Children of Immigrants Longitudinal Survey

in Four European Countries

CILS4EU

Sociometric Fieldwork Report

Wave 2 – 2011/2012

Version: 2.3.0

Year: 2016

Citation: Kruse, Hanno, Markus Weißmann and Konstanze Jacob. 2016. Children of Immigrants Longitudinal Survey in Four European Countries. Sociometric Fieldwork Report. Wave 2 – 2011/2012, v2.3.0. Mannheim: Mannheim University.

Contents

Introduction

The sociometric module in the second wave of the CILS4EU data aims at capturing the social structure surrounding the respondents in their school environment in the school year 2011/12.[1]

Compared to other network-related measurements, like ego-centered networks, sociometric questionnaires are able to capture only those social relations that are held within one predefined context, in this case the classroom. This restriction, however, provides certain advantages: All members of the context can be surveyed, thus providing information about the complete structure of the network, including relations present between actors, those that are absent and the position of each actor in the network relative to the others. Also, information about characteristics of the alters (those that are nominated) relies on self-reports and does not depend on reports by ego (the nominating person), thus providing higher data validity.

In order to broadly capture the social embeddedness of adolescents in their classrooms respondents reported different types of relations:

- best friend(s)
- most popular classmates
- classmates that they often spend time with outside of school
- classmates that they sometimes do homework with
- classmates whom their parents know
- classmates whose parents get together with / call their own parents
- classmates that are sometimes mean to them (in the Netherlands only).

All seven types of relations have already been surveyed in the first wave, which provides the chance to get a detailed image of how social relations in the classroom evolve over time. However, two points should be acknowledged when taking a longitudinal perspective on the classroom networks: 1) For obvious reasons, the sociometric module of wave 2 only applied to respondents attending a school in wave 2 where in-school-surveys could be realized. Longitudinal network data is therefore not available for those classrooms where students had already finished their education after wave 1 (for example, most low-track schools in Germany). 2) In various classrooms there had been composition changes between the first and second wave (especially so in the Netherlands). In classrooms with

[1] The sociometric module in wave 2 is in many regards identical to that of the first wave, which has been described at length in the sociometric fieldwork report wave 1. Nevertheless, in this second's wave report we again included all basic information concerning the sociometric module. Many sections of this report are therefore identical to the sociometric fieldwork report wave 1.

substantially large variation it would be erroneous to assume the second wave sociometric data as being longitudinal info but rather as cross-sectional network information at a second point in time.

Data collection

The sociometric module was generally issued in the classroom context, implying that students could only account for the relations to their classmates. The survey procedure was as follows: First, classroom-specific nomination lists were defined containing respondents' nomination possibilities, with personal socio-IDs being assigned to each classmate. In the Netherlands and Sweden, all students attending one of the surveyed classes in the second wave, in other words the complete sociometric gross sample of wave 2[2], could be nominated. In various German and English classrooms non-participating classmates (refusals or absent students) were crossed off the list, so that students could nominate participating classmates only. For more information refer to the subsection "Nomination of absent students".

Respondents received a sociometric questionnaire (see figure 1 for the first five questions) and were then asked to report all those socio-IDs from the nomination list that corresponded to classmates for whom the questioned type of relation applied. If the question did not refer to any of their classmates they were asked to write down a big "X" in the empty space.

[2] For more information about the sociometric gross sample of wave 2, refer to the section "A note on the sociometric gross sample of wave 2".

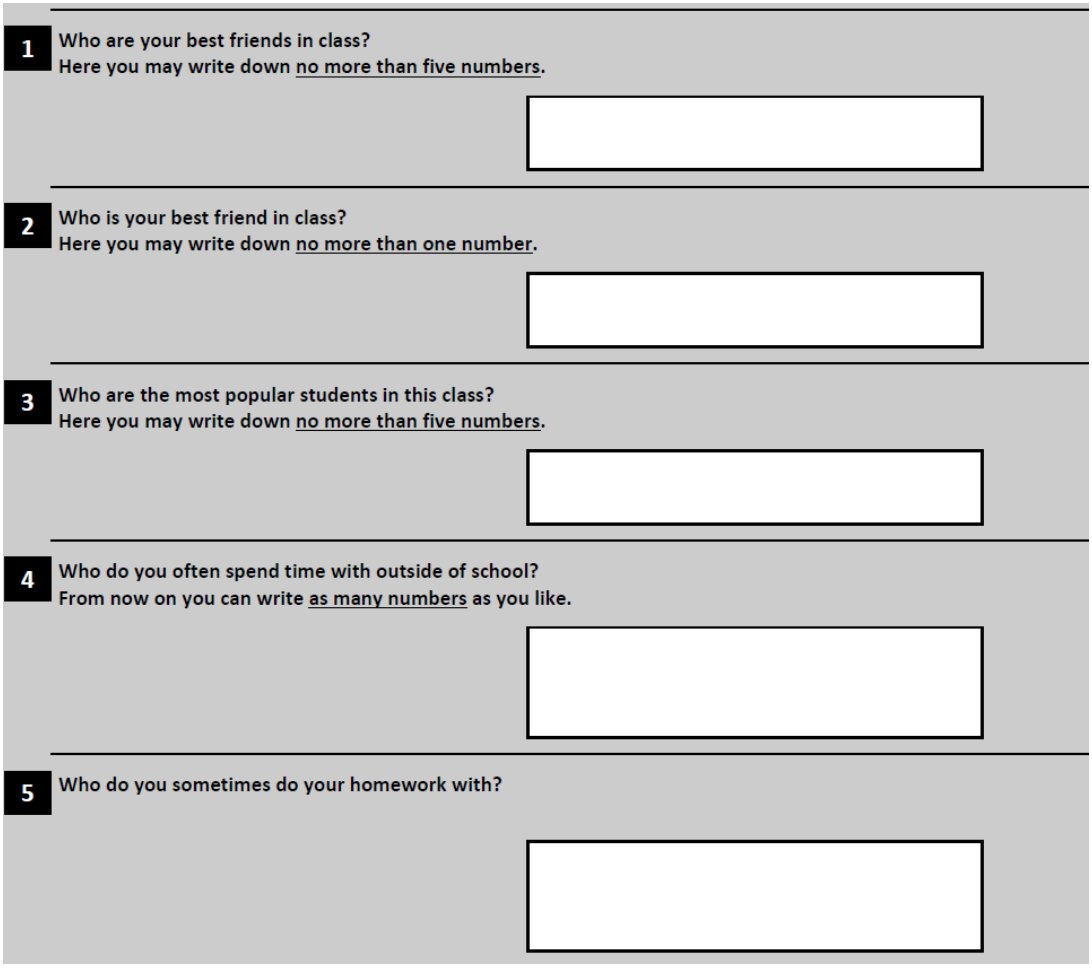


Figure 1: Sociometric Questionnaire Wave 2, Questions 1-5

Structure of the sociometric data

For each item in the sociometric questionnaire, the dataset contains several nomination variables indicating the youthids of the classmates that were nominated. The maximum number of these nomination variables is either restricted to one (*y2_bf_1*, see question 2 in figure 1), to five (*y2_bfs_**, *y2_pos_**, see questions 1 and 3 in figure 1) or it is unrestricted and thus determined by the empirical maximum number of nominations (applies to all other questions, for example question 5 in figure 1). If students made more nominations than allowed for we kept only the first answers they reported.

To illustrate the structure of the nomination variables we provide an example of a classroom consisting of seven students in figure 2. Panel (a) displays the variable structure of the five best friend nominations, the other panels indicate how these data would translate into a sociomatrix (panel (b)) and into a sociograph (panel (c)).

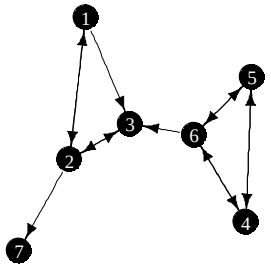
Figure 2: Exemplary Classroom Network Data

youthid	*y1_bfs_1*	*y1_bfs_2*	*y1_bfs_3*	*y1_bfs_4*	*y1_bfs_5*
01	02	03	-88	-88	-88
02	01	03	07	-88	-88
03	02	-55	-88	-88	-88
04	05	06	-55	-88	-88
05	04	06	-88	-88	-88
06	03	04	05	-88	-88
07	-77	-77	-77	-77	-77

(a) variable-structure

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(b) sociomatrix



(c) sociograph

For each sociometric question we added another variable which indicates whether a respondent explicitly reported having no relation of the specific kind asked about. This was the case if the respondent wrote down a big "X" in the respective empty space. The names of these variables all end with "_0". For example, *y2_bfs_0* indicates if a respondent reported having one or more best friends in the classroom (actual nomination(s) made), having no best friends ("X" in the empty space) or whether it was a case of item nonresponse (nothing filled in). The variables ending with "_0" are coded as follows:

0 = at least one classmate nominated
1 = relation does not apply to any classmate
-55 = other missing
-66 = question not asked
-88 = no answer.

Missing values in the nomination variables

Within the nomination variables four different types of missing values can occur: "no answer" (-88), "not applicable" (-77), "question not asked" (-66), and "other missing" (-55). They are described in further detail in the following.

No answer (-88)

In many cases respondents did not nominate the maximum possible number of classmates. For example, instead of nominating their five best friends in the classroom they only nominated two (see student "01" in figure 2). For these cases all "empty" nomination variables are assigned to the missing category "no answer" (-88).

Not applicable (-77)

As mentioned before, there are respondents who nominated no classmates at all on one or more questions, either because they stated not having any relations such as those asked ("_0"-variable has

value of 1) or because of item non-response ("_0"-variable has value of -88). In both cases, the missing category of all their nomination variables is then "not applicable" (-77).

In the exemplary classroom from figure 2, student "07" reported having no best friends in the classroom (wrote down an "X" in the respective empty space), which leads to a value of -77 in his or her own nomination variables. Since he or she participated in the survey, he himself or she herself could be nominated by others, as was done by student "02".

Question not asked (-66)

The question "Who is sometimes mean to you?" (*y2_mean1_**) was only asked in the Netherlands, in all other countries these variables are set to "question not asked" (-66). Also, questions about parents' social relations (*y2_pa1_*, y2_pa2_**) could not be asked in one German federal state, leading also to codings of -66.

Other Missing (-55)

Several answers were considered as implausible and therefore assigned to the "other missing" category (-55). This category includes:

- all nominations that do not correspond with an existing youthid in the sociometric gross sample of wave 2: Several IDs were nominated that are not part of the gross sample of wave 2.[3]
- all nominations that correspond with more than one existing youthid in the sociometric gross sample: In some cases in England, one single socio-ID was assigned to more than one person in the same class. These nominations were coded -55.
- all self-nominations (exception: question about most popular students): Some respondents nominated themselves, which is logically impossible for most of the sociometric questions.
- all nominations that a respondent made more than once: Some respondents nominated the same classmate more than once within the same question. Only the first of these multiple nominations was kept.
- further nominations that could not be linked to a valid youthid: In some cases, respondents did not report a socio-ID but real names or nicknames that could not be linked to actual classmates.
- all nominations whose respective "_0"-variable had a value of -55 (only in England).

[3] For country-specific information about the sociometric gross sample refer to the section „A note on the sociometric gross sample".

Please note that for England, socio-IDs of absent classmates could not be distinguished from invalid socio-IDs (see also "A note on the sociometric gross sample"). We therefore recoded them to -55, as well. The variable *y2_absnomsG* informs about which English classrooms had been subject to these changes (see also "Nomination of absent students").

Additional variables

Besides the nomination variables outlined above, the dataset contains several variables that provide additional information about the structure of the survey units or about respondents' nomination behavior. This information thereby refers to different levels of aggregation; it can be school-specific, classroom-specific, survey-unit-specific or individual level information.[4] In the following, we lay out all types of variables in greater detail.

Structure of the survey units

Survey unit ID (y2_svyunit_c)

The variable *y2_svyunit_c* identifies the survey unit in which respondents could make their nominations. All respondents who could nominate each other in the sociometric module of wave 2 share the same value. Figure 3 displays different ways of how survey units (grey-shaded areas) overlap with the classroom boundaries from wave 1, exemplified by two classes consisting of three students each. For the majority of the sample, survey unit definitions correspond to the case displayed in panel (a), where the survey units are equal to the classroom context. Here, nominations are restricted to classmates only (as indicated by the arrows).

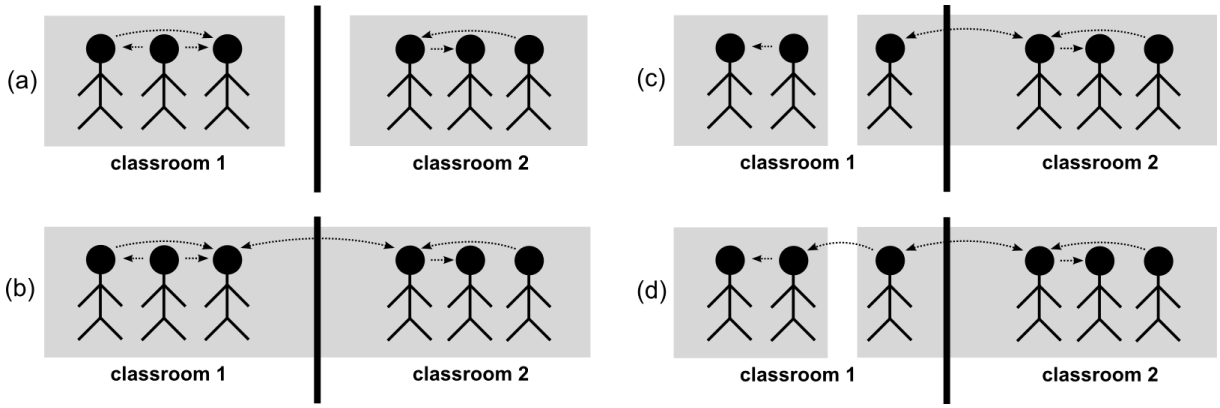


Figure 3: Exemplary Survey Unit Structures

[4] Note, that the levels of aggregation cannot necessarily be thought of as being hierarchical, as there are cross classifications on all levels.

In all Dutch schools, 57 English schools, 27 Swedish schools, and two German schools, however, *y2_svyunit_c* differs from respondents' classroom context of wave 1, which indicates that for these cases nomination possibilities were not exclusively bounded by the originally sampled classroom context from wave 1. This could be the case due to several reasons:

- the classroom context from the first wave does not exist anymore in the second wave. This is especially the case for schools in the Netherlands. Here, due to the institutional rules of the Dutch educational system, survey unit compositions changed considerably from wave 1 to wave 2.[5]
- merged nomination list of two classrooms (see figure 3b): Two classrooms were surveyed together and respondents could therefore nominate not only their own classmates but also those in the other class.
- respondents moved to another survey unit (see figure 3c): Some students took part in the sociometric survey in a classroom other than their own. They are therefore part of a different survey unit than their classmates, which implies that they can only nominate and be nominated by students in the other survey unit.

Survey unit bound to wave 1 classroom? (y2_classid_matchG, school level)

The flag variable *y2_classid_matchG* indicates whether all survey units in the respective school are defined in accordance with the classroom boundaries as sampled in wave 1 (corresponding to the situation in panel (a)). The variable is coded as follows:

0 = Survey units in school are not all bound to wave 1 classrooms
1 = All survey units in school are bound to wave 1classrooms (*classid=y2_svyunit_c*).

Different nomination list (y2_outnomsG, individual level)

For 18 respondents in 15 English survey units and 10 respondents in 5 survey units in Germany nomination lists differed from those of the others in their survey unit. Besides the students within their own survey unit they could also nominate students participating in other survey units. The flag variable *y2_outnomsG* indicates all survey units where this is the case. It is coded as follows:

0 = Nomination list bound to survey unit
1 = Nomination list exceeds survey unit.

[5] In the majority of the Dutch cases, wave 2 classrooms entailing at least one student that participated in wave 1 were interviewed as a whole, leading to many out-of-sample interviews (for further information refer to the fieldwork report wave 2). However, in 6 schools, participants from wave 1 were regathered in their wave 1 class setting to be interviewed in wave 2. In other words, students in the latter case shared a classroom in wave 1 and not in wave 2, but were interviewed as if they were in the same class in wave 2. The variable *y2_classcons*, directly stemming from Dutch fieldwork, is added as additional information to inform about whether students were interviewed proximately in their wave 1 setting (not a 'real' class, i.e., y2_classcons=1) or not.

The reason for this deviation is that a number of those respondents who moved to a survey unit different from their own classroom context were able to nominate within their survey unit as well as among their own classroom members. The resulting survey unit structure is displayed in panel (d) of figure 3.

Nomination of absent students (y2_absnomsG, survey unit level)

As mentioned before, in the Netherlands and Sweden, all students from the sociometric gross sample of wave 2 could be nominated even if they were absent during the in-school surveys or opted-out to participate. For Germany and England, the procedure differed between schools or even survey units within schools, implying that in some cases (England: N(survey units) = 23; Germany: N(survey units) = 21) absent students were excluded from nomination lists. In England, nominations of absent students could not be distinguished from invalid nominations, which is why we recoded them to -55 (see also "Missing values in the nomination variables").

We assume that respondents were allowed to nominate absent students if in any of the sociometric questions an absent student was actually nominated (in England: if a nomination did not refer to a participating student). The flag variable *y2_absnomsG* contains this information. It is coded as follows:

0 = absent classmates cannot be nominated
1 = absent classmates can be nominated.

Table 1: Country differences in the survey unit structure

	England	**Germany**	**The Netherlands**	**Sweden**
Survey unit ≠ classroom context (*y2_classid_matchG=0*)	Yes, in 57 schools	Yes, in 2 schools	Yes, in all schools	Yes, in 27 schools
Different nomination lists in survey unit (*y2_outnomsG=1*)	Yes, in 15 survey units	Yes, in 5 survey units	No	No
Nomination of absent students possible (*y2_absnomsG=1*)	Not in 23 survey units	Not in 21 survey units	Yes	Yes

Nomination behavior

Share of participants (y2_p_participantsG, class level)

The validity of classroom network data highly depends on the share of network members who actually report their nominations. As participation rates differ substantially between the classes sampled in wave 1, we included a variable to identify the share of wave 1 classroom members participating in the sociometric module in wave 2, *y2_p_participantsG*. This information is therefore only available for schools where all survey units are wave 1 class-specific (i.e., if *y2_classid_matchG=1*).

The share is calculated as the number of classroom members who participated in the sociometric survey of wave 2 divided by the total number of classroom members in the sociometric gross sample of wave 2.

Share of invalid nominations (y2_p_invalidnomsG, survey unit level)

In order to get an impression of the relative frequency of invalid nominations in a survey unit we included the variable *y2_p_invalidnomsG*. It informs about the relative share of invalid nominations out of all nominations made in a survey unit that had to be recoded to "other missing" (-55).

Please note that invalid nominations do not necessarily suggest that there have been mistakes in the surveying process. Rather, these may also be the result of self-nominations or double nominations.

Respondent never nominates (y2_nnominatorG, individual level)

Some respondents did not make any nominations throughout any of the seven sociometric questions (eight in NL), and also never explicitly reported having no ties at all. They are referred to as never-nominators, with the corresponding flag variable *y2_nnominatorG* identifying these respondents. It is coded as follows:

0 = at least one nomination made
1 = no nomination made in any sociometric question.

Respondent is never nominated (y2_nnomineeG, individual level)

Furthermore, some students were never nominated by others in their survey unit on any of the questions (never-nominees). The flag variable *y2_nnomineeG* can be used in order to identify these cases. It is coded as follows:

0 = nominated at least once
1 = never nominated by a classmate on any sociometric question.

A note on the sociometric gross samples of wave 2

In general, the four country-specific sociometric gross samples of wave 2 consist of the following students:

- the complete sociometric gross sample of wave 1 (i.e., all participants of the sociometric survey in the first wave and all students of the sampled classes who either refused to participate or who were absent in wave 1)
- all students newly attending a survey unit that wave 1 students are part of in wave 2.

As mentioned before, every person in the sociometric gross sample gets assigned an individual socio-ID. Eventually, all students not being part of the sociometric gross sample of wave 1 were moved to an extra "out-of-sample data set"[6], while nominations pointing to these cases are still reported.

In England, however, the socio-IDs of absent/refusing students in wave 2 were not known. We therefore had to exclude all reported socio-ID nominations not corresponding with an existing youthid, as they could not be differentiated from invalid nominations. In accordance with the general procedure (see section "Missing values in the nomination variables") they were recoded to "other missing" (-55).

In the other three countries we are left with two types of valid nominations: those that correspond to an existing youthid in the realized sample and those that correspond to a person from the sociometric gross sample of wave 2 whose interview is not available in the regular second wave dataset. Concerning the latter, the nominations themselves nevertheless carry some information, for example when analyzing respondents' outdegree (= nomination activity).

Regardless of the type of analysis carried out, we suggest to always keep in mind these differences between gross samples.

A note on the longitudinal nature of the data

As mentioned before, all seven types of relations (eight in NL) surveyed in wave 2 have already been surveyed in wave 1. Taken together, the sociometric modules of both waves therefore constitute longitudinal classroom network data that informs about how different social relations in the classroom evolve over time.

A central prerequisite for valid longitudinal network data is stability of the networks' compositions over time. This has two implications which ought to be met: 1) The survey context in both waves

[6] Please, refer to the fieldwork report of wave 2 for further information.

should be more or less identical, meaning that at both points in time actors should face (nearly) the same nomination possibilities. 2) Panel attrition should not be too high, meaning that most of wave 1 participants also participate in wave 2.

There are various reasons why one of the two conditions might not apply: For example, in schools where students usually finish school after ninth grade (applies to various low-track schools in Germany) there is simply no second wave sociometric data available. Further, there are schools where students are reassigned to newly forming classes, thus potentially leading to very different survey unit compositions in waves 1 and 2 (applies especially to Dutch schools). Beside these institutional specificities, there is, of course, another pitfall: survey dropout. Admittedly, survey dropout is a point of concern for any type of data collection. In the case of sociometric data, however, the consequences of survey dropout can be especially severe: A non-participating student does not only imply one interview less but potentially also a restriction in nomination possibilities of the others in his/her survey unit.

The laid-out additional variables help to assess to what extent both laid out conditions are met in the different classroom networks. For example, the variable *classid_matchG* in both waves informs about whether survey units have been class-specific in both waves. If this is the case, we can assume that the survey context has been more or less identical in both waves (despite a number of students that might have newly entered or left the class in wave 2). The variable *p_participantsG* of both waves then additionally informs about how much of the classroom context is actually covered by the data, as well as the level of panel attrition between the two waves (difference between *p_participantsG* of wave 1 and wave 2).

Given that there are various reasons for a lack of stability in the networks' compositions over time we would expect that only a fraction of the surveyed networks might be suitable for longitudinal analyses. And indeed, this seems to be the case, as the following exemplary sample restriction shows (see table 2). Before turning to the example, however, please note that its sole focus is the data's applicability in terms of longitudinal network analysis. Other issues, such as measurement differences between schools and survey units are not taken into account.[7]

In wave 1 we realized sociometric interviews in 938 survey units. Merging the sociometric data of waves 1 and 2 and keeping only those cases participating in both waves yields our baseline population that we refer to as the "balanced sample". Doing so leads to a substantial decrease in sample sizes

[7] For examples of how to account for these measurement differences refer to the sociometric fieldwork report of wave 1.

(N(survey units) = 576). As laid out earlier, this is mainly due to Dutch schools, where the surveying contexts changed completely between the two waves. In a first step we restrict this sample to all cases where both surveys had been conducted within the wave-1-class context, which leaves us with 449 survey units. In a second step we then consider what portion of the classroom was actually covered in each of the two waves. We hereby restrict the sample to those survey units that covered at least 75% of the respective classroom in both waves. Again, this leads to a substantial reduction in sample size, with Germany showing the largest decrease. Overall, we thus end up with 221 survey units whose network data seems well-suited to be analyzed longitudinally.

Table 2: Examples of how to restrict the sample for longitudinal network analysis

	England **N(survey units)**	**Germany** **N(survey units)**	**The Netherlands** **N(survey units)**	**Sweden** **N(survey units)**	**TOTAL** **N(survey units)**
Wave 1 survey units	200	269	220	249	938
Balanced sample over the two waves	127	203	0	246	576
+ wave 1 class remains the survey context	61	200	0	188	449
+ at least 75% of class participated in both waves	42	68	0	111	221