

Methodological guidelines

Update: 29.12.2013

How to define the size of the initial sample for national surveys which contribute to EUROSTUDENT V

Since the quality of the data delivered to EUROSTUDENT is essential for the relevance and reliability of the project, you should pay special attention to a proper sampling and an adequate sample size.

A relatively small sample is needed for the participation in EUROSTUDENT V because we need data only on the national level of each country. At the same time, we should achieve enough cases in order to represent the focus groups (see below) sufficiently. This paper provides some hints on how to calculate the **minimum initial sample size** (not the sampling as this highly depends on the specific higher education systems) needed from each country. However, increasing the sample size might make sense, on the one hand because a larger sample usually provides better data quality and is therefore more reliable. On the other hand this enables you to make additional analyses of the situation in your country, for example deeper insight into the situation of smaller subgroups not relevant in the EUROSTUDENT context.

When planning your sample, you should focus on the target group of EUROSTUDENT V, which was defined as follows:

- All students in your country, i.e. national and foreign students who are pursuing their studies for a degree in the country of the survey¹, except students on leave and excluding students on incoming and outgoing credit mobility.
- Full-time and part-time students by status.
- Students in all ISCED 5, 6 and 7 programmes, regardless of their character as general or professional, as long as the programmes are considered higher education in your national context.²
- All higher education institutions offering programmes considered "normal". In many cases this means only public, non-specialist institutions of higher education.
- All national degrees corresponding to ISCED 5, 6 and 7 (e.g. BA, MA, traditional diploma, Lizentiat, national degrees in medicine. Short courses only if they are based on ISCED 5.)
- Distance students who study at a "normal" higher education institution, i.e. excluding institutions solely for long distance students like open universities, Fernuniversität Hagen and similar.

¹ Widening the focus of EUROSTUDENT towards students who obtained their prior education in another country but the country of the survey is a major change in the definition of the standard target group. We have thereby taken into account the increasing mobility throughout Europe on all educational levels.

² EUROSTUDENT V now refers to the new ISCED 2011 classification. For further information see <http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx>

Besides a proper sampling in terms of equal chance for every student to be sampled³, you have to have enough questionnaires returned from each subgroup of interest for EUROSTUDENT. The focus groups used for the calculation of indicators in EUROSTUDENT V are as listed below:

Very important subgroups of students: The following groups are more or less the standard comparative groups, used for the calculation of almost all core indicators for EUROSTUDENT V. To deliver comparable data to EUROSTUDENT V, you need a minimum number of at least 50 questionnaires returned from each of these groups:⁴

- Female students
- Male students
- Students up to 21 years
- Students aged 22-24 years
- Students aged 25-29 years
- Students older than 30 years
- Bachelor students
- Master students
- Other types of degree programmes on ISCED level 5, 6, 7
- Groups of students by type of institution according to your national HE system (e.g. University vs. University of Applied Science vs. Teacher Training College)
- Groups of students by broad fields of study (ISCED one-digit level of detail)
- Students without HE background (parents' educational attainment ISCED 0 to 4)
- Students with HE background (parents' educational attainment ISCED 5 to 8)

Less important subgroups of students: These groups are of special interest in different subtopics of EUROSTUDENT. To be able to provide all indicators used in EUROSTUDENT, you should have a minimum number of 50 returned questionnaires also from these groups:

- Students with children
- Students with impairments
- International students
- Students with migration background
- Groups of students with different access routes to HE (especially APR and the like)
- Students with prior experience on the labour market
- Students who interrupted their HE educational career (between entering HE and graduation)
- Students who interrupted their educational career (between graduating from secondary school and entering HE)
- Full-Time students (by formal status)
- Part-Time students (by formal status)
- Students in distance learning courses
- Students living with parents

³ For instructions on sampling see <https://eurostudent.his.de/wiki/images/e/e6/Sampling.pdf>

⁴ This applies to groups present at all in your country. E.g. if you have no students older than 30, you do not need this group in your sample.

- Students maintaining own households
- Students living in student accommodations, i.e. dormitory or halls of residence
- Students from study locations with up to 100.000 inhabitants
- Students from study locations with 300.000 to 500.000 inhabitants
- Students studying in the capital city
- Students receiving public student support
- Students not receiving public student support
- Non-working students
- Working students
- Students working 1-5hrs/wk
- Students working more than 5hrs/wk
- Students with study period abroad (enrolment)
- Students with other study related experiences abroad (e.g. internship, summer school)

Calculation of the minimum sample size needed

For the calculation of the sample size, we differentiate between the **initial** and the **actual sample size**. The initial sample size is the number of questionnaires you should send out to the students, while the actual sample size (the product of initial sample size by the response rate) is the sample size, you use for analysis.

The key numbers for the presented calculation are

- a given response rate of **20%** (conservative estimate)
- a minimum of **50** questionnaires per subgroup is needed for analysis (actual sample size).

This means in order to attain an *actual* sample size of 50 questionnaires, you need an *initial* sample size of $50 \cdot 1/0,2 = 250$ **questionnaires send out for each subgroup**.

When calculating the initial sample size, you have to take several characteristics of the national higher education system into account. For example, the different types of institutions, different degree types, different shares of full- and part time students, gender composition by field of study and so on. As these characteristics differ in the participating countries, it is not possible to provide a formula that fits all.

Instead, we will provide you a very simple formula to calculate an **approximate value** of the **minimum initial sample size** needed. This formula is based on **type of higher education institution** and **degree** studied for.

Why do we focus on the type of institution and the type of degrees? In most European countries, different HEIs offer different programs or fields of study (e.g. teacher training colleges). These different programs attract a diverse student body. For example, students attending a private or a public university or an university of applied science (college) differ, e.g. by their social background. Moreover, some HEIs offer certain degrees, like M.D., which also contain implicit information about the students, whilst we assume that the Bologna-Structure of degrees (BA, MA) is of special interest

for international comparisons. A second reason is that you may draw your sample utilising HE institutions or information about the HE institution and program attended by students. Therefore you should consider these two characteristics as the “basis” for your initial sample size.

In any case, you have to do a proper random sampling based on the real data of your student population considering the subgroups listed above. However, for a **first and rough calculation** of the initial sample size, the following will do:

Take the number of different degree programmes (e.g. Bachelor, Master, national degree (e.g. Lizentiat)) per type of higher education institution (private universities, public universities...) in your country and multiply it by 2.000. Thus you have enough questionnaires in order to provide data on most of the very important subgroups listed above – unless a certain subgroup is very small in your country. In such a case, you should add questionnaires for oversampling that group. However, this formula gives you only an approximated value of the initial sample **size**. It is not a substitute for a proper random sampling!

Table 1: Rough formula to calculate a minimum initial sample size

Type of Institution	Type A	Type B	Type C	Type D	Sum
Type of Degrees	#	#	#	#	#

→ x different types of programmes * 2.000 = # minimum initial sample size

→ expected return rate 20% = # actual sample size

Example 1: Country with a differentiated HE system

Type of Institution	Public Universities	Public Univ. of Applied Sciences	Private Universities	Teacher Training Colleges	Sum
Type of Degrees	BA, MA, Dipl.	BA, MA	BA, MA	BA, Dipl.	9

→ 9 different types of programmes * 2.000 = 18.000 minimum initial sample size

→ expected return rate 20% = 3.600 actual sample size

Example 2: Country with a homogeneous HE system

Type of Institution	Public Universities	---	---	---	Sum
Type of Degrees	BA, MA	---	---	---	2

→ 2 different types of programmes * 2.000 = 4.000 minimum initial sample size

→ expected return rate 20% = 800 actual sample size⁵

As mentioned above, you have to consider actual data about your student population or – in absence of data – use assumptions, for a proper random sampling. If you want to provide all indicators

⁵ NB: The actual sample size needed is only partly dependent on the absolute number of students. In our example, the second countries total student population may be even greater than in the first example. But as the second one is assumed to be more homogenous, the smaller sample will do.

needed for your country, you have to ensure that you have at least 50 questionnaires from each of the groups listed above for analysis. E.g.: If we assume that 5% of the students have been enrolled in a foreign country (“mobile students”)⁶ we need an initial sample of 5.000, of which 1.000 will participate (20% return rate) so we will end up with 50 mobile students in the realized sample. In such a case, an initial sample of 5.000 is the minimum – regardless of the shape of your higher education system (if 10% of the students were mobile, a sample size of 2.500 would suffice). Hence, when you do your actual sampling, you have to consider such assumptions for all “very important” subgroups listed above. The general formula above cannot be used to decide on the number of questionnaires sent to each type of degree per type of institution. Instead, you have to sample your students carefully to make sure each of the subgroups is covered by at least 250 questionnaires (by an assumed return rate of 20%).

Usually questionnaires will *not* be received (and responded) evenly by all groups, therefore some groups need to be oversampled according to the real number of enrolled students. Groups difficult to cover in your sample might be e.g. international or mature students, as they often do not consider themselves part of the target population. Any oversampling has to be corrected in the final data set by weighting the data.⁷ So, if your budget is limited, you should invest in a very detailed planning of your sample to use your resources as efficient as possible while still getting enough questionnaires for each group of analysis. In other words, **the more limited your budget is, the more attention you should pay to the sampling of your student population!**

You may also use the formula above to calculate your sample the other way round: Fix your sample size according to your resources and then check with the rough formula presented above which groups you can analyse and what kind of indicators you can calculate. For example, if your budget allows for sending out 20.000 questionnaires, you can estimate how many questionnaires you can expect from each subgroup according to the specifications of your HE-System.

Invitations sent via email

For EUROSTUDENT V, we recommend to use online surveys.⁸ If you send the invitations to the survey via email, you should consider increasing the sample size anyway, because this may be done without any additional costs.

Please do not hesitate to contact us (eurostudent@ihs.ac.at or eurostudent@dzhw.eu) if you have any questions or need further assistance.

⁶ Be aware that we are surveying students not graduates. Therefore the rate of mobile students is relatively low, because it includes beginners too, which did not yet have the chance to be mobile.

⁷ Further weighting (post stratification) is usually needed, because we have a different share of non-responses in different groups. For instructions see

https://eurostudent.his.de/wiki/images/b/be/EV_Coding_Instructions.pdf and

https://eurostudent.his.de/wiki/images/e/ea/Data_cleaning.pdf

⁸ See https://eurostudent.his.de/wiki/images/4/4e/Instructions_Survey_Monkey.pdf for instructions and access to a model questionnaire, which can be easily adapted to your national context.