as persons who are themselves, or have at least one parent, born in a non-Western country. Non-western countries are countries in Africa, Latin-America, Asia (excluding Indonesia and Japan), and Turkey (CBS, 2010).

The stratification of schools is based on the proportion of students with an immigrant background attending third grades in the school year 2008/2009. Figures for the immigrant proportion for the school year 2010/2011 (i.e., the year in which the fieldwork started) were not available yet. The reason for using immigrant proportion within the third grade population in 2008 instead of immigrant proportion within the first grade population in 2008 (who would attend third grade at the time of the fieldwork) was that only for third grades we were able to implicitly stratify by school track. The figures we received from the Dutch Central

Funding of Institutions Agency CFI (Centrale Financiën Instellingen) did not allow us to implicitly stratify by school track in the first grade. In the first grade, almost all students attend the so-called “brugklas”, a bridging year. The “brugklas” is a combination of two adjacent school tracks. Based on the school results in this “brugklas”, a decision is made whether the student will attend the lower or the higher track in subsequent school years.

Sweden

Schools were stratified according to the proportion of students in grade 7 the previous school year (i.e. the target cohort) with “foreign background”, defined as the students themselves or at least one biological parent born in a “non-Western” country. The definition of “West” comprises: Australia, Belgium, Canada, Denmark, Faroe Islands, Finland, France, Germany, Greenland, Ireland, Iceland, Liechtenstein, Luxembourg, Monaco, Netherlands, New Zealand, Norway, Austria, Sweden, Switzerland, United Kingdom, United States, i.e., Western Europe (excluding South and East Europe), the U.S., Canada, New Zealand, and Australia.

2.3.1.2.2 *Implicit stratification*

In addition to the explicit stratification described above, other implicit stratification characteristics like *region* or *school type* were used in some of the countries. This was done to ensure the representation of all different school types or all regions in one country's school sample. Furthermore, implicit stratification and the specific ordering procedure within the explicit strata of the sampling frame is helpful to ensure a maximum of comparability of replacement schools, that are selected if a sampled school refuses to participate (cf. 2.3.1.3). Table 4 shows the different implicit stratification characteristics used in the different countries.

Table 4: Implicit stratification characteristics

England	<ul style="list-style-type: none"> • Region (i.e. the 8 administrative Government Office Regions in England) • School type <ul style="list-style-type: none"> - Religious – any denomination - Selective – and not religious - Academy – and not religious or selective - Community – and not religious or selective - Foundation – and not religious or selective - Voluntary Aided – and not religious or selective - Voluntary Controlled – and not religious or selective - City Technology College – and not religious or selective
Germany	<ul style="list-style-type: none"> • Region (i.e. 15 Federal States, excluding Bavaria) • School type <ul style="list-style-type: none"> - Hauptschule - Realschule - Gymnasium - Integrierte Gesamtschule (comprehensive schools) - Waldorfschule - Förderschule - Schulen mit mehreren Bildungsgängen (schools with several tracks)
Netherlands	<ul style="list-style-type: none"> • school type <ul style="list-style-type: none"> - vmbo-basis/vmbo-kader/vmbo-mbo2 - vmbo-gemengd/vmbo-theoretisch - havo; - vwo/English track/International baccalaureate
Sweden	<ul style="list-style-type: none"> • None

2.3.1.2.3 *Ordering within the Sampling frame*

All schools in the explicit strata were sorted according to their (approximate) immigrant proportion. If implicit stratification was used, then the sorting by immigrant proportion was done within each implicit stratum using a serpentine approach (cf. Joncas 2008). Such a serpentine approach was also used for the second (if applicable) implicit stratifiers.

Table 5: Sort Order for Implicit Strata within one exemplary explicit strata

Implicit Stratifier 1: School type	Implicit Stratifier 2: Region	Implicit stratum (School type in Region)	Sort order of “immigrant proportion” in this implicit stratum
Lower	Region A	Lower – Region A	High to Low
Lower	Region B	Lower – Region B	Low to High
Lower	Region C	Lower – Region C	High to Low
Middle	Region C	Middle – Region C	Low to High
Middle	Region B	Middle – Region B	High to Low
Middle	Region A	Middle – Region A	Low to High
Higher	Region A	Higher – Region A	High to Low
Higher	Region B	Higher – Region B	Low to High
Higher	Region C	Higher – Region C	High to Low

As can be seen from the example in Table 5 with two exemplary implicit stratifiers “school type” and “region”, all schools in one explicit stratum are firstly ordered according to their first implicit stratifier, i.e. school type (here from lower to higher school tracks). Within the schools of one specific track, schools are then ordered according to their second implicit stratifier, i.e. region the schools are located in (e.g. alphabetically; in this example three hypothetical regions A, B and C). This is done by using the so called serpentine approach: for the lower tracks, schools are ordered according to the Region alphabetically from A to C, while the schools in the middle track are ordered from C to A. This results in (number of school types) x (number of regions) different implicit strata. Within these implicit strata, the schools are then ordered according to the immigrant proportion in school. Again, the serpentine approach is used in the different implicit strata.

2.3.1.3 Selection of the schools

In the following two sections the rationale behind choosing a specific number of schools within each explicit stratum as well as the selection process of the schools within the strata are described.

2.3.1.3.1 Number of schools to be chosen per stratum

After the stratification of the schools according to the available information on (approximate) immigrant proportion and possible further implicit stratification criteria, schools were selected out of the different strata. The number of schools that have to be selected per stratum depends on the desired number of students sampled in the different strata. The number

of schools is then simply a function of number of classes selected per school and average number of students in class as well as the average expected response rates on students' level per stratum.

The allocation of the desired total number of 4,000 students over the different strata in each country is the result of a twofold consideration: (1) Oversampling of students with an immigrant background and (2) sampling of a most effective sample, while ensuring comparability between the countries. The first goal leads to a sampling strategy where we oversampled students (and therefore schools) in the third and especially fourth stratum (students in schools with 30% or more students with an immigrant background) to ensure to get an adequate number of immigrant background students in the sample. To achieve the second goal, we followed the considerations proposed in the Neyman allocation approach (cf. Groves et al. 2004: 117), with the general strategy to maximise the effectiveness of the sample with respect to the immigrant background sub-sample, as this is the most relevant group of the survey.

The Neyman allocation deals with the problem how to allocate the sampling units (students) over the different strata to achieve an as effective as possible sample – that is a sample with the smallest sampling variance. We will present the two main general considerations to uncover the important design principle proposed by the Neyman allocation approach:

- Like in proportionate sampling, one should allocate more of the sample to one stratum if the stratum in the population is large.
- However, if there is more variability among the elements in one stratum, then one should allocate also more of the sample to this stratum.

Given these two key considerations, the Neyman allocation approach requires some information about the distribution of the population over the different strata as well as the variance of characteristics in the different strata, which are of interest for the given survey. The data from the PISA survey can be a very helpful source of such information, as this is a representative sample. As such, information about the distribution of 14-15 year old students in schools with different immigrant proportions is available. Furthermore, by using PISA data,

it is possible to assess the variance of the PISA language test results as one of the key outcome variable in the integration process in each of the four strata. As PISA covers all of the CILS4EU countries, it is probably the best data source to explore the question how to allocate the sampling units over the different strata.

By using PISA data, it can be shown that immigrant background students are found predominantly in the stratum with 10–30% immigrant proportions, and the variance of the immigrant background students' language test scores is also highest in this stratum. This holds true for all of the four participating countries, with more or less differences. Given this finding and following the key considerations of the Neyman allocation approach, we allocated the largest proportion of the sample (32.5%) to this stratum. In contrast, for the first stratum, there were only very few students with an immigrant background, with more homogenous achievement scores. As such, the smallest proportion of the sample is allocated to this stratum (17.5%).

Because of the oversampling considerations, we decided to allocate 50% of the sample equally to the third and fourth stratum. Based on these considerations, the distribution of students over the strata shown in Table 6 resulted. From this

Table 6: Stratification over strata

Stratum	Proportion of students
1	17.5
2	32.5
3	25.0
4	25.0
Total	100

distribution, the number of schools needed in each stratum was calculated using the estimated average number of students expected in a school class, assuming the selection of two classes per school, as well as the expected response rates on school level per stratum. According to these figures, target numbers on school level in each stratum were defined. These figures are displayed in Table 7.

Table 7: Target numbers of schools per stratum in each country (%)

Stratum	England	Germany	Netherlands	Sweden
1	24 (17.1)	20 (13.7)	18 (16.5)	28 (16.7)
2	46 (32.9)	40 (27.4)	34 (31.2)	53 (31.5)
3	35 (25)	36 (24.7)	31 (28.4)	43 (25.6)
4	35 (25)	50 (34.2)**	26 (23.9)	44 (26.2)
Total	140 (100)	146 (100)	109 (100)	168 (100)
Indep.*	23 (14.1)	-	-	-
Total	163 (100)	-	-	-

* Independent schools provide no information about proportion of non-White British pupils and are therefore listed in a separate stratum

** The large number of schools targeted in Germany in the immigrant dense stratum is due to the fact that comparable small classes are located in this stratum according to recent research projects.

The concrete number of cases achieved on school and student level in each country can be found in Table 13 in section 2.4.2. The process of selecting the schools and the replacement schools is described in the next chapter.

2.3.1.3.2 Selection of schools and the handling of non-respondents

After the decision was made concerning how many schools to select in the different explicit strata, the school random selection process started. The basis for the school selection was the sampling frame, where all eligible schools were listed and ordered according to the implicit stratification variables. Furthermore, within the so defined implicit strata, schools were again sorted according to the explicit stratification criterion, i.e. immigrant proportion or its proxies. In addition, schools were represented in the sampling frame according to their size, measured by number of students in the respective grade level. The representation of schools according to their size is necessary when sampling with probability proportional to size (PPS) is intended.

According to the number of schools to be selected out of the different explicit strata and the total number of schools per stratum, one can calculate a sampling interval to select the schools. Using this sampling interval and a random start point within the first sampling interval in a specific stratum, one can use the random start-constant interval approach to draw the sample of schools to be sampled – the so called first or initially sampled schools.

Non-response is a common problem in social research, not only general in population surveys, but also in school surveys (Sturgis et al. 2006). To account for non-responding schools in our survey, we implemented a replacement strategy, which is also followed in other large scale school surveys, like PISA and TIMSS (Adams 2003). The purpose of this replacement strategy is that participating schools are matched to non-participating schools according to the implicit and explicit stratification criteria in order to select similar schools as replacements for the non-participating schools. The “matching process” is done by ordering the schools according to their implicit and explicit stratification criteria in each explicit stratum as described in section 2.3.1.2.3. To select a comparable replacement school (at least with respect to the implicit and explicit criteria), one can simply select a school located

next to the initially sampled school, still ensuring a random sample, even with replacement schools (Adams 2003).

Table 8 illustrates the replacement procedure. Due to the stratification and the sorting within the different strata, replacement schools are very similar to the sampled schools with respect to immigrant proportion, school type and region (the explicit stratification criterion as well as the two implicit stratification criteria in our example).

Table 8: Example for replacement school selection – excerpt of a hypothetical sampling frame

Status	School-ID	Implicit stratum	Immigrant proportion
...
Replacement 4	22046	Lower – Region 1	0,32
Replacement 2	12579	Lower – Region 1	0,31
Sampled School	81644	Lower – Region 1	0,31
Replacement 1	60254	Lower – Region 1	0,30
Replacement 3	13235	Lower – Region 1	0,30
...

All countries but Sweden applied the replacement strategy described in this paragraph. In contrast to the procedure in the other countries, Statistics Sweden sampled 168 schools distributed over the strata as described in Table 7, refraining from any replacement strategy. As Sweden is the only country where no implicit stratification was implemented, the application of the replacement strategy wasn't seen as essential as in the other countries. This holds especially true as the explicit stratification and the usually low non-response rates of schools in comparable surveys in Sweden ensured the representation of schools with different ethnic compositions in our sample.

2.3.1.4 Recruitment of schools

The beginning and the duration of the school recruitment process in each country are displayed in Table 9, together with the persons responsible for the school contacts. For Germany, the school contacts were interrupted for six weeks during the summer breaks in the respective Federal States. The same holds true for the Netherlands, where no schools were contacted during the summer, autumn, Christmas and spring break.

Table 9: Beginning, duration and procedure of the school contact work

Country	Beginning of school contact work	End of school contact work	Covering letter sent to school before contact	Mode of contact/contact persons
England	September 2010	July 2011	no	Telephone contact by <i>NatCen's</i> telephone Unit Personal contacts by interviewers/test administrators themselves
Germany	May 2010	November 2010	yes	Telephone contact by student assistants, <i>University of Mannheim</i>
The Netherlands	June 2010	March 2011	yes	Telephone as well as personal contacts by student assistants, <i>University of Tilburg and Utrecht</i>
Sweden	November 2010	May 2011	yes	Telephone and email contact by <i>Statistics Sweden</i> after initial cover letter

In England, no cover letter was used. Instead, a £500 incentive was offered at the first telephone contact to all English schools in order to increase participation rates. Together with the telephone contacts, some schools were issued to interviewers to recruit face-to-face in cases where telephone contact had not been achieved after many attempts. Overall, 61 hard to reach schools were issued to field without being recruited by telephone; by the end of fieldwork class sessions were completed at 11 of these schools, yielding a slightly higher response rate of 18.0% compared to the overall response rates reported in Table 11.

In Germany, the Netherlands and Sweden, the sampled schools received a covering letter before the first personal contact in order to facilitate school recruitment. However, those Dutch schools that were contacted in the last weeks of approaching the schools (from about spring break onwards) received no letter due to time constraints. Furthermore, from January 2011 onwards, all Dutch schools in the stratum with the highest immigrant proportion that could still be approached were offered a financial incentive of 1,000€ for participation in Wave 1 (and already for the subsequent Wave 2). In total, twelve Dutch schools were offered an incentive. Five schools agreed to participate after the incentive was offered. Like in England, some school principals or contact persons in the Netherlands were contacted face-to-face in order to achieve their consent. Overall, ten schools in the stratum with the highest immigrant proportion were approached in a face-to-face contact at the beginning of

January 2011. Unfortunately, only one school agreed to participate after the face-to-face contact.

However, it was not always hard to convince schools to take part in the survey in the Netherlands. 18 Dutch schools that were not sampled participated in the survey voluntarily in addition to the randomly selected schools. These schools participated for several reasons (e.g., it was demanded by a school that was sampled that an additional track participated). In all the following tables reporting response rates and sample sizes, the figures are reported for the randomly selected schools only.

In Sweden, a covering letter was sent to all sampled schools in two versions: One addressed to the school principal and one to the concerned teacher. Each participating class received a remuneration of 1,000 SEK. This was transferred to the school principal after the students' completion of the survey. The principal decided whether to allocate this money to the class directly to the school as a whole.

The remarkably long duration of school contacts listed in Table 9 is due to the replacement strategy followed in the survey. Contacting and trying to persuade a single school to take part in the survey was very time-consuming and sometimes lasted several weeks. However, drawing and contacting replacement schools was only possible if the initially sampled school (or the previous replacement school) finally refused to participate. So, after a refusal, the time-consuming task of school recruitment started again for the replacement school. The partially high non-response rates on school level with the need of contacting several replacement schools for one initially sampled school (cf. Table 11) contributed to the long duration of school recruitment.

2.3.2 *Class Sample Selection*

The selection of classes was the second step in the three stage sampling process. The standard solution for this step was to randomly select two classes within the target grade in each school. If only one or two classes are available in the selected school, then these classes are selected by default. Due to different reasons, deviations from this standard procedure resulted in all countries.

England

In 6 English schools non-standard sampling methods were applied. Random selection of the school classes can't be guaranteed here. Furthermore, in 6 schools, the two sampled classes were interviewed together, without any possibility to separate the information afterwards. These classes within a school are listed as one school class in the data. Implications from this specificity for sociometric analyses can be found in the sociometric fieldwork report (Kruse/Jacob 2016).

Germany

In five German schools, random selection of classes wasn't possible: In two schools, the headmaster was not willing to let one of the randomly selected classes take part in the survey, and a specific class was chosen that was interviewed instead. In another two schools, headmasters only agreed to participate in the survey under the condition that the school will be allowed to choose two specific classes to participate. Thus, no random class selection took place at all. Finally, three instead of the two aimed classes participated in one school. In total, 264 out of the 271 (97.4%) tested classes were randomly selected.

Netherlands

In order to reach the aimed number of children with immigrant background based on the sampling scheme (n=1,500), schools belonging to stratum 4 were initially asked to participate with three classes. The reason for this deviation are the expected low response rates on school level in this stratum combined with the overall low number of immigrant dense schools in the Netherlands, exacerbating the replacement strategy in stratum 4. As these efforts turned out to be still insufficient to reach the target of 1,500 immigrants, the procedure was changed in October 2010. Once a school belonging to stratum 3 or 4 agreed to participate, the attempt was made to have as many classes as possible participating.

In 23 schools, the headmaster or the responsible teachers were not willing to let the randomly selected classes take part in the survey. Instead, specific classes were chosen by

the headmaster/teacher. In total, 177 out of the 222 tested classes (79.7%) were randomly selected.

Sweden

A random sample of 2 classes for each participating school was drawn with the following two exceptions. If the number of students in the school's two smallest classes did not sum to at least 40, then more than 2 classes were drawn. No deviations from random sampling have been reported for Sweden.

Table 10 provides an overview over the number of participating classes in each country over the different strata, as well as the average number of participating classes and the total number of participating classes.

Table 10: Class sample selection

		Number of schools with n participating classes					Total number of participating classes	Av. № of participating classes
		n=1	n=2	n=3	n=4	n=5		
England	Strata 1	0	19	0	0	0	38	2.0
	Strata 2	0	32	0	0	0	64	2.0
	Strata 3	0	24	0	0	0	48	2.0
	Strata 4	0	21	0	0	0	42	2.0
	Indep.	0	11	0	0	0	22	2.0
	Total	0	107	0	0	0	214	2.0
Germany	Strata 1	4	15	0	0	0	34	1.8
	Strata 2	2	38	0	0	0	78	2.0
	Strata 3	6	30	0	0	0	66	1.8
	Strata 4	6	42	1	0	0	93	1.9
	Total	18	125	1	0	0	271	1.9
Netherlands	Strata 1	2	11	1	1	1	36	2.3
	Strata 2	6	22	6	1	0	72	2.1
	Strata 3	4	16	3	2	3	68	2.4
	Strata 4	5	11	2	2	1	46	2.2
	Total	17	60	12	6	5	222	2.2
Sweden	Strata 1	0	18	0	1	0	40	2.1
	Strata 2	3	40	0	0	0	83	1.9
	Strata 3	2	30	0	1	0	66	2.0
	Strata 4	7	26	1	0	0	62	1.8
	Total	12	114	1	2	0	251	1.9
Total	Strata 1	6	63	1	2	1	148	2.0
	Strata 2	11	132	6	1	0	297	2.0
	Strata 3	12	100	3	3	3	248	2.0
	Strata 4	18	100	4	2	1	243	1.9
	Indep.	0	11	0	0	0	22	2.0
	Total	47	406	14	8	5	958	2.0

In contrast to the non-randomly selected, voluntarily participating schools in the Netherlands described above, non-randomly selected classes as well as the students attending these classes are represented in all subsequent tables. However, non-randomly selected classes (and students) in all countries (see above) can be identified by the flag variable [*random*] in the data set (cf. CILS4EU 2016a).

2.3.3 *Sample Selection within classes*

2.3.3.1 *Students Sample Selection*

Within the sampled classes, all students were selected. However, within-school exclusions were possible if students were mentally or physically disabled in a way that made completion of the questionnaire impossible. In addition, students with insufficient language skills in the questionnaire/test language were also excluded from the in-school survey. Table 1 in section 2.2 provides an overview over the exclusions within schools. As can be seen from these figures, there were only very few – if at all – exclusions within selected schools and grades in all of the four countries.

The problem of non-response is not only an issue on school level, but also on student level. Although one can assume that the willingness to cooperate in the survey during regular school hours is comparable high, as no extra time for the students is involved, students had the possibility to refuse to participate in the study. Furthermore, as target population of CILS4EU are children of under age, the parents were also able to decline the participation of their child in the survey. A more restrictive procedure was applied in Germany: here, all students needed an *active parental consent* (and in some States also an additional explicit *students' consent*) before they were allowed to take part in the survey. In addition to these refusals, non-response on student level also occurred because of absent students during the survey date.

2.3.3.2 *Parents Sample Selection*

In addition to the students selected for the survey in the sampled grade, we also aimed to complete interviews with the parents of the children taking part in our survey. Therefore, we asked the participating students to provide their parents with a self-completion questionnaire. However, we didn't specify whether the parental interview should be completed by the mother or the father, but left this totally to the students and their families. We included information in the questionnaire to identify ex post whether the mother, the father or somebody else completed the interview.

2.3.3.3 *Teacher Sample Selection*

For each sampled class we aimed to survey also a teacher to provide insights in the learning environment of the students within schools. As the teacher being mostly involved in the everyday school life of each sampled class seems to be the person best suited for this task, we aimed to recruit the class teacher to take part in a teacher questionnaire. However, due to practicability reasons, the teacher being present during administering the questionnaires participated in the teacher survey – which was in many cases actually the form teacher (cf. Table 19).

2.4 Response rates and sample sizes

The following sections provide an overview of the response rates in the different countries on school, class, students, parent and teacher level, as well as the number of replacement schools needed in each country (2.4.1), the number of participating schools, classes, students, parents and teachers (2.4.2) and the composition of the student and parent sample (2.4.3).

2.4.1 *Response rates*

According to the replacement strategy on school level described in section 2.3.1.3.2, one can calculate participation rates on school level before and after the school replacement. The

school participation rate before replacement ($PR_{schoolBR}$) is calculated as the ratio of all participating and initially sampled schools in one country ($n_{schoolsp}$) and the number of all originally sampled and initially contacted schools, which is the targeted number of schools (i.e. the number of sampling points, $n_{samplingp}$).

$$PR_{schoolBR} = \frac{n_{schoolsp}}{n_{samplingp}}$$

The school participation rate after replacement ($PR_{schoolAR}$) is calculated as the ratio of all participating schools in one country, which is the sum of all initially sampled and participating schools ($n_{schoolsp}$) and of all participating replacement schools ($n_{schoolrp}$), and the number of all originally sampled and initially contacted schools, which is again the targeted number of schools (i.e. the number of sampling points, $n_{samplingp}$).

$$PR_{schoolAR} = \frac{n_{schoolsp} + n_{schoolrp}}{n_{samplingp}}$$

The class participation rate (PR_{class}) is calculated as the ratio of the number of all initially sampled and participating classes ($n_{classsp}$) and the number of all eligible and sampled classes in one school, which is the sum of all participating and non-participating sampled schools classes ($n_{classsp} + n_{classnp}$).

$$PR_{class} = \frac{n_{classsp}}{n_{classsp} + n_{classnp}}$$

The student participation rate ($PR_{student}$) is then calculated as the ratio of all eligible and participating students enrolled in each participating school class ($n_{studentp}$) and the number of all eligible students in the respective participating school class, which is the sum of all participating and non-participating students ($n_{studentp} + n_{studentnp}$).

$$PR_{student} = \frac{n_{studentp}}{n_{studentp} + n_{studentnp}}$$

The overall participation rate before replacement ($PR_{totalBR}$) is then calculated as

$$PR_{totalBR} = PR_{schoolBR} \times PR_{class} \times PR_{student}$$

whereas the overall participation rate after replacement ($PR_{totalAR}$) can be derived from the following formula:

$$PR_{totalAR} = PR_{schoolAR} \times PR_{class} \times PR_{student}$$

As can be seen from Table 11, the participation rates especially on school level differ remarkably between the different countries. While the response rate is high in Sweden, it is considerably low in England even after the replacement strategy was applied. However, the unfavourable school participation rates in England doesn't seem to be a problem specific to CILS4EU, but has also shown to be true in other large scale school surveys as PISA or TIMSS – although not to the extent reported here.

Table 11: Response rates in the student survey: Schools, Students, Overall

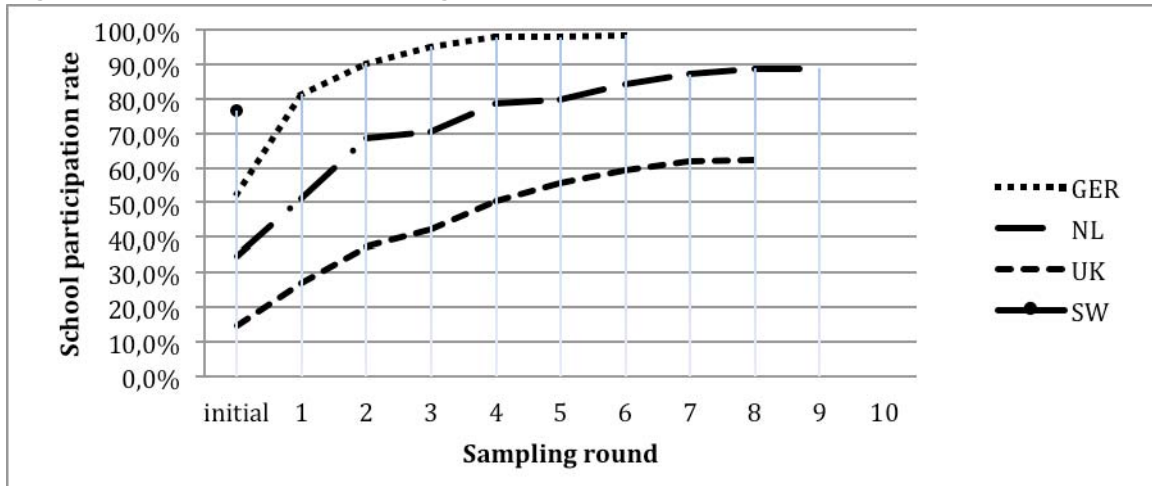
		School Participation Rate			Class Participation Rate (in %)	Student Participation Rate ⁵ (in %)	Overall Participation Rate	
		Before replacement (in %)	After replacement (in %)	av./max. №. of replac. schools			Before replacement (in %)	After replacement (in %)
England*	Strata 1	8.3	79.2	2.8/7	100	80.8	6.7	64.0
	Strata 2	10.9	69.6	3.2/8	100	79.4	8.7	55.3
	Strata 3	14.3	68.6	3.8/9	100	85.6	12.2	58.7
	Strata 4	14.3	60.0	2.0/5	100	77.0	11.0	46.2
	Indep.	30.4	47.8	0.5/2	100	80.5	24.5	38.5
	Total	14.7	65.6	2.7/8	100	80.5	11.8	52.8
Germany	Strata 1	40.0	95.0	1.1/3	97.1	86.4	33.6	79.7
	Strata 2	57.5	100	0.7/6	100	85.1	48.9	85.1
	Strata 3	66.7	100	0.6/4	100	78.8	52.6	78.8
	Strata 4	44.0	98.0	0.8/4	100	76.0	33.4	74.5
	Total	52.7	98.6	0.8/6	99.6	80.9	42.5	79.4
Netherlands	Strata 1	38.9	88.9	2.1/8	100	92.9	36.1	82.6
	Strata 2	26.5	100	2.4/10	98.6	90.9	23.8	89.6
	Strata 3	32.3	90.3	1.7/10	98.6	91.4	29.1	81.4
	Strata 4	46.2	80.8	1.7/10	80.7	89.5	33.4	58.4
	Total	34.9	91.7	2/10	94.5	91.1	30.0	78.9
Sweden	Strata 1	67.9	-	-	100	85.7	58.2	-
	Strata 2	81.1	-	-	100	86.9	70.5	-
	Strata 3	76.7	-	-	98.5	86.8	65.6	-
	Strata 4	77.3	-	-	96.9	84.7	63.4	-
	Total	76.8	-	-	98.8	86.1	65.3	-

* The reasons for the low figures on school level even after replacement are due to the fact that the response rate on school level was extremely low, with an associated expansion of recruitment and fieldwork time. Therefore, although 140 state schools and 23 independent schools were aimed for, the recruitment process was stopped after having achieved 96 state schools and 11 independent schools. As it is possible that the schools in the school sampling clusters having not agreed to participate in the study at that time the recruitment process was abandoned differ systematically from the schools participating in the study, several non-response models were run on school level by NatCen. Given these models, there were no significant school type or area level characteristics suggesting that non-response bias was not an issue with regards to measurable and collected info.

⁵ In England, 64 out of 214 (29.9%) responding classes had class size data that was either inaccurate or incomplete, therefore the student participation rate has been calculated by replacing missing class size data with the average within strata class size. Student responses are counted if they responded to the administrative questionnaire only.

The following Figure 1 demonstrates how the response rates on school level changed with each contacted replacement school over the strata. According to the figures, it is possible to compare response rates in CILS4EU to the response rates in PISA, who used two replacement schools as maximum. As already mentioned, Sweden has not applied the replacement strategy, therefore only the response rate for the initially schools are displayed.

Figure 1: Response per each sampling round



In addition to the response rate of the schools, classes and students, other important indicators of the survey are response rates on parental level as well as the participation rates for the teacher survey. The response rate for participating parents ($n_{parentp}$) of all eligible parents in the sample ($n_{parents}$) – which is simply the number of participating students $n_{studentp}$ ⁶ – can simply be calculated as

$$PR_{parent} = \frac{n_{parentp}}{n_{parents}}$$

resulting in an overall participation rate for the parents before replacement ($PR_{parentBR}$)

$$PR_{parentBR} = PR_{schoolBR} \times PR_{class} \times PR_{student} \times PR_{parent}$$

whereas the parental participation rate after replacement ($PR_{parentAR}$) can be derived from the following formula:

⁶ All persons acting as guardians for the under age children are counted as “parents”. Therefore, if a child lives with other persons than its biological parents, these persons are used for the calculation of the parental response rate.

$$PR_{parentAR} = PR_{schoolAR} \times PR_{class} \times PR_{student} \times PR_{parent}$$

The same considerations hold true for the calculation of the participation rate of the teachers, which is simply the number of teachers participated in the survey ($n_{teacherp}$) as the numerator and the overall number of eligible teachers in the sample ($n_{teachers}$) – which is the number of participating classes n_{classp} – as the denominator:

$$PR_{teacher} = \frac{n_{teacherp}}{n_{teachers}}$$

The overall participation rate for the teachers before replacement ($PR_{teacherBR}$)

$$PR_{teacherBR} = PR_{schoolBR} \times PR_{class} \times PR_{teacher}$$

whereas the teacher participation rate after replacement ($PR_{teacherAR}$) results from the following formula:

$$PR_{teacherAR} = PR_{schoolAR} \times PR_{class} \times PR_{teacher}$$

Table 12 provides an overview over these different response rates on parental and teacher level.

Table 12: Response rates in the parental and teacher survey: Schools, Students, Overall

	Parental participation Rate (in %)	Overall Participation Rate		Teacher Participation Rate (in %)	Overall Participation Rate		
		Before replacement (in %)	After replacement (in %)		Before replacement (in %)	After replacement (in %)	
England	Strata 1	42.7	2.9	27.3	73.7	6.1	58.4
	Strata 2	44.5	3.9	24.6	87.5	9.5	60.9
	Strata 3	33.0	4.0	19.4	91.7	13.1	62.9
	Strata 4	24.0	2.6	11.1	85.7	12.3	51.4
	Indep.	40.1	9.8	15.4	81.8	24.9	39.1
	Total	36.8	4.4	19.4	85.0	12.5	55.8
Germany	Strata 1	84.1	28.2	67.0	100	38.8	92.2
	Strata 2	84.5	41.3	71.9	91.0	52.3	91.0
	Strata 3	75.9	39.9	59.8	87.9	58.6	87.9
	Strata 4	69.4	23.2	51.7	91.4	40.2	89.6
	Total	78.0	33.1	62.0	91.5	48.0	89.9
Netherlands	Strata 1	85.1	30.8	70.3	86.1	33.5	76.5
	Strata 2	80.5	19.1	72.2	86.1	22.5	84.9
	Strata 3	76.5	22.2	62.3	91.2	29.0	81.2
	Strata 4	51.8	17.3	30.2	76.1	28.4	49.6
	Total	74.7	22.4	59.0	85.6	28.2	74.2
Sweden	Strata 1	63.3	36.8	-	90.0	61.1	-
	Strata 2	65.0	45.8	-	80.7	65.4	-
	Strata 3	61.0	40.0	-	86.4	65.3	-
	Strata 4	45.2	28.7	-	90.3	67.6	-
	Total	58.9	38.5	-	86.1	65.3	-

2.4.2 Sample sizes

In the following, we will present the sample sizes on school, classes, students, parents and teachers level. Table 13 provides an overview over the number of cases on the different levels in all countries; differentiated by the stratum in which the units are located.

Table 13: Sample Sizes

		Achieved sample sizes (% in brackets)				
		School level	Class level	Student level	Parental level	Teacher level
England	Strata 1	19 (19.8)	38 (19.8)	716 (18.1)	306 (21.2)	28 (17.1)
	Strata 2	32 (33.3)	64 (33.3)	1,338 (33.8)	595 (41.2)	56 (34.2)
	Strata 3	24 (25.0)	48 (25.0)	986 (24.9)	325 (22.5)	44 (26.8)
	Strata 4	21 (21.9)	42 (21.9)	918 (23.2)	220 (15.2)	36 (22.0)
	Total (State)	96	192	3,958	1,446	164
	Indep.	11 (10.3)	22 (10.3)	357 (8.3)	143 (9.0)	18 (9.9)
	Total (All)	107	214	4,315	1,589	182
Germany	Strata 1	19 (13.2)	34 (12.5)	715 (14.3)	601 (15.4)	34 (13.7)
	Strata 2	40 (27.8)	78 (28.8)	1,643 (32.8)	1,388 (35.5)	71 (28.6)
	Strata 3	36 (25.0)	66 (24.4)	1,189 (23.7)	902 (23.1)	58 (23.4)
	Strata 4	49 (34.0)	93 (34.3)	1,466 (29.2)	1,018 (26.0)	85 (34.3)
	Total	144	271	5,013	3,909	248
Netherlands	Strata 1	16 (16.0)	36 (16.2)	744 (17.1)	633 (19.4)	31 (16.3)
	Strata 2	35 (35.0)	72 (32.4)	1,463 (33.5)	1,178 (36.2)	62 (32.6)
	Strata 3	28 (28.0)	68 (30.6)	1,341 (30.7)	1,026 (31.5)	62 (32.6)
	Strata 4	21 (21.0)	46 (20.7)	815 (18.7)	422 (13.0)	35 (18.4)
	Total	100	222	4,363	3,259	190
Sweden	Strata 1	19 (14.7)	40 (15.9)	769 (15.3)	487 (16.5)	36 (16.7)
	Strata 2	43 (33.3)	83 (33.1)	1,666 (33.2)	1,082 (36.6)	67 (31.0)
	Strata 3	33 (25.6)	66 (26.3)	1,381 (27.5)	842 (28.5)	57 (26.4)
	Strata 4	34 (26.4)	62 (24.7)	1,209 (24.1)	546 (18.5)	56 (25.9)
	Total	129	251	5,025	2,957	216
Total	Strata 1	73 (15.6)	148 (15.8)	2,944 (16.0)	2,027 (17.5)	129 (15.8)
	Strata 2	150 (32.0)	297 (31.7)	6,110 (33.3)	4,243 (36.7)	256 (31.3)
	Strata 3	121 (25.8)	248 (26.5)	4,897 (26.7)	3,095 (26.8)	221 (27.0)
	Strata 4	125 (26.7)	243 (26.0)	4,408 (24.0)	2,206 (19.1)	212 (25.9)
	Total (State)	469	936	18,359	11,571	818
	Indep.(EN)	11 (2.3)	22 (2.3)	357 (1.9)	143 (1.2)	18 (2.2)
	Total	480	958	18,716	11,714	836

As can be seen from the first column in Table 13, we achieved and even surpassed our aim to sample at least 100 schools in all countries, although with slight different allocations over the strata. Subsequently, we clearly reached the aim of interviewing 4,000 students in each country (cf. column “student level” in Table 13). Section 2.4.3 goes into more detail about the composition of the sample and further describes to what extent these numbers of cases also include students with an immigrant background.

Compared to the aimed distribution of students in Table 6 one can see that we were quite successful in achieving the planned distribution of students over the strata. As comparable small class sizes in schools with a high proportion of children of immigrants is an issue in Germany, the German team decided to sample more schools in this stratum in order to overcome this problem (cf. already Table 7). When looking at the distribution of schools and students over the strata, it can be seen that this procedure was helpful to converge to the aimed distribution of students as described in Table 6.

The response rates in Table 12 already pointed out to lower response rates on parental level, which is clearly represented in the lower overall case numbers on parental level. Overall, we were able to survey between 1,589 parents (England) and 3,909 parents (Germany) in each country, with an overall achieved parental sample of 11,714 parents. In contrast to the parental sample, we were almost possible to survey all teachers in the sampled class, resulting in 836 completed teacher questionnaires in all four countries.

2.4.3 *Composition of the sample*

In this section, we will focus on the composition of the student and parent sample with respect to the immigrant status of the respondents. In doing so, we will see whether the oversampling approach described in section 2.3.1 really worked and whether we were able to obtain the aimed number of immigrant students (and their parents) in the different countries.

In the following, children with an immigrant background are defined as students who were either themselves born outside the survey country, or who have at least one parent or at least two grandparents who were born outside the survey country. Table 14 provides an overview over the composition of the student sample in terms of immigrant proportion (see for the construction of the immigrant background variable see Dollmann et al. 2014). As can be seen, we almost consistently achieved or even surpassed our aim to survey at least 1,500 (37.5%) children with an immigrant background in each country. In total, about 46% of our sample are students with an immigrant background given the definition described above (n=8,557).

Table 14: Composition of the student sample

		Students			Total
		with immigrant background (% of total)	without immigrant background (% of total)	immigrant background unclear (% of total)	
England	Strata 1	93 (12.99)	597 (83.38)	26 (3.63)	716
	Strata 2	422 (31.54)	883 (65.99)	33 (2.47)	1,338
	Strata 3	589 (59.74)	378 (38.34)	19 (1.93)	986
	Strata 4	772 (84.10)	125 (13.62)	21 (2.29)	918
	Indep.	169 (47.34)	186 (52.10)	2 (0.56)	357
	Total	2,045 (47.39)	22,169 (50.27)	101 (2.34)	4,315
Germany	Strata 1	154 (21.54)	559 (78.18)	2 (0.28)	715
	Strata 2	609 (37.07)	1,031 (62.75)	3 (0.18)	1,643
	Strata 3	664 (55.85)	520 (43.73)	5 (0.42)	1,189
	Strata 4	1,150 (78.44)	311 (21.21)	5 (0.34)	1,466
	Total	2,577 (51.41)	2,421 (48.29)	15 (0.3)	5,013
	Netherlands	Strata 1	117 (15.73)	626 (84.14)	1 (0.13)
Strata 2		362 (24.74)	1,099 (75.12)	2 (0.14)	1,463
Strata 3		474 (35.35)	863 (64.35)	4 (0.30)	1,341
Strata 4		528 (64.79)	280 (34.36)	7 (0.86)	815
Total		1,481 (33.94)	2,882 (65.73)	14 (0.32)	4,363
Sweden		Strata 1	159 (20.68)	605 (78.67)	5 (0.65)
	Strata 2	522 (31.33)	1,114 (66.87)	30 (1.80)	1,666
	Strata 3	709 (51.34)	661 (47.86)	11 (0.80)	1,381
	Strata 4	1,064 (88.01)	136 (11.25)	9 (0.74)	1,209
	Total	2,454 (48.84)	2,516 (50.07)	55 (1.09)	5,025
	Total	Strata 1	523 (17.76)	2,387 (81.08)	34 (1.15)
Strata 2		1,915 (31.34)	4,127 (67.55)	68 (1.11)	6,110
Strata 3		2,436 (49.74)	2,422 (49.46)	39 (0.80)	4,897
Strata 4		3,514 (79.72)	852 (19.33)	42 (0.95)	4,408
Indep.(EN)		169 (47.34)	186 (52.10)	2 (0.56)	357
Total		8,557 (45.72)	9,974 (53.29)	185 (0.99)	18,716

To summarize, these figures clearly demonstrate the effectiveness of our strategy to over-sample schools with high immigrant proportions in order to achieve the desired number of children with an immigrant background in our sample. In addition, Table 15 presents the composition of the achieved parental sample. The figures here are consistently lower than in Table 14 due to the lower parental response rate compared to the student's participation rate.

Table 15: Composition of the parent sample

		Parents			Total
		with immigrant background (% of total)	without immigrant background (% of total)	immigrant background unclear (% of total)	
England	Strata 1	31 (10.13)	274 (89.54)	1 (0.33)	306
	Strata 2	162 (27.23)	432 (72.61)	1 (0.17)	595
	Strata 3	169 (52.00)	156 (48.00)	0 (0.00)	325
	Strata 4	175 (79.55)	45 (20.45)	0 (0.00)	220
	Indep.	58 (40.56)	85 (59.44)	0 (0.00)	143
	Total	595 (37.44)	992 (62.43)	2 (0.13)	1,589
Germany	Strata 1	125 (20.80)	476 (79.20)	0 (0.00)	601
	Strata 2	490 (35.30)	897 (64.63)	1 (0.07)	1,388
	Strata 3	493 (54.66)	406 (45.01)	3 (0.33)	902
	Strata 4	787 (77.31)	231 (22.69)	0 (0.00)	1,018
	Total	1,895 (48.48)	2,010 (51.42)	4 (0.10)	3,909
	Netherlands	Strata 1	87 (13.74)	546 (86.26)	0 (0.0)
Strata 2		260 (22.07)	917 (77.84)	1 (0.08)	1,178
Strata 3		289 (28.17)	737 (71.83)	0 (0.0)	1,026
Strata 4		198 (46.92)	224 (53.08)	0 (0.0)	422
Total		834 (25.59)	2,424 (74.38)	1 (0.03)	3,259
Sweden		Strata 1	87 (17.86)	400 (82.14)	0 (0.00)
	Strata 2	300 (27.73)	781 (72.18)	1 (0.09)	1,082
	Strata 3	380 (45.13)	461 (54.75)	1 (0.12)	842
	Strata 4	452 (82.78)	93 (17.03)	1 (0.18)	545
	Total	1,219 (41.22)	1,735 (58.67)	3 (0.10)	2,957
	Total	Strata 1	330 (16.28)	1,696 (83.67)	1 (0.05)
Strata 2		1,212 (28.56)	3,027 (71.34)	4 (0.09)	4,243
Strata 3		1,331 (43.00)	1,760 (56.87)	4 (0.13)	3,095
Strata 4		1,612 (73.07)	593 (26.88)	1 (0.05)	2,206
Indep.(EN)		58 (40.56)	85 (59.44)	0 (0.00)	143
Total		4,543 (38.78)	7,161 (61.13)	10 (0.09)	11,714

In the parental sample, 39% of the sample are respondents where at least one parent in the family migrated himself/herself or has parents who migrated in the respective survey country, comprising 4,543 parents in total. While in Germany almost 50% of the parental sample has an immigrant background, this share is considerably lower in the other three countries.

3 Weighting⁷

As described in the previous section, the complex sampling design results in varying selection probabilities of students in the CILS4EU sample. Students in ethnic dense schools have

⁷ The whole section on weighting – with the exception of 3.4 and 3.5, which is on the basis of the “TIMSS 2003 User Guide for the International Database” (Martin 2005: 2-47f) – is mainly based on the chapter “Sampling weights and participation rates” in the ICCS Technical report (Zuehlke/Vandenplas 2011) as well as on a short manual on weights calculation kindly provided by Olaf Zuehlke from IEA-DPC (<http://www.iea-dpc.de>). Olaf Zuehlke as well as Sabine Meinck from IEA-DPC were also helpful in solving several specific problems while calculating the weights.

a higher selection probability than students in schools with a low immigrant proportion. To correct for these different selection probabilities and to obtain correct estimates of population characteristics, design weights are needed. Furthermore, non-response could also influence the estimation results. Therefore, non-response adjustment-weights are used in the survey. The calculation of both types of weights is described in the following, while the following notation is used throughout this chapter:

- H is the total number of explicit strata in a country,
- N_h is the total number of schools in the country in explicit stratum h , and n_h is the number of sampled schools in that explicit stratum, $h=1, \dots, H$.
- M_{hi} is the number of students enrolled in the target grade by the time the school sampling frame was compiled in school i in explicit stratum h , $h=1, \dots, H$ and $i=1, \dots, n_h$. M_h is the total number of students enrolled in the target grade in the explicit stratum h .

3.1 Design weights

School level

The first stage of sampling for CILS4EU was the sampling of schools in each country or explicit stratum $h=1, \dots, H$ of a country. In most cases, a systematic sample of schools was drawn, with selection probabilities proportional to the school size M_{hi} . In these cases, the school base weight of school i in stratum h is given by:

$$WGTFACI = \frac{M_h}{n_h \times M_{hi}}$$

In an explicit stratum where a census of schools was taken, the school base weight is given by $WGTFACI=1$.

Class level

In each school, two classes were randomly selected using a systematic random method.⁸ The class weight factor is given by

$$WGTFAC2S = \frac{C_{hi}}{c_{hi}}$$

where C_{hi} is the total number of classes with students enrolled in the target grade and c_{hi} is the number of sampled classes in school i in stratum h .

Non-randomly selected classes (cf. section 2.3.2) are assigned a class design weight of 1, letting the non-randomly selected classes represent themselves. The design weights of the other, randomly selected classes are then calculated in a way that these classes represent all other but not the non-randomly selected classes in the relevant grade. If all classes in schools were not randomly selected or we are not certain whether all or only some classes were not randomly selected, these classes represent all classes in the respective school.

For England, seven classes were selected although the administrative as well as the survey data wasn't clear about the total number of classes in the respective schools. Therefore, the class weight was calculated using the number of pupils in the grade divided by the average class size of schools in our sample in order to get a proxy for C_{hi} , which is actually not a natural number. Therefore, when multiplying the class weight factor $WGTFAC2S$ with c_{hi} , a non-natural total number of classes in the relevant target grade may result for some cases.

Student level

The aim was to obtain a census of all students in the classroom. Therefore, the design weight on student level is 1.

⁸ Deviations from this procedure are reported in section 2.3.2. If less or more than two classes were sampled, this will be considered in the term c_{hi} .

3.2 Adjustment weights

School level

To account for non-responding schools in the sample, a school weight adjustment factor is calculated as

$$WGTADJ1S = \begin{cases} \frac{n_h}{r_h}, & \text{for participating schools} \\ 0, & \text{for non - participating schools,} \end{cases}$$

where r_h represents the number of participating schools in explicit stratum h .

Class level

In the event of class non-response within a participating school, an adjustment factor gets computed. In order to avoid an inflation of the weights for participating students in a school with class-level non-response, the class non-response is adjusted for at the stratum level. The class weight adjustment factor is calculated as

$$WGTADJ2S = \begin{cases} \frac{c_h}{d_h}, & \text{for participating classes} \\ 0, & \text{for non - participating classes} \end{cases}$$

where c_h is the total number of sampled classes in participating schools in a stratum and d_h is the total number of participating classes within participating schools in the explicit stratum h .

Student level

The adjustment for student non-response is calculated inside each class as follows:

$$WGTADJ3S = \begin{cases} \frac{s_{hij}}{p_{hij}}, & \text{for participating students} \\ 0, & \text{for non - participating students,} \end{cases}$$

where s_{hij} is the number of students and p_{hij} is the number of participating students in class j in school i in stratum h .

In England, as already described in Footnote 5 in Table 11, 29.9% of the school classes had missing or wrong class size information, making it impossible to calculate an adjustment weight on students' level. For those cases, the average student adjustment weight

calculated for each stratum separately was used in order to replace the missing student adjustment weight in classes without eligible class size information.

3.3 Final Student Weight

The final student weight TOTWGTS is calculated as the product of the five factors WGTFACTS, WGTADJ1S, WGTFACTS, WGTADJ2S and WGTADJ3S.

3.4 House Weight

To avoid deflated standard errors when performing significance tests, the so called House Weight can be used to weight the data. It may be used when the actual sampling size is required when performing significance tests. It is calculated by scaling down the TOTWGTS to the actual sample size and is therefore TOTWGTS multiplied by the ratio of the sample size (the number of students, p) in each country divided by the sum of the weights over all students in the target grade (Martin 2005: 2-47f).

$$HOUWGT = TOTWGTS \times \frac{p}{\sum TOTWGTS}$$

The sum of HOUWGT over all participating students is then the actual sample size.

3.5 Senate Weight

To assign the same contribution to each country – regardless of a countries population size – when estimating parameters in international analyses, one may use the senate weight SENWGT. The SENWGT sampling weight is TOTWGTS multiplied by 500 divided by the sum of the weights over all students in the target grade in each country (Martin 2005: 2-47).

$$SENWGT = TOTWGTS \times \frac{500}{\sum TOTWGTS}$$

When using the SENWGT, each country has a sample size of 500.

4 Development of the instruments

This chapter illustrates the development of the international instruments for the students (4.1), parents (4.2), and teachers (4.3) and the subsequent adaption and translation of these instruments to the national questionnaire versions (4.4).

4.1 Instruments for the student survey

The student questionnaire focusses on three core dimensions of integration: structural, social and cultural integration. Furthermore, the instrument is also designed to measure key demographic and migration-specific characteristics of the students. Overall, the final student questionnaire consisted of four parts (ordered as appeared during the survey): the main student questionnaire (4.1.1), a name generator to assess the five best friends (4.1.2), an achievement test to assess verbal and cognitive competencies (4.1.3), and a sociometric measure to capture the relations in the class context (4.1.4). These instruments are described in the following.

4.1.1 *Student main questionnaire*

The main student questionnaire focusses on several constructs measuring structural, cultural, and social integration, and on the key explanatory and intervening variables. With respect to structural integration, the instrument measures school grades, tracks, transitions in the educational career, drop-outs, truancy, and economic resources, while the cultural integration part covers cultural capital, cultural practices related to the country of origin as well as to the country of destination, religion and religious activities, language usage, and ethnic identity. In addition, some topics regarding social integration, like romantic relations, weak ties, family relations, and participation in clubs etc. are included in the main questionnaire. A wide range of explanatory and intervening variables are also included in the main questionnaire, for example the socio-demographic background, return orientations, in-group identification and out-group rejection, general values, normative attitudes, and information about the migration history of the family, to name only a few.

As can be seen from this listing of many immigrant- and minority-specific topics, one difficulty was to cover all these important issues while still ensuring the feasibility of the questionnaire for students with as well as for those without an immigrant background, as no a priori categorization was possible in the chosen in-school survey mode. Furthermore, with the questionnaire being designed as a self-completion interview administered in the classroom, the inclusion of sophisticated filter instructions had to be avoided, complicating the construction of the questionnaire for the different target groups in addition. The solution to these possible problems was to construct simple and sometimes quite general questions that can be answered irrespective of the ethnic and immigration background of the students. Furthermore, whenever filters weren't avoidable, a very simple filter structure with only short and straightforward skips were used, complemented by eye-catching instructions. The master version of the students' main questionnaire can be found at www.cils4.eu.

4.1.2 *Name generator – “Five best friends”*

Social integration of students takes place in schools and classes, but not only. In order to capture social integration outside the class context, a name generator was used to assess characteristics of ego-centred networks, where the five best in- and out-school friends could be named. As there was no restriction to only name out-school friends, some of these friends could be students being named in the sociometric measure, another instrument capturing social integration within school classes (cf. 4.1.4). Characteristics that were assessed in the name generator were age and sex of the friend, his/her ethnic background, the education the person is following, the usual meeting places and the frequency of meetings and whether the student's parents know this person. Furthermore, the density of the network as well as some behavioural attributes of the friends was measured. In the Netherlands, additional questions concerning drinking and smoking behaviour were asked in this section. The master version of the instrument can be found at www.cils4.eu.

4.1.3 *Achievement test*

A scholastic achievement test was administered to assess cognitive-structural integration. The aim of the achievement test was twofold. On the one hand, we wanted to get some ob-

jective measure of language abilities in the language of the survey country. On the other hand, the focus was on cognitive skills. Both tests are described in the following.

4.1.3.1 Language test

The aim of the language ability test was to compare the verbal competencies in the language of the survey country between different groups – within and between the different countries. Therefore, a test with different language version being administrable in all of the participating countries was desired. However, this requirement posed several problems. The major issue was that within a given test with different language versions, the equivalence with respect to the level of difficulty is hard to guarantee. Furthermore, and even more restricting, the variety of international comparable tests for the age group surveyed in CILS4EU in the different countries is extremely limited. The only existing international language tests that are available (and suitable in the in-school context) in all countries for the relevant age group of 14-year old students were the PISA tests. However, these tests are much too long, but shortening them to fit in the survey would result in imprecise measures. Given these obstacles, the idea of using language tests that allow directly for cross-country comparisons was rejected. Instead, all country teams decided to use independent national language tests, with a focus on the measurement of the children's lexicon. This solution still permits comparisons between native students and students with an immigrant background in each country and there are ways to also use these national tests in the international context (e.g. by the use of achievement percentiles etc.).

As already mentioned, the focus of the language test was on the lexicon of the students in the language of the survey country. This was done by using synonym- or antonym-tests. The procedure was twofold: some country teams used existing tests, like Germany and the Netherlands, while the English and Swedish team designed a language test on their own. In Germany, a verbal subtest of a cognitive achievement test (KFT 5-12+ R, "Kognitiver Fähigkeitstest für 5. bis 12. Klassen Revision", verbal subtest V1; Heller/Perleth 2000) was used, where respondents had to find synonyms for 25 items out of 5 answer alternatives respectively. In the Netherlands, the Subtest "Synoniemen" from the "Nederlandse Intelligentietest voor Onderwijsniveau" (NIO; Van Dijk/Tellegen 2004) was used. Like in the

other tests before, synonyms for 30 words had to be found out of 5 answer alternatives. The development of the English and Swedish language test was guided by the ideas and principles of the existing language tests in the other countries. In the English test, 25 synonyms had to be found out of 5 different answer alternatives respectively. In contrast to the tests in the other countries, the Swedish language test asked for antonyms instead of synonyms. Overall, the test consisted of 30 items, with 4 answer alternatives each. Figure 2 provides an exemplary overview over the English verbal test.

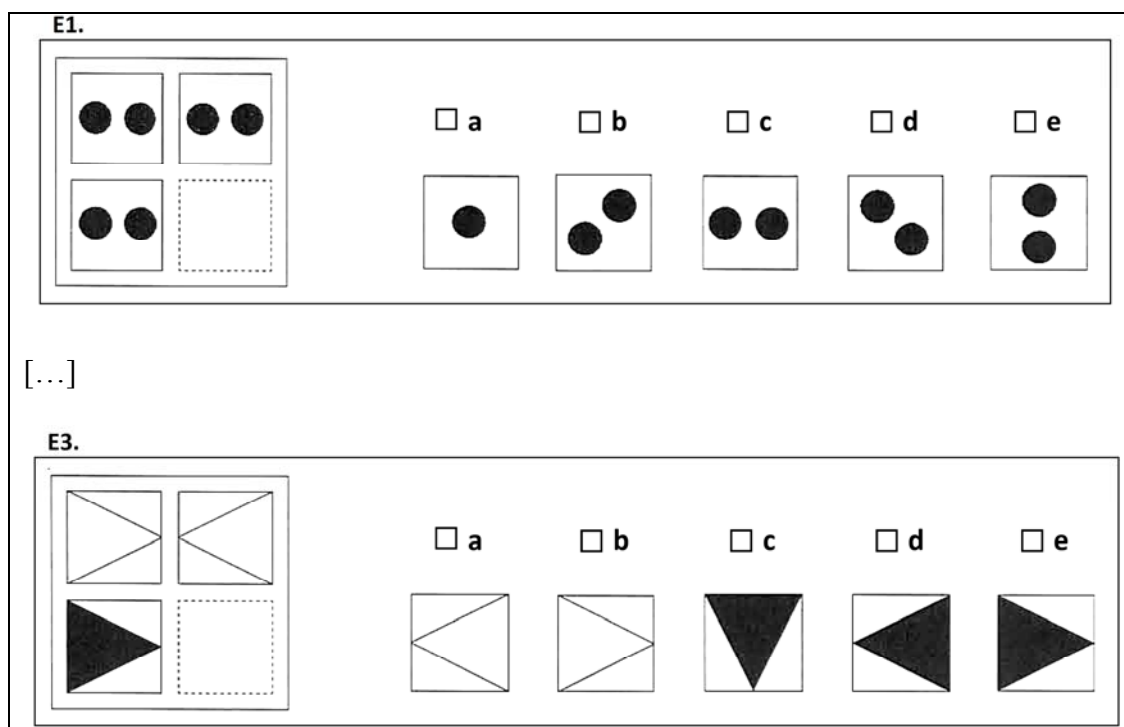
Figure 2: Example of the English Language Ability Test

B1. closed	<input type="checkbox"/> A clear	<input checked="" type="checkbox"/> B shut	<input type="checkbox"/> C finished
	<input type="checkbox"/> D fallen	<input type="checkbox"/> E tired	
B2. sick	<input type="checkbox"/> A new	<input type="checkbox"/> B fresh	<input checked="" type="checkbox"/> C unwell
	<input type="checkbox"/> D happy	<input type="checkbox"/> E tired	

4.1.3.2 Cognitive ability test

To measure the more general cognitive abilities a standard German cognitive ability test (CFT20R, “Grundintelligenztest Skala 2 Revision”; Weiß 2006) was used. This test is based on graphical problems and therefore language free and usually titled “Culturally Fair”. Only the instructions had to be adapted – with the assistance of the test’s publisher – to the different language versions in the participating countries. The advantage of using a test of this kind in all countries is that it allows for direct cross-country and within-country comparisons between different groups. Figure 3 provides an overview over the structure of the different tasks. In total, 27 of these tasks had to be solved, resulting in a raw score ranging from 0 to 27.

Figure 3: Example of the Cognitive Ability Test



4.1.4 Sociometric instrument

One of the most important contexts for building up friendships and social contacts for 14-year old students is the school and class context. Therefore, this context needs particular attention when studying social integration of immigrant background children. With the sociometric measure that complements the social integration measures described in the main student survey, it is possible to get a complete picture of the relations and friendships in the school class. Prior to completion of this instrument, a class list with all students and a corresponding and unique ID was disseminated to each student in the class. During the completion of the interview the students simply named the number of the student or students to whom the question applied, e.g. to the question “Who is your best friend in class?”. This procedure was necessary to meet the data protection regulations after which no student should write down names in the survey instrument. In total, 11 items were used in the sociometric instrument. A complete overview over the questions used in the sociometric measure (and all other instruments) can be found at www.cils4.eu.

4.2 Parental questionnaire

The student survey was designed as a self-completion interview in the school context, administered during two regular school hours. In contrast, the parental questionnaire was designed as a self-completion interview that the children should hand over to their parents or was sent to the parents for completion at home. This leads to stronger restrictions with respect to length of the parental interview, as the response rate to such surveys fundamentally depends on the length of the instrument. In order to capture intergenerational transmission processes with respect to social, structural and cultural integration, the questions being identified as key constructs on each of those dimensions and already having been asked in the students' questionnaire were included in the parental survey as well.

Examples for such key constructs on structural integration are educational aspirations for their children and educational involvement, but also indicators of their own success in the educational system and occupational attainment, where more specific and differentiated questions were asked in order to get better information compared to the students' questionnaire. The cultural integration part covers again cultural capital, cultural practices related to the country of origin as well as to the country of destination, religion and religious activities, language usage, and ethnic identity. Furthermore, some topics on social integration were assessed, like the ethnic composition of the parental friendship network or the ethnic origin of the people the parents meet in some day-to-day activities like at their workplace or at sport clubs or other associations. Like in the students questionnaire, there are also some explanatory and intervening variables included in the parental questionnaire, for example, return orientations, in-group identification and out-group rejection, general values, normative attitudes, and information about the own migration history or those of the parents, i.e. the grandparents of the child. Questions referring to the migration history, but also about educational outcomes were thereby also asked for the partner/spouse of the person completing the parental interview so that information about both parents (and all four grandparents) of the child should ideally be available – however, only if the partner/spouse lives in the same household.

Like the main student instrument, the parental questionnaire needed a design that makes it easy to complete the survey, irrespective of the ethnic and immigrant background

of the parents. Therefore, the same approach was followed for the development of the parental questionnaire as in the main student instrument: The implementation of simple and sometimes quite general questions that can be answered irrespective of the ethnic and immigration background of the parents; as few filters as possible and whenever filters weren't avoidable, a very simple filter structure with only short and straightforward skips, complemented by eye-catching instructions. The master version of the parental questionnaire can be found at www.cils4.eu.

4.3 Teachers' questionnaire

The aim of the teacher questionnaire was to get a more detailed picture of the class and school characteristics students are confronted with in their everyday school life. Therefore, compositional characteristics of the class and the school were assessed, like the ethnic, social and achievement composition. In addition, indicators of the school equipment and personnel resources were covered with the questionnaire for the teachers. Furthermore, and in order to get some insights over possible non-response bias, the class teacher was asked to provide some characteristics for absent students, however only on aggregate level due to privacy reasons. Finally, some characteristics of the teacher were assessed, like his or her age and sex, social background and achieved university degree. The master version of the questionnaire for the teachers can be found at www.cils4.eu.

4.4 Adaption and translation of the instruments

The questionnaires described above were initially developed as English master instruments. Therefore, some questions in the questionnaire had to be adapted to fit to each country's specificities, e.g. question about the educational system, but also questions dealing with monetary resources where the currency had to be adapted. Furthermore, some country team included country specific questions that were not asked in all four countries. Together with these adjustments and additional questions, the questionnaires had to be translated in Dutch, German, and Swedish before they were used in the different countries. In addition to the translations in Dutch, German and Swedish and in order to increase the response rates of the parents, the parental questionnaire was translated in several minority languages, although

not all language versions were used in all countries. In the following sections, an overview over the country specific questions and adjustments (4.4.1), a description of the translation approach (4.4.2), and a summary over the minority languages used in each country for the parental survey is provided (4.4.3)

4.4.1 Adaptions and country specific questions

The major adaptations of the international master instrument refer to two topics. On the one hand, some country teams included country specific questions that were only asked in one – or more, but not in all – countries. On the other hand, questions that were designed to be asked in all – or at least more than one – countries, needed adaptation due to the country specific context the questions referred to.

Country-specific questions that were not asked in all four countries due to national specificities, but also due to specific interest of the national research teams were restricted to the bare minimum in order to achieve an as comprehensive instrument as possible. The Codebook (CILS4EU 2016) documents these country specific questions. Adaptions that were necessary due to country specificities mainly deal with the educational system of a country. Here, questions as well as answer categories had to be adapted to suit the specific characteristics of a national educational system (cf. questions about the setting system, educational aspirations, parental education in the main student questionnaire and questions concerning educational attainment in the name generator). Similar adaptations were necessary for answer categories referring to immigrant groups. Here, the largest immigrant groups in each country had to be included in the answer categories. Furthermore, for some questions dealing with economic resources, the currency had to be adapted (cf. questions about money earned from jobs and pocket money in the main student questionnaire).

However, adaptations were not only necessary between different countries, but also within a country. In Germany, some Federal States expressed their concerns regarding specific items. Therefore, it was not possible to ask about third persons in the survey in one Federal state (e.g. no questions about parents or grandparents in the students' survey). While some questions were reformulated and posed to the students' parents in the parental ques-

tionnaire, some had to be removed. Furthermore, one Federal State objected to questions asking about deviant behaviour, like questions about smoking, drinking, stealing etc. Items dealing with these issues were excluded in the state-specific instruments. Again, the adaptations can be found in the Codebook.

4.4.2 *Translation of the instruments*

Two aims were crucial during the translation and adaptation process: The instruments should suit the cultural and linguistic peculiarities in each country and should still be comparable across countries. Different methods are possible to handle problems caused by translations. The TRAPD method was selected due to the suggested step-by-step procedure and a close collaboration of different persons fulfilling different functions. The acronym captures the five steps Translation, Review, Adjudication, Pre-testing and Documentation (Harkness 2007). The whole process is based on teamwork between translators, reviewers and an adjudicator for one language from the beginning on.

To ensure a high degree of cross-country comparability of the translated instruments, all countries were provided with detailed guidelines, which had to be followed during the whole translation process. These guidelines were based on previous research indicating typical problems and sources of errors that can occur during the translation process of questionnaires. As recommended by the European Social Survey (ESS) coordinators, TRAPD was preferred as a method to translate the instruments.

The translators were selected by their language skills and practical experiences, with two translators being recruited for each language. Both translators produced their own versions (parallel translation) of the questionnaire. After a first translation draft, the translators discussed their work and decided between the different suggestions for each item, which covers the intention of the original questions the most, but ensures a balance between original content of the questions and linguistic specificities in the survey language at the same time. This compiled version was used for the second step: the review process. The reviewers need to have at least as good translation skills as the translators but should be familiar with questionnaire design principles, as well as the study design and topic. One reviewing person

with linguistic expertise, experience in translating, and survey knowledge was seen as sufficient for each language version. The main task of the reviewer was to revise the version the translators agreed on the basis of the English master version. They assess the comparability and readability of the translated instruments and document possible problems of single items. The final decision about the translated instruments was done by the adjudicator, preferably in co-operation with reviewer and translators, but at least after discussion with a reviewer. As the adjudicator has the most responsibility, the person given these tasks must meet more requirements than the translators and the reviewer. The adjudicator is expected to understand the research subject, know about the survey design, and be proficient in the languages involved.

After the translation and adaption procedure the instruments were tested by cognitive interviews with persons of the targeted groups. The general aim of cognitive interviews is to gather information about possible problems with single questions and to compare how translated questions are interpreted and understood in the different language versions (Prüfer and Rexroth 2005). The method gives insights about respondents answering behaviour and the decision process leading to certain answers. With the transcribed and analysed results of the cognitive pre-tests members of the international team (one per country) discussed solutions for possible translation errors or translated and adapted concepts that differ from the original intention of questions. The revised versions were used for pilots under realistic survey situations, thus students' instruments were tested in class context. These class surveys were essential to ensure the quality, comprehensibility and feasibility of the instruments, instructions and filtering. Subsequent to the class survey, a part of the class was interviewed about their problems with the questionnaire and interpretations of the questions. Due to the extensive testing and adapting earlier only minor changes had to be done at this stage to finalize the instruments.

4.4.3 *Minority languages*

The CILS4EU instrument for parents was translated in the four languages of the respective survey countries as well as in another 17 additional languages to cover the most common languages in the four countries. The translation of the parental interview was seen as essen-

tial to provide parents with an immigrant background with instruments in their native tongue in order to increase the response rate on parental level. However, not all language versions were used in all countries. As in most of the countries the parental questionnaires were handed out to the students in the classroom, increasing numbers of language versions increases possible problems during the dissemination process. Therefore, the different countries only used language versions of immigrant groups that make a significant share of the immigrant population in one country. Table 16 provides an overview over the languages used in the four countries for all students. The language of the parental interview is given in the variable [p1_lang] (CILS4EU 2016).

Table 16: Minority languages used in the different countries for the parental questionnaire

	England	Germany	Netherlands	Sweden
Language versions	Bengali	English	Arabic	Arabic
	Gujerati	Italian	Turkish	Kurmanji
	Punjabi Gurmukhi	Serbian		Persian/Farsi
	Punjabi Urdu	Polish		Serbian
	Somali	Russian		Somali
	Urdu	Spanish		Sorani
		Turkish		Spanish
				Turkish

5 Fieldwork

5.1 Students' survey

The CILS4EU-team aimed for a comparable procedure when administering the different instruments of the students' survey. In order to achieve this, the international project coordination in Mannheim together with the help of the German institute being specialized in this kind of school research (IEA Data Processing Center DPC, Hamburg; <http://www.iea-dpc.de>) compiled a survey manual where the procedure was described. This survey manual was then discussed and disseminated to all project partners in the countries and was used as the basis for the student survey.

The first step of the students' survey in schools – after having sampled and persuaded the schools to take part in the survey – was to inform the parents about the planned research project. Therefore, information material about the study was disseminated to the parents via the school. As an active parental consent was necessary in Germany (cf. section 2.3.3), the parents were asked to sign a form to allow their child to take part in the study. In the other countries, families had the opportunity to opt-out. To increase the anticipated lower response rate in Germany due to the active opt-in, students were offered a 10,- Euro incentive. This incentive was then disseminated to all students after the survey was finished directly after the test session in school.

The fieldwork was conducted by Statistics Sweden in Sweden, by the National Centre for Social Research (NatCen) in the United Kingdom and the IEA Data Processing Center (DPC) in Germany. The Dutch team conducted the survey on their own, using student assistants to administer the survey. Regardless who conducted the survey, the test administrators that were responsible for the administration of the survey were trained extensively in order to ensure a smooth procedure. The test administrator was responsible for disseminating the materials to the students, leading them through the questionnaires and tests, controlling the test time for the achievement tests and finally collecting the completed survey parts and sending them back to the national research centers. In order to ensure that the dif-

ferent instruments were completed by the respective students, all instruments were identified with a unique ID, which was linked to the specific student on a class list.

The whole survey was planned to maximally last 80 minutes. Table 17 provides an overview over the time allowed for the different modules. Not shown here are additional times for breaks (mainly after the student survey) and for the time needed to disseminate, control and collect the material before and after the test session.

Table 17: Time frame for different CILS4EU student instruments

Instrument (in the order the instruments were administered)	Length
Student questionnaire	35+ minutes
5-friends questionnaire	10+ minutes
Language test	7 minutes exactly (test)
Cognitive test	7 minutes exactly (test)
Sociometric questionnaire	10+ minutes

For the administration of the sociometric instrument, there were some differences between but also within the countries with respect to the inclusion of names of absent students on the class list. In some countries, all students were included on the class list, irrespective of whether they were absent or not, while in other countries this procedure was more ambiguous. The additional report (Kruse/Jacob 2016) provides information on how the sociometric survey was implemented in all countries.

At the end of the survey time, the parental questionnaires together with a cover letter for the parents describing the project aims were handed over to the students, again with the help of the list combining IDs and students names ensuring that the parental questionnaire was handed over to the child of the targeted family. During the dissemination process, students with an immigrant background were asked about possible language usage of their parents and whether a minority language version of the questionnaire should be added in order to facilitate the completion of the parental questionnaire (cf. 4.4.3).

5.2 Parental survey

As described at the end of the previous section, the main mode for the parental questionnaire in all countries was paper/pencil. The students were asked to hand the parental instrument to that parent that is mostly involved in educational processes, but it was also possible and allowed for the parents to complete the questionnaire together. In Sweden, the parental questionnaire was sent to the parents. Table 18 provides an overview on who completed the parental interview in the different countries.

Table 18: Respondent to parental interview

	England	Germany	Netherlands	Sweden
Mother	1,303 (82.0)	2,953 (75.5)	2,452 (75.2)	2,156 (72.9)
Another female	32 (2.0)	73 (1.9)	47 (1.4)	58 (2.0)
Father	222 (14.0)	786 (20.1)	686 (21.1)	638 (21.6)
Another male	18 (1.1)	44 (1.1)	20 (0.6)	58 (2.0)
Missing/unclear	14 (0.9)	53 (1.4)	54 (1.7)	47 (1.6)
Total	1,589	3,909	3,259	2,957

While the dissemination of the parental instruments was organised at the end of the student session, the collection of the interviews was organised differently in the participating countries. In Germany, the teacher collected the interviews in sealed envelopes and send all back to the German research Center after a given time period. Some of the parents used the opportunity to send the questionnaires directly to the National research centers.

In Germany, telephone surveys were conducted in addition to the self-completion interviews. By using this mode, 1,309 out of 1,992 approached target persons of the parental survey were surveyed, resulting in a response rate of 66%.

In the Netherlands, three reminders were sent out via mail when parents did not sent back the questionnaire distributed in schools. The third reminder included a shortened version of the parental questionnaire. Non-respondents were in a last step contacted by telephone. Using the last two steps, 681 additional parental interviews were realized. Additionally, since a large number of students were expected to participate in wave 2 for the first time, a parental survey was implemented parallel to the students' survey in the second wave using identical procedures and instruments.

Swedish parental questionnaires were sent out by mail – not brought home by students. The rationale for this was that we were able to receive the students’ home addresses from the population register and thought this procedure safer because we expected a higher proportion of the questionnaires to actually reach the parents this way. The first batch of parental questionnaires was sent out in April 2011. A short version was sent to by-then non-respondents one month later. In total, 2,456 (49%) parents returned the questionnaire via mail until September 2011. Since the parental survey was not as successful as initially planned, all parents who did not return the questionnaire until this date were re-approached parallel to the second wave of data collection among the students with identical methods. The Swedish team did not extend the parental survey’ fieldwork in the first wave after the end of the students’ survey because close timing of the parental interviews to the students’ survey is seen as helpful for the parental response rate. This attempt resulted in 501 additional interviews. In total, 2,957 (59%) parents returned the questionnaire via mail.

5.3 Teacher survey

In addition to the students’ and parental interviews, an additional assessment was aimed with the form teacher of the respective class. However, it was not always possible to recruit the form teacher for the teacher. In contrast, in some cases the teacher being available during the test session was interviewed. Table 19 provides an overview about the teacher having completed the teachers’ interview.

Table 19: Respondent to teacher interview

	England	Germany	Netherlands	Sweden*
Form teacher	123 (67.6)	179 (72.2)	94 (49.5)	-
Other teacher	47 (25.8)	60 (24.2)	92 (48.4)	-
Missing	12 (6.6)	9 (3.6)	4 (2.1)	-
Total	182	248	190	-

* No information available in Sweden about which teacher completed the survey

The teacher interview was scheduled during the test session, with the aim that the test administrator collects the completed teacher survey at the end of the test session. However, this was not always possible. In contrast, some teachers completed the teachers’ questionnaire after the test session and send it back to the national research centres by post.

References

- Adams, R.J. 2003. **Response to ‘Cautions on OECD’s Recent Educational Survey (PI-SA)’**. In: Oxford Review of Education, 29: 377–389.
- CILS4EU. 2016. **Children of Immigrants Longitudinal Survey in Four European Countries. Codebook**. Wave 1 – 2010/2011, v1.2.0. Mannheim: Mannheim University.
- Dollmann, J., Jacob, K., and Kalter, F. 2014. **Examining the Diversity of Youth in Europe. A Classification of Generations and Ethnic Origins Using CILS4EU Data (Technical Report)**. MZES Working Paper 156.
<http://www.mzes.uni-mannheim.de/publications/wp/wp-156.pdf>
- Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R. 2004. **Survey Methodology**. New Jersey: John Wiley and Sons.
- Harkness, J. 2007. **ESS: Round 3 Translation Guidelines**, ESS Docs.
<http://www.europeansocialsurvey.org>
- Heller, K.A., and Perleth C. 2000. **Kognitiver Fähigkeits-Test (Revision) für 5.-12. Klassen (KFT 5-12+ R)**. Göttingen: Beltz-Testgesellschaft.
- Joncas, M. 2008. **TIMSS 2007 Sample Design**. In: Olson J.F., Martin, M.O., and Mullis, I.V.S. (eds.): TIMSS 2007 Technical Report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. pp. 77–92.
- Kornmann, R., and Kornmann A. 2003. **Erneuter Anstieg der Überrepräsentation ausländischer Kinder in Schulen für Lernbehinderte**. In: *Zeitschrift für Heilpädagogik*, pp. 286–289.
- Kruse, H., and Jacob, K. 2016. **Children of Immigrants Longitudinal Survey in Four European Countries. Sociometric Fieldwork Report**. Wave 1 – 2010/2011, v1.2.0. Mannheim: Mannheim University.
- Martin, M.O. 2005. **TIMSS 2003 User Guide for the International Database**. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Prüfer, P., and Rexroth, M. 2005. **Kognitive Interviews**. ZUMA How-to-Reihe 15. ZUMA, Mannheim.
http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/How_to15PP_MR.pdf
- Sturgis, P., Smith, P., and Hughes, G. 2006. **A study of suitable methods for raising response rates in school surveys**. Department for Education and Skills, London. (<https://www.education.gov.uk/publications/eOrderingDownload/RB721.pdf>)

- Van Dijk, H., and Tellegen, P.J. 2004. **NIO Nederlandse Intelligentietest voor Onderwijsniveau. Handleiding en Verantwoording**. Amsterdam: Boom testuitgevers.
- Weiß, R.H. 2006. **CFT-20R. Grundintelligenzskala 2 – Revision**. Göttingen: Hogrefe
- Zuehlke, O. 2011. **Sampling design and implementation**. In: Schulz, W., Ainley, J. and Fraillon, J. (eds): ICCS 2009. Technical Report. International Association for the Evaluation of Educational Achievement (IEA), Amsterdam, pp. 59–68
- Zuehlke, O., and Vandenplas, C. 2011. **Sampling weights and participation rates**. In: Schulz, W., Ainley, J. and Fraillon, J. (eds): ICCS 2009. Technical Report. International Association for the Evaluation of Educational Achievement (IEA), Amsterdam, pp. 69–87.