What deters students from studying abroad? Evidence from Austria, Switzerland, Germany, The Netherlands and Poland
## Contents

1. Status of this document ................................................................. 3
2. Project outline .............................................................................. 3
3. Target group ................................................................................. 4
4. Glossary and guidelines for preparation of variables ....................... 5
5. Guidelines for analyses and data delivery ........................................ 13
   5.1 General guidelines ................................................................. 13
   5.2 Guidelines for descriptive analyses ........................................... 14
   5.3 Guidelines for regression analyses ............................................. 14
1 Status of this document

This document is the final version of the manual describing in detail the analyses conducted in the context of the Steeplechase project (last modification: 10th May 2012). It is the result of internal conceptual work by HIS-HF and the input given through various feedback loops by the international project partners from Austria (Institute for Advanced Studies), The Netherlands (ResearchNed), Poland (Ministry of Science and Higher Education) and Switzerland (Federal Statistical Office of Switzerland). It is meant to answer all questions that have arisen in the correspondence between HIS-HF and the international project partners so far.

2 Project outline

The Steeplechase project examines which factors deter students from realising a temporary enrolment period outside of the country of their home institution. In doing so, it gives special attention to carving out differences with regard to obstacles to an enrolment period between (i) the five countries participating in the project and differences between (ii) specific groups of students within and across these countries. With regard to factors deterring students from studying abroad temporarily, a distinction is made between rather objective aspects on the one hand – i. e. selected socio-demographic and study-related characteristics of the examined students – and subjectively perceived obstacles to enrolment period abroad. The possibility to do so is given by data from five national EUROSTUDENT surveys (carried out between 2009 and 2010), which offer unique insights into the personal and study-related characteristics as well as perceived obstacles impeding students to participate in mobility experiences. The goal of the Steeplechase project is twofold: On the one hand, it shall contribute to the ongoing scientific discussion on barriers to student mobility. On the other hand, it is intended to provide information that can support the design of target group-specific policies for the identification and removal of mobility barriers.

In the analyses, a basic distinction is made between three groups of students:

(A) students that have already realised a period of enrolment abroad at the time of surveying
(B) students who have not realised, but plan to realise a period of enrolment abroad in the future
(C) students who have not realised and are not planning to realise a period of enrolment abroad in the future

The analyses are structured into four sections. In a first step, the share of students who belong to groups A, B and C are examined in country comparison. Next to presenting the shares of students belonging to the three groups in country comparison, further differentiations are made for selected focus groups (e. g. female in contrast to male students, students with tertiary education background versus students without such a background). These analyses shall provide first descriptive evidence on how students’ characteristics and their obstacle perception influence the likelihood to enrol abroad temporarily.

This first evidence is tested based on a logistic regression model in the second step. The dependent variable of this model is students’ belonging to the group which has realised a temporary enrolment

---

1 It should be noted that the structure of the final publication of the Steeplechase project (Netz, Orr, Gwosć & Huß, 2012) differs from the one used for conducting the statistical analyses.
period abroad, the independent variables describe students’ socio-demographic as well as study-related characteristics. Whether this model can be included in the final report or not will depend on its fitness for the national data sets. In essence, there are two shortfalls of the logistic regression model with the realised enrolment periods as dependent variable. Firstly, the national student surveys cannot take account of the fact that students might realise enrolment periods abroad after having been surveyed. Secondly, only a small proportion of the national student populations have realised foreign enrolment periods at the time of being surveyed, which implies that the dependent variable is strongly skewed. For this reason, it is difficult to obtain good models predicting the likelihood that a student with certain characteristics belongs to one of the groups captured through the dependent variable. This does not imply that the model does not have any explanatory power for the purposes of the Steeplechase project. Although it will most likely not be suited for predictive purposes, it can still identify factors that influence whether students realise an enrolment period or not.

The focus of the analyses shall rest on the third step, where the decision to enrol abroad is modelled for those students who have not (yet) been enrolled abroad. As the decision to enrol abroad is queried at the same time as the student attributes (and not retrospectively as in the case of the first regression model), the second model is much stronger from a theoretical and methodical point of view. Next to introducing independent variables into the models reflecting students’ socio-demographic and study-related characteristics, indicators for students’ subjective perception of selected obstacles can be included. These obstacles were chosen based on two considerations: Firstly, they had to capture different aspects that might influence the decision whether to enrol abroad or not. Secondly, they had to be included in all national questionnaires of the five participating countries. In that these perceived obstacles can be included in the regression analyses as independent variables, a direct link can be postulated between obstacle perception and the likelihood to (plan to) enrol abroad.

To help interpret the results of the regression models – especially of the second one – further descriptive evidence shall be analysed in a fourth step, that is to say evidence on how intensely different types of students perceive the seven obstacles included in the second regression model. The descriptive data can especially be used to examine how the perception of obstacles varies between students still planning to realise an enrolment period abroad and those not planning to do so. In the Steeplechase project, it is argued that uncovering and understanding differences between types of students in terms of the obstacle perception is a central prerequisite for devising concrete measures supporting temporary enrolment periods abroad.

3 Target group

For the purpose of the Steeplechase project – and following the general conventions for EUROSTUDENT IV – students are defined as persons enrolled in higher education at ISCED level 5A at all types of “normal” higher education institutions. Excluded are students at specialised higher education institutions (e.g. universities of the armed forces) and pure distance universities (e.g. the Fernuniversität Hagen in Germany). Furthermore, the analyses focus only on resident students who have gained their prior school education in the country where they are enrolled at the time of the

---

2 An addition or alternative to performing further descriptive analyses on the perception of obstacles by different types of students would be to calculate and graph conditional effects plots, which shows how the decision to enrol abroad varies across students types and ‘levels’ of obstacle perceptions, based on the second inferential model.
survey (i.e. on both national and foreign students who received their higher education entrance qualification in the country where they were surveyed). This means that only students who have not crossed a border in order to directly enter higher education are included. In detail, the following EUROSTUDENT IV data cleaning rules have to be applied:

- Exclude students not enrolled at “normal” higher education institutions.
- Exclude students not enrolled at ISCED level 5A (such as 5B and 6).
- Exclude students who are non-residents and/or have not gained their higher education entrance qualification in the country where they were surveyed.
- Exclude students of which you do not know the sex and/or the age and/or the programme level at which they are enrolled.

In addition to these general EUROSTUDENT IV conventions, further data cleaning rules must be applied specifically for the Steeplechase analyses:

- Exclude students who have not answered the question(s) on whether they have been enrolled abroad and on whether they are planning/not planning to enrol abroad in the future.
- Exclude students of which you do not know how many semesters they have spent in higher education (exception: Switzerland).
- Exclude students who are enrolled above the 20th semester.

For the regression model predicting whether students have been enrolled abroad (1.14(m)), there is a further data cleaning rule:

- Exclude students enrolled in the first semester. (Students who are enrolled in their first semester of higher education should by definition not have had an opportunity to enrol abroad temporarily yet. Those students who started their studies in a country other than the one where they obtained their higher education entrance qualification are by definition not captured in any of the EUROSUTDENT surveys.)

Similarly, there is a data cleaning rule that only has to be applied before calculating the model predicting whether students are planning to realise an enrolment period abroad (1.15(m)):

- Exclude students who had already been enrolled abroad temporarily at the time of the survey.

4 Glossary and guidelines for preparation of variables

This glossary only specifies terms which are explicitly relevant for the Steeplechase project. In some cases, the conventions for the Steeplechase analyses differ from the ones elaborated for EUROSTUDENT IV. Since this second draft of the manual differs clearly from the first draft, we ask you to check it thoroughly! In doing so, please pay attention to the precise coding and labelling of the variables needed for the analyses. It is of utmost importance that all variables are coded and – if possible – labelled in the same manner so as for the results to be comparable across countries. To simplify the creation of the necessary variables, we have inserted the STATA commands we used to program the necessary variables in the German data set.
**Age, current:** Age of a student measured in full years at the time the national EUROSTUDENT survey was carried out. In the regression models, the current age centred around the median is used as an independent variable because it will later on ease the interpretation of the models. Please note that the variable capturing the age centred around the median has to be calculated separately for each of the two model series in case the number of observations the two models are based on differ (which will most likely be the case). It shall only be based on those observations which have no missing value on any of the variables included in the models. As it can be assumed that the likelihood to enrol abroad/plan to enrol abroad rises/decreases steadily as the age increases (which e. g. tests with the German data have shown to be true), the current age centred around the median will be introduced as continuous variable in the regression models.

**Age groups:** For the descriptive analyses, three different age groups will be distinguished (please consider the entry on “age, current” as well):

- a) students up to and including 24 years
- b) students between 25 and 29 years
- c) students who are 30 years or older

**Commands**

```plaintext
gen agegroup = age
recode agegroup (17 to 24 = 1) (25 to 29 = 2) (30 to age of oldest student in your sample = 3) (else = missing)
label variable agegroup "age groups"
label define agegroup_lb 1 "up to and including 24 years" 2 "between 25 and 29 years" 3 "30 years or older"
label value agegroup agegroup_lb
```

**Cases, number of:** In the descriptive analyses, we ask you to report weighted percentages as well as weighted absolute numbers (= n). n is the weighted valid number of cases for which the respective answer category of the questionnaire applies (e. g. in table 1.2 the value of n in the first row is the absolute weighted number of female students who have realised an enrolment period abroad). To put it differently: n is also the numerator of the fraction which calculates the percentage in the neighbouring column in the Excel file. In the regression analyses, you shall also calculate the models based on weighted data.

**Child(ren), responsibility for:** Children are any kind of persons, who depend on the student in social and economic ways, e. g. own children, adopted children, stepchildren, foster children, etc. With respect to children, only two groups of students will be considered within the Steeplechase framework:

- a) students with responsibility for child(ren) younger than 18 years (i.e. students who have at least one child which was younger than 18 years at the time of the EUROSTUDENT survey)

---

1 In order to precisely identify the age, refer to the student’s age in the month when the survey was carried out. In case you do not know in which month a student answered the questionnaire and your survey lasted for more than one month, refer to the month in which the majority of interviewees have answered. If the month of birth is not asked for in the questionnaire of your country, refer to the student’s age in the year when the EUROSTUDENT survey was carried out.
b) students without responsibility for child(ren) younger than 18 years (i.e. students who have either no children or children who were 18 years or older at the time of the EUROSTUDENT survey)

If students indicate that they have a child/children, but do not provide the age of the child(ren), please include students who are 35 or younger in category (a) and set students who are older than 35 to missing on this variable.

\textbf{Commands}
\begin{verbatim}
replace childage = 1 if child == 1 & childage == missing & age <= 35

gen child18 = missing
replace child18 = 1 if child == 1 & childage >= 18 & childage <= age of oldest child of a student in your sample
label variable child18 "responsibility for child(ren) younger than 18 years"
label define yesno_lb 1 "yes" 0 "no"
label value child18 yesno_lb
\end{verbatim}

\textbf{Contribution of family/partner/relatives to disposable monthly income:} These are financial means which the student receives from his/her parents, other relatives or the person he/she is sharing his/her life with. This category includes only \textbf{disposable income} which students received from family/partner/other relatives, but not \textbf{transfers in kind}. Transfers in kind will not be taken into account, \textbf{irrespective of whether students are living with parents or not!} This is to assure a higher degree of comparability of data between the countries (exception: Switzerland). The variable “contribution of family/partner/relatives to disposable monthly income” will be used for the descriptive analyses on the likelihood of a student to (not) realise an enrolment period abroad; more precisely, three groups will be differentiated:

a) contribution of up to and including 25%

b) contribution of over 25% and up to 50%

c) contribution of over 50%

\textbf{Commands}
\begin{verbatim}
gen incpr_s = summed income by parents, partner and relatives / total monthly income from all sources
label variable incpr_s "Contribution of family/partner to monthly income"

gen incpr_s_gr = missing
replace incpr_s_gr = 1 if incpr_s >= 0 & incpr_s <=.25
replace incpr_s_gr = 2 if incpr_s >.25 & incpr_s <=.50
replace incpr_s_gr = 3 if incpr_s >.5 & incpr_s <= 1
label variable incpr_s_gr "Contribution of family/partner to monthly income (aggregated)"
label define incpr_s_gr_lb 1 ">=0% & <=25%" 2 ">25% & <=50%" 3 ">50%"
label value incpr_s_gr incpr_s_gr_lb
\end{verbatim}

\textbf{Education (background), tertiary:} Socio-economic background of a student measured by the educational attainment of his/her parents’. It is approximated by the parents’ highest educational qualification according to the International Standard Classification of Education (ISCED 97). The highest educational attainment of either the father or the mother is taken into account. Students whose parents graduated from ISCED levels 5 or 6 are considered as students \textbf{with tertiary education}
Students whose parents graduated from ISCED levels 0, 1, 2, 3 or 4 are considered as students without tertiary education background.

**Commands**

```plaintext
gen tertback = missing
replace tertback = 1 if father or mother with degree at ISCED level 5 or 6
replace tertback = 0 if father & mother without degree at ISCED level 5 or 6

//Attention: If a student indicates that one parent has a degree at ISCED level 5 or 6 and does not know it for the other parents, the student should be treated as having a tertiary education background. If the student does not know which educational qualification his/her parents hold or one parent has an educational attainment below tertiary education level and the student does not know it for the other parent, the student should have a missing value on this variable.

label variable tertback "tertiary education background"
label value tertback yesno_lb
```

**Enrolment period abroad, realised:** This is one of the dependent variables in the Steeplechase analyses. It refers to temporary enrolment phases abroad, i.e. to temporary but regular enrolments in courses at higher education institutions outside of the country where the student was surveyed.

**Commands**

```plaintext
gen enrl = missing
replace enrl = 1 if student has realised a temporary enrolment period abroad
replace enrl = 0 if student has not realised a temporary enrolment period abroad (i.e. including no study-related experiences at all or only other types of study-related experiences such as internships, language courses, etc.)

label variable enrl "enrolment period abroad"
label value enrl yesno_lb
```

**Enrolment period abroad, planned:** This is another dependent variable in the Steeplechase analyses. No matter if students have concrete or rather general plans for an enrolment period abroad, they should be subsumed under the category “planned enrolment period abroad”. Different kinds of plans and no plans have to be aggregated (e.g. the answer categories “yes, definitely”; “yes, perhaps” or “no, not interested”; ”no, I can’t”; ...). The category “I don’t know” should also be treated as no plans for an enrolment period abroad!

**Commands**

```plaintext
gen enrl_pl = missing
replace enrl_pl = 1 if plans for enrolment period abroad
replace enrl_pl = 0 if no plans for enrolment period

label variable enrl_pl "plans for enrolment period abroad"
label value enrl_pl yesno_lb
```

**Field of study:** This variable contains the following subject groups: humanities and arts; teacher training and education science; social sciences, business and law; engineering, manufacturing and construction; (natural) science, mathematics and computing; health and welfare; agriculture and veterinary. These categories are taken from the ISCED classification. Please assign all cases which are currently classified in the category “services” to one of the other categories (e.g. if appropriate, cases from “services” could be assigned to “business”)!
Commands

gen teach = missing
replace teach = 1 if degree is a teaching degree
replace teach = 0 if degree is not a teaching degree
label variable teach "teacher"
label value teach yesno _lb

gen field = missing
replace field = 1 if teach == 0 and all subjects of humanities & arts
replace field = 2 if teach == 1 or all subjects of teacher training & education science
replace field = 3 if teach == 0 and all subjects of social sciences, business & law
replace field = 4 if teach == 0 and all subjects of engineering, manufacturing & construction
replace field = 5 if teach == 0 and all subjects of science, mathematics & computing
replace field = 6 if teach == 0 and all subjects of health & welfare
replace field = 7 if teach == 0 and all subjects of agriculture & veterinary
label variable field "field of study"
label define field_lb 1 "humanities & arts" 2 "teacher training & education science" 3 "social sciences, business & law" 4 "engineering, manufacturing & construction" 5 "science, mathematics & computing" 6 "health & welfare" 7 "agriculture & veterinary"
label value field field_lb

Housing form: For the Steeplechase analyses only two basic forms of housing will be considered:\n
   a) students who are living with parents
   b) students who are not living with parents

In case that there are mixed forms of housing which include also living with parents (e.g. students who live together with parents and partner, or students who live with parents and children, etc.), please always assign those forms of housing to the category “with parents”.

Commands

gen housing = missing
replace housing = 1 if type of housing == with parents
replace housing = 0 if all other types of housing
label variable housing "housing situation"
label define housing_lb 1 "living with parents" 0 "not living with parents"
label value housing housing_lb

Income, disposable: Disposable income is financial means which the student has at his/her disposal; that means he/she is free to choose what to spend it for. This comprises any money in cash or bank deposits the student can use for monthly spending. Total disposable income encompasses financial means from the following sources: a) contribution (provision) from family/partner/other relatives, b) financial support from public sources, c) self-earned income through paid job, d) savings, e) other

---

4 In the EUROSTUDENT context, it is differentiated between five forms of housing: a) with parents, b) alone, c) with partner/child(ren), d) with (an)other person/s (not mentioned before) and e) student hall. The period of time refers to the student’s study term/semester. Vacation periods or any other non-study periods are excluded. Living with parents means living with those persons who are/were in charge of the student, i.e. own parents, step-parents, foster parents, guardians, etc. or other relatives than parents. If the student spent his/her time with more than one set of parents during his/her youth it should be referred to those he/she spent most time with. The category not living with parents comprises the categories “alone”, “with partner/children” and “with (an)other person/s”.

---
sources. That means total disposable income is the result of question 3.5 of the EUROSTUDENT IV questionnaire. Transfers in kind are not included in disposable income. This variable would have been an alternative for self-earned income, which is to be used in the regression models. However, as Switzerland includes transfers in kind in total student income (which are thus also included in the total disposable income), we use self-earned income as a proxy for financial (in)dependence in the regression models.

**Commands**

```plaintext
gen totinc = sum up all kinds of income (in Euro)
label variable totinc "total disposable income"
```

**Income, self-earned:** Self-earned income is income which the student receives from gainful employment. For the descriptive and regression analyses four income groups are distinguished:

1. **No self-earned income**
2. Amount of self-earned income is within first tercile of income distribution (income ≤ lowest 33.3% of income distribution)
3. Amount of self-earned income is within second tercile of income distribution (33.3% < income ≤ 66.6% of income distribution)
4. Amount of self-earned income is within third tercile of income distribution (income > lowest 66.6% of income distribution)

Category a) contains all students who answered the respective questions and who report to have no income from gainful employment. The first tercile then contains only those students who report values for self-earned income > 0.

For the regression analyses, dummies for each of the four categories mentioned above are needed. The analyses for Germany have shown that using dummies instead of the (centred) total income from job proves more beneficial, as there are many students without any income from job whose mobility propensities are partially very heterogeneous.

**Commands**

```plaintext
xtile incjobt = generate 3 income quantiles
replace incjobt = 0 if incjob == 0
lab var incjobt "Income from job in income terciles"
lab def ijt_lb 0 "no income from job" 1 "1st tercile" 2 "2nd tercile" 3 "3rd tercile"
lab val incjobt ijt_lb

//Attention: Please also generate dummy variables for each of the four categories mentioned above.

tab incjobt, gen(incjobt)
```

**Migration background:** Within the Steeplechase framework only two types of migrant students will be considered:

1. Students with migration background (i.e. 1st and 2nd generation migrant students plus students from the category ‘other’)
b) students without migration background.5

**Commands**

```plaintext
gen migr = missing
replace migr = 1 if foreign/changed nationality or two nationalities or father/mother with foreign/changed nationality
replace migr = 0 if local nationality, no change of state and both parents with local nationality
label variable migr "migration background"
label value migr yesno
```

**Obstacle to an enrolment period abroad:** The obstacles to an enrolment abroad are measured on a 5-point scale, which ranges from “no obstacle” or “not at all” to “big obstacle” or “very strongly”, depending on the formulation of the respective question in the national questionnaires. Please ensure that the obstacle variables are coded in such a way that the value 1 means “no obstacle” or “not at all” and the value 5 means “big obstacle” or “very strongly”. For the descriptive analyses dummy variables have to be generated, which summarise the two highest answer categories (i.e. according to the suggested coding the values 4 and 5). In the regression analyses, the obstacle variables with Likert scale of measurement from 1 to 5 will be used. The obstacles considered in the Steeplechase analyses and the labelling we suggest for them are:

- a) expected additional financial burden (obsfin/obsfin2)
- b) separation from partner, child(ren), friends (obssep/obssep2)
- c) expected delay in progress of studies (obsdel/obsdel2)
- d) expected problems with the recognition of the results achieved abroad (obsrec/obsrec2)
- e) insufficient skills in foreign languages (obslang/obslang2)
- f) difficulty in getting information (obsinfo/obsinfo2)
- g) low benefit for studies at home (obsben/obsben2)

**Commands**

```plaintext
gen obsfin = variable on respective obstacles in your data set

//Attention: Please check whether the variables have to be recoded in your data set!

label variable obsfin "expected additional financial burden"
label define obs_lb 1 "no obstacle/not at all" 2 “2” 3 “3” 4 “4” 5 “big obstacle/very strongly”
label variable obsfin obs_lb

gen obsfin2 = obsfin
recode obsfin2 (4/5 = 1) (1/3 = 0) (else = missing)
label variable obsfin2 "expected additional financial burden (dummy)"
label define obs2_lb 1 "big obstacle (4+5)” 0 "no big obstacle (1+2+3)”
label value obsfin2 obs2_lb
```

5 If a student and also both of his/her parents were born in the country of the study programme, he/she is considered to be a non-migrant student. If a student was not born in the country of the study programme and neither were both of his/her parents, he/she is considered to be a 1st generation migrant student. If a student was born in the country of study programme and at least one of his/her parents were born abroad, he/she is considered a 2nd generation migrant student. Finally, if both of the student’s parents were born in the country of study programme but their child was born abroad, this case is classified in the category ‘other’ in the EUROSTUDENT context but as 1st generation migrant student for the Steeplechase analyses. In case the place of birth of either the student or one of the parents is not known, and the place of birth for the person of whom it is know is that of the country of the study programme, the student should be set to missing for this variable.
Please repeat this procedure for all seven obstacles listed above.

**Semester/Study year:** A period of six months at a higher education institution. A semester includes the lecture period as well as the semester break. In case the total number of semesters in higher education was not directly asked for in the questionnaire, the following formula can be used to calculate them: ((calendar year of the EUROSTUDENT survey minus calendar year of entering HE for the first time) multiplied by 2) minus any periods of interruption. Enrolment periods abroad and holiday semesters are no interruptions and must, therefore, be counted as semester(s) in higher education.

In the descriptive analyses, the following three groups shall be considered:

a) students up to and including the 4th semester  
b) students between 5th and 8th semester  
c) students above 8th semester

**Commands**

```bash
gen semgroup = missing
replace semgroup = 1 if sem >= 1 & sem <= 4
replace semgroup = 2 if sem >= 5 & sem <= 8
replace semgroup = 3 if sem >= 9 & sem <= student with the highest number of semesters in your sample
label variable semgroup "Semester (grouped)"
label define semgroup_lb 1 "up to 4th semester" 2 "between 5th and 8th semester" 3 "above 8th semester"
label value semgroup semgroup_lb
```

For the regression analyses, we use dummy variables for the first ten study years, i.e. we aggregate students in the first and the second semesters, in the third and the fourth semesters, etc. Please note that in the case of the model for realised enrolment periods, the first study year comprises only students in the second semester, as students in the first semester are excluded categorically (see section 4).

**Commands**

```bash
gen sj_1 = missing
replace sj_1 = 1 if sem == 1 | sem == 2
replace sj_1 = 0 if sem != 1 & sem != 2

gen sj_2 = missing
replace sj_2 = 1 if sem == 3 | sem == 4
replace sj_2 = 0 if sem != 3 & sem != 4
etc.
```

**Transfers in kind:** Within the EUROSTUDENT framework transfers in kind are considered to be living costs and study-related costs that are paid by parents/partner or others for the student (e.g. a student is not living with his/her parents and the parents pay the rent for their collegiate child directly to the landlord; in this case the financial support is intangible to the student). Note: For the Steeplechase project transfers in kind will not be taken into account, irrespective of whether students are living with parents or not (exception: Switzerland). Instead, for the descriptive and regression
analyses we refer only to the **disposable income** of students. This is to assure a higher degree of comparability of data between the countries (only Switzerland will deviate from this convention).

**Transition time:** Period of time between receiving the higher education entrance qualification and entering higher education for the first time. See entry for **Delayed transition student**.

**Universities/other types of higher education institutions:** Higher education institutions offering programmes at ISCED level 5A can be differentiated in several ways. It can be distinguished, for instance, between public and private institutions or between universities, universities of applied sciences and further types. For the Steeplechase project, the only differentiation is between universities (no matter if public or private) and other types of higher education institutions (e.g. universities of applied sciences, etc.).

**Commands**

```plaintext
gen hei = missing
recode hei = 1 if university
recode hei = 0 if all other types of higher education institutions
label variable hei "type of HEI"
label define hei_lb 1 "university" 0 "other type of HEI"
label value hei hei_lb
```

**Weighted valid n:** For the descriptive analyses we ask you to deliver the absolute data (=n) as weighted valid data, i.e. missings and "I don't know"-answers are excluded (see also **Cases, number of**). An exception is the variable on plans for enrolment period abroad which includes "I don't know"-answers under the zero value.

**Weighted valid percent:** For the descriptive analyses we ask you to calculate the percentages as weighted valid percentages. This means the weighted number of cases on which a percentage is based do not include missings or "I don’t know"-answers. An exception is the variable on plans for enrolment period abroad which includes "I don’t know"-answers under the zero value.

## 5 Guidelines for analyses and data delivery

### 5.1 General guidelines

Next to this document, the Excel data delivery files form part of the manual for the Steeplechase analyses. We ask you to fill out every empty cell in these documents surrounded by a black frame. To avoid confusion, the data delivery file has the same structure as in earlier versions, that is to say it contains a first set of sheets on enrolment as well as plans for an enrolment period abroad (1.1 to 1.15) and a second set of sheets on students’ perception of obstacles to an enrolment period abroad (2.1 to 2.8). The regression models can be found in sheets 1.14(m) and 1.15(m). The latter model is an attempt to combine the perception of obstacles to an enrolment period abroad with the actual propensities to plan an enrolment period abroad.
On each Excel sheet, we have listed which conventions have to be applied before calculating the data and instructions that should be followed whilst calculating the respective figures. These remarks should be understood as points we considered as particularly important for a certain topic. They should not be understood as a comprehensive list of rules to be applied for the calculations. Instead, they always have to be read in conjunction with the specifications made in sections 3, 4 and 5 of this document.

Your tasks in a nutshell:

- Before carrying out the calculations, please make sure that all your calculations refer to the Steeplechase target group specified in section 3 of this document.
- Also, please make sure that all variables you use are coded in the manner we have suggested in section 4 of this document. Where you deem it necessary, please make transparent which decisions you have taken in preparing your data for the analyses (e.g. please inform us where you have aggregated response categories, dropped cases and how you treated missing values).
- Please follow the topic-specific “conventions to be applied before calculating” and “instructions for calculations” you can find in the Excel sheets.
- Please indicate on each Excel sheet where data are of limited comparability in your view and where they cannot be delivered (e.g. because a specific information was not collected in your survey). At the end of the data delivery process, each framed cell of the Excel file should be either filled with a number or be marked in dark grey if data cannot be delivered.

5.2 Guidelines for descriptive analyses

The descriptive analyses shall provide an overview of general differences in terms of both the likelihood to (not) enrol abroad and to (not) plan to enrol abroad as well as the perception of obstacles (i) between countries and (ii) between selected groups of students within and across countries. Other than previously communicated, these figures shall not only be calculated for student group C, but for groups A and B as well (see section 2 for a definition of student groups A, B and C). Descriptive figures shall be calculated for all variables that are included in the regression analyses.

Once the requested descriptive figures are delivered, they will be imported through standardised formulas in Excel master sheets we have programmed. These master sheets will automatically test whether the observed differences between countries and groups of students are statistically significant. To this end, the formula presented at the meeting in Hannover in December 2011 will be used. In order for our master sheets to function, it is crucial that all figures are entered in the cells with the black frames; the position of these cells should not be changed.

Your tasks in a nutshell:

- Please prepare the data and calculate the respective figures for data sheets 1.1 to 1.13 as well as 2.1 to 2.8.

5.3 Guidelines for regression analyses

While the descriptive analyses are meant to provide some first suggestive evidence, the regression analyses are intended to determine which weight different factors actually have in deterring students from realising and from planning an enrolment period abroad. With regard to the regression analyses,
major changes have been made. As the quality of the regression model with the perceived obstacles as dependent variables where of limited quality in the three countries who have calculated them, we have decided to drop them entirely. Instead, we have decided to slightly expand the descriptive analyses on the obstacles (they are now captured for both students without enrolment who are planning and who are not planning to enrol abroad temporarily).

We are now calculating only two regression models, one with enrolment vs. no enrolment and one with plans vs. no plans for an enrolment period abroad as dependent variable. Especially the second model should be of good quality, as a (i) number of theoretical assumptions of regression analyses are met (cause – i. e. student attributes and obstacle perception – are captured before the expected effect (enrolment abroad) instead of retrospective capturing of the effect), (ii) we have an opportunity to introduce the variables capturing the financial situation and the obstacles to an enrolment period abroad as independent variables in the regression and (iii) our dependent variable is not as skewed as in the case of enrolment abroad as well as no enrolment and no plans, respectively. Those countries capturing the perception of obstacles to an enrolment period abroad for students without enrolment planning an enrolment period abroad through separate questions, will have to merge the requested information in single variables. Should there be problems in this process, please let us know.

Both regression models we suggest to calculate are logistic. Other than previously communicated, we only ask you to report us odds ratios (and no standard errors) for the individual independent variables. In case your statistical software package reports logged odds instead of odds ratios by default, please make sure you convert them correctly.

**Your tasks in a nutshell:**

- Please prepare the data and calculate the respective figures for data sheets 1.14(m) and 1.15(m). In doing so, please note the following.
- Please make sure that the individual model steps in model 1.14(m) and 1.15(m), respectively, are based on the same number of weighted observations, i. e. exclude all observations which have a missing value on any of the independent variables included in the last step of the models!
- Please deliver us the exact criteria to appraise the fitness of the models we have listed in the Excel sheets (Pearson’s chi-squared test, the Pseudo R2 according to Nagelkerke and the Adjusted Count R2). Should you not know how to report one of these criteria of model fitness, please let us know.
- Please perform at least some regression diagnostics, i. e. check whether the relation between the independent variable and the dependent variable is more or less linear (which should be the case even in logistic regressions). Also, please eliminate drastic outliers with a too strong influence on the regression coefficients.
- Please indicate for every regression coefficient whether it is significant or not. Please differentiate between significance at 5%, 1% and 0.1% level. Insert one plus/minus sign (+/-) in the fields where coefficients are significant at 5% level, two plus/minus signs (++/-) in the fields where coefficients are significant at 1% level and three plus/minus signs (+++/---)where coefficients are significant at 0.1% level. One or more plus signs stand for an odds ratio above the value of one, while one or more minus signs stand for an odds ratio below the value of one.
• If possible, you can deliver us a log file of your regression analyses next to inserting the requested information in the Excel sheets. This will make it easier for us to interpret the models in such a way that we are sensitive to their strengths and potential shortfalls.