

# Technical Report

## RECENS “Wired into Each Other: Network Dynamics of Adolescents in the Light of Status Competition, School Performance, Exclusion and Integration”

Project K 81336 of the Hungarian Scientific Research Fund (OTKA)

2010-2013

Version: 13 July 2022

### Table of contents

<b>1. Study description .....</b>	<b>3</b>
1.1. Aims and novelty.....	3
1.2. Funding.....	3
1.3. Research design .....	3
1.4. The student questionnaire.....	4
1.5. The class head teacher questionnaire .....	4
1.6. Publications from the dataset.....	5
1.7. Giving references.....	5
1.8. Contributors.....	5
1.9. Further information .....	6
<b>2. Sample description .....</b>	<b>7</b>
2.1. Sampling and participating schools.....	7
2.2. Gender composition of the sample .....	9
2.3. Ethnic composition of the sample .....	13
2.4. Participation rates .....	17
2.5. Panel change I – classroom turnover.....	19
2.6. Panel change II – Student turnover .....	23
<b>3. Datasets, codebooks, and questionnaires – an overview .....</b>	<b>26</b>
3.1. Datasets.....	26
1.1.1. Student individual variables dataset .....	26
1.1.2. Student social networks dataset.....	26
1.1.3. Class head teacher dataset .....	27
3.2. Codebooks .....	27
3.3. Questionnaires .....	27

3.4. List of available datasets, codebooks, and questionnaires .....	28
<b>4. Publications from the dataset .....</b>	<b>29</b>
4.1. Research articles in English.....	29
4.2. Research articles in Hungarian .....	30
4.3. Books in Hungarian.....	30
4.4. Defended PhD theses.....	30
<b>5. First results from the dataset.....</b>	<b>31</b>
5.1. Ethnic segregation .....	31
5.2. Academic performance.....	33
5.3. Status competition and peer relations.....	33
5.4. Dynamics and co-evolution of positive and negative ties .....	35
5.5. References .....	37
<b>6. Research ethics and data security .....</b>	<b>39</b>
6.1. Informed consent from schools, parents, and students.....	39
6.2. Anonymity of respondents .....	40
6.3. Anonymized feedback to teachers.....	40
6.4. Data security .....	40
6.5. Data access .....	40
6.6. Ethics board approvals .....	41
<b>7. Data collection procedures.....</b>	<b>42</b>
7.1. Study planning.....	42
7.2. Data collection: preparations.....	42
7.3. Data collection: in the field .....	43
7.4. Data processing and cleaning .....	43
<b>8. The study budget .....</b>	<b>44</b>
<b>Appendix .....</b>	<b>45</b>

## **1. Study description**

The project “Wired into Each Other” is a three-year longitudinal social network study conducted by the Research Center for Educational and Network Studies (RECENS). RECENS was founded at the Corvinus University of Budapest in 2010. It has been hosted by the Hungarian Academy of Sciences since 2012. This study involved the collection of a unique, large-scale dataset about the evolution of interpersonal relations and various individual behaviours and attitudes in more than 40 student communities from Hungary over a period of three years.

### **1.1. Aims and novelty**

The study can be viewed as a multi-purpose research endeavour, which aimed at a) developing novel survey measures of informal social networks and b) gaining novel insight into the social processes shaping adolescent communities. In scope of this, the RECENS team has developed a multi-item network questionnaire about peer relations in more than 30 different aspects (contact, affection, trait and behaviour perceptions, status and role attributions, bullying, etc.).

Using this measurement tool, we collected data of unprecedented depth about the multidimensional nature of social processes in school communities. So far, the dataset has allowed researchers to study the mechanisms behind status competition, group formation, ethnic integration (with focus on the Roma minority group), bullying and victimization, school performance, substance use, and so on.

### **1.2. Funding**

The funding for “Wired into Each Other” was provided by the Hungarian Scientific Research Fund (OTKA) under a regular funding scheme for basic research (reference no. K 81336). The principal investigators in the project were Tamás Bartus and Károly Takács.

### **1.3. Research design**

The data collection involved 44 high-school classrooms situated in seven schools and four towns in Hungary.

In the sample design phase, schools were selected based on

- a. their location (capital, large and smaller towns from one region),
- b. the training programmes they offered (vocational, technical, grammar), and
- c. their estimated ethnic composition (with the aim of maximizing variability between the classrooms in this dimension).

In each of the schools that agreed to participate in the study, all classrooms in the 9<sup>th</sup> (first high-school) grade of 2010-2011 were surveyed at four distinct time points over three years:

- 1) two months after the beginning of high school (October-November 2010),
- 2) half a year later (April 2011),
- 3) a year after the second wave (April 2012), and
- 4) a year after the third wave (April 2013).

#### **1.4. The student questionnaire**

In all four survey waves, students were asked to fill out a questionnaire that contained relational and individual background questions. The questionnaires were pen-and-paper and self-administered during one class period of 45 minutes previously designated by the school. The actual time to respond to the list of questions varied greatly between and within classrooms, roughly between 25 and 50 minutes (sometimes students voluntarily sacrificed a part of their break to finish answering). In every classroom and wave, at least two trained members of the RECENS team were present while students completed the questionnaires. This policy was to 1) generate personal contact with the participants, answer their questions on the spot, and thus elicit more honest responses; 2) ensure that the students answered the questions on their own, without peer pressure; 3) ensure that the completed questionnaires cannot be accessed by anyone from and around the school who might personally know the respondents.

One of the innovations implemented in this study is the large set of social network items: we asked students to nominate their peers from their own classroom in over 30 relational dimensions. These included different types of social contact (e.g. in school, outside of school, in free time), affective relations (from friendship to enmity), peer perceptions of personal traits (e.g. shy, aggressive), abilities (e.g. smart), dyadic roles (e.g. trust), social roles in the class (e.g. leader, organizer), bullying (physical and verbal, from bully and victim perspectives), and also ethnicity (perceptions of Roma ethnicity). The dataset thus allows researchers to utilize a wealth of information about the strongly multiplex informal relations among students.

Besides the network items, the student questionnaires also aimed to collect information about important individual characteristics. These involved a series of questions about students' family background (e.g. economic situation of the household, available material and cultural goods, access to services) and their ethnic self-identity (Hungarian, Roma, other minority and mixed identities). Further, we asked them about their attitudes towards school achievement, their studying habits and goals, extracurricular activities, alcohol consumption and smoking, and various other behaviours and attitudes.

#### **1.5. The class head teacher<sup>1</sup> questionnaire**

Apart from the questions asked from the students, in each of the four waves the head teachers *of the classrooms in the sample* were asked to complete a separate questionnaire. Whenever possible, these were administered in the form of a structured interview with a previously trained member of the RECENS team, in the school and on the same day the student data was collected. This allowed us to learn additional qualitative information about the studied classrooms. If a teacher was not available to fill out the questionnaire during a personal interview, they could also complete it and send it to RECENS directly by mail or e-mail shortly after the data collection.

The teacher questionnaire focused on three topics relevant to our study: the background of the class head teacher (e.g. age, subject), specific circumstances of the class (e.g. seating order), and their perceptions about students in the class (ranging from who is absent a lot to who the teacher sees as smart, popular or a clique leader in the community). The information collected in the teacher survey

---

<sup>1</sup> By “class head teacher”, we refer to a teacher who is responsible for managing the academic matters of a given classroom of students. This role is sometimes referred to in the literature as “homeroom teacher” or “headmaster”.

can be valuable in determining important contextual factors that may have affect the social dynamics of the classrooms in the sample. Furthermore, teacher perceptions about students provide an additional perspective to studying classroom performance and the school-related attitudes of students.

(Here we note that due to the lack of resources, teacher questionnaires were only aimed at class head teachers who are just one of the many teachers students interact with and learn from at this stage of the Hungarian education system. Nonetheless, head teachers are responsible for attending to the administrative and certain personal matters of their class. They are supposed to be closely involved with their students, and they also often have a considerable impact on the classroom community.)

### **1.6. Publications from the dataset**

The data collected in scope of the “Wired into Each Other” project has been used for various purposes to date. Publications involve articles in prestigious social psychological and sociological journals (e.g. *Social Psychology Quarterly*, *Journal of Research on Adolescence*, *Social Networks*, *Sociological Science*, *European Sociological Review*), presentations and posters at international sociological (ASA), social psychological (EASP), and social networks conferences (Sunbelt, EUSN, NetSci, ICCSS). Moreover, the dataset has provided the empirical basis for PhD projects of several RECENS members at the University Oxford and the Corvinus University of Budapest (CUB). The project has also contributed to a number of MA theses in sociology at the CUB, an MSc thesis in statistics at Oxford, and an MSc thesis in data science at the University of Manchester. The number of scientific outputs generated by the study is luckily constantly growing.

### **1.7. Giving references**

The presented dataset is publicly available for academic use, including research and education. Commercial use of the data is not allowed. In case you publish, present, teach, etc. any part of the dataset or results of analyses based on it, please use the following reference:

Vörös, A., Boda, Zs., Néray, B., Pál, K., Kisfalusi, D., J., Samu, F., Vit, E., Radó, M., Habsz, L., Csaba, Z., Lőrincz, L., Mandácskó, E., Panyik, B., Varga, K., Mezei, G., Makovi, K., Boldvai-Pethes, L., Havelda, A., Bartus, T., & Takács, K. (2022). *RECENS Wired into Each Other: Network dynamics of adolescents in Hungarian secondary schools (2010-2013)*. [Data Collection]. Colchester, Essex: UK Data Service.

### **1.8. Contributors**

At different stages, the following people have significantly contributed to the success of the study (in chronological order): Károly Takács, Tamás Bartus, Zsófia Boda, Judit Pál, Kinga Makovi, Bálint Néray, László Lőrincz, Zoltán Csaba, András Vörös, Roland Reiner, Eszter Mandácskó, Barbara Panyik, Dorottya Kisfalusi, Kinga Varga, Márton Marosi, Gabriella Mezei, Viktória Szalai, Sára Horlai, Anna Vancsó, Beatrix Tóth, Hanna Kónya, Tamás Ribárszki, Eszter Vit, Flóra Samu, Laura Boldvai-Pethes, Anikó Havelda, Márta Radó, Lilla Habsz. We also thank many others who occasionally supported the work of the group!

## **1.9. Further information**

Further important details about the project are discussed below. These include the following:

- 2 Sample description
- 3 Datasets, codebooks, and questionnaires
- 4 Publications from the dataset
- 5 Summary of first results
- 6 Ethics, privacy, and data security
- 7 Description of data collection procedures
- 8 Overview of project budget

## **2. Sample description**

### **2.1. Sampling and participating schools**

The studied sample of classrooms was selected through a two-stage cluster sampling procedure. First, schools were sampled based on expert considerations (see below). Second, all 9<sup>th</sup> grade (first secondary-school year) classes were selected within the sampled schools. As a result, the sample initially consisted of every 9<sup>th</sup> grade student attending any of the selected schools in November 2010.

The sample is a targeted sample, therefore, it is not representative of Hungarian secondary schools or classrooms. However, the selection of schools was based on a number of considerations. First, the sample consists of schools that are located in economically diverse areas of the country (e.g. capital vs. smaller towns). Second, despite the small number of schools, all three training programmes of the Hungarian secondary education system (grammar, technical, vocational), are represented in the sample by multiple classrooms. Third, the studied classrooms are strongly heterogeneous in ethnic composition. The variation in these three, non-independent dimensions within the sample makes interesting comparative studies possible (of course, keeping in mind that the sample is non-representative).

The main reason for choosing 9<sup>th</sup>-grade student communities was that they were formed at the beginning of the 2010-11 academic year, and so subsequent changes in community structure and individual variables are interpretable compared to an initial state. In addition, 9<sup>th</sup>-grade students are at the beginning of their secondary-school years, which makes it possible to study the evolution of their classroom social networks over several years (although quite a few classrooms change substantially in their composition due to drop-outs). After the first data collection in the beginning of 9<sup>th</sup> grade (autumn 2010), further survey waves were conducted in the second semester of 9<sup>th</sup> grade (spring 2011), in the second semester of 10<sup>th</sup> grade (spring 2012) and in the second semester of 11<sup>th</sup> grade (spring 2013).

Our research is unique considering the creation of a large (40-classroom) and heterogeneous longitudinal network database. Heterogeneity primarily refers to a variation in regions, training programmes, and ethnicity. In addition, the dataset contains measures for over 30 different dimensions of interpersonal relationships, and so makes possible the analysis of network evolution from many different aspects. Tables 1-3 provide an overview of the number of classrooms and students by schools and training programmes.

**Table 1.** *Number of participating classrooms by school*

School id	Classroom training programmes	Participating classes 1 <sup>st</sup> wave	Participating classes 2 <sup>nd</sup> wave	Participating classes 3 <sup>rd</sup> wave	Participating classes 4 <sup>th</sup> wave
1000	Grammar	5	5	5	5
2000	Technical, Vocational	10	10	10	12
3000	Grammar, Technical, Vocational	7	7	6	4
4000	Grammar, Technical	4	4	4	4
5000	Technical, Vocational	6	6	5	5
6000	Grammar	4	4	4	4
7000	Technical, Vocational	8	8	7	4
Total		44	44	41	38

**Table 2.** *Number of participating students by school*

School id	Number of students 1 <sup>st</sup> wave	Number of students 2 <sup>nd</sup> wave	Number of students 3 <sup>rd</sup> wave	Number of students 4 <sup>th</sup> wave
1000	166	161	164	165
2000	315	307	246	259
3000	245	248	188	110
4000	124	123	106	97
5000	179	172	138	145
6000	136	136	133	124
7000	259	230	180	80
Total	1425	1377	1155	980

**Table 3.** *Number of classrooms by training programme and school*

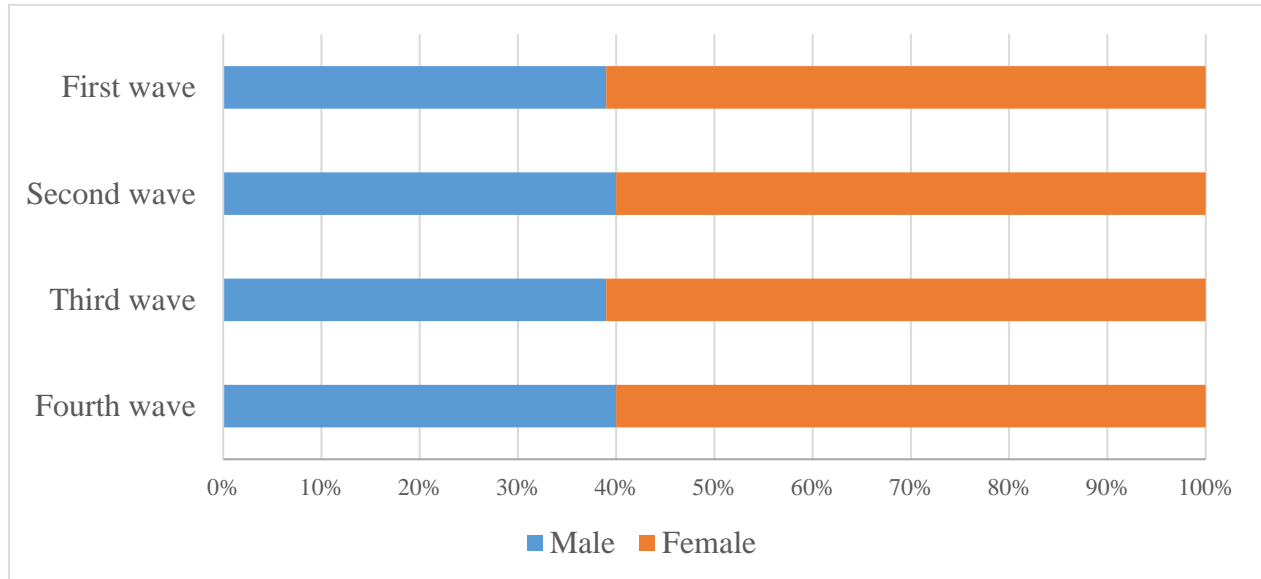
School id	Grammar	Technical	Vocational	Total	Town pseudonym
1000	5	0	0	5	Bigtown
2000	0	4	6	10	Bigtown
3000	3	1	3	7	Smalltown1
4000	3	1	0	4	Smalltown2
5000	0	2	4	6	Smalltown2
6000	4	0	0	4	Capital
7000	0	4	4	8	Capital
Total	15	14	14	44	



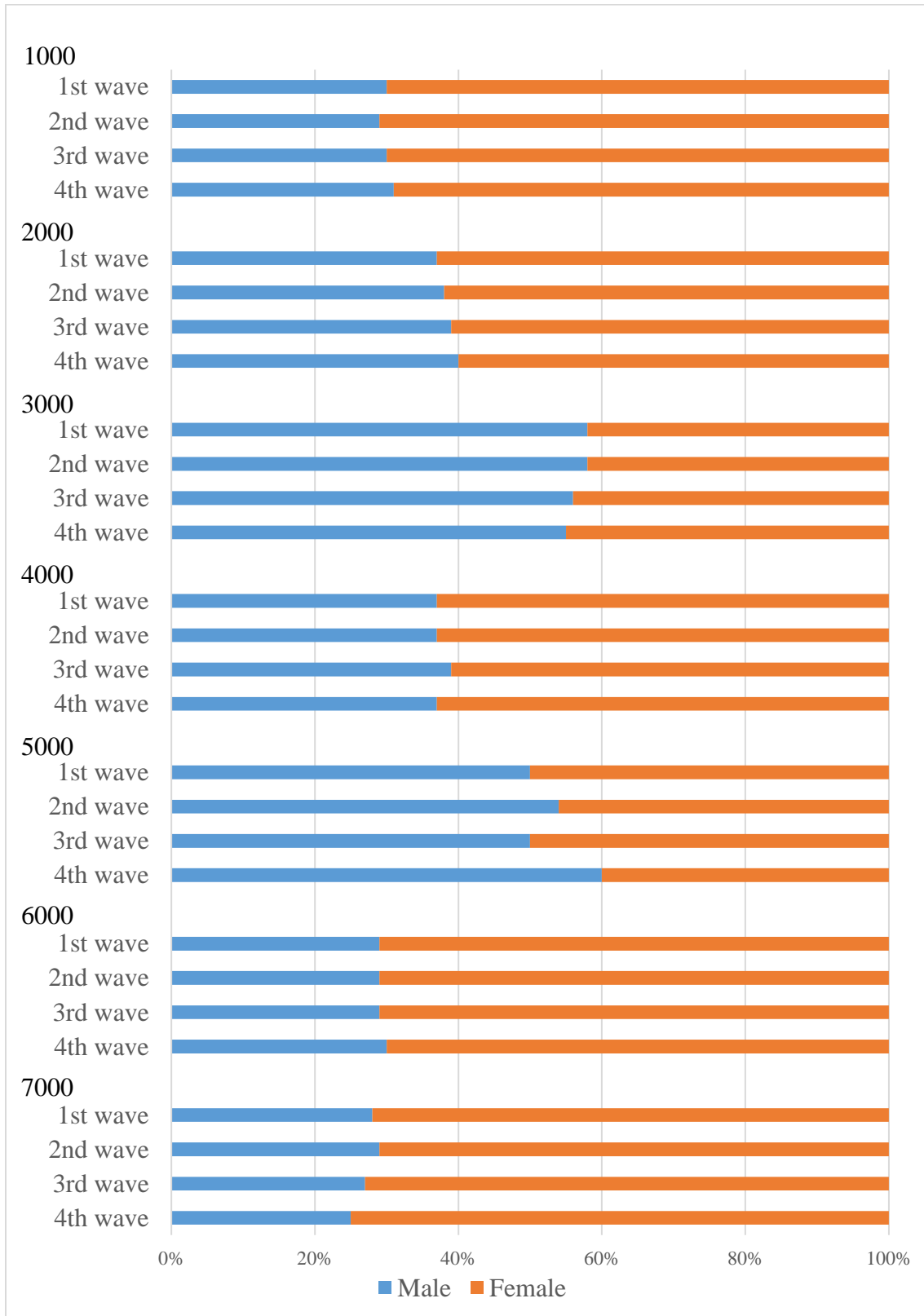
## 2.2. Gender composition of the sample

The proportion of female students in the sample was 61% in the first wave and 60% in the fourth wave (Figure 1). While the overall gender composition was stable, the ratio of male and female students changed in some of the schools over time (Figure 2). Figures 3.1-3.4 show comparisons between schools in each wave.

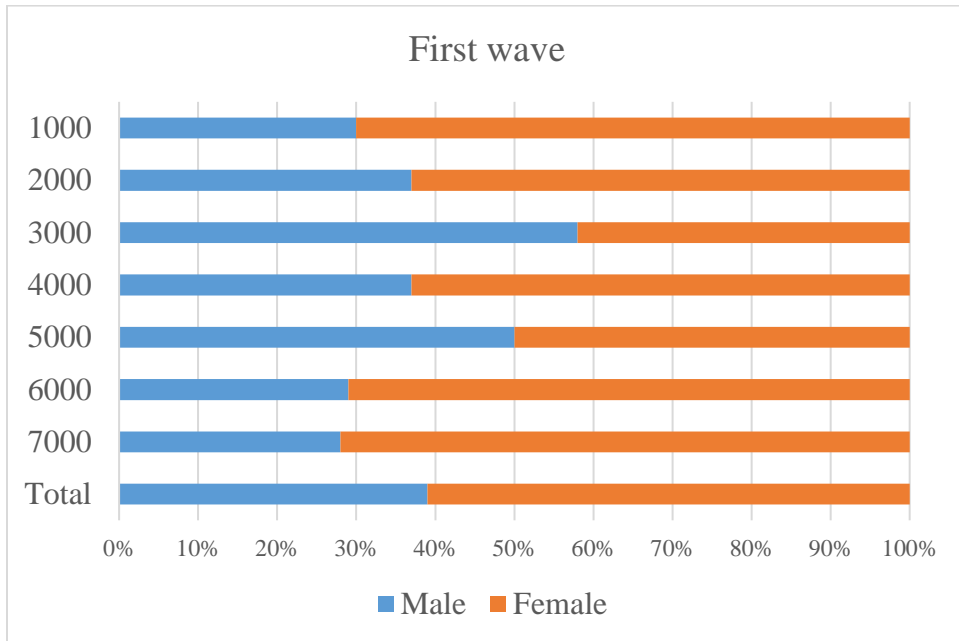
**Figure 1.** *Gender composition over time in the entire sample*



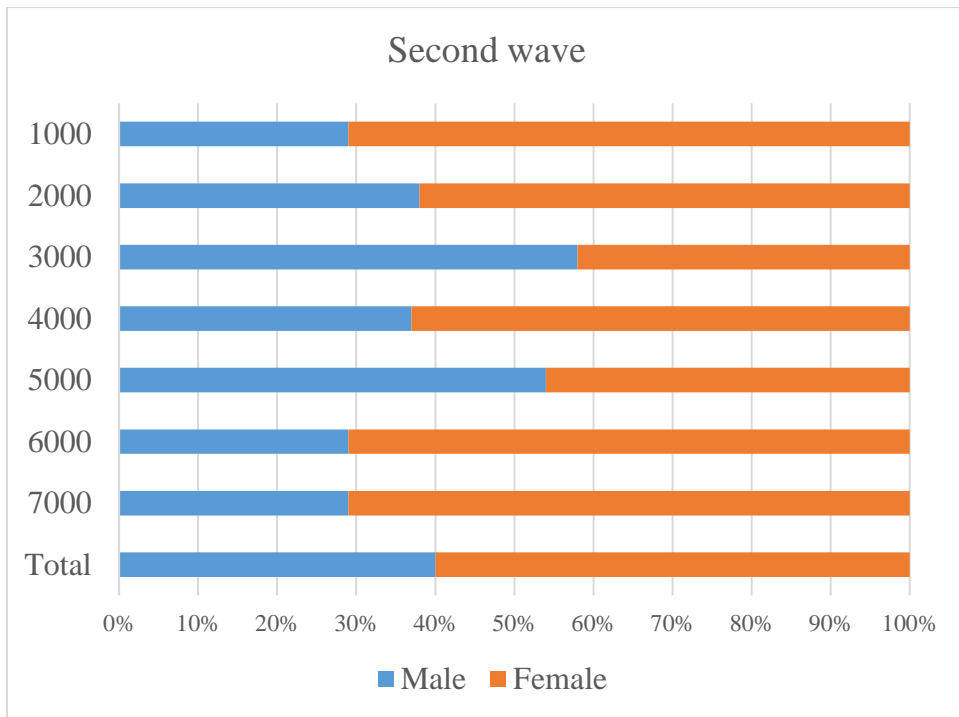
**Figure 2.** Gender composition over time by school



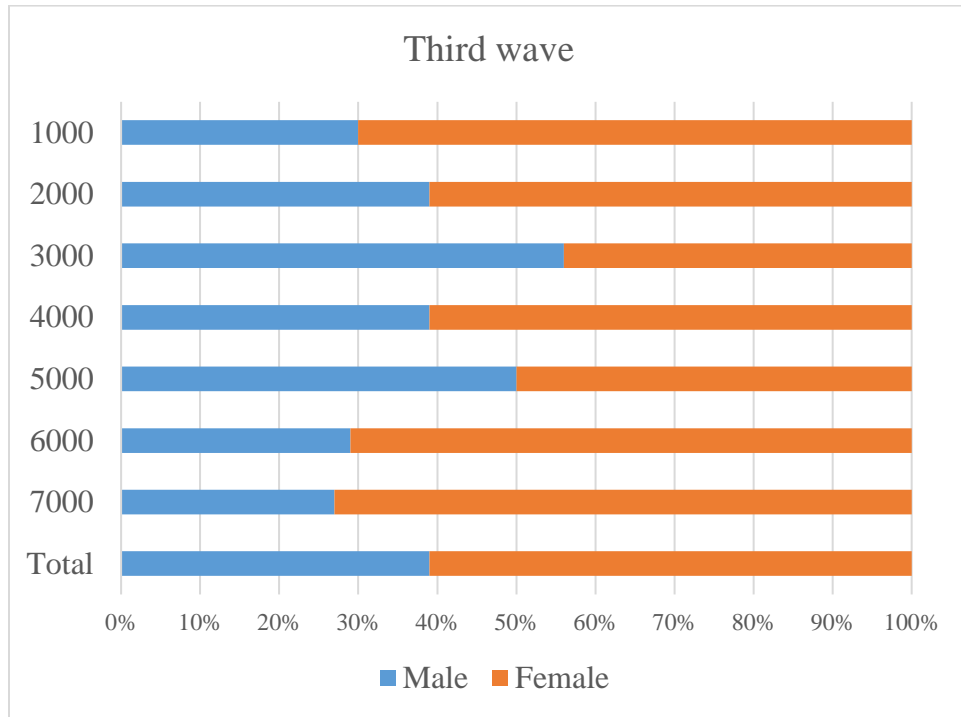
**Figure 3.1.** Gender composition by school in wave 1



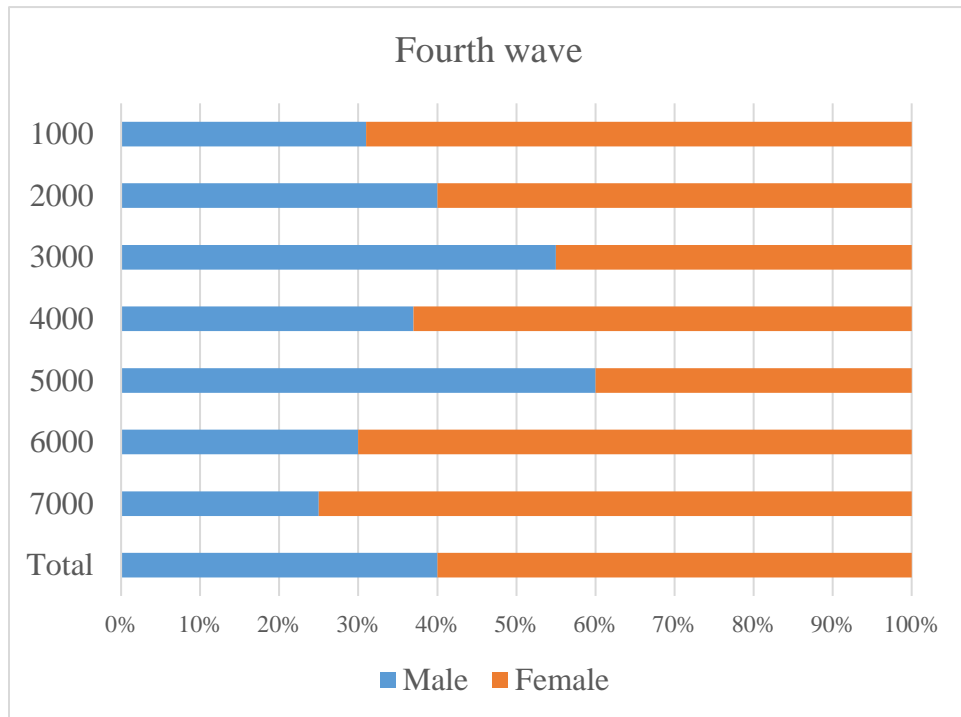
**Figure 3.2.** Gender composition by school in wave 2



**Figure 3.3.** Gender composition by school in wave 3



**Figure 3.4.** Gender composition by school in wave 4

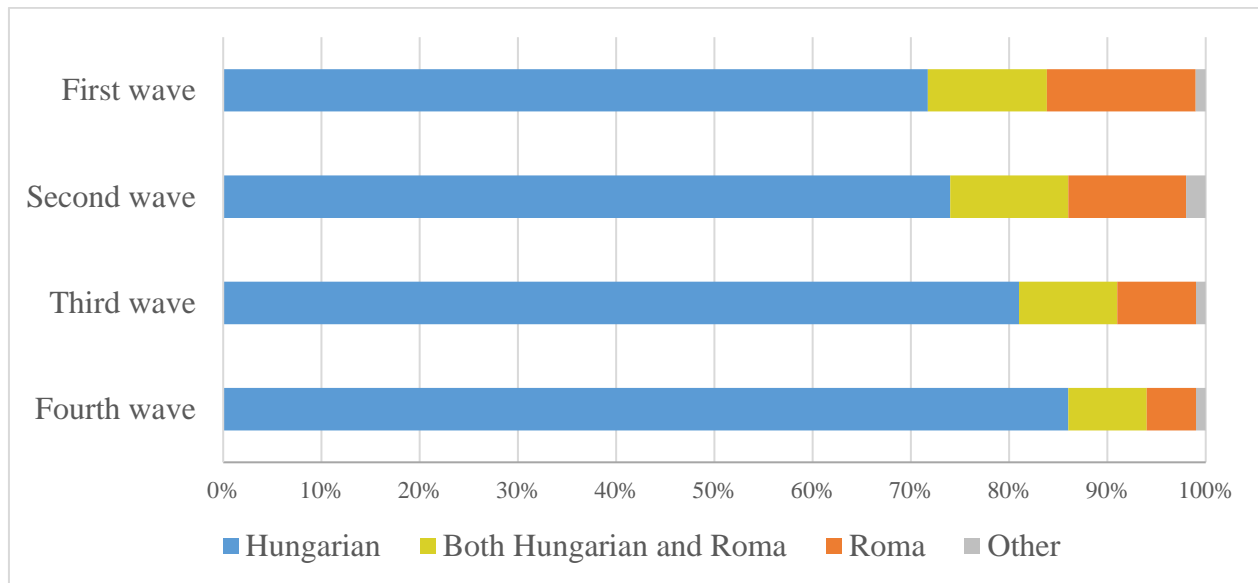


### 2.3. Ethnic composition of the sample

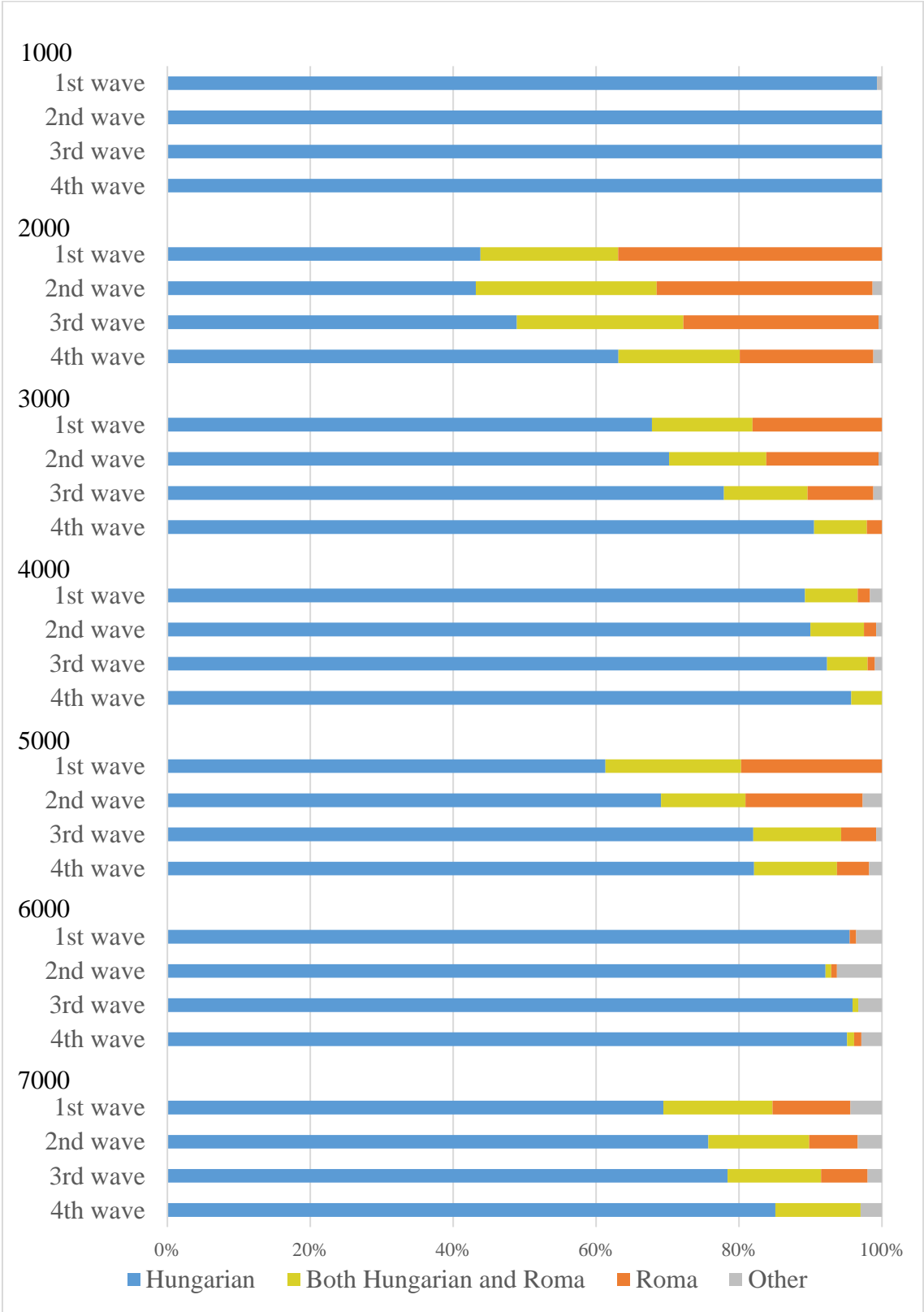
The proportion of self-reported Roma students in the entire sample was 15% in the first wave and 5% in the fourth wave (Figure 4). The decrease is likely to be related to ethnic biases in school (and sample) drop-out, a widely observed phenomenon in the Hungarian secondary education system. In line with the sampling goals, we can see a mixed picture of ethnic composition at the school level: 0-37% of the participating students identified as Roma across the seven schools, considering all four waves (Figure 5, Figures 6.1-6.4). At the classroom level, the percentage of students who identified as Roma ranged between 0% and 60% (see Appendix).

It was also possible for students to choose in the questionnaires that they have a Hungarian-Roma mixed-ethnic identity. Similar to the proportion of students with Roma identity in the entire sample, the proportion of students with a mixed identity showed a decline over time. However, the extent of the decrease is smaller in this case: from 12% in the first wave to 8% in the fourth (Figure 4). The proportion of mixed-ethnic identity students ranged between 0-25% in schools (Figure 5, Figures 6.1-6.4) and between 0-47% in classrooms (see Appendix).

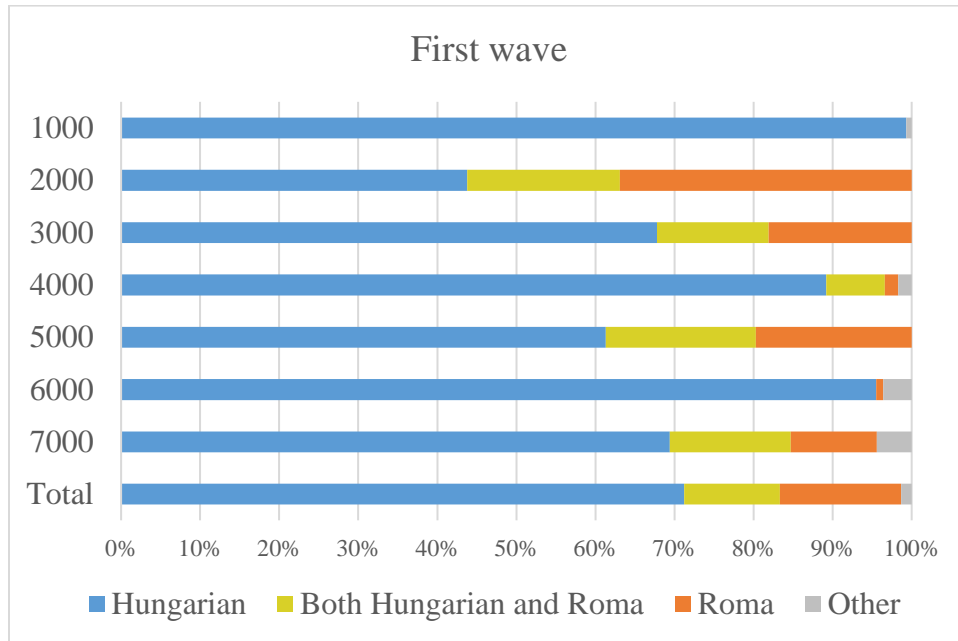
**Figure 4.** *Ethnic composition over time in the entire sample  
(based on self-reported ethnic identity)*



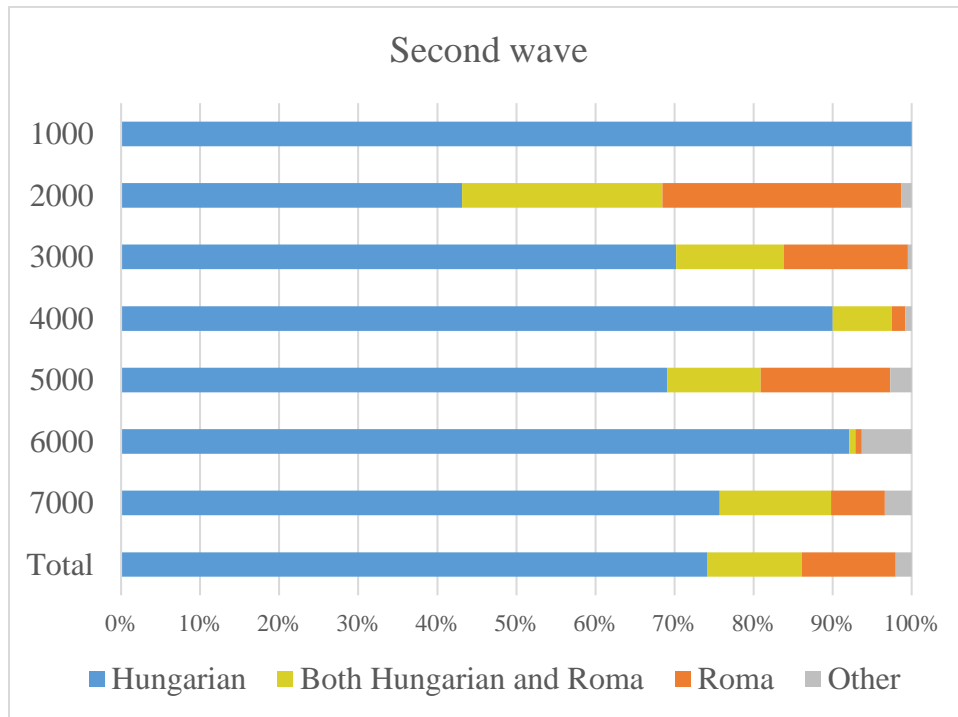
**Figure 5.** Ethnic composition over time by school (based on self-reported ethnic identity)



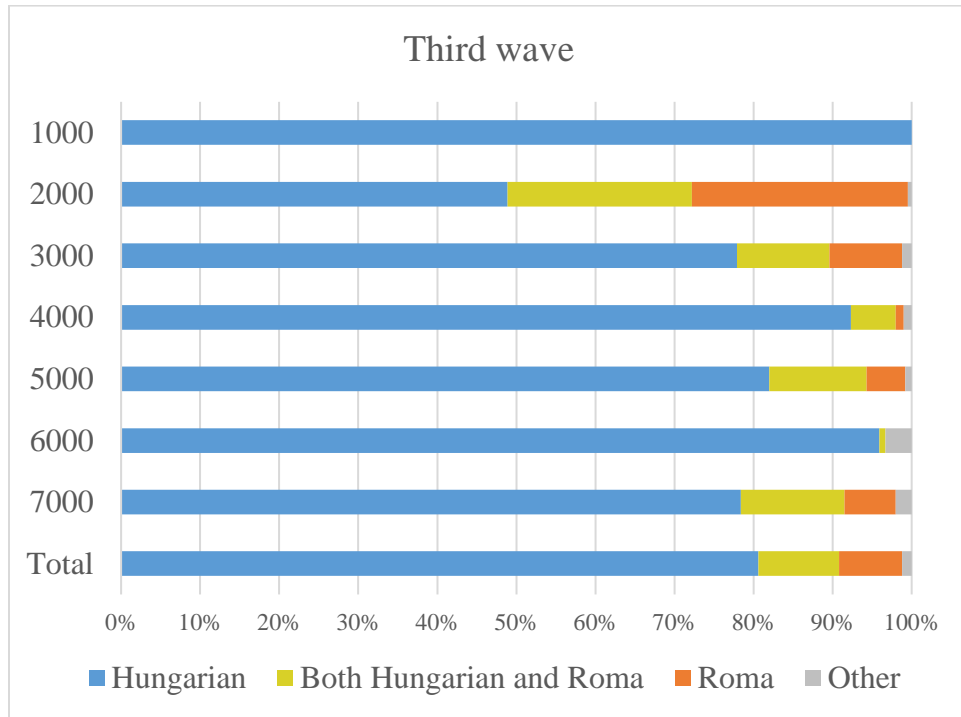
**Figure 6.1.** *Ethnic composition by school in wave 1 (based on self-reported ethnic identity)*



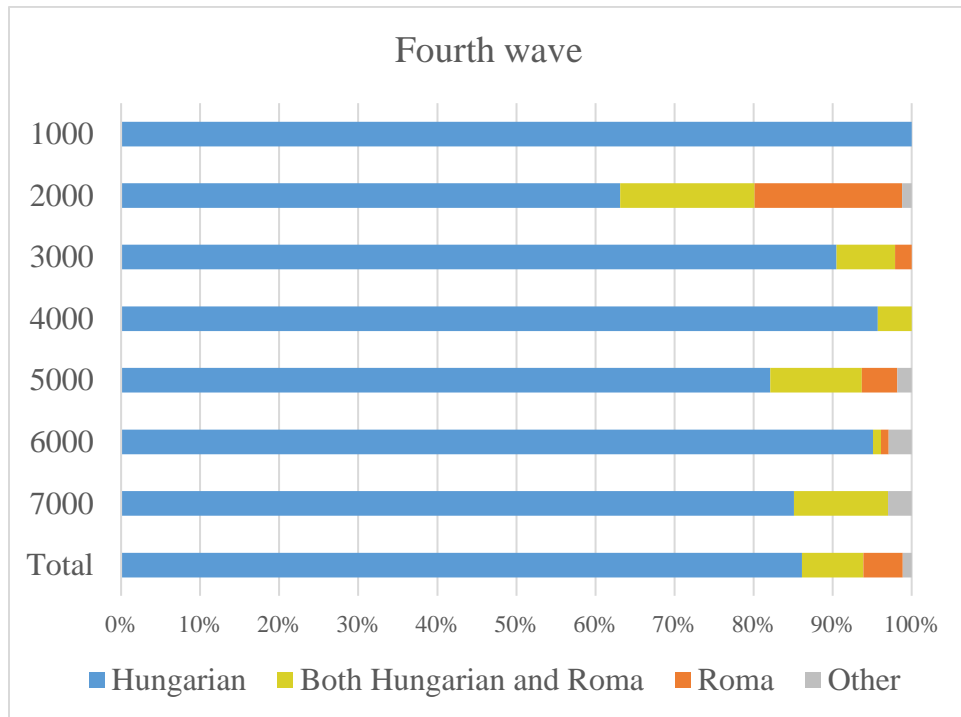
**Figure 6.2.** *Ethnic composition by school in wave 2 (based on self-reported ethnic identity)*



**Figure 6.3.** *Ethnic composition by school in wave 3 (based on self-reported ethnic identity)*



**Figure 6.4.** *Ethnic composition by school in wave 4 (based on self-reported ethnic identity)*





## 2.4. Participation rates

At the school level, the participation rates ranged between 71% and 100% across all waves, while for classes they ranged between 41% and 100% (Tables 4.1-4.2). Non-responses in large part came from students being absent from school at the time of data collection and in a smaller part from opt-outs.

The case of absent students highlights an important aspect of the group-boundary problem in our study: in some classrooms, especially in vocational classes, a number of students were officially attending school, but never or rarely showed up physically. That is, they were possible on their way to drop out from the given school or the entire education system. Apart from the severity of the school drop-out problem, it is also an open question whether these students should be counted as members of their respective classroom communities or not. We decided to keep them in the dataset, because the social connections with them reported by their classmates may be of significance for the development of the class community. What this means is that the participation rates reported below are pessimistic in the sense that some classes with low rates may still be used in meaningful analyses.

**Table 4.1.** *Participation rates by school*

School id	Participation rates 1 <sup>st</sup> wave	Participation rates 2 <sup>nd</sup> wave	Participation rates 3 <sup>rd</sup> wave	Participation rates 4 <sup>th</sup> wave
1000	93%	94%	98%	96%
2000	79%	78%	74%	70%
3000	87%	79%	88%	89%
4000	100%	99%	99%	99%
5000	84%	71%	90%	82%
6000	91%	96%	92%	85%
7000	80%	79%	86%	90%
Total	86%	83%	88%	85%

**Table 4.2. Participation rates by classroom**

Classroom id	Participation rate 1 <sup>st</sup> wave	Participation rate 2 <sup>nd</sup> wave	Participation rate 3 <sup>rd</sup> wave	Participation rate 4 <sup>th</sup> wave
1100	94%	100%	94%	97%
1200	100%	91%	97%	92%
1300	100%	100%	100%	94%
1400	77%	79%	100%	97%
1500	97%	100%	100%	100%
2000	79%	70%	100%	50%
2100	81%	96%	85%	83%
2200	94%	92%	52%	81%
2300	67%	79%	74%	-
2400	97%	94%	89%	85%
2500	77%	67%	70%	-
2600	66%	68%	70%	-
2700	60%	63%	65%	-
2800	91%	90%	78%	-
2900	88%	71%	62%	-
3100	97%	100%	96%	83%
3200	100%	91%	100%	88%
3300	86%	97%	81%	100%
3400	97%	83%	100%	88%
3500	73%	58%	-	-
3600	74%	63%	89%	-
3700	81%	64%	66%	-
4100	100%	94%	100%	100%
4200	100%	100%	100%	100%
4300	100%	100%	97%	100%
4400	100%	100%	100%	96%
5100	82%	74%	87%	96%
5200	88%	75%	-	77%*
5300	53%	0%**	81%	-
5400	89%	84%	94%	87%
5500	91%	80%	100%	93%
5600	90%	83%	76%	59%
6100	94%	97%	100%	96%
6200	100%	100%	86%	83%
6300	73%	85%	85%	73%
6400	97%	100%	97%	94%
7100	100%	80%	100%	100%
7200	68%	79%	-	-
7300	97%	86%	89%	92%
7400	85%	81%	76%	80%
7500	64%	65%	91%	-
7600	73%	76%	84%	-
7700	77%	75%	88%	-
7800	77%	90%	70%	93%
9100	-	-	-	65%
9200	-	-	-	41%
9300	-	-	-	38%
9400	-	-	-	85%
9500	-	-	-	83%

9600	-	-	-	38%
9700	-	-	-	100%
9800	-	-	-	65%

\* The “reappearance” of class 5200 is explained below.

\*\* Classroom 5300 was not surveyed in the second wave due to organizational difficulties.

## 2.5. Panel change I – classroom turnover

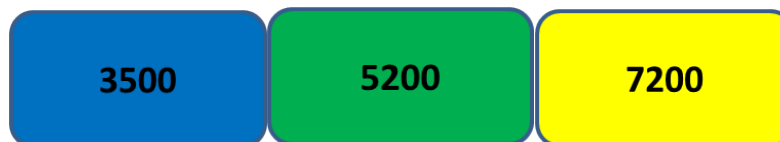
While there were 44 participating classes in the first wave, this number decreased to 38 by the fourth wave. Before the start of each academic year, some schools need to or decide to reorganize their classroom structure. This is most often, but not solely, motivated by some classes becoming too small. (In certain cases, the minimum class size is regulated by law, but organizational reasons may also facilitate restructuring small classrooms.) The primary reasons for shrinking classrooms sizes are 1) students switching between training programmes and thus moving to another class and 2) students dropping out from school (either leaving to another secondary school or dropping out from the system). This phenomenon led to some turnover in our sample on the level of classrooms.

Class mergers and class dissolutions took place in the breaks between academic years: between the second and third and the third and fourth waves. Altogether, 15 classes that existed in wave 1 and 2 were dissolved by wave 4. In addition, 8 new classes were created after wave 3. These also mostly contain students from classrooms that existed earlier, but with such a high level of mixing (a couple of students from many classrooms) that they can be considered as new communities (classrooms 9100-9800).

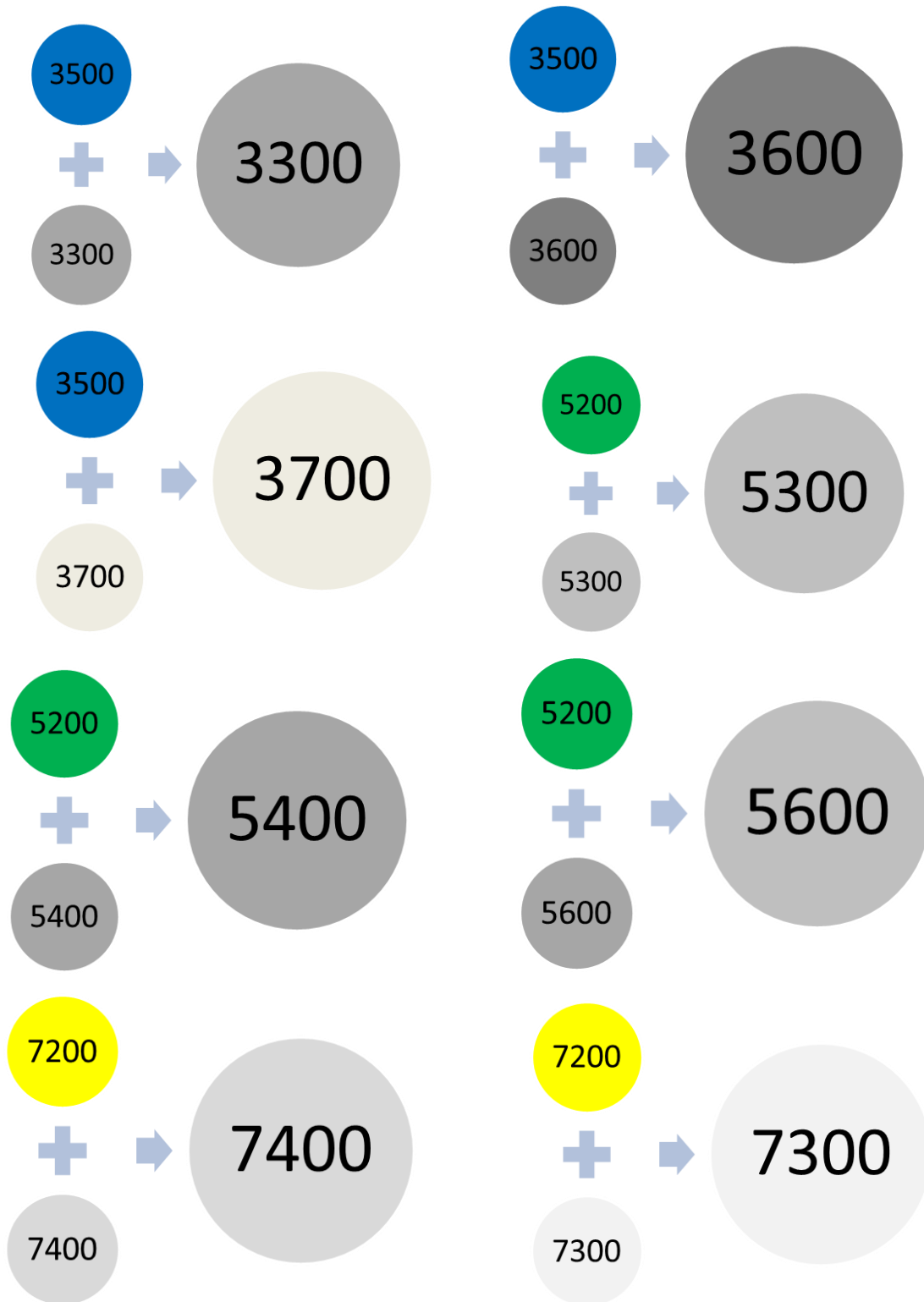
15 dissolved and 8 new classes would leave us with 37 classrooms in wave 4 ( $44 - 15 + 8$ ), so there is one more case to account for. Class 5200 was dissolved after wave 2 (Figure 7), and a significant part of this classroom joined class 5300 in wave 3 (Figure 8). However, many students originally from 5300 dropped out from our sample (left the school or system) after wave 3, and so 5300 dissolved (Figure 9). Due to this composition change, former members of class 5200 became the overwhelming majority in 5300 by wave 4. As a result, we decided to label this community in wave 4 as 5200 as its composition is closest to what once was class 5200.

Just like in the case of classes 5200 and 5300, Figures 7-10 below help to reconstruct the fate of each classroom that did not exist with a relatively stable student body throughout the study period.

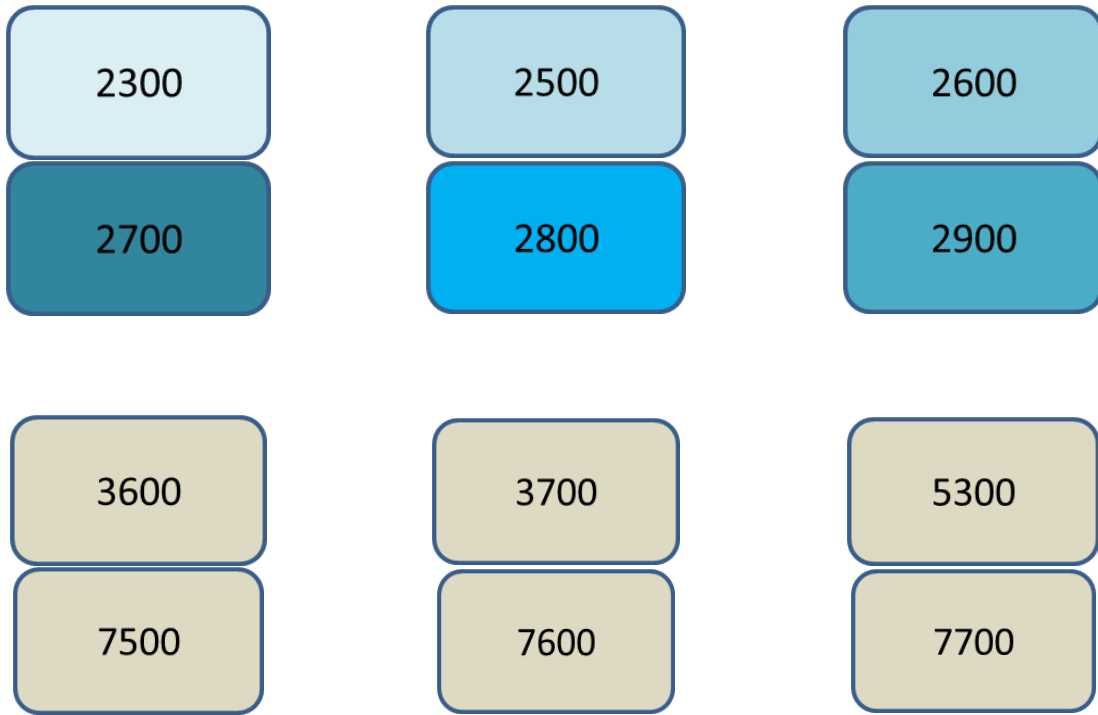
**Figure 7.** Dissolved classes between the second and third waves



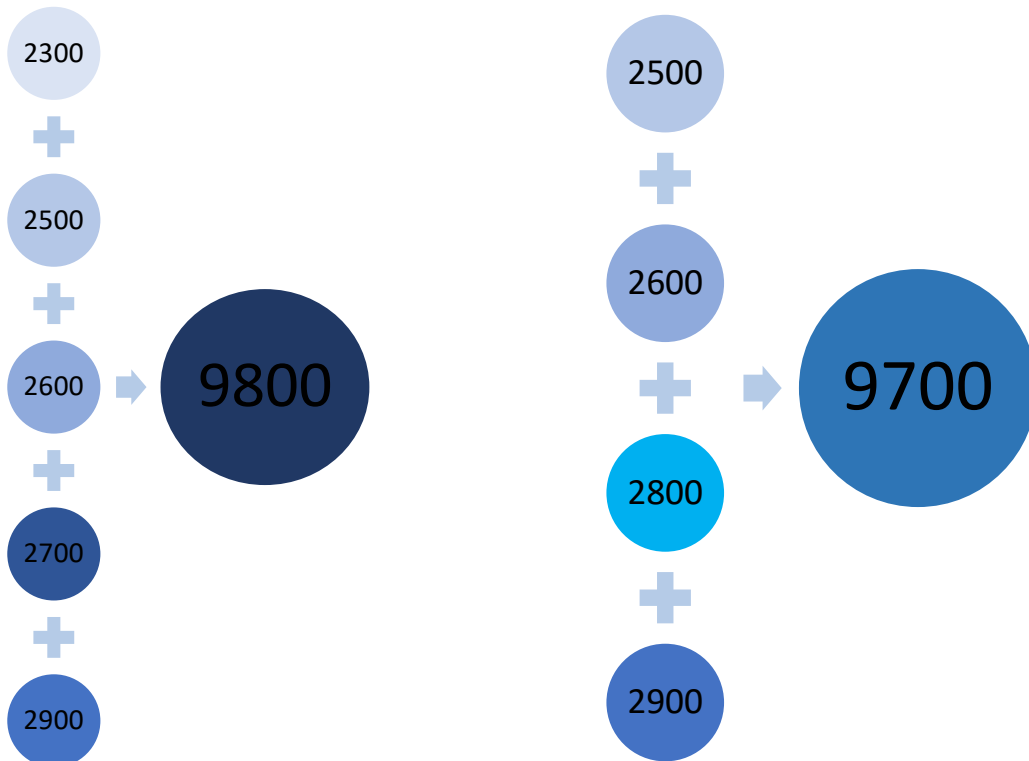
**Figure 8.** *Class mergers between the second and third waves*



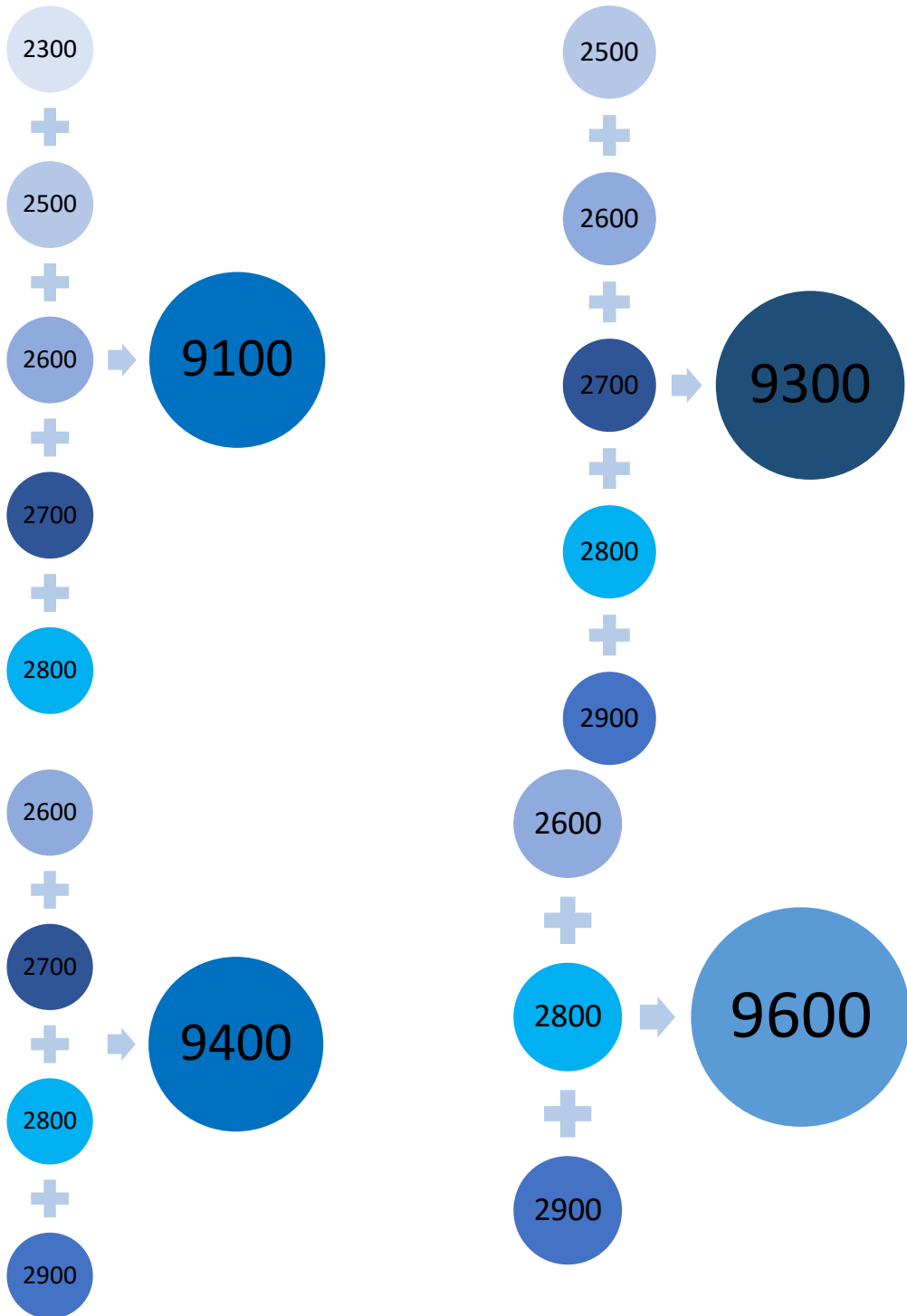
**Figure 9.** Dissolved classes between the third and fourth waves



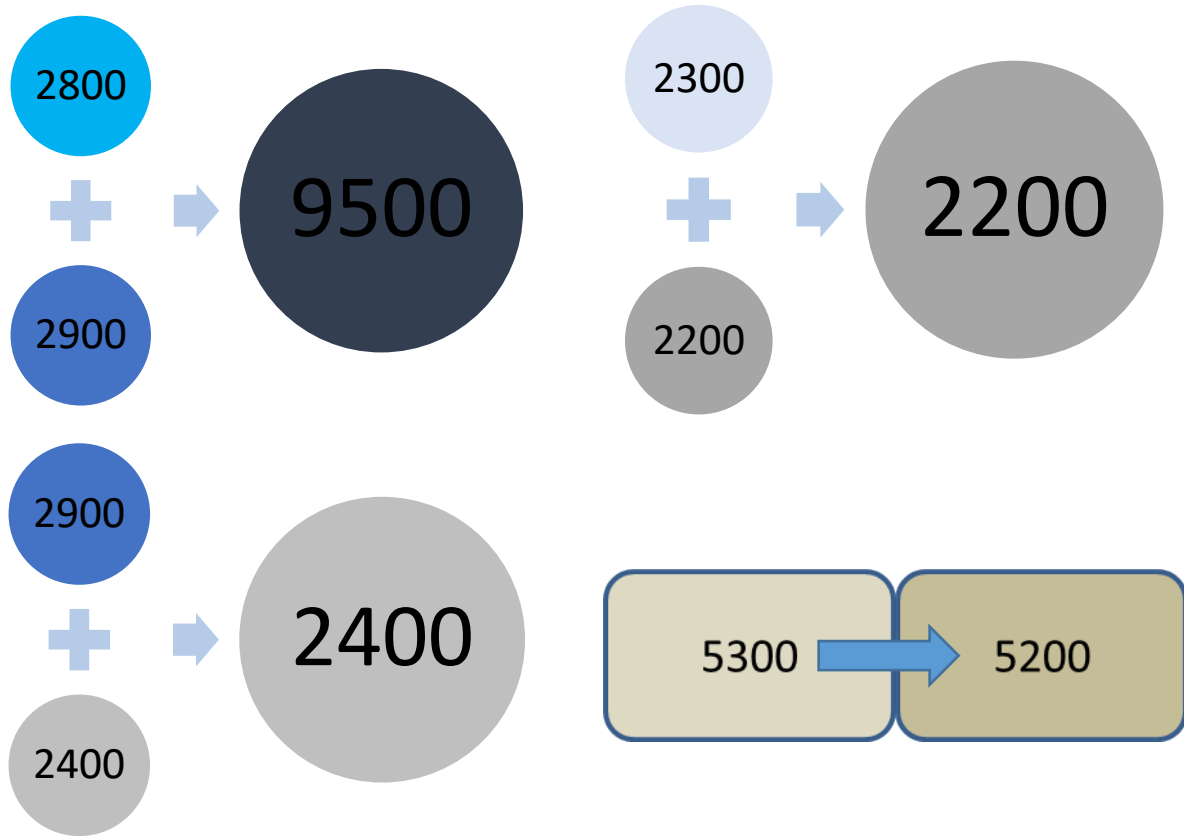
**Figure 10.** Class mergers between the third and fourth waves



(Fig. 10 continued from previous page)



(Fig. 10 continued from previous page)



## 2.6. Panel change II – Student turnover

Figures 11 and 12 provide an overview of student turnover rates per classroom over the four waves (cells coloured lighter reflect low turnover / high stability). The columns of Figure 11 show the proportion of students in each classroom in wave 2, 3, and 4, respectively, who were members of the same classroom already in wave 1. Higher proportions mean that very few of the original members of a class left (transferred to another classroom, school or dropped out from the education system), and in this sense the classroom had a stable core. Low proportions signal that the original set of students in a class were basically replaced by new students.

Figure 12 is very similar to Figure 11, but it provides a slightly more general view on sample stability. This figure shows how many of the original student body of each classroom was still in the sample (but not necessarily in the same class) in waves 2, 3, and 4.

**Figure 11.** Classroom stability: proportion of first-wave participants who were still in their initial classroom in waves 2, 3, and 4

Class id	Stability between wave 1 and 2	Stability between wave 1 and 3	Stability between wave 1 and 4
1100	0.97	0.94	0.91
1200	1.00	0.94	0.94
1300	0.97	0.92	0.92
1400	0.94	0.91	0.91
1500	0.93	0.93	0.93
2000	0.86	0.41	0.38
2100	0.82	0.75	0.57
2200	0.84	0.61	0.48
2300	0.93	0.53	
2400	0.93	0.69	0.52
2500	0.89	0.46	
2600	0.86	0.51	
2700	0.74	0.49	
2800	0.91	0.78	
2900	0.88	0.50	
3100	1.00	0.97	0.97
3200	1.00	0.91	0.88
3300	0.92	0.70	0.54
3400	0.97	0.83	0.67
3500	0.97		
3600	0.91	0.46	
3700	0.97	0.54	
4100	1.00	0.94	0.94
4200	1.00	0.97	0.94
4300	0.97	0.76	0.59
4400	0.97	0.47	0.47
5100	0.97	0.61	0.48
5200	0.92		0.35
5300	0.89	0.47	
5400	1.00	0.57	0.34
5500	0.94	0.63	0.51
5600	0.74	0.32	0.26
6100	1.00	0.97	0.70
6200	1.00	1.00	0.97
6300	1.00	0.91	0.85
6400	1.00	0.86	0.86
7100	0.94	0.69	0.41
7200	0.90		
7300	0.84	0.52	0.29
7400	0.91	0.67	0.42
7500	0.73	0.55	
7600	0.85	0.55	
7700	0.87	0.55	
7800	0.83	0.57	0.40



**Figure 12.** *Sample stability: Proportion of first-wave participants who were still in the sample (but not necessarily in the same class) in waves 2, 3, and 4*

Class id	Stability between wave 1 and 2	Stability between wave 1 and 3	Stability between wave 1 and 4
1100	0.97	0.94	0.91
1200	1.00	0.94	0.94
1300	0.97	0.92	0.92
1400	0.94	0.91	0.91
1500	0.93	0.93	0.93
2000	0.93	0.41	0.38
2100	0.93	0.82	0.64
2200	0.94	0.68	0.55
2300	0.93	0.57	0.40
2400	0.97	0.79	0.59
2500	0.89	0.49	0.37
2600	0.91	0.54	0.46
2700	0.77	0.49	0.37
2800	0.91	0.78	0.69
2900	0.91	0.63	0.47
3100	1.00	0.97	0.97
3200	1.00	0.91	0.88
3300	0.95	0.76	0.57
3400	0.97	0.92	0.67
3500	0.97	0.32	0.03
3600	0.91	0.46	0.06
3700	0.97	0.54	0.00
4100	1.00	1.00	1.00
4200	1.00	0.97	0.94
4300	1.00	0.79	0.62
4400	0.97	0.50	0.50
5100	0.97	0.64	0.52
5200	0.96	0.35	0.38
5300	0.95	0.53	0.53
5400	1.00	0.60	0.34
5500	0.94	0.66	0.57
5600	0.74	0.32	0.26
6100	1.00	0.97	0.70
6200	1.00	1.00	0.97
6300	1.00	0.91	0.85
6400	1.00	0.86	0.86
7100	0.94	0.69	0.41
7200	0.90	0.32	0.32
7300	0.90	0.58	0.32
7400	0.91	0.70	0.42
7500	0.73	0.55	0.00
7600	0.88	0.58	0.00
7700	0.90	0.58	0.00
7800	0.83	0.57	0.40

### 3. Datasets, codebooks, and questionnaires – an overview

This section provides a brief overview of the datasets, codebooks, and questionnaires from the “Wired into Each other” study of RECENS. Here we introduce all the data structures and documentation necessary to work with the dataset. Before analysing the data, we suggest reading this introduction, then studying the questionnaires and codebooks.

#### 3.1. Datasets

The data collection during the study has resulted in three interlinked datasets, which are available to researchers for scientific use. These contain all the recorded information from all four waves of the study. In the following, we briefly introduce the content and format of the three datasets.

##### *1.1.1. Student individual variables dataset*

Besides the network items, the student questionnaires also aimed to collect information about important individual characteristics. These involved a series of questions about students’ family background (e.g. economic situation of the household, available material and cultural goods, access to services) and their ethnic self-identity (Hungarian, Roma, other minority and mixed identities). Further, we asked them about their attitudes towards school achievement, their studying habits and goals, extracurricular activities, alcohol consumption and smoking, and various other behaviours and attitudes.

**Format.** This dataset is available as a single file, in text-based (csv) format. Variables can be linked to the other two datasets by student IDs. Consult the student questionnaires and codebooks for variables and coding.

##### *1.1.2. Student social networks dataset*

One of the innovations implemented in this study is the large set of social network measures: we asked students to nominate their peers from their own classroom in over 30 relational dimensions. These included different types of social contact (e.g. in school, outside of school, in free time), affective relations (from friendship to enmity), peer perceptions of personal traits (e.g. shy, aggressive), abilities (e.g. smart), dyadic roles (e.g. trust), social roles in the class (e.g. leader, organizer), bullying (physical and verbal, from bully and victim perspectives), and ethnicity (Roma perceptions). This dataset allows researchers to utilize detailed information about the strongly multiplex informal relations among students.

**Format.** This dataset is a collection of numerous text files (csv). Each file contains an adjacency matrix for a given network item in the questionnaire from a specific classroom in a given wave. For example, the file “Wave 1\1000\1100\14\1100\_14\_6.csv” contains the adjacency matrix in wave 1 of school 1000’s classroom 1100 for item 6 in question block 14 (see the questionnaires). The file contains row and column headers, which are the student IDs – all the students who attended class 1100 at the time of wave 1. By looking into the codebook of the network data, we can see that item 14\_6 shows student reports about spending free-time with others. This means that the value in row 1101 and column 1102 represents whether student 1101 reported spending some of his free-time with student 1102. From the codebook, we can also learn that the free-time network

in wave 2 was item 4 in question 10. This means that the data on this variable for class 1100 should be in the file “Wave 2\1000\1100\10\1100\_10\_4\_2h.csv” (notice the “\_2h” suffix, which refers to wave 2). In the third wave, this network is in the file “Wave 3\1000\1100\17\1100\_17\_4\_3h.csv”, and so on.

In sum, the network variables can be linked to the other two datasets by student IDs. Consult the student questionnaires and codebooks for network items and their coding. Be aware that the numbering of items may be different in different waves.

### *1.1.3. Class head teacher dataset*

The teacher questionnaire focused on three topics relevant to our study: the background of the class head teacher (e.g. age, subject), specific circumstances of the class (e.g. seating order), and their perceptions about students in the class (ranging from who is absent a lot to who the teacher sees as smart, popular or a clique leader in the community). The information collected in the teacher survey can be valuable in determining important contextual factors that may have affect the social dynamics of the classrooms in the sample. Furthermore, teacher perceptions about students provide an additional perspective on classroom performance and the school-related attitudes of students.

**Format.** This dataset is available as a single file, in text-based (csv) format. Although originally the respondents were classroom head teachers, the dataset has been reformatted so that each row refers to a student in a given head teacher’s classroom. For example, the year of birth variable for student 1101 in the teacher dataset refers to the year of birth of the head teacher of student 1101; similarly, the variable “smart” for student 1101 shows whether the head teacher of this student thinks that he/she is smart. We merged the data in such a format from each classroom into a single data table. As a result, variables in the teacher dataset can be linked to the other two datasets by student IDs. Consult the teacher questionnaires and codebooks for variables and coding.

## **3.2. Codebooks**

For each of the three datasets of the study, detailed codebooks are available. These provide basic information about every variable collected in the four questionnaire waves. Most importantly, the codebooks present the variable names by which each item in the questionnaires can be found in the individual, network, and teacher datasets. All three datasets have their separate codebook documents, along with summary tables for a more organized overview of the questionnaire items.

## **3.3. Questionnaires**

The original questionnaires (translated to English) are publicly available. These include four waves of student questionnaires and four waves of classroom head teacher questionnaires. The set of items used in the different waves are very similar, but a few items were added or removed between waves due to space constraints (see the codebooks for specifics). The student questionnaires are the source for the student individual and student network datasets. All information collected by the teacher questionnaires are stored in the head teacher dataset.

**Important!** When using any of the more specialized items from the questionnaires (e.g. the full set of perception network questions) in your own research, please remember to cite the study, as shown in section 1 above.

### **3.4. List of available datasets, codebooks, and questionnaires**

The following datasets, codebooks, and questionnaires are available along with this report:

- A. Datasets
  - 1. Student individual variables dataset
  - 2. Student social networks dataset
  - 3. Class head teacher dataset
- B. Codebooks
  - 1. Student individual variables codebook
  - 2. Student social networks codebook
  - 3. Class head teacher variables codebook
- C. Questionnaires
  - 1. Student questionnaires
  - 2. Class head teacher questionnaires

## 4. Publications from the dataset

### 4.1. Research articles in English

Boda, Zsófia and Néray, Bálint (2015). Inter-Ethnic Friendship and Negative Ties in Secondary School. *Social Networks*, 43: 57-72.

Boda, Zsófia. (2018). Social influence on observed race. *Sociological Science*, 5, 29-57.

Boda, Zsófia (2019). Friendship bias in ethnic categorization. *European Sociological Review*, 35(4), 567-581.

Boda, Zsófia, Néray, Bálint, and Snijders, Tom A.B. (2020). The Dynamics of Interethnic Friendships and Negative Ties in Secondary School: The Role of Peer-Perceived Ethnicity. *Social Psychology Quarterly*, 83(4), 342-362.

Estévez, J. L., Kisfalusi, D., & Takács, K. (2022). More than one's negative ties: The role of friends' antipathies in high school gossip. *Social Networks*, 70, 77-89.

Grow, André, Takács, Károly, and Pál, Judit (2016). Status Characteristics and Ability Attributions in Hungarian School Classes: An Exponential Random Graph Approach. *Social Psychology Quarterly*, May. DOI: 10.1177/0190272516643052.

Kisfalusi, Dorottya (2016): The Quality of Inter- and Intra-Ethnic Friendships among Roma and Non-Roma Students in Hungary. *Corvinus Journal of Sociology and Social Policy*, 7(1): 3-26.

Kisfalusi, Dorottya (2018): Ethnic Classification among Secondary School Teachers and Students in Hungary. *Corvinus Journal of Sociology and Social Policy*, 9(1): 35-54. DOI: 10.14267/CJSSP.2018.1.02

Kisfalusi, D.; Takács, K., and Pál, J. (2019). Gossip and Reputation in Adolescent Networks. In: Giardini, F. and Wittek, R.P.M. (eds.): *Oxford Handbook on Gossip and Reputation*, Oxford University Press, 359-379. Chapter DOI: 10.1093/oxfordhb/9780190494087.013.19

Kisfalusi, Dorottya, Pál, Judit, and Boda, Zsófia. (2020). Bullying and victimization among majority and minority students: The role of peers' ethnic perceptions. *Social Networks*, 60, 48-60.

Lőrincz, László (2016). Interethnic dating preferences of Roma and non-Roma secondary school students. *Journal of Ethnic and Migration Studies*. DOI: 10.1080/1369183X.2016.1160769

Pál, Judit, Stadtfeld, Christoph, Grow, André, and Takács, Károly (2016). Status Perceptions Matter: Understanding Disliking among Adolescents. *Journal of Research on Adolescence*, 26(4), 805-818. DOI: 10.1111/jora.12231

Stadtfeld, Christoph, Takács, Károly, and Vörös, András. (2020). The emergence and stability of groups in social networks. *Social Networks*, 60, 129-145.

Vörös, András. and Snijders, Tom A.B. (2017): Cluster analysis of multiplex networks: Defining composite network measures. *Social Networks*, 49: 93-112. DOI: 10.1016/j.socnet.2017.01.002

Vörös, András, Block, Per, and Boda, Zsófia (2019). Limits to inferring status from friendship relations. *Social Networks*, 59, 77-97.

#### 4.2. Research articles in Hungarian

Kisfalusi, D., & Takács, K. (2018). A pletyka és a reputáció összefüggései középiskolai osztályközösségekben. *Szociológiai Szemle*, 28(1), 83-104.

Lőrincz, László (2014). A magyar középiskolások párválasztási preferenciái: a roma – nem roma heterogenitás hatása. *Szociológiai Szemle*, 24(2).

Mandácskó, Eszter, Panyik, Barbara (2013). Osztályon belüli népszerűség mérései közötti különbség a dohányzási szokások esetében, In: Juhász Péter (szerk.), *Közgazdász diáktudós – Hatékonyság és imázs*, pp. 105-120. Budapesti Corvinus Egyetem és Alinea Kiadó, Budapest

#### 4.3. Books in Hungarian

Néray, Bálint, Vörös, András (eds.) (2013). Behálózott Iskolák: Iskolai hálózatkutatás egy kelet-magyarországi kisvárosban. Budapest: L'Harmattan Kiadó. (Széchenyi Füzetek XXI.) (ISBN:978-963-236-717-0).

#### 4.4. Defended PhD theses

Boda, Zsófia (2016). [\*Friendship Based on Race or Race Based on Friendship? The Co-Evolution of Friendships, Negative Ties and Ethnic Perceptions in Hungarian School Classes\*](#). University of Oxford, Nuffield College.

Kisfalusi, Dorottya (2016). [\*Interethnic Relations among Roma and Non-Roma Students in Hungary\*](#). Corvinus University of Budapest, Doctoral School of Sociology. DOI 10.14267/phd.2016022

Néray, Bálint (2017): [\*Relational Integration as The Analysis of Friendship, Negative Ties and Ethnic Identity Among Adolescents\*](#) Corvinus University of Budapest, Doctoral School of Sociology. DOI 10.14267/phd.2017009

Pál, Judit (2016). [\*Status and Negative Ties: A Longitudinal Network Study among Adolescents\*](#). Corvinus University of Budapest, Doctoral School of Sociology. DOI 10.14267/phd.2016010

Vörös, András (2016). [\*The Emergence of Multiple Status Systems in Adolescent Communities: A Multiplex Network Theory of Group Formation\*](#). University of Oxford, Nuffield College.

## 5. First results from the dataset

*(The following parts are excerpts from the project final report)*

The main aim of the research project was to describe and explain segregation of friendships within school classrooms. We assumed that ethnic integration is related to the formation of friendship, negative and romantic ties; and that the problem of ethnic integration in the classroom cannot be understood without paying attention to the interrelated dynamics of social networks, status, and performance. We also realized that our classification of pupils into ethnic groups should rely not only on self-identification but also on the perception of the classmates. We also presumed that status competition typically intensified the segregation of friendship ties and might also lead to the social exclusion of disadvantaged pupils, or alternatively, to the social exclusion of the best performing students. To test this hypothesis, we examined the relationship between the network structure of classes and the potential emergence of conflicting parallel status-hierarchies. Our panel dataset allowed us to analyse the aforementioned questions.

Our results suggested that analysing the evolution of peer relations could explain deeper the phenomena of social exclusion and ethnic segregation. We managed to show that the perception of classmate's ethnicity could influence who befriended with whom, or who hated who. However, our outcomes also pointed out that bullying and victimization occurred as often between as within ethnic groups. Moreover, we were able to demonstrate the importance of popularity, especially the positive and negative influence of perceived popular peer in regard of school achievement, disliking relations and deviant behaviour. Our results validated how stereotypes related to different attributions such as being female or Roma could impact the perception of student's performances. As our aim was to analyse the co-occurrence of negative and positive networks in order to better understand the structure of multiple networks, we analysed multidimensional networks together. Results highlighted that the closure of friendship triads might be occurred partly due to mechanisms that operate across different networks, i.e. friendship, liking, dislike and enmity.

In the following section, we present these results in a more detailed way categorized by research topics. The first summarizes research sub-projects about the interrelation of Roma ethnicity, romantic partner selection and negative links as diverse dimensions of segregation. The section offers summaries of papers' outcomes about academic achievement and school aspirations. The third section demonstrates outcomes of investigations about how status enhancement and status perception could result segregation through the analysis of conflicted ties. Finally, the fourth section presents conclusions of multiplex analysis of different networks.

All of the abovementioned results have been already presented on national and international conferences. Besides, several papers will be submitted in 2014 to international and Hungarian journals.

### 5.1. Ethnic segregation

The core question of our project pertains to the implications of social relations for ethnic segregation. There is a vast amount of literature documenting the harmful effects of school segregation on the scholastic performance and mobility chances of members of ethnic minorities (Kemény and Havas, 1996; Havas, Kemény and Liskó 2002; Havas and Liskó 2005; Kertesi and Kézdi, 2005; Hanushek and Wössmann, 2006; Brunello and Checci, 2007, Kézdi and Surányi,

2008). It has only been recently recognized, however, that even in integrated schools friendships ties are typically highly segregated (Moody, 2001), thus integrated education does not necessarily implies integration at the level of primordial social order (Coleman, 1990). For this reason, we examined the relationship between ethnic integration, on the one hand, and the structure of peer relations, such as romantic, friendship and negative relations, on the other hand.

First, we examined whether (ethnic) group composition have a direct effect even on (ethnic) preferences themselves. The first wave data of the research was used for analysis. Ethnicity (Roma and non-Roma) was based on self-assessment. Three levels of preferences were measured: (1) Perceived norm of dating with someone with Roma origin, (2) Individual attribution of physical attractiveness of each classmates, (3) Individual preference for dating each classmates. Multi-level regression models were used for analysis: two-level linear model in case of the norms, three-level logistic models (level 1: tie, level 2: individual, level 3: class) in case of the attractiveness and preference for dating. Increasing share of Roma students in the class was found associated with increasing acceptance of dating with Roma students. Additionally, increasing share of Roma students was associated with increased attributed physical attractiveness of Roma classmates. Both norms and perceived physical attractiveness affected preference for dating and share of Roma students did not have an independent effect after controlling these.

Second, we investigated several different aspects of inter-ethnic relationships, mainly focused on friendships and negative ties between secondary school students. Friendships and negative ties were modelled using cross-sectional Exponential Random Graph Models for 16 classrooms separately, and then individual models were summarized using meta-analysis. Our results suggested that non-Roma students tended to dislike those whom they perceived as Roma, regardless of their self-declared ethnicity. On the other hand, Roma students were likely to send friendship nominations towards their perceived Roma classmates if these also declared themselves as Roma, and negative nominations if these declared themselves as non-Roma. This supported our idea that different ethnicity concepts might influence friendships and negative ties in different ways, and that inconsistencies in someone's ethnic categorization might play an important role in social rejection. Students perceived as Roma but declaring themselves as non-Roma might seem to their Roma peers as “traitors” of their “original” ethnic group.

Third, we analysed whether bullying had been occurred more common between same-ethnic students or between students of different ethnic background. As minority students might be victims of bullying behaviour particularly frequently if their cultural norms differ from the majority culture, we expected that bullying occurs more often between than within ethnic groups, and Roma students become more often victimized than non-Roma students. We used Exponential Random Graph Models where we could control for both attribute variables (e. g. socio-economic status, gender) and network configurations (e. g. reciprocity, transitivity, star-like structures, etc.). Results showed that our data did not support the aforementioned hypotheses: in the majority of the analysed 18 classes, ethnicity of the students did not have significant effect on the prevalence of bullying and victimization; and bullying occurs as often between as within ethnic groups.

Fourth, we suggested that there was a connection between self-identification, classification by peers and relational integration with classmates. In our study, we therefore hypothesized that students who identified themselves as Roma but were classified as non-Roma by a significant proportion of their peers are more likely to have a higher amount of positive interethnic relations than those who are classified as Roma by their peers. We also assumed that as the rate of peers



classifying the respondent as Roma decreased, Roma students were more likely to change their self-reported identity towards non-Roma than towards Roma in time. Multilevel regression models of 35 classes supported our first hypothesis: after controlling for socio-demographic factors, we found significant negative correlation between the rate of positive interethnic relations and the proportion of classmates who classified the respondent as Roma. However, contrary to our second hypothesis, we found that as the rate of peers classifying the respondent as Roma increases, Roma students are more likely to change their self-reported identity towards non-Roma than towards Roma in time.

## **5.2. Academic performance**

The other core question of our project concerned the role of peer influence in academic achievement. Academic achievement has been examined by social scientists for decades either due to its proved importance in individual life chances and future career, or as the most important part of the mechanism of reproduction of elite. Moreover, peer influence is also essential to understand this phenomenon, based on accepted values and norms at the school class and its cliques, social rewards and sanctions used in the community and special network effects, such as the direct impact of friendships and adversarial ties. In this section, we briefly summarize results on the relationship between academic achievement and peer relations.

To understand how academic performance in group context, first, we designed a rational choice model, and we proposed an agent-based simulation for academic performance in group context. Performance was divided to two main components: the first one contained every, relatively constant effect which made an individual more or less capable to perform at school, and the second one was a special kind of reservation price function for these levels, which was dynamically influenced by the network and different network ties. In the model and the simulation, performance also affected the network, related to the concept of homophily and the utility of having well-performer friends. The teacher also played a role in the process as he/she either could or could not dynamically adopt to the given level of performance with his/her requirements. As a result, it was found that differences in the effects of certain parameters could cause different levels of network segregation, such as different average performance and different homogeneity in school achievement in the class.

Second, we examined empirically how popular peers and friends' educational plans effect students' individual school aspirations. For defining who was popular and unpopular, we used the definition of sociometric and peer-perceived popularity (LaFontana and Cillessen, 1999; Moody et al., 2011), while friendship was determined as "who is your best" friend. For measuring academic achievement, we used answers of question about school aspirations, as we did not have proper data on students' notes. Using multilevel and multinomial logistic regression models in a cross-sectional data, we found a positive correlation between friends and their own school aspirations. Results confirmed that popular peers have a significant and positive effect on students' school aspirations, especially sociometrically popular peers (e.g. well-liked) had bigger effect on individual school aspiration than perceived popular peers.

## **5.3. Status competition and peer relations**

One of the major innovations of the project was to place status competition in a network perspective that has been rarely done before (Gould, 2002). In order to gain a deeper understanding of group

behaviour and dynamics of adolescents, we examined the interrelations among status competition, social networks and deviant behaviour, such as smoking.

First, we explored the connection between individual personality traits and friendship nominations. Our model took into consideration the similarity of peers in each potential dyad by background dimensions. Due to homophily, it could be hypothesized that the probability of a friendship nomination was larger in more similar dyads. Besides, based on the principal role of homophily in tie formation (McPherson et al., 2001), we assumed that the impact of personality traits on friendship choice within similar dyads should be smaller if there were fewer similar individuals in the community to choose as friends. Therefore, in order to estimate personality effects in friendship choice, we needed to explore the individual background characteristics on which homophily were based in a community. Gender based homophily and personality effects were tested in two classrooms from the dataset. Results showed evidence for homophily in both groups. Hypotheses about the effects of personality were partly confirmed. In one of the classes, personality did not matter for girls in making friends. However, when ethnicity was accounted for, it appeared that personality traits have an impact on Roma-Roma and non-Roma-non-Roma friendships. This confirmed that certain personality traits might indeed have a role in friendship choices, although homophily or other superior network mechanisms might alternate or even suppress their effects.

Second, we analysed the relation the effect of high status on smoking behaviour. We suggested that popular students smoke more than unpopular ones. Furthermore we claimed that students with larger friendship networks have a significant impact on their friends' smoking habits. We tested these questions by a comparative cross sectional analysis on the first two waves of the dataset .To test these assumptions; we applied two-level multinomial logistic regression models. We controlled for gender, school performance, alcohol consumption and school type. Important to emphasize that a grammar school student smoke not as much as a vocational school student and in grammar school, smoking is not necessarily makes students popular. The results suggested that the size of friendship networks and popularity matters in smoking habits, however, various types of popularity affect it differently. Based on other research and articles we checked the two types of popularity such as sociometric and perceived popularity. Furthermore, to receive more precise results on the formation of smoking habits in high school classes we used two more approach of the perceived popularity such as direct and indirect respect. The results showed significant effect of the perceived popularity on smoking habits.

Third, we allocated with the theory of in-group favouritism (Berger et al., 1972). Our starting point was that experimental research suggests that the status of the social categories that individuals belong to affects the abilities that others attribute to them. We highlighted that in this context existing theories might not apply and that testing them is complicated by the fact that group members' attributions are not statistically independent. To deal with this problem, we suggested the use of exponential random graph models as a novel way for studying attributions in small groups. Our results suggested that status differences could affect ability attributions, even in enduring groups. In particular, Roma pupils might be confronted with intelligence attributions that are in line with societal level stereotypes during their face-to-face interactions in class contexts. In some classes, these attributions might only be encountered during interactions with members of the Hungarian majority, but not with members of the Roma minority. In other classes, the reproduction of stereotypes might be more ubiquitous and might occur during interactions with members of both the Hungarian majority and the Roma minority. In those classes it was most likely that the self-perception of Roma pupils negatively affected and social inequality is most strongly

reproduced. Our results also suggested that future research would benefit from taking potential non-independence of attribution processes in small group contexts into account.

Fourth, we argued that existing disliking network patterns and perceived status-related frustration can explain the existence and the formation of disliking ties. Using the Stochastic Actor-Oriented Model (SIENA) on three waves of data among 9<sup>th</sup> and 10<sup>th</sup> graders in ten classes, results showed that reciprocal disliking ties, those actors who were once indicated as “black sheep” were more likely to be nominated as disliked. Inconsistency between direct and indirect perceived status measures also lead to the formation of disliking ties, however when these ties were reciprocated, we did not see the same impact.

#### **5.4. Dynamics and co-evolution of positive and negative ties**

The last major innovation of the project was that unlike previous studies that concentrated only on friendship ties in describing status dynamics, we postulated that negative ties and romantic relations were also highly important for status competition and for interrelated problems of school segregation, social exclusion, and low performance. As positive friendship ties have proven to be highly relevant for social development, status and behavior of adolescents (Moody, 2001; Hallinan and Williams, 1989; Mouwe and Entwisle, 2006), negative relations such as disliking and bullying, might as well be important in many aspects. Disliking, bullying, and hate were examples of negative interpersonal relations that could also be captured in social network terms (Salmivalli et al., 2008). Network patterns of negative relations, especially handled together with friendship relations could help to explain core problems of segregation, status competition, exclusion and conflict in school classes. Determining the key mechanisms about the structural constraints on negative ties and their dynamics and constructing a theory of negative ties based on these fundamentals would be a major theoretical contribution of this project. So far, we analyzed how dynamics of negative ties related to positive ties. We also examined the co-evolution of friendship and linking relations. Last, but not at least we propose new methodological innovation, namely dimension reduction methods to identify the interdependence of different relational dimensions.

First, we managed to show that transitive closure in friendship networks could potentially be regarded, at least in part, as a result of social influence in friendship selection. More generally, how an individual's friends felt about a certain member of the group might increase or decrease the desirability of making friends with him or her. Therefore, it was hypothesized that the more friends with positive attitudes towards a certain peer will increase the probability of the individual befriending him or her. This relationship was expected to hold over and above the single-network triadic closure mechanism in friendship networks which is therefore controlled for in the analysis. For the analysis, the Stochastic Actor-Oriented Model (SIENA) was applied, which made controlling for confounding actor-level, dyadic and triadic effects possible. Results confirmed the presence of multiple network triadic closure effects in the examined classrooms. However, the observed patterns differed from group to group: in one example, students tended to like the friends of their friends, while in another peers liked by friends were also more attractive. The differences might be due to classroom-level characteristics, such as structural features of the friendship network, the stage of the evolution of relationships in the classroom, gender composition, etc., which shall be investigated in future work. The results highlighted that the closure of friendship triads may be partly due to mechanisms that operate across multiple networks, i.e. friendship, liking, dislike and enmity.

Second, we investigated the role of negative ties in single and cross-network transitive closure. Single- and cross-network transitive closure in friendship and liking networks of small, face-to-face groups might be considered, in part, as a result of a status maintenance mechanism. In line with this logic, it can be hypothesized that some of the negative ties observed in a group will emerge following the creation of a positive tie: an individual may come to dislike either a former friend who 'turned away' from him/her or the one who 'stole' this friend's attention. We tested these predictions using the Stochastic Actor-Oriented Model (SIENA), which makes controlling for confounding actor-level, dyadic and triadic effects possible. Results suggest that there may be two parallel tendencies for forming relations in triads: there is a tendency to make friends with friends of friends, but it is also more likely than random to dislike them. This result pointed out that staying neutral in a group might sometimes be difficult, and either positive ties or enmity were likely to emerge.

Although, we did not specify in the research plan, however we went to the direction of understating group formation through the analysis of co-evolution of friendship and liking relations. As we differentiated between these two emotional ties in our questionnaire, we aimed to understand how weak ties –in this case - could influence the formation of friendship ties, which is the best measure to catch up social influence as social processes. We hypothesized that being members of school communities, adolescents and their social behaviour were not only influenced by their close friends, but also by the larger peer groups that surrounded them. Therefore, in order to understand the evolution of friendship networks in schools, we had to take into account the impact of weaker positive relationships between students. Using the first two waves of the dataset, we distinguished between two types of positive ties: friendship and liking. The Stochastic Actor-Oriented Model (SIENA) allowed us to analyse the joint evolution of the two networks by modelling the probabilities of two threshold crossings: from no tie to liking and from liking to friendship. We explored the common trajectories of network dynamics in the analysed communities. The results indicated that in some classrooms there were tendencies against liking friends of friends: becoming friends or not having a positive relation with them was more probable than this option. This highlighted those affective relations within groups of students tended to become strong and positive (friendship) or deteriorate towards negative ties (neutrality, dislike), while simply liking group members was not a state of equilibrium.

Furthermore, we tried to understand how friendship groups could be identified through shared role attributions. Through the concept of multiplex structural equivalence, we explored the multivariate similarities between class members with regards to which peers they nominate as occupying certain social roles in the community. The structure of positive affections (friendship and liking ties) within the identified groups was then explored. Based on the density and connectedness of within-group networks, the analysis reveals different types of student groups which may capture certain forms of friendship groups. We identified the friendship clique, the liking clique, the friendship group, and the friendship circle as distinct subgroup types. Further analysis should focus on testing the role of these formations in the evolution of classroom communities.

The new methodological innovation of the project suggested that multiplex network data, information on several network layers in a given group, provides researchers an opportunity to study social processes in depth, and to answer questions about the interdependence of different relational dimensions. Although some multivariate network methods (e.g. ERGM, SIENA) made it possible to jointly analyse multiple network dimensions, modelling becomes impossibly complex when the investigation focuses on more than a few, say more than three or four, network layers. In

these cases, dimension reduction methods might be applied to obtain a manageable set of variables. Drawing on existing statistical methods and measures, we proposed a strategy to reduce the dimensions of multiplex network data measured in multiple groups. We achieved this by clustering the networks based on their pairwise similarities and constructing composite network measures as combinations of the items in each resulting cluster. The procedure was demonstrated on a random subsample of 18 classrooms. Starting from 24 perception networks, we arrived at a solution of three clusters which we labelled as positive traits, negative traits and social role attributions. Though our procedure did not rely on an explicit statistical model, it presented a useful and flexible approach for dimension reduction in multiplex networks. Following such an approach may aid researchers in defining complex network measures and may also provide some theoretical insights into multiplex social mechanisms.

## 5.5. References

- Berger, J., Bernard, P. Cohen, & Zelditch, Jr. M. (1972). Status Characteristics and Social Interaction. *American Sociological Review*, 37: 241-255.
- Boda, Z., & Néray, B. (2014). Inter-Ethnic Friendship and Negative Ties in Secondary School.
- Brunello, G., & Checci, D. (2007). Does School Tracking Affect Equality of Opportunity? New International Evidence, *Economic Policy*, 22 (52), 781–861
- Coleman, J. S., & al. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Dept. of Health, Education, and Welfare, Office of Education.
- Coleman, J. S. (1990). *The Foundations of Social Theory*. Cambridge (Mass.), The Belknap Press.
- Gould, R. V. (2002). The Origins of Status Hierarchies: A Formal Theory and Empirical Test. *American Journal of Sociology* 107: 1143-1178.
- Hallinan, M. T. & Williams, R. A. (1989). The Stability of Students' Interracial Friendships. *American Sociological Review* 52: 653–64.
- Havas, G, Kemény I. & Liskó I. (2002): *Roma gyerekek az általános iskolában*. Új Mandátum kiadó, Budapest.
- Havas, G., & Liskó, I. (2005). Szegregáció a roma tanulók általános iskolai oktatásában. Felsőoktatási Kutatóintézet, Kutatás közben, 266. kötet.
- Kemény I. & Havas G. (1996). Cigánynak lenni. In: Andorka R., Kolosi T. & Vukovich Gy. (eds.): *Társadalmi Riport 1996*, TÁRKI-Századvég, Budapest.
- Kertesi G. & Kézdi G. (2005). Általános iskolai szegregáció I-II. *Közgazdasági Szemle*, 52(4, 5): 317-356, 462-480.
- Kézdi, G., & Surányi, E. (2009). *A Successful School Integration Program. Roma Education Fund*. Working Paper.
- Lafontana, K. M., & Cillessen, A. H. N. (1999). Children's Interpersonal Perceptions as a Function of Sociometric and Peer-Perceived Popularity. *The Journal of Genetic Psychology*, 160(2), 225–242.
- Lusher, D., Koskinen, J., & Robbins, G. (2013). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press
- McPherson, M., Smith-Lovin, L., & Cook, J.M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27: 415-444.

- Moody, J. (2001). Race, School Integration, and Friendship Segregation in America. *American Journal of Sociology*, 107(3), 679–716.
- Moody, J., Brynildsen, W. D., Osgood, D. W., Feinberg, M. E., & Gest, S. (2011). Popularity Trajectories and Substance Use in early Adolescence. *Social Networks*, 33(2), 101–112.
- Mouwe, T., & Entwisle, B. (2006). Residential segregation and interracial friendship in school, *American Journal of Sociology*, 112 (2): 394-441
- Ripley, R. M., Snijders, T. A., Boda, Z., Vörös, A., & Preciado, P. (2013). *Manual for RSIENA*. University of Oxford: Department of Statistics, Nuffield College.
- Salmivalli, C., Antti, K., & Poskiparta, E. (2008). Development, evaluation, and diffusion of a national anti-bullying program, chapter submitted for publication in the Handbook of youth prevention science, edited by B. Doll et al.
- Snijders, T. A. B., van de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1), 44–60.

## **6. Research ethics and data security**

We aimed to ensure that the data collection procedures applied in this project comply with research ethics standards in the social sciences. To this end, we acquired informed consent from all parties involved in the data collection (schools, parents, students). We implemented stringent measures to ensure the anonymity of our respondents and the privacy of their responses, both while in the field and when processing and storing digitized data. Our procedures as described here were approved by ethics boards and educational institutions (see below for details).

### **6.1. Informed consent from schools, parents, and students**

In order to adhere to research ethics principles and to ensure high participation and low drop-out rates over the three-year study period, we decided to follow a rigorous policy of open and honest communication with all parties involved in the data collection process: with schools, parents, and students.

First, as part of the selection of the sample, head teachers of the chosen secondary schools were thoroughly informed about the planned research several months prior to the start of the data collection. Once they expressed willingness to participate, in agreement with the teacher body, more information about the actual data collection procedures was communicated. In addition, up-to-date questionnaires were sent to schools for approval before every survey wave. School principals were ensured that they may partially or completely withdraw their school from the study at any time.

Second, with the approval and assistance of the school head teachers and teachers, parents were contacted several weeks before each survey wave. They received an information brochure explaining the aims and methods of the study and the involvement it would require from participating students. They could also view and comment on the latest questionnaire at the school head teacher's office. Further, parents were encouraged to ask for additional information about the study from the teachers of their children's classroom, and to get in touch with the researchers by e-mail or phone at any time in the survey period.

Before every survey wave, parents received a consent form along with information brochures. If they did not wish their child to participate in the data collection, they had to return the signed form to the head teacher (opt out). Families with more than one child attending the sample schools had to declare consent for each student separately. Parental consent was necessary since most of the students in the sample were underage (<18) at the time. The consent procedure preceded every survey wave in the study.

Third, while in the field at the four survey waves students were informed by the researchers about the aims of the study, how their responses are anonymized and handled, and how their privacy is protected. In addition, it was clearly communicated that they can deny answering any questionnaire item or they can choose to opt out from the given survey wave or the entire project if they wish so, even if their parents granted consent.

## **6.2. Anonymity of respondents**

Several steps were made in the study preparation stage, in the field, and during data processing to ensure the anonymity of students and the privacy of their responses. Participants were randomly assigned 4-digit identification numbers (IDs) prior to the first wave of the data collection. Since the study was both relational and panel, it was crucial for the researchers to be able to link students with their IDs, by means of a key containing both student names and IDs, throughout the project. As a basic rule, only the researchers had access to the entire key list, and it was deleted after the final cleaning of the dataset.

In the field, students received (and were reminded of) their IDs classroom by classroom in each data collection wave. Upon getting the empty questionnaires, students were asked to write their own ID on them, but not their names. Although the questionnaires did contain the ID of the respondent (cf. network panel survey), the completed questionnaires were personally collected by the researchers and put into an envelope in front of the students. The classroom envelopes were sealed after the last questionnaire was collected. They were only opened again in the research offices when they were prepared for recording, cleaning, and processing. In every case, researchers and trained research assistants were present in the classrooms during data collection to answer questions and ensure that respondent anonymity and privacy were respected (both by fellow students and teachers).

## **6.3. Anonymized feedback to teachers**

As part of maintaining a good working relation with the sample schools and their teachers, we provided them high-level feedback in the form of information booklets and short overview presentations held by the PI at each institution. There were two feedback rounds during the project: one after wave 1 (first summary) and one after wave 4 (project overview). The booklets and presentations contained only aggregated information about classrooms to ensure that teachers would not be able to identify individual students. We shared information that was least likely to affect teaching practices in the studied classrooms (e.g. classroom composition in student background variables, an overview of friendship networks). The information packages were tailored to each participating school, that is, teachers in one school did not learn about results from other participating schools.

## **6.4. Data security**

All digitized data was stored using secure cloud storage services. Only the researchers participating in the project have access to the raw, uncleaned data files and other sensitive materials (e.g. contact information for schools). The original paper-based questionnaires are being stored to date in a secure office at the Hungarian Academy of Sciences. Only the project team members have access to the physical data.

## **6.5. Data access**

During the entire data processing and cleaning procedure, up to the point when the database was fully anonymized by the disposal of the student-ID keys, only researchers participating in the project had access to the collected data. Partial exceptions were made to allow already cleaned and anonymized parts of the dataset to be used for research by other academics, upon request. With the



data cleaning and documentation tasks complete, we are happy to make all materials openly available, according to our obligations to the funding agency (please cite when presenting/publishing the data or results from it as described in section 1).

## **6.6. Ethics board approvals**

The data collection procedures were approved by the following institutional ethics boards:

- Institute of Sociology and Social Policy, Corvinus University of Budapest;
- Social Sciences & Humanities Inter-divisional Research Ethics Committee, University of Oxford (approval ref. no. SSD/CUREC1A/12-130).

## **7. Data collection procedures**

### **7.1. Study planning**

The project was preceded by a long preparation and planning phase. This involved a pilot study in two Hungarian high schools, which was conducted in the course of the 2009-2010 academic year. The pilot, which was funded separately, provided a good testing ground for the questionnaire items that were included in the “Wired into Each other” surveys. Besides that, the core team of researchers acquired invaluable experience in communicating with schools and respondents and handling various issues that may arise in the field.

The preparations for the data collection started in January 2010 (about 10 months before the first survey wave). Due to budget constraints, the selection of a representative sample of schools in any sense was not a viable option. Instead, we aimed at maximizing the variation of classrooms in ethnic composition, since ethnic integration was one of the main research topics of the project. Sampling based on ethnic background was done using regional and school-level statistics, where available. Other aims of the sample selection were to ensure that all high-school training programmes (vocational, technical, grammar) appeared in the sample, and to have a set of schools which were geographically clustered to optimize data collection costs. More descriptive information about the sample is provided in section 2 above.

The questionnaires were compiled in the first half of 2016 based on the experience collected in the pilot study. The research group at the time was already of a considerable size (around 10 people), and with a large variety of research interests within the broad topic of the project. This resulted in long structured debates over what should be included in the questionnaires and how exactly items should be formulated. Beside the substantive arguments, we also did pretests with a few children to evaluate how much the questions were intelligible and meaningful for adolescents in the targeted age group. Further information about the questionnaires is provided in section 3 above

The research plans and questionnaires were approved by the funding body before the actual data collection started. Further, we acquired two additional ethics approvals during the project period, one from the University of Oxford and one from the Corvinus University of Budapest – these found that our ethics procedures met the highest standards for social science data collections.

### **7.2. Data collection: preparations**

After the selection of the preliminary sample, the compilation of the questionnaires, and the ethical approval of the funding body, the research group made contact with the head teachers of the selected schools. In case the head teachers were positive about their school’s participation in the study, we thoroughly informed the teacher body and the parents of selected classrooms about the nature of the study. This step was repeated before every survey wave. Both the schools and individual students had the opportunity to opt out, from a single wave or from the entire study. Before each survey wave, the core research team assembled field guides and trained a selected set of interviewers to conduct the data collection.

### **7.3. Data collection: in the field**

In the field, at least two trained researchers or interviewers working in the project were present in each surveyed classroom during the data collection. Before starting to fill out the questionnaires, students were informed that their answers would be kept strictly confidential and that they can choose to refuse answering any or all of the questions, regardless of consent from their parents. Researchers also ensured that students did not share their answers with anyone else by any chance. As soon as a student finished answering, his or her questionnaire was placed in an envelope, without anyone looking into it (the cover pages did not contain any information about the identity of students). After all questionnaires had been handed in, the envelope was closed by the researchers present, and was not opened in the presence of students, teachers or anyone outside of the research group.

### **7.4. Data processing and cleaning**

After the collection of the questionnaire data in each wave, at least one researcher reviewed every single questionnaire in order to clarify unclear responses based on previously set rules (e.g. two answers for a multiple choice question should be coded as invalid/missing data), and to assess the general quality of the questionnaire data (e.g. looking for visual patterns in responses, which may suggest random or dishonest responses).

Following the first round of hand-coding and quality assessment, the answers were digitized by a professional data recording team in the first wave and by an automatic scanning procedure in later waves. In all cases, the digitized data was submitted to a second round of rigorous tests: a comparison with the paper-based questionnaires to see how much error was introduced by the recoding; any errors discovered were corrected in the dataset. This process was followed by further usual data cleaning and recoding steps.

## 8. The study budget

Cost type (in Euros)		2010	2011	2012	2013	Total
1	Personnel costs	3,195	3,109	2,996	2,731	12,031
2	Indirect personnel costs (taxes & contributions by the employer)	501	569	549	505	2,125
3	Material costs	3,947	5,548	3,728	3,597	16,820
4	Investments (equipment)	0	0	0	0	0
<b>5</b>	<b>Total costs (1+2+3+4)</b>	<b>7,643</b>	<b>9,226</b>	<b>7,273</b>	<b>6,833</b>	<b>30,976</b>
	<i>Exchange rate (HUF/EUR)*</i>	<i>275.41</i>	<i>279.21</i>	<i>289.42</i>	<i>296.92</i>	-

*\* based on the yearly average rates of the Hungarian Central Bank*

Cost type (in Hungarian Forints)		2010	2011	2012	2013	Total
1	Personnel costs	880,000	868,000	867,000	811,000	3,426,000
2	Indirect personnel costs (taxes & contributions by the employer)	138,000	159,000	159,000	150,000	606,000
3	Material costs	1,087,000	1,549,000	1,079,000	1,068,000	4,783,000
4	Investments (equipment)	0	0	0	0	0
<b>5</b>	<b>Total costs (1+2+3+4)</b>	<b>2,105,000</b>	<b>2,576,000</b>	<b>2,105,000</b>	<b>2,029,000</b>	<b>8,815,000</b>

## Appendix

First wave: Ethnic distribution of schools and classes based on self-reported ethnic identity

School id	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)
1000	99.3	0.0	0.0	0.7
2000	43.8	36.9	19.3	0.0
3000	67.8	18.1	14.1	0.0
4000	89.3	1.7	7.4	1.7
5000	61.3	19.7	19.0	0.0
6000	95.5	0.9	0.0	3.6
7000	69.4	10.9	15.3	4.4
Total	71.2	15.3	12.1	1.3

Class id	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)
1100	100.0	0.0	0.0	0.0
1200	100.0	0.0	0.0	0.0
1300	100.0	0.0	0.0	0.0
1400	100.0	0.0	0.0	0.0
1500	95.7	0.0	0.0	4.3
2000	21.7	56.5	21.7	0.0
2100	38.1	28.6	33.3	0.0
2200	48.3	41.4	10.3	0.0
2300	50.0	31.3	18.8	0.0
2400	76.0	20.0	4.0	0.0
2500	37.5	37.5	25.0	0.0
2600	28.6	52.4	19.0	0.0
2700	31.6	42.1	26.3	0.0
2800	39.3	32.1	28.6	0.0
2900	59.3	29.6	11.1	0.0
3100	96.3	0.0	3.7	0.0
3200	87.9	6.1	6.1	0.0
3300	79.3	3.4	17.2	0.0
3400	87.9	0.0	12.1	0.0
3500	24.0	52.0	24.0	0.0
3600	60.9	21.7	17.4	0.0
3700	27.6	51.7	20.7	0.0
4100	93.8	0.0	6.3	0.0
4200	97.1	0.0	2.9	0.0
4300	90.6	0.0	6.3	3.1
4400	78.9	5.3	13.2	2.6
5100	69.2	19.2	11.5	0.0
5200	42.9	38.1	19.0	0.0
5300	70.0	10.0	20.0	0.0
5400	67.9	10.7	21.4	0.0
5500	92.3	0.0	7.7	0.0
5600	26.9	38.5	34.6	0.0
6100	96.3	3.7	0.0	0.0
6200	96.9	0.0	0.0	3.1
6300	95.5	0.0	0.0	4.5
6400	93.5	0.0	0.0	6.5
7100	70.4	11.1	14.8	3.7
7200	95.0	0.0	5.0	0.0
7300	70.4	18.5	7.4	3.7
7400	78.3	8.7	4.3	8.7
7500	68.4	15.8	10.5	5.3
7600	50.0	5.0	40.0	5.0
7700	57.1	14.3	23.8	4.8
7800	65.4	11.5	19.2	3.8

Second wave: Ethnic distribution of schools and classes based on self-reported ethnic identity

School id	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)
1000	100.0	0.0	0.0	0.0
2000	43.1	30.2	25.3	1.3
3000	70.2	15.7	13.6	0.5
4000	90.0	1.7	7.5	0.8
5000	69.1	16.4	11.8	2.7
6000	92.1	0.8	0.8	6.3
7000	75.7	6.8	14.1	3.4
Total	74.2	11.9	11.9	2

Class id	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)
1100	100.0	0.0	0.0	0.0
1200	100.0	0.0	0.0	0.0
1300	100.0	0.0	0.0	0.0
1400	100.0	0.0	0.0	0.0
1500	100.0	0.0	0.0	0.0
2000	37.5	43.8	12.5	6.3
2100	39.1	34.8	26.1	0.0
2200	50.0	37.5	12.5	0.0
2300	50.0	31.8	18.2	0.0
2400	67.9	14.3	17.9	0.0
2500	52.4	33.3	14.3	0.0
2600	27.3	36.4	36.4	0.0
2700	20.0	40.0	40.0	0.0
2800	37.0	25.9	37.0	0.0
2900	40.9	13.6	36.4	9.1
3100	96.4	0.0	3.6	0.0
3200	90.6	3.1	6.3	0.0
3300	68.6	14.3	14.3	2.9
3400	88.2	0.0	11.8	0.0
3500	27.8	38.9	33.3	0.0
3600	55.0	25.0	20.0	0.0
3700	33.3	50.0	16.7	0.0
4100	100.0	0.0	0.0	0.0
4200	97.1	0.0	2.9	0.0
4300	96.9	0.0	3.1	0.0
4400	73.0	5.4	18.9	2.7
5100	85.0	5.0	10.0	0.0
5200	46.7	40.0	13.3	0.0
5300	-	-	-	-
5400	77.4	6.5	9.7	6.5
5500	89.3	3.6	3.6	3.6
5600	18.8	50.0	31.3	0.0
6100	93.3	3.3	3.3	0.0
6200	91.2	0.0	0.0	8.8
6300	96.3	0.0	0.0	3.7
6400	88.6	0.0	0.0	11.4
7100	78.3	4.3	13.0	4.3
7200	95.2	4.8	0.0	0.0
7300	83.3	4.2	12.5	0.0
7400	75.0	8.3	8.3	8.3
7500	70.6	11.8	17.6	0.0
7600	68.2	9.1	18.2	4.5
7700	65.0	15.0	15.0	5.0
7800	69.2	0.0	26.9	3.8



Third wave: Ethnic distribution of schools and classes based on self-reported ethnic identity

School id	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)
1000	100.0	0.0	0.0	0.0
2000	48.9	27.3	23.3	0.6
3000	77.9	9.2	11.7	1.2
4000	92.4	1.0	5.7	1.0
5000	82.0	4.9	12.3	0.8
6000	95.9	0.0	0.8	3.3
7000	78.4	6.5	13.1	2.0
Total	80.6	8.0	10.2	1.2

Class id	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)
1100	100.0	0.0	0.0	0.0
1200	100.0	0.0	0.0	0.0
1300	100.0	0.0	0.0	0.0
1400	100.0	0.0	0.0	0.0
1500	100.0	0.0	0.0	0.0
2000	13.3	60.0	26.7	0.0
2100	47.8	30.4	21.7	0.0
2200	69.2	23.1	7.7	0.0
2300	64.7	23.5	11.8	0.0
2400	82.6	8.7	8.7	0.0
2500	43.8	18.8	37.5	0.0
2600	33.3	33.3	33.3	0.0
2700	17.6	35.3	47.1	0.0
2800	57.1	23.8	19.0	0.0
2900	43.8	25.0	25.0	6.3
3100	92.6	0.0	3.7	3.7
3200	93.8	6.3	0.0	0.0
3300	92.0	0.0	8.0	0.0
3400	93.8	0.0	6.3	0.0
3600	52.2	21.7	26.1	0.0
3700	29.2	33.3	33.3	4.2
4100	93.8	0.0	6.3	0.0
4200	100.0	0.0	0.0	0.0
4300	89.3	0.0	7.1	3.6
4400	84.6	3.8	11.5	0.0
5100	74.1	11.1	14.8	0.0
5300	91.7	0.0	8.3	0.0
5400	88.2	2.9	8.8	0.0
5500	90.9	3.0	3.0	3.0
5600	56.3	6.3	37.5	0.0
6100	93.8	0.0	3.1	3.1
6200	93.5	0.0	0.0	6.5
6300	96.4	0.0	0.0	3.6
6400	100.0	0.0	0.0	0.0
7100	84.6	3.8	7.7	3.8
7300	80.0	4.0	12.0	4.0
7400	84.6	7.7	7.7	0.0
7500	85.0	5.0	10.0	0.0
7600	81.0	9.5	4.8	4.8
7700	54.5	9.1	36.4	0.0
7800	76.9	7.7	15.4	0.0

Fourth wave: Ethnic distribution of schools and classes based on self-reported ethnic identity

School id	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)
1000	100.0	0.0	0.0	0.0
2000	63.2	18.7	17.0	1.2
3000	90.5	2.1	7.4	0.0
4000	95.7	0.0	4.3	0.0
5000	82.1	4.5	11.6	1.8
6000	95.2	1.0	1.0	2.9
7000	85.1	0.0	11.9	3.0
Total	86.1	5.0	7.8	1.1

Class id	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)
1100	100.0	0.0	0.0	0.0
1200	100.0	0.0	0.0	0.0
1300	100.0	0.0	0.0	0.0
1400	100.0	0.0	0.0	0.0
1500	100.0	0.0	0.0	0.0
2000	42.9	42.9	14.3	0.0
2100	58.8	23.5	17.6	0.0
2200	61.9	9.5	28.6	0.0
2400	86.4	0.0	13.6	0.0
3100	95.5	0.0	4.5	0.0
3200	92.6	7.4	0.0	0.0
3300	87.0	0.0	13.0	0.0
3400	87.0	0.0	13.0	0.0
4100	92.9	0.0	7.1	0.0
4200	97.1	0.0	2.9	0.0
4300	100.0	0.0	0.0	0.0
4400	91.3	0.0	8.7	0.0
5100	92.0	0.0	8.0	0.0
5200	70.4	7.4	18.5	3.7
5400	94.7	0.0	5.3	0.0
5500	88.5	3.8	3.8	3.8
5600	60.0	13.3	26.7	0.0
6100	82.6	4.3	4.3	8.7
6200	100.0	0.0	0.0	0.0
6300	95.8	0.0	0.0	4.2
6400	100.0	0.0	0.0	0.0
7100	86.7	0.0	6.7	6.7
7300	90.0	0.0	5.0	5.0
7400	84.2	0.0	15.8	0.0
7800	76.9	0.0	23.1	0.0
9100	18.8	50.0	25.0	6.3
9200	57.1	14.3	28.6	0.0
9300	50.0	12.5	37.5	0.0
9400	81.5	7.4	11.1	0.0
9500	75.0	15.0	10.0	0.0
9600	0.0	50.0	0.0	50.0
9700	72.7	27.3	0.0	0.0
9800	53.8	30.8	15.4	0.0

First wave: Ethnic, gender and age distribution by schools

School id	% of boys	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)	Median year of birth
1000	29.5	99.3	0.0	0.0	0.7	1995
2000	36.5	43.8	36.9	19.3	0.0	1995
3000	58	67.8	18.1	14.1	0.0	1995
4000	37.1	89.3	1.7	7.4	1.7	1995
5000	50.3	61.3	19.7	19.0	0.0	1995
6000	28.7	95.5	0.9	0.0	3.6	1995
7000	27.8	69.4	10.9	15.3	4.4	1995
Total	40.5	71.2	15.3	12.1	1.3	1995

Second wave: Ethnic, gender and age distribution by schools

School id	% of boys	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)	Median year of birth
1000	28.6	100.0	0.0	0.0	0.0	1995
2000	38.1	43.1	30.2	25.3	1.3	1995
3000	58.1	70.2	15.7	13.6	0.5	1995
4000	36.6	90.0	1.7	7.5	0.8	1995
5000	54.1	69.1	16.4	11.8	2.7	1995
6000	28.7	92.1	0.8	0.8	6.3	1995
7000	28.7	75.7	6.8	14.1	3.4	1995
Total	40.5	74.2	11.9	11.9	2	1995

Third wave: Ethnic, gender and age distribution by schools

School id	% of boys	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)	Median year of birth
1000	30.5	100.0	0.0	0.0	0.0	1995
2000	39	48.9	27.3	23.3	0.6	1995
3000	56.4	77.9	9.2	11.7	1.2	1995
4000	38.7	92.4	1.0	5.7	1.0	1995
5000	50	82.0	4.9	12.3	0.8	1995
6000	28.6	95.9	0.0	0.8	3.3	1995
7000	27.2	78.4	6.5	13.1	2.0	1995
Total	40.6	80.6	8.0	10.2	1.2	1995

Fourth wave: Ethnic, gender and age distribution by schools

School id	% of boys	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)	Median year of birth
1000	30.9	100.0	0.0	0.0	0.0	1995
2000	40.2	63.2	18.7	17.0	1.2	1995
3000	54.5	90.5	2.1	7.4	0.0	1995
4000	37.1	95.7	0.0	4.3	0.0	1995
5000	60	82.1	4.5	11.6	1.8	1995
6000	29.8	95.2	1.0	1.0	2.9	1995
7000	25	85.1	0.0	11.9	3.0	1995
Total	40.5	86.1	5.0	7.8	1.1	1995

First wave: Ethnic, gender and age distribution by classes

Class id	% of boys	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)	Median year of birth
1100	34.4	100.0	0.0	0.0	0.0	1995
1200	32.4	100.0	0.0	0.0	0.0	1995
1300	22.2	100.0	0.0	0.0	0.0	1995
1400	34.3	100.0	0.0	0.0	0.0	1995
1500	24.1	95.7	0.0	0.0	4.3	1995
2000	48.3	21.7	56.5	21.7	0.0	1994
2100	14.8	38.1	28.6	33.3	0.0	1995
2200	29	48.3	41.4	10.3	0.0	1995
2300	46.7	50.0	31.3	18.8	0.0	1995
2400	37.9	76.0	20.0	4.0	0.0	1995
2500	51.4	37.5	37.5	25.0	0.0	1995
2600	37.1	28.6	52.4	19.0	0.0	1995
2700	25.7	31.6	42.1	26.3	0.0	1995
2800	34.4	39.3	32.1	28.6	0.0	1995
2900	37.5	59.3	29.6	11.1	0.0	1995
3100	51.7	96.3	0.0	3.7	0.0	1995
3200	23.5	87.9	6.1	6.1	0.0	1995
3300	94.6	79.3	3.4	17.2	0.0	1995
3400	61.1	87.9	0.0	12.1	0.0	1995
3500	59.5	24.0	52.0	24.0	0.0	1995
3600	100	60.9	21.7	17.4	0.0	1995
3700	13.5	27.6	51.7	20.7	0.0	1995
4100	47.1	93.8	0.0	6.3	0.0	1995
4200	45.7	97.1	0.0	2.9	0.0	1995
4300	41.2	90.6	0.0	6.3	3.1	1995
4400	21.1	78.9	5.3	13.2	2.6	1995
5100	21.2	69.2	19.2	11.5	0.0	1995
5200	100	42.9	38.1	19.0	0.0	1995
5300	100	70.0	10.0	20.0	0.0	1994
5400	48.6	67.9	10.7	21.4	0.0	1995
5500	60	92.3	0.0	7.7	0.0	1995
5600	0	26.9	38.5	34.6	0.0	1994
6100	18.2	96.3	3.7	0.0	0.0	1995
6200	32.4	96.9	0.0	0.0	3.1	1995
6300	36.4	95.5	0.0	0.0	4.5	1995
6400	27.8	93.5	0.0	0.0	6.5	1995
7100	25	70.4	11.1	14.8	3.7	1995
7200	38.7	95.0	0.0	5.0	0.0	1995
7300	35.5	70.4	18.5	7.4	3.7	1995
7400	18.2	78.3	8.7	4.3	8.7	1995
7500	33.3	68.4	15.8	10.5	5.3	1995
7600	15.2	50.0	5.0	40.0	5.0	1994
7700	25.8	57.1	14.3	23.8	4.8	1994
7800	31.4	65.4	11.5	19.2	3.8	1995

Second wave: Ethnic, gender and age distribution by classes

Class id	% of boys	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)	Median year of birth
1100	35.5	100.0	0.0	0.0	0.0	1995
1200	32.4	100.0	0.0	0.0	0.0	1995
1300	20	100.0	0.0	0.0	0.0	1995
1400	30.3	100.0	0.0	0.0	0.0	1995
1500	25	100.0	0.0	0.0	0.0	1995
2000	48.1	37.5	43.8	12.5	6.3	1994
2100	12	39.1	34.8	26.1	0.0	1995
2200	26.9	50.0	37.5	12.5	0.0	1995
2300	50	50.0	31.8	18.2	0.0	1995
2400	38.7	67.9	14.3	17.9	0.0	1995
2500	55.6	52.4	33.3	14.3	0.0	1995
2600	38.2	27.3	36.4	36.4	0.0	1994
2700	28.6	20.0	40.0	40.0	0.0	1995
2800	36.7	37.0	25.9	37.0	0.0	1995
2900	40	40.9	13.6	36.4	9.1	1995
3100	51.7	96.4	0.0	3.6	0.0	1995
3200	22.9	90.6	3.1	6.3	0.0	1995
3300	94.4	68.6	14.3	14.3	2.9	1995
3400	61.1	88.2	0.0	11.8	0.0	1995
3500	60.5	27.8	38.9	33.3	0.0	1995
3600	100	55.0	25.0	20.0	0.0	1995
3700	17.9	33.3	50.0	16.7	0.0	1995
4100	47.1	100.0	0.0	0.0	0.0	1995
4200	44.4	97.1	0.0	2.9	0.0	1995
4300	39.4	96.9	0.0	3.1	0.0	1995
4400	21.6	73.0	5.4	18.9	2.7	1995
5100	23.5	85.0	5.0	10.0	0.0	1995
5200	100	46.7	40.0	13.3	0.0	1995
5300	100	NA	NA	NA	NA	NA
5400	52.6	77.4	6.5	9.7	6.5	1995
5500	65.7	89.3	3.6	3.6	3.6	1995
5600	0	18.8	50.0	31.3	0.0	1994
6100	18.2	93.3	3.3	3.3	0.0	1995
6200	32.4	91.2	0.0	0.0	8.8	1995
6300	36.4	96.3	0.0	0.0	3.7	1995
6400	27.8	88.6	0.0	0.0	11.4	1995
7100	23.3	78.3	4.3	13.0	4.3	1995
7200	39.3	95.2	4.8	0.0	0.0	1995
7300	34.5	83.3	4.2	12.5	0.0	1995
7400	16.1	75.0	8.3	8.3	8.3	1995
7500	38.5	70.6	11.8	17.6	0.0	1995
7600	17.2	68.2	9.1	18.2	4.5	1994
7700	28.6	65.0	15.0	15.0	5.0	1994
7800	34.5	69.2	0.0	26.9	3.8	1995



Third wave: Ethnic, gender and age distribution by classes

Class id	% of boys (%)	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)	Median year of birth
1100	38.7	100.0	0.0	0.0	0.0	1995
1200	36.1	100.0	0.0	0.0	0.0	1995
1300	23.5	100.0	0.0	0.0	0.0	1995
1400	30.3	100.0	0.0	0.0	0.0	1995
1500	23.3	100.0	0.0	0.0	0.0	1995
2000	53.3	13.3	60.0	26.7	0.0	1994
2100	11.1	47.8	30.4	21.7	0.0	1995
2200	28	69.2	23.1	7.7	0.0	1995
2300	43.5	64.7	23.5	11.8	0.0	1995
2400	48.1	82.6	8.7	8.7	0.0	1995
2500	47.8	43.8	18.8	37.5	0.0	1995
2600	40.7	33.3	33.3	33.3	0.0	1995
2700	34.6	17.6	35.3	47.1	0.0	1995
2800	48.1	57.1	23.8	19.0	0.0	1995
2900	42.3	43.8	25.0	25.0	6.3	1995
3100	53.6	92.6	0.0	3.7	3.7	1995
3200	21.9	93.8	6.3	0.0	0.0	1995
3300	93.5	92.0	0.0	8.0	0.0	1995
3400	62.5	93.8	0.0	6.3	0.0	1995
3600	100	52.2	21.7	26.1	0.0	1995
3700	21.1	29.2	33.3	33.3	4.2	1995
4100	50	93.8	0.0	6.3	0.0	1995
4200	42.9	100.0	0.0	0.0	0.0	1995
4300	41.4	89.3	0.0	7.1	3.6	1995
4400	23.1	84.6	3.8	11.5	0.0	1995
5100	25.8	74.1	11.1	14.8	0.0	1995
5300	100	91.7	0.0	8.3	0.0	1995
5400	50	88.2	2.9	8.8	0.0	1995
5500	61.8	90.9	3.0	3.0	3.0	1995
5600	28.6	56.3	6.3	37.5	0.0	1995
6100	18.8	93.8	0.0	3.1	3.1	1995
6200	33.3	93.5	0.0	0.0	6.5	1995
6300	39.4	96.4	0.0	0.0	3.6	1995
6400	21.9	100.0	0.0	0.0	0.0	1995
7100	23.1	84.6	3.8	7.7	3.8	1995
7300	28.6	80.0	4.0	12.0	4.0	1995
7400	17.6	84.6	7.7	7.7	0.0	1995
7500	36.4	85.0	5.0	10.0	0.0	1995
7600	28	81.0	9.5	4.8	4.8	1994
7700	32	54.5	9.1	36.4	0.0	1995
7800	30	76.9	7.7	15.4	0.0	1994

Fourth wave: ethnic, gender and age distribution by classes

Class id	% of boys	Hungarian (%)	Roma (%)	Both Hungarian and Roma (%)	Other (%)	Median year of birth
1100	37.9	100.0	0.0	0.0	0.0	1995
1200	36.8	100.0	0.0	0.0	0.0	1995
1300	23.5	100.0	0.0	0.0	0.0	1995
1400	32.4	100.0	0.0	0.0	0.0	1995
1500	23.3	100.0	0.0	0.0	0.0	1995
2000	50	42.9	42.9	14.3	0.0	1994
2100	8.7	58.8	23.5	17.6	0.0	1995
2200	37	61.9	9.5	28.6	0.0	1995
2400	38.5	86.4	0.0	13.6	0.0	1995
3100	55.2	95.5	0.0	4.5	0.0	1995
3200	25	92.6	7.4	0.0	0.0	1995
3300	91.3	87.0	0.0	13.0	0.0	1995
3400	57.7	87.0	0.0	13.0	0.0	1995
4100	50	92.9	0.0	7.1	0.0	1995
4200	41.2	97.1	0.0	2.9	0.0	1995
4300	39.1	100.0	0.0	0.0	0.0	1995
4400	20.8	91.3	0.0	8.7	0.0	1995
5100	35.7	92.0	0.0	8.0	0.0	1995
5200	100	70.4	7.4	18.5	3.7	1995
5400	47.8	94.7	0.0	5.3	0.0	1995
5500	60.7	88.5	3.8	3.8	3.8	1995
5600	37	60.0	13.3	26.7	0.0	1995
6100	20.8	82.6	4.3	4.3	8.7	1995
6200	34.3	100.0	0.0	0.0	0.0	1995
6300	42.4	95.8	0.0	0.0	4.2	1995
6400	18.8	100.0	0.0	0.0	0.0	1995
7100	17.6	86.7	0.0	6.7	6.7	1995
7300	37.5	90.0	0.0	5.0	5.0	1995
7400	12	84.2	0.0	15.8	0.0	1995
7800	35.7	76.9	0.0	23.1	0.0	1994
9100	15.4	18.8	50.0	25.0	6.3	1995
9200	35.3	57.1	14.3	28.6	0.0	1994
9300	81	50.0	12.5	37.5	0.0	1995
9400	70.6	81.5	7.4	11.1	0.0	1994
9500	33.3	75.0	15.0	10.0	0.0	1994
9600	76.9	0.0	50.0	0.0	50.0	1994
9700	9.1	72.7	27.3	0.0	0.0	1994
9800	21.7	53.8	30.8	15.4	0.0	1995

Sample stability: % of students chosen in the sample in the first wave who were still part of the sample in the fourth wave – based on original classes (students who changed class within the sample due to class mergers or other reasons are included).

Class id	Class size in the first wave (students)	Stability rate between the first and the second wave	Stability rate between the first and the third wave	Stability rate between the first and the fourth wave
1100	32	97%	94%	91%
1200	34	100%	94%	94%
1300	36	97%	92%	92%
1400	35	94%	91%	91%
1500	29	93%	93%	93%
2000	29	93%	41%	41%
2100	28	93%	81%	78%
2200	31	94%	68%	55%
2300	30	93%	57%	50%
2400	29	97%	79%	55%
2500	35	89%	49%	46%
2600	35	91%	54%	46%
2700	35	77%	49%	46%
2800	32	91%	78%	78%
2900	32	94%	66%	59%
3100	29	100%	97%	97%
3200	34	100%	91%	88%
3300	37	95%	76%	62%
3400	36	97%	92%	75%
3500	37	97%	32%	24%
3600	35	91%	46%	34%
3700	37	97%	54%	46%
4100	17	100%	100%	100%
4200	35	100%	97%	94%
4300	34	100%	79%	62%
4400	38	97%	47%	47%
5100	33	97%	64%	52%
5200	26	96%	35%	-
5300	19	95%	53%	47%
5400	35	100%	60%	34%
5500	35	94%	66%	54%
5600	31	74%	32%	26%
6100	33	100%	97%	70%
6200	34	100%	100%	97%
6300	33	100%	91%	85%
6400	36	100%	86%	86%
7100	32	94%	69%	41%
7200	31	90%	32%	32%
7300	31	90%	58%	35%
7400	33	91%	70%	45%
7500	33	73%	55%	55%
7600	33	88%	58%	58%
7700	31	90%	58%	58%
7800	35	83%	57%	40%

Stability rates: sample stability compared to the second wave (second wave=100%)

Class id	Class size in the second wave (students)	Stability rate between the second and the third wave	Stability rate between the second and the fourth wave
1100	31	97%	94%
1200	34	94%	94%
1300	35	94%	94%
1400	33	97%	97%
1500	28	100%	100%
2000	27	52%	52%
2100	25	92%	88%
2200	26	73%	58%
2300	28	61%	54%
2400	31	84%	58%
2500	36	47%	44%
2600	34	59%	47%
2700	35	54%	51%
2800	30	83%	83%
2900	35	60%	57%
3100	29	97%	97%
3200	35	91%	89%
3300	36	75%	61%
3400	36	94%	78%
3500	38	34%	26%
3600	35	49%	37%
3700	39	51%	44%
4100	17	100%	100%
4200	36	97%	94%
4300	33	79%	61%
4400	37	49%	49%
5100	34	65%	53%
5200	24	38%	-
5300	18	50%	44%
5400	38	58%	34%
5500	35	71%	60%
5600	23	43%	35%
6100	33	97%	70%
6200	34	100%	97%
6300	33	91%	85%
6400	36	86%	86%
7100	30	73%	43%
7200	28	36%	36%
7300	29	69%	41%
7400	31	77%	52%
7500	26	77%	77%
7600	29	62%	62%
7700	28	64%	64%
7800	29	69%	48%

Stability rates: Sample stability compared to the third wave (third wave=100%)

Class id	Class size in the third wave (students)	Stability rate between the third and fourth wave
1100	31	97%
1200	36	100%
1300	34	100%
1400	33	100%
1500	30	100%
2000	15	100%
2100	27	96%
2200	25	64%
2300	23	96%
2400	27	78%
2500	23	96%
2600	27	85%
2700	26	88%
2800	27	100%
2900	26	96%
3100	28	100%
3200	32	97%
3300	31	74%
3400	32	81%
3600	27	74%
3700	38	68%
4100	16	100%
4200	35	97%
4300	29	79%
4400	26	92%
5100	31	71%
5300	16	88%
5400	36	61%
5500	34	76%
5600	21	52%
6100	32	72%
6200	36	97%
6300	33	94%
6400	32	97%
7100	26	58%
7300	28	68%
7400	34	62%
7500	22	100%
7600	25	100%
7700	25	100%
7800	20	70%

Class stability between the first and the fourth wave: % of members of the original class (first wave) who were members of the class in the 2nd, 3rd or 4th wave as well (students who changed class within the sample due to class merger or other reasons are not included)

Class id	Class size in the first wave	Stability rate between the first and the second wave	Stability rate between the first and the third wave	Stability rate between the first and the fourth wave
1100	32	97%	94%	91%
1200	34	100%	94%	94%
1300	36	97%	92%	92%
1400	35	94%	91%	91%
1500	29	93%	93%	93%
2000	29	93%	41%	38%
2100	28	93%	82%	64%
2200	31	94%	68%	55%
2300	30	93%	57%	-
2400	29	97%	79%	59%
2500	35	89%	49%	-
2600	35	91%	54%	-
2700	35	77%	49%	-
2800	32	91%	78%	-
2900	32	91%	63%	-
3100	29	100%	97%	97%
3200	34	100%	91%	88%
3300	37	95%	76%	57%
3400	36	97%	92%	67%
3500	37	97%	-	-
3600	35	91%	46%	-
3700	37	97%	54%	-
4100	17	100%	100%	100%
4200	35	100%	97%	94%
4300	34	100%	79%	62%
4400	38	97%	50%	50%
5100	33	97%	64%	52%
5200	26	96%	-	-
5300	19	95%	53%	-
5400	35	100%	60%	34%
5500	35	94%	66%	57%
5600	31	74%	32%	26%
6100	33	100%	97%	70%
6200	34	100%	100%	97%
6300	33	100%	91%	85%
6400	36	100%	86%	86%
7100	32	94%	69%	41%
7200	31	90%	-	-
7300	31	90%	58%	32%
7400	33	91%	70%	42%
7500	33	73%	55%	-
7600	33	88%	58%	-
7700	31	90%	58%	-
7800	35	83%	57%	40%