Methodical guidelines:

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

A relatively small sample is needed for the participation in EUROSTUDENT IV. The reason for this is that we only need data on the national level of each country and we only compare large subgroups of students (e.g. male – female). This paper provides you with hints on how to calculate the **minimum sample size** (not on the actual sampling as this depends too much from the specific higher education systems) needed from each country. However, it may make sense to increase the sample size, on the one hand because a larger sample usually provides better data quality and is therefore more reliable, on the other hand, because this would enable you to make additional analysis of the situation in your country, for example deeper analysis of smaller subgroups not of relevance in the EUROSTUDENT context.

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

- Resident students. Resident students are students who have finished their prior education (school) in the respective country regardless of their nationality. (Not citizenship, which may be different.)
- Full-time and part-time students by status. (Not by study intensity, which may be different and will be included in the analysis of the data.)
- Students in ISCED 5A-programmes
- All higher education institutions offering programmes at ISCED 5A and considered “normal”. In many cases this means only public, non-specialist institutions of higher education.
- BA, MA and all national degrees corresponding to ISCED 5A (E.g. traditional diploma, Lizentiat, national degrees in medicine. Short courses only if they are based on ISCED 5A)
- Distance students that study at a “normal” higher education institution, i.e. excluding institutions solely for long distance students like open universities and similar.

For drawing a sample of your students, you have to ensure that you have enough questionnaires returned from each subgroup of interest for EUROSTUDENT. Since not all indicators needed in EUROSTUDENT IV are defined yet, we provide you a list of subgroups used in EUROSTUDENT III:

<u>Very important</u> subgroups of students. You should ensure to have a minimum number of questionnaires (at least 50) returned from each of these groups:

- Male students
- Female students
- Male first year students
- Female first year students

- Groups of students by type of institution (e.g. University vs. University of Applied Science vs. Teacher Training College)
- Groups of students by "ownership" (public HE-Institution vs. other Types of HE-Institutions)
- Students from low educational background (Father ISCED 0, 1, 2)
- Students from high educational background (Fathers ISCED 5, 6)
- Full-Time students (by formal status)
- Part-Time students (by formal status)
- Bachelor students
- Master students
- Other Types of degree programmes on ISCED 5A
- Students aged 18-24 years
- Students aged 25+ years
- Students living with parents
- Students not living with parents
- Working students

Less important subgroups of students. To be able to provide all indicators used in EUROSTUDENT, you should also ensure a minimum number of returned questionnaires from these groups:

- Students with children
- Students from study locations with less than 100 thousand inhabitants
- Students from study locations with more than 500 thousand inhabitants
- Students living in own lodging/sublet/private flat
- Students living in student halls
- Students aged 20 years
- Students aged 21 years
- Students aged 22 years
- Students aged 23 years
- Students aged 24 years
- Students aged 25 years
- Students aged 26 years
- Students aged 27 years
- Bachelor students not living with their parents
- Students from high education background not living with their parents
- Students from low education background not living with their parents
- Receivers of state support for students not living with their parents
- Non-working students
- Students working 1-5hrs/wk
- Students working 6-10hrs/wk
- Students working 11-15hrs/wk
- Students working more than 15hrs/wk
- Students in engineering studies
- Students in humanities/arts

- Students by year of study (1st, 2nd, 3rd, 4th, 5th year)
- Students with study experience abroad ("mobile" students)
- Students with low education background, who have not been abroad
- Students studying engineering, who have not been abroad

For the calculation of the sample size, you may assume the following:

- Return rate: 20% (conservative)
- A minimum of 50 questionnaires per subgroup is needed for analysis.

That means, the *initial* sample size for each subgroup should be 250 questionnaires.

Calculation of the minimum sample size needed

The calculation of the sample size has to consider several characteristics of the national higher education system. For example the different types of institutions, different degree types, different shares of full- and part time students, gender segregation by field of study and so on. Hence, it is not possible here to provide a formula that fits all countries.

Instead, we will provide you with a very simple formula that allows you to calculate an **approximated value** of the **minimum sample size** needed. 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 sample size, the following will do:

Take the number of different degree programmes (e.g. Bachelor, Master, Lizentiat) per type of higher education institution (private universities, public universities...) in your country and multiply it with 2.000. That will provide you with enough questionnaires to be able 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 sample **size**. It is not a substitute for a proper random sampling as such!

Table 1: Rough formula to calculate a minimum 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% = # realised sample

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 * 2000 = 18000 minimum initial sample size

➔ expected return rate 20% = 3600 realised sample

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 * 2000 = 4000 minimum initial sample size

➔ expected return rate 20% = 800 realised sample

Why do we focus here on the type of institution and the type of degrees? We assume that students attending a private or public university or an UAS are different, e.g. by their social background. Moreover, we assume that the Bologna-Structure of degrees (BA, MA) is of special interest for international comparisons. That's why we regard these two characteristics as the "basis" for any sampling.

As mentioned above, you have to consider real data about your student population or – in absent of data – use assumptions, for a proper random sampling. If you want to provide all indicators needed for your country, you have to ensure that you have at least 50 questionnaires for analysis from each of the groups listed above. E.g.: If we assume that 5% of the students have been enrolled in a foreign country ("mobile students")[1] we need to have an initial sample of 5000, of which 1000 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 5000 is the minimum – regardless of the shape of your higher education system (if 10% of your students were mobile, a sample size of 2500 would do it). Hence, when you do your real sampling, you have to consider such assumptions for all the subgroups listed above. Thus, the general formula above cannot be used to decide, how many questionnaires should really be sent to each type of degree programme per institutional type. Instead, you have to carefully sample your students to ensure that each of the subgroups listed above receives at least 250 questionnaires.

[1] Be aware that we are surveying students not graduates. Hence the rate of mobile students is relatively low, because it includes beginners as well who did not yet have the chance to be mobile.

However, questionnaires will usually *not* be sent equably to all groups, but some groups have to be oversampled according to the real number of enrolled students ("quota sample"). This oversampling has to be corrected in the final data set by weighting the data.[2] Hence, 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 gaining enough questionnaires for each group of analysis. Or in other words, as more limited your budget is, as more you should pay attention on the sampling of your student population. **Please do not hesitate to contact us if you need further assistance with that!**

An Alternative

You may also use the considerations above to calculate your sample the other way around: Fix your sample size according to your resources and then check with the here presented rough formula on what level you can do analysis and what kind of indicators you can calculate. For example, if your budget allows you to send 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

If you have chosen to do an online survey and if you have the possibility to send the invitations for the survey by email, you should consider increasing the sample size, because this may be done with nearly no extra costs.

[Version: October 2009]

[2] Further weighting (post stratification) is usually needed, because we have a different share of non-responses in different groups.