



Lifelong
Learning
Programme



BMB
Bundesministerium
für Bildung



This project has been funded with support from the European Commission and the Austrian Federal Ministry of Education.

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

A randomised control trial of efficacy of a case management approach on the early school leaving rate of 14-17 year olds in Spain and Italy: Aftermath lessons learnt

P4 - Centre for Social Innovation (ZSI)



1. Introduction

Although randomised control trials – like the one implemented in the evaluation of the Jump@School pilot – are regarded as the most powerful research designs to establish causality as well as effectiveness of interventions as has been proven by the evaluation of the Jump@School intervention, they are not without challenges. Such evaluations are demanding as they require early and detailed planning coupled with the fact that they are quite complicated to design and implement especially when implemented by different people in different contexts (need for both standardisation and context sensitivity). Needless to say, implementing such an evaluation in the framework of funded projects with tight schedules complicates the situation even further.

This report presents the reflection from an evaluation point of view on the challenges faced as well as the lessons learnt in the implementation of this evaluation design.

The report is divided into five main broad areas that form the framework of randomised control trials:

1. **Evaluation design:** Objective, intervention, target group, financial resources, time resources and funder requirements.
2. **Participants:** Eligibility criteria, sampling method and randomisation.
3. **Apparatus/survey methodology:** Hypothesis, appropriateness for target group (including suitability, pretesting and translations), setting of data collection (including location, time and mode), comparability and supplementary analyses.
4. **Procedure:** Ethical review, informed consent, anonymisation, blinding and roles and responsibilities
5. **Analysis:** Data quality, data delivery and presentation of results

Each section presents a short introduction then explains how the specific aspects were handled in Jump@School. Following this, a learning or ‘tip’ for future studies is provided.

For detailed information on the Jump@school project, the experimentation, the design and results of the experimental evaluation, please consult the website www.jumpatschool.eu. Here you will also find the final publication, the operational model of intervention and the report on evaluation design & data collection instruments.

2. Evaluation design

The Jump@School intervention implemented a ‘two-group pretest-posttest design’ whereby in each implementation school, the eligible sample was randomly assigned to either the intervention group – this group was provided with the services and support of the Jump@Operator (the intervention) – or the control group – this group was similar to the

Results of the impact analysis in brief...

- *There is no firm evidence that the intervention impacted the grades that students received at the end of the school year. Small differences in favour of the intervention group were obtained in two schools.*
- *Such a mixed picture is mirrored with regard to the impact of the intervention on the soft skills of the students. In some items the intervention group significantly improved compared to the control group while in others the change was in the opposite direction. In many items no change is observable.*

intervention group on a number of variables with the underlying difference being that they did not receive the intervention. Both groups filled in the same questionnaire at two points in time (pretest and posttest) and their grade point averages (GPAs) were collected at four points in time: June/July 2015 (baseline), December/January 2016 (pretest), June/July 2016 (posttest) and December/January 2017.

The challenges or lessons learnt in the section below represent questions regarding the design of the evaluation itself including the target group, the objective of the study, the logic model of intervention as well as the time and financial resources available which are in essence requirements or restrictions imposed by the funding organisation of the project.

2.1 (In)direct relationship of dependent and independent variable

Impact analyses resulting from experimental designs (designs that include an element of randomisation) are usually straightforward. The results of such evaluations are obtained by measuring the treatment/causal effect. This is the comparison of the observed difference on the dependent variable after an intervention and the difference that would have been observed had the intervention not taken place (the counterfactual scenario). As a result, the independent and dependent variable need to have a direct cause-effect relationship.

The main dependent variable of the Jump@School study was the grade point average (GPA) while the activities of the intervention in combination were the main independent variable. The intervention was made up of three main parts: for each participant, at least four individual counselling sessions, a support action on self-esteem and motivation and two workshops whose topic varied from school to school depending on the expertise of the Jump@Operators as well as the evaluation of the collective needs of the students.

Other than specific individual sessions that might have dealt with topics that directly affect the GPA for example tuition on a specific school subject or the 'learning to learn' workshops carried out in Spain, the other activities of the intervention appear to have only had an indirect relationship with the outcome variable – the GPA. This to some extent explains the marginal changes that were detected on this variable.

The GPA was chosen as the main independent variable because it provided an independent measure that overcame the limitations of self-assessment questionnaires such as: threat to internal validity, practice effects, regression to the mean, self-report retrospective data, social desirability and cognitive dissonance. In addition, because it is an interval-scaled variable, it provided the possibility of complex statistical analyses and models such as multiple regressions.

Next to the GPA, the items of the 'soft' questionnaire which reflected the objectives of the intervention more accurately or in other words, which were more directly linked to the dependent variable were also considered as dependent variables. The items of the 'soft' questionnaire covered the following seven dimensions: school motivation & valuing learning in school, engagement with learning, self-confidence with learning, commitment to complete an education, anxiety & uncertainty control, self-regulation & control and withdrawal.

The results from these items varied from country to country and even between schools. Some of the items show significant improvements in the expected direction, while others show significant changes in the opposite direction to the expected and in others the status quo is maintained. This suggests that the intervention was not as standardised as intended, and this was exacerbated by the fact that the workshops or the intensity of the sessions for instance

were not standardised. It is difficult enough to standardise an intervention being implemented in just one place by two different professionals; however, having four sites with two professionals implementing the intervention per site convolutes the situation even further. The implementation in the four different implementation sites needed to be as standardised as possible for the true impact of the intervention to be detected. A fidelity study of the implementation was conducted in the implementation sites to shed more light on what aspects of the intervention elicited changes and the respective changes (see the operative model).

The considerable numbers of items showing no change from pretest to posttest or which show no difference between the groups questions whether the instrument was ideally suited to measuring the targeted outcomes of the intervention. Due to time restrictions and the sequence of the activities of the project, the soft questionnaire, which was the main evaluation instrument of the experiment, had to be developed parallel to the development of the logic model of intervention. Ideally these two activities should have taken place one after the other with the soft questionnaire being primarily based on the intended outcomes of the intervention outlined in the logic model of intervention.

→ *Ensure you choose dependent variables that have a direct relationship with the independent variables*

→ *Defining an intervention and evaluating it through such a thorough method like the counterfactual impact analysis within a short time (3 years) is very challenging. A short funded project cycle should focus on either the development of an intervention or its evaluation. It would be recommendable:*

- *either to have a longer project duration where the intervention is defined first and then the evaluation model is developed and implemented afterwards;*
- *or if only limited time is available, like in most funded projects, to implement and counterfactually evaluate an existing measure that has not yet been thoroughly evaluated.*

2.2 The content of the intervention

On one hand, as thoroughly discussed in different settings during the extent of the project and also as different studies have shown, early school leaving is a complex phenomenon with no simple, 'one-size fits all' explanation. On the other hand a complex evaluation method such as counterfactual impact analysis is best applied to a simple problem and straightforward intervention.

Considering the complexity of early school leaving and the fact that its causes are very individual and vary from student to student, it is generally difficult to come up with a simple enough measure that can tackle the problem; let alone a simple enough measure that can be evaluated easily and in a straightforward manner using a counterfactual method. In the development of the intervention, the Jump@School consortium embraced the fact that there is no single solution for early school leaving and consequently developed quite a complex intervention that dealt with as many risk factors as possible integrating both individual and group activities. In discussions with different stakeholders after the pilot as well as different investigations carried out ex-post, the validity of the Jump@School intervention as a pedagogical approach was confirmed.

During the Jump@Operator exchange seminar in Valencia in Spain, it was quite clear that there were country differences as pertains their attitude towards different aspects of the intervention suggesting that both were equally important: the Spanish Jump@Operators

emphasised the importance of the individual meetings while the Italian Jump@Operators emphasised the importance of the group activities.

Furthermore, an expert discussion in Austria in the framework of the exchange seminars¹, confirmed the importance of integrating both the individual sessions and the workshops. Specifically they pointed out circumstances under which workshops work well:

- At strategically selected points in time when the need has been determined.
- When the topic is relevant or useful to the target group.
- When the workshops are differentiated according to different target groups – even among the ‘at risk students’ as defined by Jump@School, there could still be differences among the participants.

→ A right balance should be struck between the aim of the intervention and that of the evaluation model. For counterfactual impact analyses, the intervention should be designed in such a clear, simple fashion that the relationship between the cause and effect is straightforward. However, to combat complex problems such as early school leaving, more multifaceted interventions may be better suited.

2.3 Lag between treatment and detection of effect

To be able to measure the effects of an intervention, it is necessary for the outcome to mature and become measurable within the intervention period. In essence, the ‘lag between the time of treatment and the time that the effect becomes detectable should be shorter than the treatment period’ (Glennerster & Takavarasha 2013).

A possible explanation for the results from the Jump@School pilot is that the effects needed more time, than the length of the pilot, to become visible and measurable. This is especially true in relation to the dimensions of the soft questionnaire which covered soft skills like self-esteem and motivation. In comparison to the improvement of hard skills like the development of students’ abilities for instance in mathematics which is considerably straightforward and the effects could be considered immediate, the change in soft skills is more complicated and less immediate. There may also be a time-lag between skill improvements and participants’ becoming aware of such improvements, as in the field of soft skills this requires some experiences and opportunities of self-reflection.

In the case of the GPA, the results showed that there was no firm evidence that the intervention impacted the students’ grades. However, unlike for soft skills as described above, the outcomes related to the GPA would be expected to be more immediate. Nevertheless, to control for this, the GPA from the end of the first semester of the school year 2016/2017 was collected. In three out of the four schools, the average GPA declined from the previous measurement (at the end of the school year 2015/2016); in the first school in Spain, it marginally improved. The analysis of this fourth measurement revealed tactical grading especially in Italy; assessments tests harsher mid-year to possibly encourage students to work harder through the rest of the term. As a result, the results unfortunately cannot prove or disprove this hypothesis. The grades from the end of the school year 2016/2017 would be most compatible.

→ Plan enough time even after the implementation of the pilot for the effects to set-in and even become measurable.

¹ The ‘Austrian expert roundtable’ referred to severally in this report, means a discussion in the framework of the Jump@School ‘exchange seminars’ held in Austria with the programme managers of courses for the completion of compulsory education in the Vienna VHS (adult education centre) in June 2017.

→ *Other than only including measures of self-assessment in the evaluation design, also include ‘hard/observable’ facts like in the GPA in Jump@School or other data sources which can be collected without direct involvement of the participants therefore allowing follow-up.*

2.4 Setting of implementation (within school)

In general, interventions evaluated using counterfactual designs require their implementation not to affect the daily routine of the participants - all aspects should remain the same with the only difference being the injection of the intervention for the intervention group.

Due to budgetary and time restrictions as well as schools requirements and long school days (in Spain until 15:00 hours; in Italy until 13:30 hours), the Jump@School intervention took place both during the normal school hours and where possible, outside the normal school time. As a result of this fact, it cannot be ruled out that this could have had a negative effect on the students’ grades because this means that the participants in the intervention could have missed out some important straightforward academic activities (like missing lessons) which could have contributed to lower grades at the end of the semester or the control group performing better than the intervention group (see the analysis of the GPA).

→ *Make the intervention long enough and plan enough time before the implementation to coordinate with the relevant stakeholders as much as possible to ensure that the daily routine of the participants is not affected much and no unintended effects are generated.*

2.5 Timing of the intervention

As described above the implementation of an intervention measured using experimental methods should not affect the ‘normal life’ of the participants. The Jump@School pilot was implemented in Italy in the second semester of the school year 2015/2016. This translated to the second and third trimester of the same academic year in Spain. This period can be considered a difficult time because students are focussed on preparing for the end of year examination (confirmed in discussions with schools in Italy and by reports from national researchers). As the whole intervention was compressed into 5 months, this meant that quite a lot of activities were implemented during a short amount of time; all these activities were additional to the students’ usual activities considerably increasing their workload and pressure, and could have led to the unexpected results.

Similarly, the effect caused by the time of year could have affected the results as explained by the Austrian expert roundtable:

- At the beginning of the year most pupils are very motivated; while towards the end of the year, they become less motivated. As the intervention and the posttest were carried out in the second half of the school year, this could have played a role in the results.
- At the same time, the opposite could be true. For example in Austria, in the September courses whose exams are in spring, the attentiveness and concentration of the students increase because they know they have to pass the exams. In the spring or summer semester it is more difficult to keep the attention of the pupils.
- This phenomenon could be true for items of the soft questionnaire such as ‘When I get too much homework, I just don’t do it’ and ‘When I’m in class, I often think about other unrelated things’

The Jump@School intervention was designed as a preventative measure of early school leaving. The fact that the pilot was implemented in the second part of the academic year, suggests that it lost its ‘preventative power’. Both the Jump@School project (where absences had to be considered in the sampling) and other studies (e.g. Olson, 2014 and Hoff et al, 2015) have shown that many students drop out of school by not returning for the second part of the academic year.

As the intervention in Spain was implemented in two cohorts and the second cohort received the intervention during the first half of the academic year (2016/2017), it was hoped that the results of these analyses would provide more insights or confirmation or rejection of these hypotheses. Unfortunately due to the small sample size of the second cohort in Spain, it was not possible to prove or disprove this hypothesis as multi-level analysis could not be carried out.

→ *When planning the timing of the intervention consider the timings of the ‘usual’ activities of the participants and the likelihood of critical transitions over the time cycles of the respective context.*

2.6 Dosage and length of the intervention

For counterfactual impact analyses, the dosage or the intensity of the intervention plays a crucial role. The intervention needs to be concentrated enough for the effects to be visible during the available duration. If the dosage, which highly correlates to the length of the intervention, is too low, then the results of the impact analyses will not bear the expected results. Concretely, the Jump@School intervention was implemented over a period of 5 months to different intensities: in Spain for example, all the activities of Jump@School were restricted to 45 minute slots. In Italy, the situation was somewhat more flexible. This duration was further shortened by delays caused by for example obtaining parental consent, school holidays and strikes.

Furthermore, the GPA was considered as the main dependent variable and although one of the main determinants of early school leaving, it is definitely not the only determinant. For the Jump@School consortium, the improvement of the self-esteem and motivation of the students was also one of the key objectives, hence the inclusion of the compulsory support-action on self-esteem and motivation as part of the activities of the intervention. The expert roundtable in Austria also highlighted that working on the self-esteem and motivation of pupils is extremely important, and the inclusion of this compulsory workshop into the Jump@School logic model of intervention was considered a good starting point. Such actions would need to be followed up by concrete, long-term activities and individual plans as well as interventions that also involve pupils’ network. However, such an intervention will often exceed the time and budgetary framework of EU projects.

Moreover, the intervention was based on human interaction: between the students and the Jump@Operators. Like in all forms of human interaction, development of trust or bonding needs time. The Jump@Operators, thanks to their skills and experiences with young people, managed to coin trust relationships with the students within a relatively short time. Considering at the same time that the target group represents students at risk of early school leaving, many of whom may have had years of negative school experiences, less than half a year of intervention despite positive relationships with the Jump@Operators may not be enough to translate to improved grades.

→ *The dosage of the intervention should be high enough for the effects to mature and become measurable within the experimentation period.*

→ *Allow for enough lead time before the intervention to deal with delays or other complications that may arise.*

2.7 Funder requirements

The Jump@School project is a funded project under the LifeLong Learning programme of the European Commission. The project concept was developed as a response to a specific call on social policy experimentation with the aim of supporting the ‘implementation of innovative policy solutions at institutional level to reduce early school leaving, in line with the priorities set out in ‘Europe 2020’ and ‘ET 2020’ (European Commission, 2013).’ For the project to be considered, a sample size had to be defined already in the project application phase before the intervention was designed, the schools were chosen and the local context analysed. This goes against the idea of experimentation where it is of utmost importance to consider whether the intervention is likely to generate impacts that are capable of being detected statistically (the idea of the minimum detectable effect) because the intervention was developed at the beginning of the project implementation phase. Considering the indirect relationship of the independent and dependent variables discussed previously (see above), a larger sample size could have enabled the detection of the true impact of the intervention.

The call specification defined the maximum available budget per project as well as the maximum duration of its implementation. The predefined sample size was one of the aspects that formed the basis of the project budget calculation. In funded projects, both the budget and duration are rather inflexible. As can be seen from the sections above, coupled with the complexity of designing and implementing social policy experimentation, these restrictions can cause numerous challenges that have a domino effect on both the results and the assessment to what extent the results are measurable. Many experiences have shown that strategies against early school leaving, i.e. appropriate interventions, although widely known, mostly fail due to a lack of resources especially financial, time and personnel. Paradoxically, robustly assessing the impact of an innovative social intervention may require experimentation beyond the circumscribed context of the generic 3-year project. Some interventions may need to be incorporated into a school’s or education system’s longer-term processes of change before their efficacy can be assessed.

→ *Consider experimental impact analyses in funding mechanisms that are flexible.*

3. Participants

The participants of the Jump@School randomised control trial or social policy experiment were high school students at risk of early school leaving. This target group was chosen based on the definition of early school leavers of the European Commission’s thematic group on early school leaving (2013): early school leavers are ‘those young people (**18-24 years old**) who leave education and training with **only lower secondary education or less**, and who are no longer in education and training.’ As the Jump@School intervention was a preventative measure, then the participants needed to be younger than 18. As only high schools or secondary schools were included in the experiment, the participants would have attained a maximum of lower secondary education.

The section below describes the challenges and lessons learnt specifically relating to the selection of the target group (eligibility criteria), sampling and randomisation including available data and indicators.

3.1 Definition of the target group

An experiment cannot begin without locating participants who are willing and eligible to participate. If the number of eligible participants is lower than anticipated this can reduce its power dramatically. In Jump@School, it was difficult to find schools big enough and willing to participate to start with. In addition, not all students sampled were willing to participate at all (non-takers) or for the full duration (drop-outs) (see section below on post-assignment attrition).

Once suitable schools had been located, to determine the 'at risk' students, the evaluation team decided on three main indicators based on a thorough literature review: Grade point average (GPA), attendance and disruptive behaviour. These were the closest indicators to define the target group considering that other than the age and gender of the students, no other socio-demographic indicators were accessible due to questions of data protection.

As the pilot was to be run in the same way and concurrently in three different countries, there was need for standardisation. Information received from local researchers indicated that several schools in the participating countries did not collect information on attendance and disruptive behaviour in a format which can be used to identify students at-risk of early school leaving. Therefore for the selection process of at-risk students, sampling and randomisation, the Jump@School project relied solely on the school performance of students measured in terms of their grade point average (GPA). The performance benchmark used was based on research by Allensworth/Easton (2007) which indicate that students with a GPA of 2.0 or less (American system) at the end of their first year of high school should be considered at risk of dropping out. As only the GPA of the current academic year was accessible, for Jump@School, the benchmark was relaxed to comparable GPAs of 2.0 or **at the end of the current school year** (June/July 2015).

Early school leaving is a very complex phenomenon that is not just influenced by grades, but often at times is influenced by other social characteristics such as family situation, income and education of parents, et cetera. As a result of such characteristics, pupils could bring along a 'package' for instance social problems which develop over the course of the year affecting the results. Due to a lack of this data, such effects could not be controlled for in the analysis.

→ *As early school leaving is such a complex phenomenon, schools should draw on existing research to continually document data on its predictors such as social characteristics, GPA, absences and disruptive behaviour to monitor the students and react in a timely fashion to any signs of risk of early school leaving and also contribute to further scientific analysis.*

3.2 Post-assignment attrition

Post-assignment attrition refers to loss of data caused by the abandonment of a trial by participants after their random assignment to different conditions. All attrition lowers statistical power and treatment-correlated attrition of participants from conditions threatens internal validity in randomised experiments because the uniformity of the intervention and control group will not carry over to the posttest.

In Italy, the Jump@School pilot was included as part of the schools' extra-curricular activities (POF - Piano dell'Offerta Formativa). This meant that only the consent of the school directors was required - There was no need to get parental consent for each student to participate in the intervention or in the control group. The situation in Spain was more complicated. In one school, students from both the intervention and control groups required parental consent while in the other school, only the intervention group participants' required parental consent. This resulted, especially in Spain, in the self-selection of the intervention group: Many of the selected participants failed to get parental consent and therefore many of them never started the intervention (non-takers). The hypothesis here is that the participants who failed to get the parental consents were poorer at school than those who managed to get it. However, the data failed to prove this hypothesis: in one Spanish school those with the worst GPA were more likely to take part in the intervention while in the other Spanish school, the opposite was true.

Although Italy experienced lower levels of non-takers, it experienced higher drop-out rates than in Spain: More students started some activity related to the experiment and at the same time more students stopped their participation before the end of the trial. There are two colliding hypotheses here: The first is that the poorer students decided to drop-out of the intervention to concentrate on their school work as the intervention took up a lot of their time and also to some extent took place during normal classes, time which could have otherwise been used for studying.

The other hypothesis is that the poorer students were more likely to stay in the intervention until the end because the intervention provided them with exciting activities compared to normal school work and taking part in the intervention helped them to escape normal classes. This is because activities of the intervention, especially the workshops, enabled the students to improve themselves be it learning new methods of learning, improving their self-confidence etc. through fun and exciting ways like fishing or photography. This is a completely different way of learning compared to normal school work.

With regard to the hypotheses mentioned in the paragraphs above, the data gives a mixed picture. In two schools, the poorer students carried through with the intervention, in one the better students did, and in the last one, the poorer students started the programme but then dropped out.

Some options to overcome the problem of attrition include:

- **Timing of consent:** Getting consent before random allocation → danger of lack of resources due to overconsumption.
- **Replacing drop-outs:** When participants drop out or refuse to take part, they are replaced with new participants who may even be randomly selected from the same pool of applicants → threat to internal validity because of selection bias.

In the case of Jump@School, it was decided to collect the GPAs of all students who were selected to take part in the trial regardless of whether they were non-takers or drop-outs at two points in time: The beginning (first term of the school year 2015/2016) and at the end (end of year 2015/2016) of the intervention to allow intention-to-treat analyses. 'This analysis preserves the benefit of random assignment for causal inference but yields an unbiased estimate only about the effects of being assigned to treatment, not of actually receiving treatment. The inference yielded by the intent-to-treat analysis is often of great policy interest, because if a treatment is implemented widely as a matter of policy, imperfect treatment implementation will occur (Shadish et al. 2002).'

Other than the intention-to-treat analysis with the main dependent variable, the GPA, per-protocol analyses (analysis according to receipt of treatment) as well as treating non-participants differently in the analysis was possible. For the ‘soft facts’ only the per-protocol analyses were possible.

→ *Including ‘hard’ facts or data that can be collected independently of the participants allows the possibility of ‘intention-to-treat’ analyses in the case of post-assignment attrition which would have otherwise not be possible if one only relied on the results of the self-assessment questionnaires.*

4. Apparatus / Survey methodology

The main instruments for the impact evaluation of the Jump@School pilot were the questionnaires on ‘hard facts’ and on ‘soft facts’. The ‘hard’ questionnaire covered: GPA and if possible absences, disruptive behaviour and number of negative grades. The ‘soft’ questionnaire covered the following seven dimensions whose improvement was considered the goal of the intervention:

1. School motivation & valuing learning in school
2. Engagement with learning
3. Self-confidence with learning
4. Commitment to complete an education
5. Anxiety & uncertain control
6. Self-regulation & control
7. Withdrawal

To complement these instruments by offering a formative evaluation of the intervention, the following instruments were also implemented:

- **‘Individual development plan agreement’ (IDPA):** A sort of binding contract between the **Jump@Operators** and the **students** both showing their commitment to achieving the jointly formulated goals.
- **‘Individual monitoring process journal’:** Completed by the **Jump@Operators** after every follow-up session, to document the pupils’ paths
- **‘Workshop feedback form’:** Filled in by the **students** to evaluate the group activities.
- **‘Workshop reflection form’:** Completed by the **Jump@Operators** to evaluate the group activities.
- **‘Experiment monitoring template’:** A monthly report from the national **researchers** to monitor the implementation of the trial.

Although the tools for the formative evaluation listed above posed an extra strain on the target group (either students, Jump@Operators or researchers), they were considered important to give the evaluation team an insight on how far the activities of the intervention were being implemented according to the logic model of intervention or as to how far standardisation of the intervention was successful and provide timely solutions to any problems faced in the implementation.

The challenges or lessons learnt with regard to the survey methodology itself, the appropriateness of the instruments to the target groups, the setting of the data collection and external validity or comparability are described below.

4.1 Appropriateness of questionnaires for students

All the methods of directly surveying the students were through questionnaires. Due to the fact that the Jump@School intervention was tailor-made, no existing questionnaires could be used to measure its exact objectives. As a result, the soft questionnaire was developed on the basis of a review of relevant literature. From existing validated instruments, questions that appeared to be most promising in terms of their scientific quality as well as their content regarding the Jump@School intervention were selected. The Jump@School methodology report as well as the first draft of the questionnaire was peer reviewed by the Italian Institute of Evaluation (IVAL).

Considering the target group in question, the appropriateness of the use of questionnaires for this target group is questionable because of its 'test-like' nature or its similarity to exam conditions which can bring out school-related negative experiences like anxiety. However, at the same time considering the task at hand was to experimentally test the effect of the intervention, other methods like qualitative case studies could not have been implemented here - the use of a questionnaire was therefore a prerequisite.

Furthermore, all the questions required the students to answer the items on a 5-point Likert scale 'strongly disagree' to 'agree strongly' with one exception of 'very unlikely' to 'very likely'. The answering scale was changed from the initially preferred (from a statistical point of view) 7-point Likert scale to a 5-point Likert scale from the observations of the cognitive pretesting (see report on 'Results of the cognitive pretesting and the corresponding changes to the evaluation instrument (2017)'). The 5-point Likert scale was considered more appropriate as it was not only more comprehensible, but it also provided the respondents with a middle or neutral option ('neither agree nor disagree' or 'neither likely nor unlikely'). As no further cognitive pretesting were carried out to assess the appropriateness of the 5-point Likert answering scale on the target group, it cannot be ruled out that it was still difficult for this target group to fill out the questionnaires with no assessment on its degree of agreement to this level.

In addition, to determine the effectiveness, the strengths and weaknesses of the soft questionnaire, cognitive pretesting of the English version of the questionnaire were carried out in Austria with young people fitting the description of the Jump@School target group. The fact that English was not the mother tongue of most of the interview partners was considered an advantage because it determined whether the questions were formulated simply enough for non-native English speakers. Unfortunately due to resource restrictions, it was not possible to carry out cognitive pretesting on the translated questionnaires (Italian and Spanish). This would have been key in further optimising the soft questionnaires. Nevertheless, measures were taken to ensure that the quality of the questionnaire was still high after the translation process: Translation by researchers and quality check by local project managers and third parties.

→ *Checking the quality of the questionnaire before its administration is key. This can be done through peer-review and wherever possible, cognitive pretesting should be carried out.*

→ *Translated questionnaires should also go through a quality check either through review by a third-party or cognitive interviews or both.*

4.2 Limitations of self-assessment questionnaires

Self-assessment questionnaires have their advantages for example they provide the views of the respondents directly rather than proxy questionnaires where these views are reported by someone else. However, self-assessment questionnaires are not short of limitations. The soft questionnaire filled in by both the control and intervention group participants of Jump@School is such a self-assessment questionnaire. These limitations are captured below.

4.2.1 Threat to internal validity and cognitive dissonance

Self-assessment questionnaires which are completed by participants several times may pose a threat to internal validity which is 'the extent to which one can say that the variation observed in the dependent variable is due to the variation in the independent variable' (Verma, 2016). As each Jump@School participant was required to fill in the same questionnaire twice within a span of five to six months, the pretest in particular could have familiarised them with the testing situation or with the intended objectives of the study. This could in turn have influenced their answering patterns distorting the results.

Considering the age group of the participants and that there was a five to six-month time lapse between the measurements, it is unlikely that the participants would have been able to base their scores in the posttest on the pretest.

To control for this bias when interpreting results, the soft questionnaire was not relied on solely but in addition results were also reported on a more objective measure – the GPA. The results of the impact analysis were further supplemented by qualitative focus groups with different stakeholders like school directors, teachers and students after the pilot to get a subjective view of the success of the intervention.

→ *Supplement the analyses of self-assessment questionnaires with objective measures (e.g. GPA) as well as qualitative analyses (e.g. ex-post focus groups with involved parties).*

4.2.2 Social desirability

Another limitation of self-assessment questionnaires is social desirability which describes the tendency of respondents to answer questions in a manner that will be viewed favourably by others (Phillips & Clancy, 1972). As the Jump@School intervention relies on the close collaboration between students and Jump@Operators, it was a possibility that the participants would exaggerate their answers in the soft questionnaire or workshop feedback form or provide socially desirable responding because they believed that they were assessing the performance of the Jump@Operators. To reduce this bias, respondents were assured of an anonymous administration of their data (especially through participant codes) which usually results in a person not feeling directly and personally involved in the answers he or she is giving. Furthermore, the researchers also explained to the participants that these questionnaires did not act as assessments of the Jump@Operators' work per se.

→ *Assure participants of anonymous administration of their data to elicit honesty and avoid social desirability.*

4.2.3 Regression to the mean

Another limitation when observing repeated measurements from the same subjects is regression to the mean. Extreme scores on the pretest (either very high or very low) are likely to be followed by less extreme scores closer to the true mean on subsequent measurements; by chance regardless of the treatment (Field & Hole, 2003). Many of the

results of the analysis of the soft questionnaire showed a negative effect of the treatment. A common cause of this is that during the first measurement the students could have had an unrealistic idea of their abilities, attitudes and so on and as a result, chosen the extreme answer options. It could be the case that through the intervention, the students' ability to self-reflect improved and as a consequence, they were able to reflect more and gain a more realistic picture of their abilities, knowledge etc. resulting in poorer assessments in the posttest showing a negative change whereas in essence it is a positive change that the students were able to reflect more. A third measurement could have proved or disproved this notion.

→ *The random allocation of participants in experimental designs limits the effects of regression to the mean. Nevertheless, it should always be controlled for.*

4.3 Nature of pre-post measurements to elicit reflection

Jump@School implemented a pre-post design as explained in the sections above. This means that the participants were measured using the same data collection instrument, the 'soft' questionnaire, just before the start of the intervention and right after. Before-and-after measurements are often characterised by reflection or so to say a learning effect. At the beginning, many participants are likely to unconsciously exaggerate their abilities. However, after certain interventions, like the one of Jump@School where needs assessment is integrated, participants are likely to reflect not only on their needs, but also on their abilities, wishes and attitudes. As a consequence, the results of the pre-measurements are likely to be higher than those of the post-measurements which more correctly reflect the participants' abilities. As a result, although it could not be proven in the framework of this experimentation, this could explain some of the significantly negative results of the intervention group. In this case, these are not negative results per se but rather positive results, as it could be argued that the intervention elicited more reflection, honesty and therefore more realistic assessments of the participants' abilities. A third point of measurements would have possibly revealed the 'true' effect of the intervention in this case.

Correspondingly an intervention like Jump@School could have a similar effect against mobbing where interventions elicit an increase in the feeling of being mobbed, but after some time, the true effects of the intervention start coming to light with the decrease in the feeling of being mobbed. This could be true for instance for the statement: 'I feel like I don't belong to this school' which observed a statistically negative result. Due to the exposure of a new learning environment through the Jump@School intervention, the students' immediate feeling could have been that of not belonging to the school. It cannot be ruled out however that once the students had found a way of integrating what they had learned in Jump@School in their normal school learning environment, they could have felt that they belonged more.

→ *When interpreting the results of impact analyses take into account effects of the evaluation model itself e.g. learning effect caused by before-after measurements. Wherever possible, include a third measurement point for the self-assessment questionnaire to control for this.*

4.4 Timing of the posttest

The administration of the posttest coincided with the end of the school year where either the students are preparing for their end of year examinations or they are in the process of taking them. Subsequently, in this time students are usually characterised by heightened levels of stress and anxiety which were presumably increased by additionally having to complete the

posttest. This could not only have influenced the results of the posttest where more negative results than expected were observed.

→ *In future studies, such factors should be taken into account and assessments or interventions in general should be planned in a way that they do not increase the stress levels of the target group.*

4.5 Varying exam assessments depending on time of year

Several studies (e.g. Docan, 2006 and Cullen et al., 1975) have shown that different examinations are assessed according to different degrees of severity depending on the time of year. For example, mid-year examinations tend to be assessed harsher by teachers so as to motivate or push students to study harder for the end of year exams. As already mentioned, both the mid and end of year grades of the students were collected during the lifetime of the experiment. To determine the eligible sample, the GPAs at the end of the school year 2014/2015 were collected (baseline). The GPAs from the assessments in the middle of the school year 2015/2016 were then regarded as the pretest GPAs while those at the end of the same academic year were regarded as the posttest. Due to this fact, it was decided to present only the results of the comparison of the end of year examinations; the baseline GPAs and the posttest GPAs in the evaluation report. The pretest GPAs were considered in analysis done with different regression models but only bore negligible differences.

It is not only enough to measure the medium-term effects of the intervention, but also to make a fitting comparison of the mid-year results. Hence it was decided to collect the GPAs in December 2016/January 2017 (at the middle of the academic year 2016/2017). The inclusion of the GPAs from mid-school year 2016/2017 in the analysis (see report on impact analysis) revealed tactical grading hence proving this hypothesis. In Italy, especially, there is a clear trend that mid-year exams are graded harsher than the end of year ones. As a result, assessments from different times of year are not directly compatible and should be handled cautiously.

→ *When interpreting the results of impact analyses that use the GPA as a dependent variable, remember to consider different assessments depending on time of year and therefore ensure that the results of the most similar examinations can be compared e.g. end of year with end of year rather than end of year with mid-year.*

4.6 Maturation

Another factor that could have influenced the responses of the participants with regard to the soft questionnaire is maturation. The age group of the participants was between 14 and 17 and the two measurements were taken five to six months apart. For such an age group many changes in attitudes occur within relative short periods of time and therefore prediction of possible outcomes can become quite challenging. However, the fact that all the participants grew older by six months between the two measurements is a control for the changes due to maturation.

Although this factor does not play a role in the results of the comparisons of the differences between the changes in the control group and that of the intervention group from the pretest to the posttest, this phenomenon can be observed in a number of results where both groups change in the same direction and to a similar extent from the pretest to the posttest and should therefore be considered when interpreting the results. Some of the items where this ageing phenomenon could have played a role include:

- ‘When I get too much homework, I just don’t do it’ – When young people become older, they are more autonomous and can more readily decide whether they for example do their homework or not.
 - The results on this item could also be related to absences. High absences mean an accumulation of homework which could lead to the pupils being overwhelmed and therefore decide not to do it altogether.
- ‘I often feel like giving up on school’ – Shows that one does not always have the ambition to pull through and the older one becomes, the more one becomes ‘tired’ of school and ready to do something else.
- ‘In school I work only hard enough to receive a passing grade’

→ *Maturation is an important factor when carrying out studies with young people. Even though theoretically in a social experiment all participants would have aged the same from pretest to pretest, this factor can be useful when interpreting group results from pretest to posttest.*

4.7 Consumption of other interventions

In many schools and especially in countries where the rate of early school leaving is relatively high, there has been an increase in support and offers for students to encourage them to complete their education. These include for example introduction of school psychologists, social workers, work experience and many more. The Jump@School intervention was implemented in two countries with some of the highest rate of early school leaving in Europe: Italy (about 18% - 25% in Sardinia - in 2011 and 14% in 2016) and Spain (about 26% in 2011 and 19% in 2016)².

Although no concrete data or information is available, from the changes in early school leaving rate from 2011 to 2016, it can be assumed that before and during the implementation of the Jump@School intervention, these schools implemented their own interventions to combat early school leaving. Unfortunately, within the framework of this project, it was not possible to control whether the control or intervention group participants were exposed to other offers either just before or during the Jump@School intervention. In any case, if a majority of the participants (both control and intervention group) were exposed to interventions aiming at the same goal as Jump@School; reduction or rather prevention of early school leaving, it could be an explanation of the unexpected results of Jump@School – where on many items, the difference in the changes between pretest and posttest between the two groups was not significant.

→ *In as much as possible, schools should document whether and which offers outside of the normal school system each student is taking part in. This information is not only very valuable for researchers carrying out impact analysis of interventions within schools but it could be helpful for teachers and school directors to determine what works and what doesn’t in the specific context.*

5. Procedure

Randomised controlled trials require early and detailed planning and are quite complicated to design and administer especially in the framework of funded projects. Programme managers may face significant challenges in implementing them correctly; and where they are implemented in different places at the same time, standardisation of intervention delivery is a

² Source: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=edat_lfse_14&lang=en

big issue. Some of the main aspects that fall into the procedure of designing and implementing social policy experiments like the one implemented in Jump@School include: obtaining ethical approval from an ethical commission as well as informed consent forms from the Schools, the participants and/or their parents, anonymisation, blinding and distribution of the different roles and responsibilities.

In the sub-sections that follow the experience and challenges faced when implementing the Jump@School intervention in this respect are highlighted.

5.1 Ethical review

Due to the fact that designing and implementing an experimental design was part of the call requirements under which the Jump@School project was funded coupled with the restrictions of a programme-funded project (e.g. on time, funds and operational flexibility), neither the evaluation design nor the logic model of intervention underwent a formal ethical review. This does not imply that no ethical considerations were made in the process. On the contrary, the project consortium through management board meetings or online meetings often discussed ethical aspects of the experimentation and gave them high importance. The evaluation design for example fulfilled the three basic ethic principles that should be followed when carrying out scientific research for human participants as set out in the Belmont report (Commission for the Protection of Human Subjects, 1979):

- **Respect for persons:** individuals are autonomous agents free to choose their participation or lack of participation in a study.
 - Jump@School adhered to this principle in the sense that participation in the study was voluntary and each participant had the right to leave the study at any time.
- **Beneficence:** an experiment should not knowingly harm the participants and should seek to maximise benefits and minimise harm.
 - The Jump@School intervention was based on the 'truancy' programme from Germany which had already been successfully implemented for a number of years and therefore was not expected to be harmful to the participants.
- **Justice:** there should be fairness in distribution of the benefits and harms of a study ('hence the need to recruit participants fairly' (Shadish, et al., 2002, p. 281). A person should not be denied something that would be of benefit to him or her without good reason, and a burden should also not be unduly imposed on a person.
 - Jump@School imposed random sampling and random allocation of the participants and as a result satisfies the first part of the requirements of this principal.
 - Where the Jump@School project fell short is with regard to the second part of the requirements for this principle – denial of the intervention. This is described in more detail in the following paragraph.

→ *Experimental models pose many ethical questions whose discussions and possible solutions should be at the forefront of all the activities.*

5.1.1 Withholding treatment/intervention from the control group

The evaluation design described above requires withholding the treatment/intervention from a fraction of the eligible sample, namely the control group, in order to provide the counterfactual scenario. The counterfactual approach considers the difference of the

outcomes of having benefitted from a programme (the changes in the intervention group attributed to the intervention) with those of having not benefitted from the programme, all else being equal (the changes in the control group in the duration of the intervention without receipt of treatment/intervention).

Interventions in social policy experimentations are usually well-intended and are expected to benefit the recipients. The Jump@School intervention model was developed based on an existing model – the ‘truancy’ programme in Germany – which is well established. As a result, it was ethically questionable to withhold this intervention, which was expected to be successful, from a vulnerable target group (pupils at risk of early school leaving). According to diverse literature and based on examples of other programmes that have implemented a similar evaluation model, there are some solutions to this problem, however not without consequences:

- Offering all participants an intervention prior to randomisation → this may reduce the chance of finding a treatment effect
- Treatment-on-demand (TOD) → bears the risk of self-selection (participants self-select and determine into which group they fall)
- Phase-in: everyone receives the treatment, but entry into treatment is just delayed for a randomly selected ‘control’ group → time span for intervention duration is reduced and highly dependent on available resources

The above options were considered for Jump@School. However, the mere fact that the project is funded under very strict guidelines with a fixed budget and timeline, these options were considered unsuitable. Nevertheless, from the project application phase, the project partners considered possible solutions. As a result, public bodies were involved in the project from the onset and were especially involved in the implementation of the pilot. The idea was to try and win their commitment or consideration to continue with the programme even after the end of the pilot. In Sardinia, this proved successful in that the region of Sardinia is already considering implementing some aspects of the Jump@School intervention in an existing programme aimed at preventing early school leaving.

→ When using a counterfactual approach, consider ways of overcoming ethical concerns especially those regarding withholding treatment from the onset and throughout the project.

On a similar note, due to the high rates of post-assignment attrition in Spain i.e. the loss of participants after the random assignment (about a third of the selected participants in each group), the project consortium took an ethical stand by deciding to run the experiment in the Spanish schools in two rounds to give more participants a chance of benefitting from the Jump@School intervention.

5.2 Informed consent

Informed consent is widely regarded as an essential ethical requirement when running randomised controlled trials like the one implemented in the Jump@School project. In addition, the Jump@School consortium considered informed consent as well as the signing of the individual development plan agreement (IDPA) as an indication of the commitment of the participants to carrying out the activities involved in the Jump@School intervention.

In reality however, the process of obtaining informed consent is very complex and long-winded. This was very clear in Spain where it was the requirement of the schools to get parental consent for each participant (except of the control group participants in one of the

two schools) before the commencement of the study. Due to the nature of experiments and the fact that the participants are required to carry on as normal with the only difference being the injection of the intervention in the experimental group, it was not possible to share all the details of the study with for example the parents and the students. They were provided with information which was somewhat vague and this could have caused them to be suspicious of the project and hence either taking very long to provide the consent or refusing to give the consent at all. The fact that the Jump@Operators and local researchers were new to the schools and that participating in the intervention meant students missing some classes, further complicated this process and resulted in high rates of post-assignment attrition (see section 3.2 above).

To overcome the problem of post-assignment attrition caused by non-takers i.e. refusal to take part in the study before its commencement, many studies suggest obtaining consent prior to randomisation. This obviously poses the danger of ‘overconsumption’, however, depending on the resources available, a second random sampling before randomisation would be necessary to choose only the number of participants that can be supported by the resources available. Consequently, already at the point of obtaining consent, participants should be informed that the resources are limited and consenting to participate in the study does not automatically mean participation.

→ *Consider obtaining consent before random allocation of participants to reduce post-assignment attrition. However, if participation is limited, to avoid over-consumption make sure that all actors are informed that informed consent does not automatically mean participation.*

→ *As the process of obtaining informed consent can be quite long, enough time should be dedicated to these activities to ensure that the implementation or duration of the intervention itself is not jeopardised*

5.3 Blinding

Blinding is a characteristic of randomised control trials aiming at increasing validity of the results by reducing detection and performance bias. Blinding occurs through concealments of some aspects of the experiment for example group allocation.

To a certain extent, blinding goes against the ‘respect for persons principle’ (Commission for the Protection of Human Subjects, 1979) which requires participants in a study to be informed about what the study is about, what exactly they will be involved in and to provide a signed consent form agreeing to participate in the study. However, in some cases where providing the participants with all the information of the study would risk the validity of the experiment, the informed consent can be waived altogether or the participants can be given partial information (Glennerster & Takavarasha, 2013). In such a case according to the Belmont report, no more than minimal risks undisclosed to the participants and an adequate plan to debrief the participants should be put in place (Commission for the Protection of Human Subjects, 1979).

In Jump@School, partial blinding was considered a necessary element because lack of blinding risked the validity of the results. Other than the project consortium, only the school directors were given an insight into all the aspects of the experimentation. It was important to conceal:

- the experimental character of the study i.e. the existence of a control and intervention/experimental/treatment group;
- the criteria for selection – student considered at risk of early school leaving and
- the main goal of the intervention – to reduce the risk of early school leaving.

These elements were concealed because it was considered that if any of these stakeholders were informed about them, they could eventually change their behaviour or attitudes leading to distorted results. Ideally in an experimental situation, all actors continue as normal with the only difference being the introduction of the intervention. In particular, for students to be labelled ‘at risk’ could cause various reactions by students themselves, their parents or teachers: from increasing frustration to a contrary sense of ‘I’ll show them’, from increased pressure and monitoring by parents or teachers to resignation. Such labelling then may be directly harmful to the target group and also distort results.

As has been mentioned previously, the resources were limited and hence pupils were randomly selected to take part in the scheme. Those selected were to be involved to different degrees: the *intervention group*, *control group* and *non-eligible or non-sampled eligible participants*. This was the common message that was supposed to be transmitted by all persons involved in the study.

The blinding of teachers was especially problematic although the logic behind it was based on sound scientific research. Many studies and especially Robert Rosenthal’s study (1969) shows that higher expectations of teachers lead to an increase of performance – a phenomenon called the ‘Rosenthal’ or ‘Pygmalion’ effect. This means for instance that given a piece of real or false information about certain students for example, teachers may unknowingly treat them differently from the rest of the class thus positively affecting their performance. In Rosenthal’s experiment teachers were told that certain students were ‘late-bloomers’, and in Jump@School they would have received the information that students were ‘at risk of early school leaving’. However, on the site, it was noted that this well-intentioned blinding had other adverse effects: resistance from the teachers in some of the schools to the extent that some teachers directly threatened the participants that participating in the Jump@School activities could negatively affect their studies, refusing to allow students to leave class to attend the Jump@School activities, and direct conflicts with the Jump@Operators.

The experience in Jump@School with the teachers poses one of the biggest obstacles in the implementation of the pilot and at the same time presents a dilemma. On the one hand, their cooperation was considered crucial in the implementation of the Jump@School activities and on the other hand giving them too much information about the project and especially its criteria for selection of students threatens the validity of the experiment. A balanced solution for similar experiments would be to introduce the Jump@Operators and local researchers earlier into the school system which is a relatively closed system. They should have enough time to coin trusting relationships with the teachers and the school system and to get to understand and distinguish their individual roles so as to give the teachers some security that they are not taking over their job for example.

Furthermore, the question of debriefing the participants is important and was discussed several times during the management board meetings. This aspect requires a clear commitment to participant information by researchers and other involved parties that is sustained throughout the project.

→ *Blinding is important when exposing all details of the pilot risks different stakeholders changing their behaviour. At the same time, the context of the project requires careful consideration of blinding versus ensuring stakeholder collaboration. A lot of thought should be put into which stakeholders should be blinded and to what degree as well as how and when to debrief the involved parties.*

→ *If an intervention requires introduction of ‘new’ people to the closed school structure, they should have enough time beforehand to form trusting relationships and understand their individual roles to avoid any unnecessary friction.*

5.4 Personnel, their roles and resulting workload

Personnel are one of the most important inputs of a pilot project. From the onset of the project, the design required that each intervention country to have a local researcher and case managers (Jump@Operators). The idea of having different people for these two roles was to separate the data collection, which in an experimentation is quite heavy, from the activities of the intervention. Theoretically, the local researchers were therefore exclusively in charge of the data collection while the Jump@Operators were in charge of the implementation of the activities of the intervention. In Italy it was possible to have one researcher for each school, while in Spain it was only possible to have one researcher for both schools.

The training of the Jump@Operators took place a month before the official start of the project in a 1.5 day workshop. The driving force for the training was the fact that the logic model of the intervention is quite complicated as it encompasses different activities (individual, group, work with the student’s network etc.). Coupled with the facts that the nature of the intervention, the sensitivity of the evaluation design and the high requirements of data collection, the training had to not only contain information on the contents or activities of the intervention, but also background of the evaluation design and the guideline for data collection. It was quite clear from the first training in Sardinia that the Jump@Operators were quite overwhelmed with the tasks awaiting them notwithstanding the fact that they had suitable backgrounds and qualifications: The Jump@School intervention is one of a kind and they had not done this kind of work before. These resulted in some changes in the logic model such as the removal of one workshop.

By and large the Jump@Operators and researchers in the implementation sites with the support of project partners managed to develop teams with a common mission. This worked very well in terms of communication and support. Even in Spain where there were quite a few researcher changes over the lifetime of the project, the team built here was able to support the integration of the new personnel with as little loss of information as possible.

→ *When developing the logic model of intervention, also consult the people who are going to implement it e.g. in Jump@School the Jump@Operators to gauge how realistic and useful the model is. This is especially important when there is a division of labour between implementation and research roles. At the risk of stating the obvious, workloads and work peaks for all roles in the project should be anticipated realistically.*

→ *Training of the personnel involved both in the activities of the intervention but also of the evaluation aspect of the pilot allows a common vision and understanding easing the implementation of the intervention and the data collection.*

→ *Building a stable team ensures sharing of information and avoids loss of information caused by high-turnover rates.*

6. Analysis

Evaluating outcomes means to prove or disprove a causal relationship between the intervention and the outcome measured. The goal of impact evaluation is to reconstruct the counterfactual scenario. To assess the effectiveness of the Jump@School intervention, descriptive and inferential statistical techniques were used. Descriptive statistics, which form the basis of every quantitative analysis of data, help to describe the basic features of the data in the study. Inferential statistics determine if the intervention had a significant effect and eventually how big that effect was.

6.1 Data

The data collected using the different data collection instruments – more so the hard and soft facts – formed the basis of the impact evaluation analysis. Based on the pilot and project timeline, different deadlines were set for the collection and submission of the data for analysis. Unfortunately this process was often delayed due to various reasons such as complications with the schools regarding the collection of hard facts, change of local researchers, workload of local researchers, unavailability of the students during the administration of the questionnaire on soft facts and so forth. The project consortium reacted to these delays by increasing the budget in the implementation sites to make up for the longer working time than foreseen, and extending the Jump@Operator and researchers' contracts to the end of the project. This was very useful as they proved to be a useful resource especially when interpreting the results.

Furthermore, because not all the data could be delivered on time, the data cleaning efforts were increased as the same process had to be repeated numerous times. It would have been much more productive to clean all the data for example all the hard facts or all the soft facts at the same time. A corrective action that had to be taken for the extra GPAs from the first semester of the school year 2016/2017 – here no data was cleaned or analysed until all the data was received.

→ *Working contracts of the personnel directly involved with the implementation and data collection should last longer than the implementation of the pilot; ideally to the end of the project.*

→ *To save resources, data cleaning and analysis should start only when all the data has been received*

6.1 Results

The results of an impact analysis can be considered as the culmination of an experimentation. However as has been proven by Jump@School, what happens after the results is equally or even more important. Researchers often lack enough information on context for example. Sharing and discussing the results of an experimentation with the involved parties starting with the project consortium, which in the case of Jump@School was quite diverse, can go a long way in reflecting on the results and correctly interpreting them. Furthermore, the pilot availed a very rich dataset where different analyses would have been possible. Such discussions aid evaluators in getting more ideas of which other analyses should be carried out or for example which combination of variables the consortium is interested in.

The results also provided a starting point for discussion with people involved in the pilot on a local level for example the students themselves, the Jump@Operators, teachers, school

directors and volunteers. Jump@School made these discussions possible by utilising the framework of the planned 'exchange seminars'. These discussions further supplemented the interpretation of results as well as reflection on the activities of the pilot, what worked well, what didn't and what could have been done better.

→ *A workshop to discuss the preliminary results of the analyses should always be realised to reflect on the results. Other than just the participation of the project consortium in such a workshop, other experts and stakeholders for example in the Jump@School case; the Jump@Operators, researchers, teachers, participants and even parents would be very valuable.*

References

- Colosi, L. & Dunifon, R., 2006. *What's the Difference? "Post then Pre" & "Pre then Post"*. [Online] Available at: <http://www.human.cornell.edu/pam/outreach/parenting/parents/upload/What-s-20the-20Difference-20Post-20then-20Pre-20and-20Pre-20then-20Post.pdf> [Accessed on 30th August 2016].
- Commission for the Protection of Human Subjects, 1979. *The Belmont report: ethical principles and guidelines for the protection of human subjects of research*. [Online] Available at: <http://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/> [Accessed 31 August 2016].
- Cullen, F. T., Cullen, J. B., Hayhow, V. L., and Plouffe, J. T., 1975. The effects of the use of grades as an incentive. *The Journal of Educational Research*, 68, 27-279.
- Docan, N. T., 2006. Positive and Negative Incentives in the Classroom: An Analysis of Grading Systems and Student Motivation. *Journal of Scholarship of Teaching and Learning*, Vol. 6, No. 2, October 2006, pp. 21 – 40.
- Eurostat. 2017. Early leavers from education and training by sex and labour status. [ONLINE] Available at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=edat_ifse_14&lang=en. [Accessed 31 July 2017].
- Field, A. & Hole, G., 2003. *How to Design and Report Experiments*. Wiltshire: SAGE Publications Ltd.
- Hoff, N., Olson, A. & Peterson L. R., 2015. Dropout Screening & Early Warning. [Online] Available at: <http://k12engagement.unl.edu/DropoutScreening&EarlyWarning3-27-15.pdf> [Accessed 31 July 2017]
- Jump@School, 2015. [Online] Available at: <http://jumpatschool.eu/download/public/brochures/Brochure%20EN%20-%20Internal.pdf> [Accessed on 17th of June 2017)
- Kirk, R. E., 2013. Experimental Design. In: S. Publishers, Hrsg. *Experimental Design: Procedures for the Behavioral Sciences*. s.l.:s.n., pp. 23-45.
- Olson, S. L., 2014. Why September Matters: Improving Student Attendance. [Online] Available at: <http://baltimore-berc.org/wp-content/uploads/2014/08/SeptemberAttendanceBriefJuly2014.pdf> [Accessed 31 July 2017]
- Phillips, D. L. & Clancy, K. J., 1972. Some Effects of 'Social Desirability' in Survey Studies. *American Journal of Sociology*, Vol. 77, No. 5, March, pp. 921-940.
- Verma, J. P., 2016. *Repeated measures design for empirical researchers*. New Jersey: Wiley.
- Rosenthal, R.; Jacobson, L., 1968. *Pygmalion in the Classroom*. New York: Holt, Rinehart & Winston.
- Tschank, J., 2017. Results of the cognitive pretesting and the corresponding changes to the evaluation instrument. Jump@School Project.