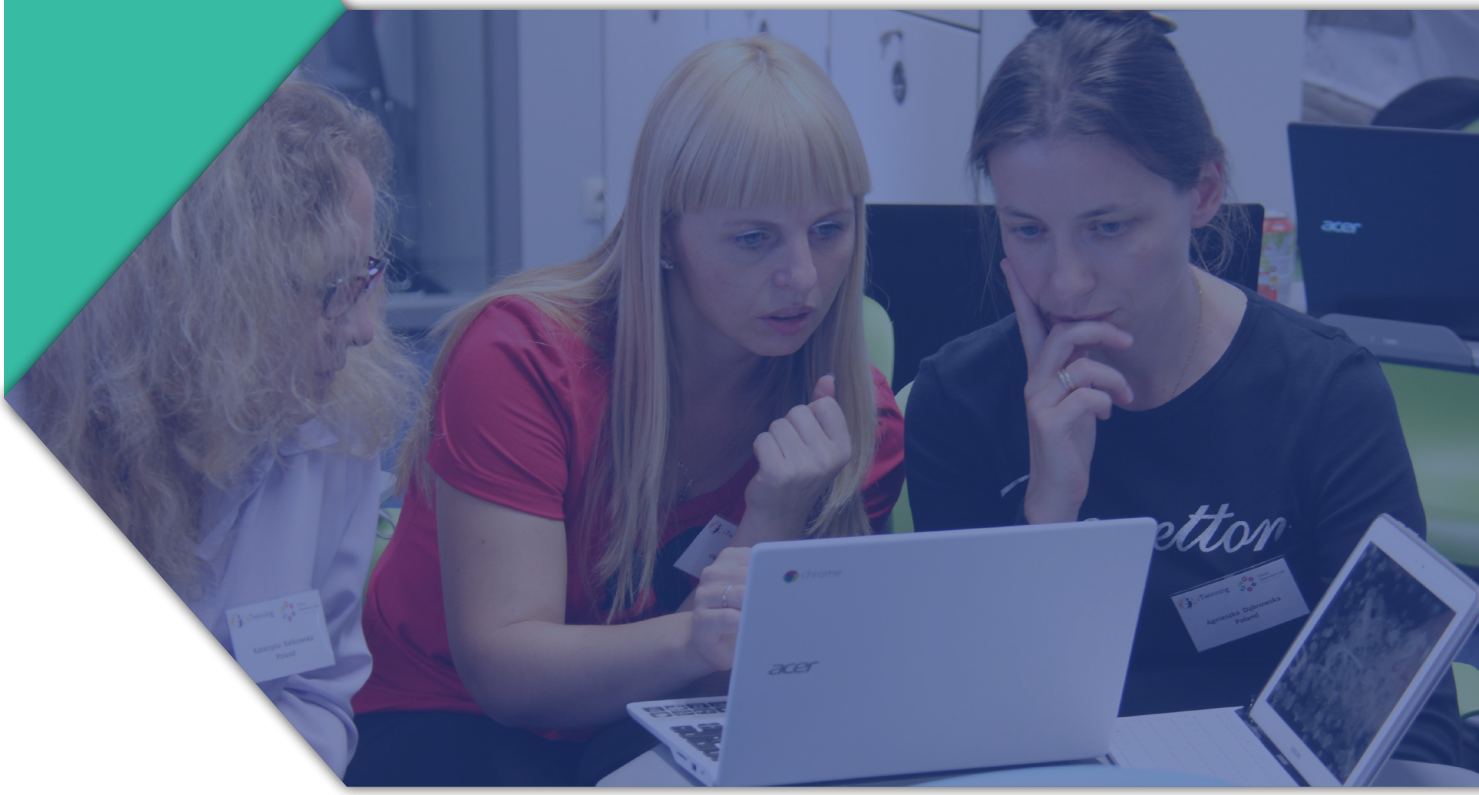
TEACHUP

TEACHUP EVALUATION REPORT



Co-funded by the
Erasmus+ programme
of the European Union

Publisher:
European Schoolnet
(EUN Partnership AISBL)
Rue de Trèves 61
1040 Brussels - Belgium
www.europeanschoolnet.org

Authors:
Davide Azzolini, *FBK - IRVAPP*
Sonia Marzadro, *FBK - IRVAPP*
Enrico Rettore, *FBK – IRVAPP*

Design and DTP:
Jessica Massini, *European Schoolnet*
Andrea Panizza, *European Schoolnet*

Please cite this publication as:
Azzolini, D.; Marzadro, S.; Rettore, E. (2020). TeachUP Evaluation Report. European Schoolnet, Brussels.

Acknowledgements:
The authors would like to acknowledge the contribution of Janet Looney and Karen Triquet. Special thanks go to Micol Beittel, Olga Gorodetskaya and Rodolfo Pezzi for their valuable research assistance. The entire TeachUP experimentation would not have been possible without the input and coordination of European Schoolnet's Patricia Wastiau, Katja Engelhardt, Benjamin Hertz, Alexandre Kirchberger, and Antoine Bilgin. Finally, the authors are also deeply indebted to all TeachUP national partners, who performed a terrific work throughout the entire policy experimentation.

Published in June 2020

This work is licensed under CC BY-NC-SA 4.0. To view a copy of this license, visit creativecommons.org/licenses/by-nc-sa/4.0

European Schoolnet
Transforming education in Europe

FONDAZIONE BRUNO KESSLER
IRVAPP
RESEARCH INSTITUTE FOR THE EVALUATION OF PUBLIC POLICIES

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF BOXES

List of acronyms

BS: Baseline Survey

CDL: Country Dialogue Labs

CPD: Continuous Professional Development

FUS: Follow-up Survey

ITE: Initial Teacher Education

KA3: Key Action 3

MOOC: Massive Open Online Courses

PT: Professional Teachers

RCT: Randomised Control Trial

SRLO: Self-Regulated Learning Online

ST: Student teachers

Glossary

Actions: specific actions implemented by personalised support agents as part of an intervention, for example sending out a reminder email, exchanging in a 1:1 session, or looking at a participant's work. Actions were shaped by the triggers which determined if a participant qualified for an intervention.

Attrition rate: percentage of TeachUP teachers who did not fill in the Follow-Up Survey.

Average treatment effect: the difference in mean (average) outcomes between units assigned to the treated group and units assigned to the control group

Baseline Survey: The online questionnaire that TeachUP teachers filled as they signed up to TeachUP. The questionnaire collected information about teachers' backgrounds, learning experience, and ICT competences. In TeachUP, these data were used to determine which teachers were to be targeted by personalised support.

Completion rate: usually the percentage of enrolled who completed the course. In the TeachUP project the rate was calculated in two different ways: among teachers enrolled in the courses as well as among teachers who actually started the courses.

Collaborative learning: topic of the third TeachUP course. A joint recognition and understanding of the nature of a problem that is being addressed together, as well as communication, negotiation and exchange towards a plan to address the problem, and finally coordinated action to carry out the plan.

Connectivism: a learning theory which posits that in a digital age where knowledge is stored and readily available, learning is less about the acquisition of knowledge and more about the process of creating connections to people and content and being able to use and navigate these connections to access the right knowledge when needed.

Contamination: the possibility that, through interaction, treated teachers pass part of the treatment on to control teachers

Controls (or Control group or comparison group): TeachUP teachers belonging to control schools/ITE organisations.

Creativity: topic of the forth TeachUP course. The ability to produce work that is both novel (or original) and useful (or valuable), and the extent to which this work is creative will depend on whether the novelty and usefulness is perceived by oneself or by the whole world.

Country Dialogue Labs (CDL): series of one-day workshops that took place at the country-level and provided opportunities for collaboration, knowledge sharing, and co-creation between Initial Teacher Education (ITE) and Continuous Professional Development (CPD) organisations and other relevant stakeholders.

Course Moderator: individual(s) who actively animate and moderate an online course. In the TeachUP project the course moderator developed an active learning community where participants support and learn from each other, animated the course portal and Facebook page, encouraged participants to find inspirations and answers from peers

Digital badge: a digital icon (= symbol or picture) or title that shows you have achieved something in an educational context. In the TeachUP project a digital badge was awarded upon successful completion of a course.

Digital certificate: a digital certificate showing that you have completed or achieved something in an educational context. In the TeachUP project a digital certificate was awarded upon successful completion of a course.

Digital competence: Digital Competence can be broadly defined as the confident, critical and creative use of ICT to achieve goals related to work, employability, learning, leisure, inclusion and/or participation in society. Digital competence is a transversal key competence which, as such, enables us to acquire other key competences (e.g. language, mathematics, learning to learn, cultural awareness). It is related to many of the 21st Century skills which should be acquired by all citizens, to ensure their active participation in society and the economy.

Effect of an intervention = Impact

European Schoolnet Academy: platform offering MOOCs for school teachers and other school

practitioners. It is run by European Schoolnet, the network of 34 European Ministries of Education, based in Brussels.

External validity: the extent to which the causal effects estimated in one context hold also in another context or in the overall population the sample was drawn from

Follow-up Survey: online questionnaire participants in TeachUP courses were asked to complete after the end of the experimentation.

Formative assessment: topic of the first TeachUP course. An assessment method which provides feedback to the students being assessed as well as the teacher organizing the assessment to modify how teaching and learning is organized. The method only becomes formative if it is actually used to change teaching and learning in response to the feedback used.

Gamification: the process of adding games or game-like elements to something (such as a task) so as to encourage participation.

Heterogeneity: systematic variations in the effect of the intervention. To investigate this aspect, the sample was split into subgroups defined by some relevant individual characteristics measured at the BS.

Impact of the intervention = Effect

Internal validity: the extent to which the estimated effects are truly causal effects rather than reflecting pre-existing differences between the group of beneficiaries of a given policy intervention (the so-called treatment group) and the group of non-beneficiaries (i.e the control group).

Intervention: two meanings: i) component of the personalised support mechanism. There were nine interventions each consisting of a set of triggers and actions; and ii) more generally, synonyms of policy intervention.

KA3: Key Action 3 provides grants for a wide variety of actions aimed at stimulating innovative policy development, policy dialogue and implementation, and the exchange of knowledge in the fields of education, training and youth, under the Erasmus+ programme of the European Commission. Two main instruments are managed through specific calls for proposals: Initiatives for policy innovation giving support to forward-looking cooperation projects on policy developments, and European policy experimentations led by high level organisations and public authorities to stimulate innovative policies and prepare their implementation.

Learning Analytics: measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.

Learning theories: theories which explain the processes by which humans learn

Course moderator: individual(s) who actively animate and moderate an online course. In the TeachUP project the course moderator developed an active learning community where participants support and learn from each other, animated the course portal and Facebook page, encouraged participants to find inspirations and answers from peers.

MOOCs: Massive Open Online Courses are free online courses available for anyone to enroll. Accordingly, they are designed for scalability and can accommodate large numbers of participants.

Natural completion (or, start) rate: competition (start) rate in the absence of the intervention, i.e. observed in the control group

Outcome(s): characteristic(s) that a given intervention is expected to change

RCT (Randomised Control Trial): a study in which subjects are allocated at random (by chance alone) to either treatment (who receive an intervention) or control groups (who don't) in order to test claims of causal relationship.

Peer assessment: In peer assessment, a collaborative learning technique, students evaluate their peers' work and have their work evaluated by peers. Often used as a learning tool, peer assessment gives students feedback on the quality of their work, often with ideas and strategies for improvement. At the same time, evaluating peers' work can enhance the evaluators' own learning and self-confidence. Peer involvement personalizes the learning experience, potentially motivating continued learning. When used in grading, peer assessment can give the instructor needed information on student performance. Especially for large online classes, it may

allow inclusion of assignments where students' creative work could not be graded reliably through automation or efficiently by teaching staff.

Personalised learning: topic of the second TeachUP course. It refers to a way of learning where the learner has greater ownership over their learning so as to shape the focus and style of learning according to their needs and to achieve greater learning outcomes. It actively engages learners in the learning process and lets them make key decisions about what and how they learn.

Personalised support: support provided to course participants which considers the participants' profile and progression in a course. It is designed to help participants successfully access and benefit from the course contents and community. It was the treatment in the TeachUP policy experimentation, so it was offered only to participants in the treatment group.

Personalised support agents: make interventions for those most in need, i.e. those most likely to drop-out. They monitor and support individual participants by proactively contacting participants to offer them a 1:1 session, to highlight useful learning material, to answer questions, or to develop a learning plan.

Policy experimentation: initiative that helps ministries and government departments test new ways to solve policy problems within a limited scale, and within a set timeframe

Regression Model: statistical model that estimate the relationships between a dependent variable and one or more independent or predictors variables

Reliability: extent to which a measure accurately measures what it is intended to measure

Sampling Strata: in TeachUP, groups of sampling units defined according to the geographical location that have been used to maximise the geographical representativeness of the sample.

Scalable online learning environments: The term "scalable online learning environments" refers to any environment that is designed in such a way that there is no practical, technical, or other limit to the number of learners in the environment. While such environment has the potential to accommodate "massive" numbers of learners, it does not necessarily do so. Massive open online courses or MOOCs would be considered a typical example of such an environment – even though the use of the term "massive" could be misleading in this context as numbers of learners are not necessarily high in numbers. Another example would be mobile learning applications like Edupills or Babbel or social media environments like a Facebook Group.

Scalable online courses: As "scalable online learning environments" but just referring to courses.

Self-regulated learning: is cyclical process by which when faced with a learning goal, students are actively planning, monitoring and adapting (performing), and reflecting on their learning strategies in order to succeed. This involves the establishment of goals and then selecting the right strategies by which to reach them. In doing so, students self-regulate their metacognition (thinking about one's thinking), behaviour (planning, monitoring, and evaluating), and motivation to accomplish the task at hand.

Social constructivism: a learning theory which posits that learners actively attempt to create meaning from experience and that the process of creating meaning is a social process shaped by interactions with others.

Study Buddy: a learning partner who can support the learning process.

Target group/targeted: all TeachUP teachers who were identified as "in-need" of tutoring based on static or dynamic triggers. Only those belonging to the treatment group actually received the intervention.

Targeting: when only a fraction of the teachers, i.e. those most in-need, in the treatment group received the personalised support

TeachUP teachers: Student teachers and professional teachers who signed up to TeachUP and enrolled in at least one course.

Teacher: either student or professional teacher.

Treated (or Treatment group or Experimental group, or test group): Participants belonging to treated schools/ITE organisations.

Treatment: is what the experiment wants to test. In TeachUP was the personalised support

Treatment compliance: percentage of units assigned to the treatment group that actually made use of the intervention offered

Trigger: determined which course participants were eligible for the personalised support, for example a lack of online learning experience. Each trigger had an associated action which was addressing the specific characteristics of the trigger

EXECUTIVE SUMMARY

The Teach-UP project provides interesting insights for policy-makers, teacher trainers, course providers, and anyone interested in finding ways to effectively and efficiently scale up online teacher training. This report provides an extensive and technical account of the evaluation carried out in the framework of the TeachUP policy experimentation. All the other outputs produced in the framework of the policy experimentation can be found on the web page teachup.eun.org.

The policy problem

Online courses have become an increasingly appetible alternative to traditional training for policymakers willing to upskill the teacher workforce. Online provided training allows reaching high numbers of participants in a short period of time and in a flexible way. However, with little or no personalised guidance, learners with little experience in online learning, limited digital competences, or low self-regulated learning competence can be easily overwhelmed by the scale, diversity and flexibility offered by online training. All this easily translates into the **low participation and completion rates** that are typically observed.

The TeachUP policy experimentation

The TeachUP policy experimentation seeks to advance knowledge on what specific support measures increase success in the delivery of online courses for teachers. It did so by **developing and testing an innovative personalised support model** for online courses, which offered a direct and **personalised support** infrastructure that helped participants navigate through the course contents and community.

Four new online courses on topics linked to the key competences for 21st century teachers were developed and addressed to teachers in **Initial Teacher Education** (ITE) and **Continuing Professional Development** (CPD) in nine European Member States (Austria, Estonia, Spain, Greece, Hungary, Lithuania, Malta, Portugal, Slovakia) and one neighbouring country (Turkey).

TeachUP evaluation questions and evaluation design

The TeachUP policy experimentation addressed three main evaluation questions:

1. Does online personalised support increase student teachers and professional teachers' participation in online courses?
2. Does online personalised support improve Self-Regulated Learning Online Competences (SRLO)? If so, does it improve them directly or indirectly (through increased online course experience?)
3. Does peer assessment represent a viable approach to assess learning achievements in online courses?

To provide a robust answer to the first two evaluation questions, the TeachUP policy experimentation features a **randomised controlled trial**. The trial involved **4,090 professional teachers (PTs) and student teachers (STs)** in lower secondary education. To investigate the third question, instead, TeachUP compared the peer and expert assessments of 106 randomly selected course works from the third TeachUP course.

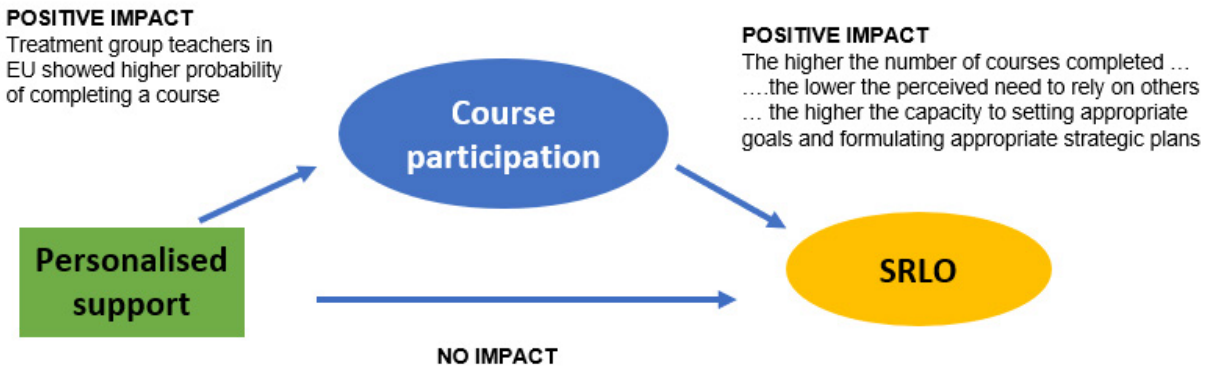
Main findings

The TeachUP online personalised support proved to be an effective approach to increase course participation among **professional teachers in EU Member States**. Teachers **receiving the TeachUP personalised support showed a 10 percentage points higher probability of course completion than control**

teachers, who received no form of support. Beyond being highly statistically significant, this impact is also substantial as it raised completion rates from 32 to 42%. The overall impact on course completion was a result of both support interventions making people complete the courses they started and interventions making teachers actually start the courses they enrolled in.

What made a difference was not the concrete support provided, but the message itself because the 1:1 online sessions that were offered by personalised support agents were taken up by a very small fraction of teachers. Messages likely contributed to **reproduce in a virtual class what happens in a physical class**, letting participants feel that someone was watching over them. This could have taken place through several mechanisms such as prompting; external observation; and guidance.

However, **personalised support did not work equally well for all**. Among student teachers in the EU, personalised support had an impact on completion rates only among those students who joined TeachUP with some past experience in online learning (i.e., teachers who completed at least 1 course per year in the past three years). Also, personalised support had no impact on professional and student teachers in Turkey, who saw no improvement in course completion rates thanks to the intervention. A number of factors (e.g. different teacher profiles, different patterns of course participation, different implementation of the personalised support) could be responsible for these heterogeneous results, but the evaluation does not allow reaching any conclusive answer.



The evaluation also provided evidence of a **positive effect of course participation on teachers' SRLO** as measured via a survey in the months after the last course. The more courses a teacher completed within the TeachUP policy experimentation, the higher his/her ability in setting her learning goals and the lower her propensity to seek others' help to solve problems encountered during the online course. Hence, by increasing course participation, personalised support could also have indirect effects on teachers' SRLO, helping them become more independent learners.

Concerning **assessment**, the lesson plans were overall considered as very good by both peers and experts. However, peers, on average, gave higher scores than experts. This difference was not very big but still statistically significant. The assessments provided by peers on the same teacher's plan were generally consistent, even if there was some variability. Peer feedback on teachers' plans was typically less detailed, less constructive and slightly more positive than feedback provided by experts. However, teachers agreed equally with the assessment received from peers as well as the experts and found them fair and useful.

Policy implications

As regards participation in online training courses, three key recommendations for teacher education and policy making emerge from the TeachUP research results:

1. **Send personalised messages:** Personalising messages is likely to have played an important role in the impact of the messages on course participation. Accordingly, finding means to generate messages that include information about the participants' profile and progress is likely to achieve positive results with regard to participation and completion. Personalising messages does not have to be a complex process requiring substantial data collection. In other words, it is not clear from the TeachUP evidence that a substantial degree of personalisation is required to achieve the results recorded.
2. **Reach out to latecomers.** The TeachUP results show that contacting those who have not yet started a course can have a significant positive impact on course completion (see section 5.2 for more details). Accordingly, finding a mechanism to reach out to those enrolled in a course but failing to start is worthwhile.
3. **Reach out to newcomers.** Research indicates that previous experience of online learning is a determining factor in non-completion and this was confirmed for teachers and student teachers in the TeachUP experimentation. Reaching out to learners new to online learning therefore promises high returns with regard to participation and completion. The reach-out should include some personalisation and an offer of support (given that it is not possible to say if the result would have been the same without these two elements.)

Peer assessment in TeachUP appears as a viable option for assessment, and may support the scalability of large online courses. Key recommendations emerging from the research are to:

1. **Boost assessment cultures** - Training for course participants on how to provide and receive feedback, in general and in online settings in particular, should be envisaged. The benefits of peer assessment both in online teacher training and with students should be further promoted. Finally, guidance to peer learning activities needs to be accessible and engaging.
2. **Emphasise the role of the assessor and of peer dialogue** – Assessing the work of others was identified as a particularly enriching learning experience. To strengthen this aspect, learning opportunities for peer assessors, that are relevant for their own professional development, should be emphasised. Further, assesses could be provided with a possibility to respond to the feedback received, thereby creating a dialogue between peers, that is based on their concrete course works (addressing e.g. respective teaching beliefs). Ideally, this dialogue could also include their actual implementation of new practices they learned about during the course with their students. To that end, the online course would need to allow for sufficient time to try out new practices in the classroom and provide feedback opportunities for participants. Guidance on how to provide effective feedback (specificity of suggestions, tone, and so on), would also be valuable.
3. **Enhance the reliability of peer assessment** – The assessments provided by three peers were generally consistent with ratings provided by experts (demonstrating inter-rater reliability). This underlines the importance of providing a well-designed assessment tool, such as a rubric, setting out clear standards and criteria with descriptors and exemplars of work at different performance levels. Inter-rater reliability may be further enhanced through a training on the use of the assessment tools, and to ensure a shared understanding of performance levels. Other approaches may include: opportunities for peers to assess a subset of the assignment and discuss any discrepancies before assessing the remainder of the assignment: and/or combining peer assessment with expert assessment for a small random selection of the course assignments. Moreover, further research to better understand significant variability would be useful in ensuring the quality of the assessment tool, identifying needs for further training, and helping to reinforce the peer assessment as an appropriate approach for large-scale online learning.

Key areas for further research include the mechanisms behind an effective personalised support and the further exploitation of predictive analysis to improve personalisation of online instruction and support, as well as how to further improve peer assessment mechanisms, in particular in view of the possible accreditation of online courses. Questions to explore in further research are put forward in section 9.3.

1. INTRODUCTION

1.1. The policy problem

Over the last decade, teachers have faced an increasing number of complex challenges in their daily work in the classroom, caused by changes in education policy as well as more general trends in society. Some of these challenges, such as the increasing digitalisation of classrooms, require teachers to fundamentally innovate and adapt their practices. **Teachers require training and support**, if they are to achieve such transformation of their practice. Initial teacher education and continuous professional development (CPD) are crucial in addressing these changes. However, many teachers in OECD countries struggle to access and benefit from the training they need, with very few, if any, improvements registered over the last decade (OECD 2009; 2014; 2019).

Online courses have become an increasingly appetible alternative to traditional training for policymakers willing to upskill the teacher workforce. Online provided training allows reaching **high numbers of participants** in a **short period of time** and in a **flexible way.** Considering that a large majority of teachers in Europe take CPD during their leisure time (European Schoolnet, and University of Liege 2013; Wastiau et al. 2013),[1] online training is more compatible with the teachers' workload and time management.

However, with little or no personalised guidance, learners with low self-regulated learning competence, little experience in online learning, or limited digital competences can be easily overwhelmed by the scale, diversity and flexibility offered by online training. All this easily translates into the **low completion rates** that are typically observed in scalable online courses like MOOCs (Jordan 2015) as well as in online courses delivered by post-secondary education institutions (Lee and Choi 2011). But, before that, this also affects **participation rates**, which are not very high either. In the OECD countries, very few teachers have taken part in induction that includes online courses and seminars (23%) or online activities (20%) and only 36% teachers attend online courses or seminars as part of their professional development (OECD 2019).

In the literature, there is a large consensus about the need to improve the instructional design and the support mechanisms of online courses in order to increase participation and success rates for every learner. Overall, **robust evidence on what works is limited and mostly concerns interventions aimed at increasing retention rates.** Nonetheless, some practices and program features have been identified as promising in the literature. These are summarised in Box 1.1. A more detailed literature review is included in Appendix A.

1 According to the "Survey of Schools", in 2011 around 70% of students at all grades were taught by teachers who have engaged in personal learning about ICT in their own time (European Schoolnet, and University of Liege 2013, p. 89).

Box 1.1 *Promising solutions to increase retention in online courses*

- **Providing advising and guidance support** through technology-mediated coaching services to assist students with goal setting, time management, and other support (Bettinger and Baker 2011; Briggs and Spaulding 2018)
- **Providing orientation programs** that introduce students to the online classes prior to enrolment (Bawa 2016)
- **Using data-driven technology (artificial intelligence, machine learning, and predictive analytics) to provide learners with personalized and equitable support** like improved advising, course scheduling, personalised instruction, etc. (Briggs and Spaulding 2018).
- **Ensuring "live" interaction in computer mediated communication** through friendly media tools that foster well-structured interactions and more social interaction between peers and students-teachers (Bawa 2016).
- **Structuring classes to favor collaborative learning** through the development of instructional strategies that enhance students' guidance, such timeline for feedback, feedback rubric, etc. (Bawa 2016).
- **Enhancing instructors training to teach in a virtual class** through ad hoc training and practice (Bawa 2016).

Among the promising practices listed in Box 1 there are the development of online support mechanisms and the use of predictive analytics to personalise support and instruction. These are two distinctive features of the TeachUP tested model. Providing **suitable 1:1 user support** in the context of scalable online courses is however very challenging (Goel & Polepeddi, 2019). Given that the majority of support queries consist of similar, ever-repeating questions, a possible solution is to outsource 1:1 support to a **machine**. Alternatively, support to learners is considered as a task of the **community**, with course providers encouraging learners to support each other using gamification systems, study buddy concepts, or other approaches that generate an active supportive community. Many course providers also use **moderators** and/or teaching assistants who offer support to users. Unfortunately, neither of these solutions are fully satisfactory in practice. Using the community to offer support only works if the person in need of support has the capacity to engage with that community. Similarly, most machine-based solutions still require knowledge of how to "communicate" with the machine in order to receive satisfactory support, quickly alienating those who lack this capacity (Scott, 2016). Using moderators or teaching assistants to address user support queries can work well depending on the scale of the learning environment but requires substantial amounts of time from those involved.

The TeachUP policy experimentation entered this field by seeking to advance knowledge on what specific support measures could work to increase success in the delivery of online courses for teachers (Figure 1.1). It did so by **developing and testing an innovative personalised support model** for online courses, which offered a direct and **personalised support** infrastructure that helped participants navigate through the course contents and community.

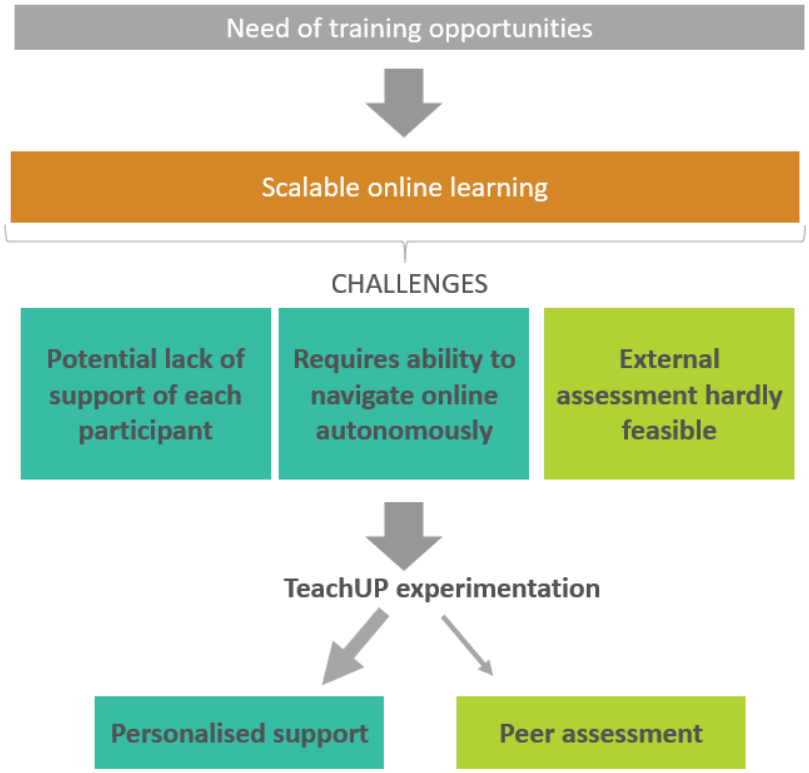


Figure 1.1 The policy problem addressed in the TeachUP policy experimentation
TeachUP pilots and tests personalised support and peer assessment in scalable online courses for teachers

A further critical challenge that was also addressed in TeachUP - although not being the main focus of the policy experimentation - concerned **assessment in online courses**. Traditional forms of assessment (e.g., performed by instructors or external experts) become more difficult and cost intensive in scalable environments. Peer assessment represents a popular alternative to expert assessment in online scalable courses. Literature on peer assessment offers contrasting views (Sadler and Good 2006). The perceived low reliability of peer assessment (i.e. extent to which the assessment measures learning consistently over time and across different learners and evaluators) and validity of peer assessment (i.e. extent to which the assessment accurately measures what it is intended to measure) are the two main reasons for resistance (Falchikov and Goldfinch 2000). Peer assessment reliability and validity very much depends on specific context; very little is known about peer assessment in online training for teachers. In this context, TeachUP aims to understand whether course participants can give a valid assessment of each other's work and, consequently, whether participants perceive the peer assessment process as a valid and valuable form of exchange that offers useful input into how to improve their work.

1.2 Evaluation questions and research setup

The TeachUP policy experimentation addressed three main evaluation questions:

1. Does online **personalised support increase student teachers and professional teachers' participation** in online courses?
2. Does online **personalised support improve Self Regulated Learning Online Competences** (SRLO)? If so, does it improve them directly or indirectly (through increased online course experience?)

Figure 1.2 The TeachUP's theory of change
Personalised support is expected to increase course participation and SRLO

3. Does **peer assessment** represent a viable approach to assess learning achievements in online courses?

Within the TeachUP policy experimentation, four **new online courses** on topics linked to the key competences for 21st century teachers were developed and addressed to teachers in Initial Teacher Education (ITE) and Continuing Professional Development (CPD) in nine European Member States (Austria, Estonia, Spain, Greece, Hungary, Lithuania, Malta, Portugal, Slovakia) and one neighbouring country (Turkey).

To provide a robust answer to the first two evaluation questions, the TeachUP policy experimentation features a **randomised controlled trial**. The trial involved 3,777 professional teachers (PTs) and student teachers (STs) in lower secondary education. To investigate the third question, instead, the research focused on a subsample of TeachUP students and professional teachers.

The report is structured as follows. Section 2 describes the four courses and the main features of the TeachUP personalised support mechanism. Section 3 describes the evaluation design with regard to the sampling, the data collection and the randomisation. A statistical profile of the TeachUP teachers involved in the experimentation is provided in section 4. Section 5 presents the results of the impact evaluation on course participation. Section 6 provides an analysis of the process, while section 7 and 8 look at the results on SRLO and learning assessment, respectively. Finally, the main lessons and the policy implications are discussed in section 9.

2. THE TEACHUP POLICY EXPERIMENTATION

The TeachUP project was designed as a randomised experiment (with a test and control group) so as to offer conclusive evidence of the **potential impact of integrating a personalised support infrastructure in scalable online courses for teachers and student teachers**. To do so the project developed and ran a course series consisting of four short online courses where this infrastructure was implemented.

In this section we provide an insight of the policy experimentation. We first describe the context in which the personalised support was implemented (subsection 2.1) we then illustrate what the TeachUP personalised support model is (subsection 2.2).

2.1. The TeachUP course series

Four online courses in the languages of the partner countries were developed as part of the project and delivered during the school year 2018/2019 in sequential order in the months of November 2018, January, March and May 2019. Courses lasted **3.5 weeks** and addressed four distinct topics (i.e., formative assessment, personalised learning, collaborative learning, creativity) related to the changing role of teachers and students in classrooms and provided practical tools and new pedagogical methods (Figure 2.1). Each course then finished with the production of a lesson plan for the learners' context, incorporating the ideas gathered during the course.

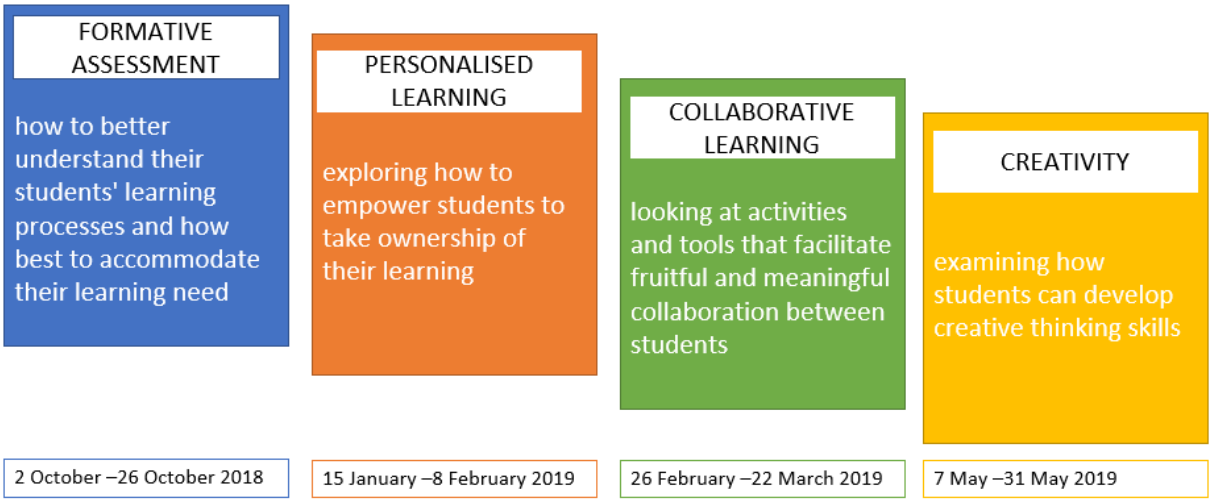


Figure 2.1 *Contents and training goals of the four TeachUP courses*
TeachUP teachers were offered four courses, delivered sequentially throughout school year 2018/2019

The four courses closely followed an instructional design based on social-constructivist and connectivist learning theories (for a full description see "Implementing Personalised Support in Scalable Online Courses" TeachUP report).

The courses drew from a range of materials including classroom observation videos, teacher and student interviews, screencasts or short practice-focussed researcher presentations. Throughout the courses, participants were required to transfer their learning to a course output in the form of a lesson plan. This work was then assessed via a peer assessment activity at the end of the courses. In order to fully benefit from the course experience, participants were encouraged to engage in the course community that was built up as the course progressed on the course platform via the forum and embedded web 2.0 tools as well as via a dedicated Facebook group. The links between different

channels were established via dedicated course moderators who actively linked up participants and content across the growing network of activity.

The experimentation consisted in offering a direct and personalised support to help course participants benefit more from the content and community. This support was offered by dedicated personalised support agents only to a group of participants (randomly selected, i.e. the test group) and only to those most in need according to some criteria that will be clarified below.

The courses were first delivered in the school year 2018-2019 in the national languages of the ten participating countries and from December 2019 these are freely available as **Open Educational Resources (OER)** on the European Schoolnet Academy.[2]

The course series was organised according to a natural progression of pedagogical approaches with formative assessment being the starting point for personalised learning and collaborative learning and creativity being approaches that can support a personalised learning approach.

Importantly, in order not to alter the set of incentives that teachers usually have when engaging with online training, no formal accreditation was offered to the teachers for completing the courses. At the end of each course all successful participants received a digital badge and certificate. Those who completed the entire course series were awarded an additional badge and certificate acknowledging their achievement.

2.2. The TeachUP personalised support model

The aim of the personalised support implemented in TeachUP was to help participants successfully complete the courses and develop their ability to do so as independent learners. Support Interventions were focused on four areas: content-related feedback, development of self-regulated learning competences, practical support on how to proceed, motivating participants (Figure 2.2).

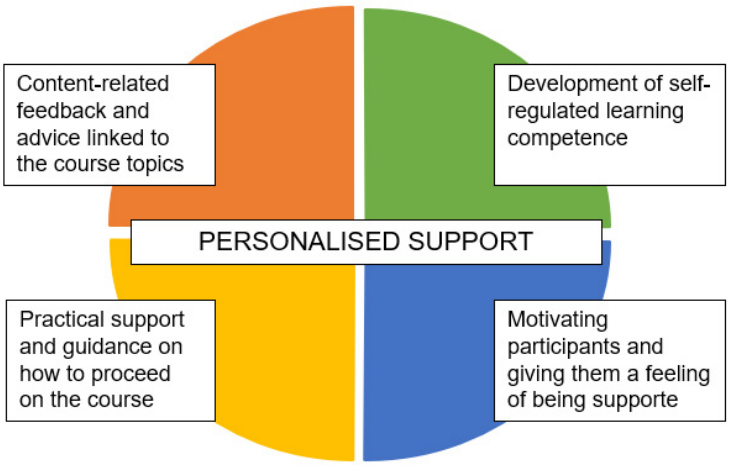


Figure 2.2 *Focus areas of the TeachUP personalised support*

A key feature of the TeachUP online support was **personalisation**. Personalisation was achieved in two main steps: 1) **identification of each teacher's needs**, 2) **provision of a support matching the identified needs**.

2 http://teachup.eun.org/moocs;jsessionid=2443060159A12D8A189ACEE0F4E7DCAC

The first step was meant not only to identify **if a teacher was in need** of support at all, but also **which particular needs she/he had**. Teachers' needs of support were predicted on the basis of **teachers' profiles** and **teachers' behaviours**.

Teachers' profiles were determined with a **Baseline Survey**, which all teachers had to fill in in order to participate in TeachUP and which collected teachers' professional background information as well as information on teachers' past experience with online learning, self-regulated learning online, views on online learning, digital competencies and teaching beliefs and practices (see section 3.4 for more details). Teachers behaviours instead were collected through the course platform data, which provided **automatically generated analytics** on participant actions (or inactions) (see again Section 3.4 for details).

With each of these two sources, a set of **triggers** was identified building upon available research results showing which characteristics or actions of participants correlate with course participation.

Baseline Survey data were exploited to identify four types of "in-need" teachers:

1. teachers with low levels[3] in at least two of the following indicators: **beliefs** about effectiveness of online learning, **expectations** of likelihood to take online courses in the future or about course completion and activity, or teachers with an **education level** lower than a Master degree;
2. teachers with low **SRLO** (i.e., indicates low competence in at least 2 of the self-regulated learning components);
3. teachers with no **experience of online courses** or low ability and **confidence with online learning technologies**;
4. for course 4 only, teachers were targeted as "in-need of support" on the basis of a predictive model of course completion, which exploited a set of baseline characteristics that was good predictors of that risk in previous courses (i.e. motivation[4], English proficiency, subject of teaching, age, previous experience, internet access, gender, level of education).[5]

Course platform data instead allowed identify five additional types of needs:

1. Teachers who didn't start 5 days after module launch,
2. Teachers who hadn't submitted their work for the peer review activity 2 days prior to the deadline,
3. Teachers who made two or more support requests via the contact form of the course platform within a period of 1 week and never visited the course FAQ page;
4. Teachers who indicated dissatisfaction/confusion with the feedback provided by at least 2 out of 3 peers in the peer review activity
5. Teachers who visited less than 70% of required module sections 1 week after module launch.

Each trigger had an associated **action** which was **addressing the specific characteristics of the trigger**, for example an email highlighting resources that can help to succeed in online learning and an offer for a 1:1 video call to conduct a "walk-through" of the course interface. Actions were adapted for each participant taking into account their profile, for example by adding references to the participant's profile into the email and by highlighting in the video call content on the course which might be of particular relevance to the participant.

A total of **nine interventions**, consisting of triggers and actions, were created for the experiment (Table 2.1).

3 Multiple choice options were used to assess the degree of agreement with respect to a set of statements concerning different aspects related to online learning. Low level means that teachers disagree, strongly disagree or slightly disagree with these statements.

4 Motivation was measured using 12 items on the reason why users decided to participate in the TeachUP project. It includes the following items: General interest in topics covered in the offered courses; Relevant to my job; For personal growth and enrichment; For fun and challenge; To meet new people online; To experience an online course; To earn a certificate/statement of accomplishment/badge or credits; Course offered by a prestigious institution; To take the courses with colleagues/friends; To get involved in a relevant European Project.

5 Further details on this predictive model are provided in section 6.3.

These interventions were based on research results showing which characteristics or actions of course participants have an impact on the likelihood of course completion (see "Implementing Personalised Support in Scalable Online Courses" TeachUP report).

Table 2.1 *TeachUP personalised support model: triggers and actions of the nine interventions*

Interventions			
#	Trigger	Action	Aim
1	User indicates low levels in at least 2 of the following indicators: - belief about effectiveness at online learning, - expectations of likelihood to take online courses in the future, - Education level lower than Masters	In course 1-3 the agent contacted the user with a personalised message that included an offer for a video call, template and guidance for succeeding in online learning using a learning plan. In course 4 the agent contacted the user with a general message offering support of any kind at any point during the course should the user like to reach out.	Course Completion
2	User indicates low competence in at least 2 of the following self-regulated learning components: - Goal setting - Task strategies/time management - Help seeking - Self-evaluation - Elaboration	In course 1-3 the agent contacted the user with a personalised message that included general guidance on the importance of SRLO, an example of how to develop the SRLO components, as well as an offer for a video call to discuss specific SRLO strategies. In course 4 the agent contacted the user with a general message offering support of any kind at any point during the course should the user like to reach out.	Course Completion
3	User has low experience of online courses or reports low ability and confidence with online learning technologies	In course 1-3 the agent contacted the user with a personalised message that included an offer for a video call to "walk-through" the course interface and tools or to identify a set of questions they have about the course. The agent would then provide answers to these questions. In course 4 the agent contacted the user with a general message offering support of any kind at any point during the course should the user like to reach out.	Course Completion
4*	Weighted mix of low motivation, low English proficiency, subject of teaching, age, previous experience, internet access, gender, and level of education	Support agent contacted the user with a general message offering support of any kind at any point during the course should the user like to reach out.	Course Completion
5	User has not started 5 days after module launch.	Support agent contacted the user with a personalised message reminding of the importance to stay on track and encouraging the user to keep going as well as offering advice on how to plan their learning time and self-motivate.	Course Start
6	User has not submitted their work for the peer assessment activity 2 days prior to the deadline	Support agent contacted the user with a personalised message about the deadline, offering tips on how to complete the work and where to find support, including the possibility for a video call prior to the deadline, in order to answer any final questions.	Course Completion
7	A user has made two or more support requests via the contact form of the course platform within a period of 1 week and has never visited the course FAQ page.	Support agent contacted the user with a personalised message highlighting ways how users can find answers to their questions, including finding peer support but in particular highlighting the FAQ page and the support section of the forum.	Course Completion
8	User indicates dissatisfaction/confusion with the feedback provided by peers in the peer assessment activity	Support agent provided feedback to the work submitted.	Course Satisfaction
9	User has visited less than 70% of module sections 1 week after module launch	Support agent sent a personal message offering support to help user benefit more from the course content, including a possibility to book a 1:1 session to discuss how to use the content.	Course Completion

*Course 4 only

Interventions addressed key elements of the instructional design and course timeline which were more complex and potentially problematic for learners (e.g. the final submission of a course product and associated peer assessments). Moreover, interventions were also shaped by the technical possibilities of which data could be collected by the course platform. This meant that not all elements highlighted as important in the literature could be accounted for in the final interventions. The exact determinants of each intervention with triggers, actions, and aims are identified in Section 6.

Personalised support was carried out by **support agents**, who were selected within each of the participating countries as knowledgeable and experienced in the course topics as well as in the field of online learning and self-regulated learning. Agents had a course cohort of a maximum of 100 participants assigned to them. They conducted spot checks at specific times before and during each course to see which participants of their cohort were eligible for the support offer. These checks were done via an online tool which displayed lists of eligible participants by interventions. The tool also allowed support agents to see the whole set of a participant's responses to the baseline survey, offering a detailed overview of the participant's profile, beliefs, and confidence levels. The tool received data from the baseline survey as well as from the course platform. Support agents used the tool to identify those qualifying

for an intervention and to "research" about the participant in question so as to personalise the support offer.

Unlike the moderators of the courses -- which, as in many other online courses and also in TeachUP, were concerned with developing an active learning community where participants support and learn from each other -- personalised support agents supported selected individual learners. In theory, rather than respond to the queries of individual learners, the **moderator** encourages other participants to respond to the query. In practice, moderators nevertheless often answer learner queries directly, but they do so usually in a generic way, without considering the learner's background or profile, information to which a moderator does not normally have access to. The following table provides an overview of the moderator role and the personalised support agent role. These roles should not be seen as opposing each other but rather as a complementary way to manage a course community.

Table 2.2 *The different roles of moderators and personalised support agents*

Moderator Role	Personalised Support Agent Role
Main Aim	
Develops an active learning community where participants support and learn from each other.	Supports individual learners to benefit from the course content and community.
Target Audience	
All course participants.	Individual course participants.
Example Tasks	
Encourages participant engagement and discussion, for example by posting a question or statement in the forum or social media Highlights generally useful resources in forum or social media	One-to-one sessions with learner, answering learners' questions about content, process, or technology and taking into account the learners' background and progress Highlights specific resources relevant to the learner's profile
Posts reminders about deadlines to all participants on forum or social media	Reminds learner about upcoming deadlines with personal message
Highlights notable participant contributions in forum or social media as a form of recognition, motivation, and example	Helps to develop a personal learning plan for the learner
Organises a synchronous session available to all participants, such as a webinar, Twitter chat, or Teachmeet	Proactively contacts the learner to offer support
"Likes" participants' posts	Shares personalised advice on how learner can self-regulate learning

3. THE EXPERIMENTAL DESIGN

3.1. What needs to be considered when designing an impact evaluation study?

When designing the evaluation of a policy intervention, three crucial questions need to be considered. First, are the estimated effects of the intervention real causal effects? Second, would the causal effects estimated in the study be found also in other samples or contexts? And, third, do the data really measure what they were intended to measure?

The first question is a question of **internal validity** and refers to the extent to which the estimated effects are truly causal effects rather than reflecting pre-existing differences between the group of beneficiaries of a given policy intervention (the so-called treatment group) and the group of non-beneficiaries (i.e the control group). If assignment to the intervention is not random (i.e., if treated subjects are systematically different from non-treated, say they are older), one cannot safely conclude that any difference detected between groups after the intervention is because of the intervention rather than being due to pre-existing differences between the groups. To avoid this threat, the TeachUP evaluation design is based on a **Randomised Controlled Trial** (RCT). This approach is considered to be the most reliable one to estimate program's effects, i.e. to produce estimates that are not biased by pre-existing differences between treated and control subjects (also known as selection bias). Using such a design, pre-treatment differences between beneficiaries and non-beneficiaries are minimised with large samples, hence the two groups are made comparable on all observables and non-observable characteristics.

The second question relates to the so-called **external validity** and refers to the extent to which the causal effects estimated in one context hold also in another context or in the overall population the sample was drawn from (i.e., generalizability or transportability of the results). If the sample on which the evaluation is based is systematically different from the entire population, say, again, it is on average older, then it is unclear whether the estimated effects found on that sample would be found also in the entire population. This is a crucial point, as typically, policy makers want to know what the impact of a given intervention would be if it was adopted on a large scale and made accessible to the entire reference population. To maximise the generalizability of the evaluation evidence, in TeachUP an ad hoc sampling protocol (i.e. a **stratified sampling**) was implemented.

The third question is about **data** and refers to the procedures to be adopted in order to make sure that the outcome variables, i.e. the variables on which the intervention is expected to have an impact, are measured in an appropriate way. This implies that the data are valid and reliable (i.e., they measure what they are intended to and they do it consistently) and that they are collected timely and in the same way for all study participants. In this regard, the TeachUP experiment made use of a variety of data sources (surveys and learning analytics) and followed a strict data collection protocol.

This section describes the main features of the TeachUP evaluation and the approach employed to estimate the impact of the personalised support on teachers' course participation and self-regulated learning online. We first describe the **target population** (i.e., the population in which the experiment was conducted) in terms of participating countries, school education level and types of teachers involved (subsection 3.2). Second, we explain the procedure through which a **sample of teachers** was drawn out of the target population and how they were invited to participate in TeachUP and thus became **TeachUP teachers** (subsection 3.3). Third, we illustrate the **randomisation process**, i.e. the way in which TeachUP teachers were randomly assigned to the treatment or the control group (subsection 3.3). Fourth, we describe the **data collection** plan, by illustrating the sources of data used and the timing of the collection of the data (subsection 3.4). Fifth, we illustrate the **outcome variables** of the evaluation, that is to say the set of teachers' behaviours and characteristics (i.e., course participation and SRLO) that the tested intervention (i.e., the personalised support offered) was expected to change according to the theory of change (subsection 3.5). Subsection 3.6 resumes the experiment's timeline.

3.2. Countries and target population

To explore how the impact of personalised support varies according to teachers' professional experience, TeachUP involved teachers at different stages of their careers. More precisely, both students completing their training to become teachers (i.e., **student teachers** or STs) and teachers already in service in a school (i.e., **professional teachers** or PTs) were involved.

Furthermore, the experiment focused on those teaching, or expected to teach, in lower secondary schools (**ISCED 2 level**, that corresponds approximately to grades 6-9/10, or to students aged between 10 and 13).[6]

The experiment involved **ten countries**: Austria, Estonia, Greece, Hungary, Lithuania, Malta, Portugal, Slovakia, Spain and Turkey.

In each of the ten participating countries, all **publicly funded schools** of the relevant grades and all **Initial Teachers Education Organizations** (ITE) were considered eligible to take part in the project.

Because the pathways to the teacher profession vary across countries (European Commission/EACEA/Eurydice 2015), the identification of ITEs had to adapt to the different models of initial teacher training (see appendix B for further details). In eight countries, STs were targeted when attending their last year of master or other university program or course providing ITE. The exceptions were Turkey and Spain which have focused on teachers in induction.

3.3. Sample

Once the target population was defined, as illustrated in section 3.1, a sample of schools and ITE organisations was drawn and, within these organisations, all PTs and STs were invited to participate in TeachUP. In doing this, two main goals were pursued: having a large sample and having a representative sample. This section provides details to help understand how the TeachUP sampling protocol was designed and implemented.

The first goal was that of reaching at least **4,000 PTs and STs** across the ten countries.[7] Such a high number of subjects was justified by the goal of having adequate statistical power and hence being capable of producing reliable impact estimates. The target of 4,000 PTs and STs was split across the ten countries to reflect each country's size.

The second goal was to build **a representative sample** of PTs and STs, i.e. a sample of teachers that is on average comparable to the entire teacher population. This equality is a crucial condition to being able to generalize the results of the evaluation to the entire population (i.e., the so-called "external validity" of the experiment). If the sample is not representative of the population, it would not be possible to infer that the results of the evaluation could be extended to the entire population of teachers and this would of course limit the learning potential of the entire evaluation. To achieve the goal of having a representative sample, we adopted a **stratified sampling design**, which followed a 3-steps approach:

- ***First***, we collected the **complete lists** of eligible schools and ITE organisations in each country;
- ***Second***, we divided the schools and ITE organisations into **sampling strata** given by their geographical

6 Greece was the only partial exception, as teachers of primary schools were involved, but limited to the two last grades i.e. the 5th and the 6th in order to improve comparability with the other countries.

7 According to a statistical power calculation, such a sample size would have allowed us to detect as statistically significant an impact on course completion as small as 6 percentage points in each of the two populations (i.e. student and professional teachers), in case of equal distribution of units across the two.

location of the organisations. This was meant to maximise the geographical representativeness of the sample;

- ***Third***, within each of these non-overlapping strata, schools and ITE organisations were randomly selected. Here, we followed a **probability sampling approach**, which took both strata and single organisations' size into account (e.g, more schools were selected in bigger strata, and bigger schools had higher likelihood of being sampled than smaller schools). Random sampling allows retrieving a representative sample of schools and ITE organisations, i.e. that have characteristics that are on average comparable to those of the entire population of schools and ITE organisations (e.g. in terms of size and geographical location).

The sampling protocol is further described in Box 3.1.

Box 3.1 *TeachUP sampling design*

The sampling procedure was articulated in 3 consecutive steps as graphically shown by Figure 3.1.

1. **List of schools and ITE organisations:** Within each country, the complete lists of eligible schools and ITE organisations were collected (each circle in the first panel of figure 3.1 represents a school in a given country). These lists contained key information such as identification codes, geographical locations (typically Nuts-codes), and organisations' size (either number of teachers or number of students);
2. **Definition of sampling strata:** For each country, we identified sampling strata based on geographical administrative aggregations (in the second panel of Figure 3.1 different colours are used to represent schools in different areas). For small countries, we either decided not to create strata or used the type of school program instead of geographical aggregations. For ITE organisations, also, no strata were formed, because of the limited number of these organisations.
3. **Weighted random selection of schools/ITE organisations within strata:** The number of schools and ITE organisations to be sampled within each stratum depended on each stratum's relative size (i.e., a higher number of schools was sampled in bigger vs smaller strata). Then, within each stratum, **schools/organisations were randomly selected considering their size as a weight**, such that bigger schools (organisations) had higher likelihood of being drawn (the selected circles in the third panel of Figure 3.1 represent the schools with a higher probability of being drawn because of their size). An additional sample of schools/ITE organisations (**"**reserve list**"** or "oversample") was drawn within each stratum in order to replace schools/ITE organisations refusing to participate. For ITEs the sampling process had to be adapted and simplified as in some countries the number of ITEs was too small to even consider sampling, hence all of them were included in the study.

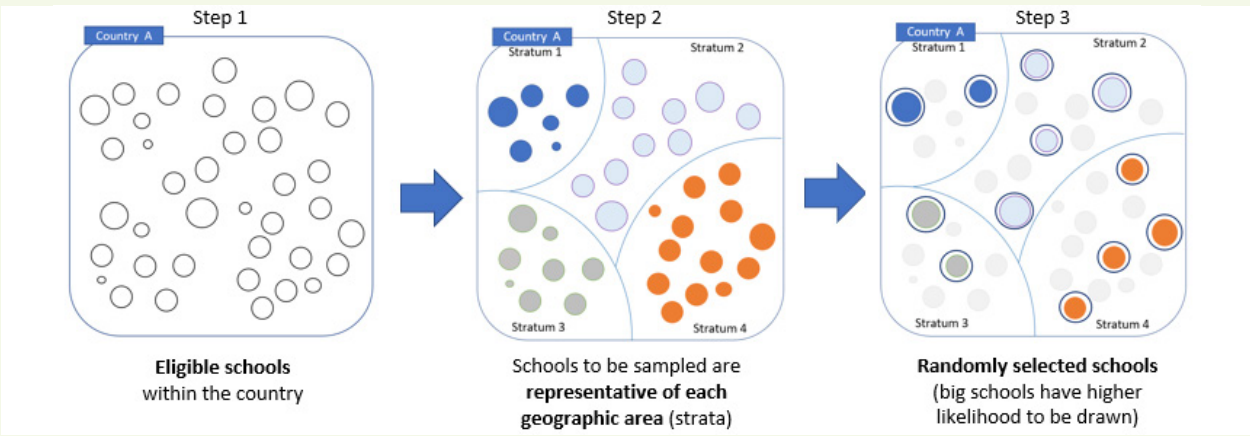


Figure 3.1 *The three steps of the TeachUP sampling*

Within the sampled schools/ITE organisations all PTs and STs were contacted and invited to participate by means of an **informative/recruiting campaign** launched in March 2018. The invitation was sent via e-mail by the National Coordinators to the directors of the TeachUP schools and ITE organisations, who then, in turn, circulated it to all teachers/student teachers belonging to their organisations.

Overall, **4,090** teachers from 511 schools and 456 ITE organisations accepted the invitation to participate in TeachUP. By signing up to the project (i.e., by filling in the Teachup Baseline Survey, see section 3.5), they became **TeachUP teachers**. Teacher recruitment followed a so-called "no one forced, no one denied" principle, i.e. within each school and ITE organisation teachers were free to sign up to the project. Response rates resulted to be low in general but very heterogeneous across groups and countries as a result of both non-response or explicit refusals and peculiar sampling procedures (more details about the sampling process are provided in Appendix B). The distribution of PTs and STs across the participating countries is shown in Figure 3.2.

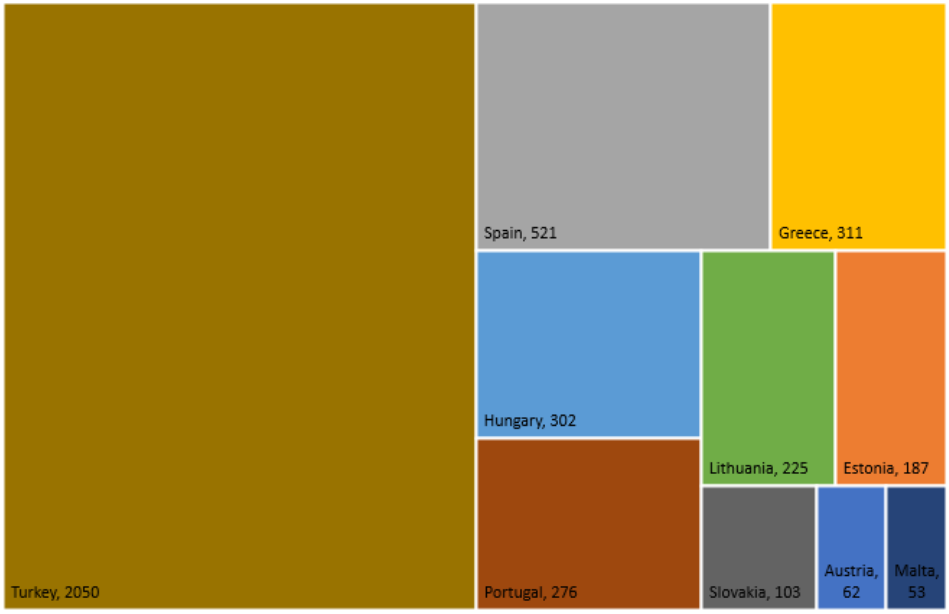


Figure 3.2 *Distribution of TeachUP teachers (N=4,090) across the participating countries*

The goal in TeachUP was to conduct a single experiment across ten countries, not ten separate experiments in each of the countries. To carry out the latter, a much bigger sample of PTs and STs would have been needed. However, thanks to the success in reaching a large sample size, it was possible to take the high contextual and teacher profile heterogeneity into due account. More precisely, the evaluation analysis focused on the four sub-groups illustrated in Figure 3.3: professional teachers in EU Member States, student teachers in EU Member States, professional teachers in Turkey and student teachers in Turkey.[8] By doing so, the TeachUP experiment was able to yield comparative estimates of the effectiveness of personalised support depending on teachers' professional stage (PTs vs STs) and institutional context (EU MSs vs Turkey).

8 Country-level estimates could be computed for Turkey only because of the large sample size obtained in this country, while they are not possible for all other EU MSs where numbers were too low.

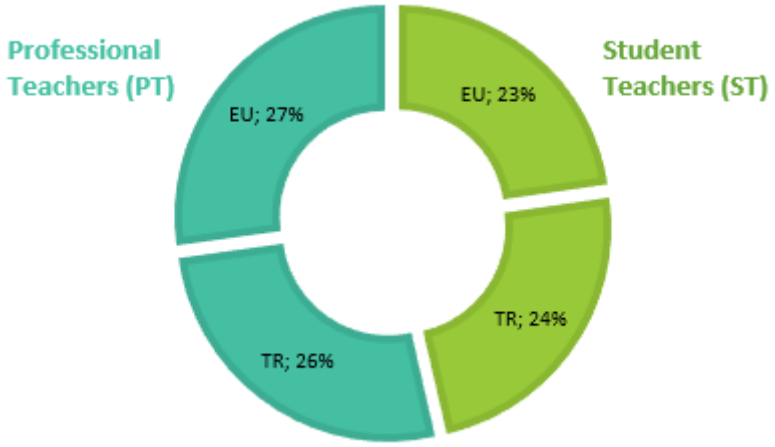


Figure 3.3 The TeachUP sample composition
PTs were a little more than STs and Turkish teachers made up half of the sample

3.4. Random assignment

The sample of TeachUP professional and student teachers was randomly split into two groups: one (the treatment group) receiving the intervention that is being tested (the personalised support in addition to the standard design of an online course), and the other (the control group) receiving an alternative (i.e the standard design of an online course).

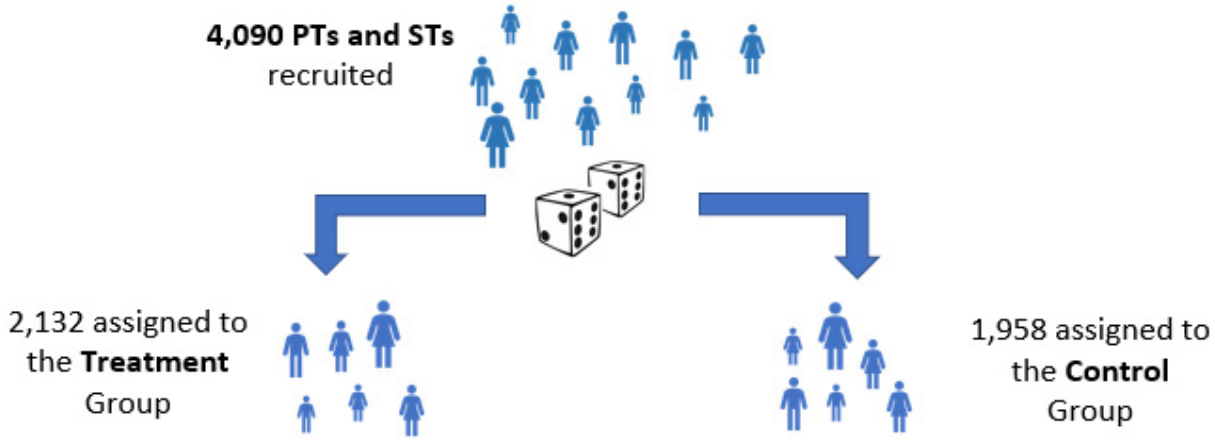


Figure 3.4 The TeachUP RCT design
About half of teachers were randomly assigned to the treatment group

As in all randomised controlled trials, also in TeachUP attention had to be paid to the experimental protocol's integrity. A first issue that had to be tackled was that of **contamination** (i.e. the possibility that, through interaction, treated teachers would pass part of the treatment on to control teachers). The threat posed by contamination in TeachUP is limited as the contents of the support is not easily transferable because it is highly personalised and support agents were instructed to interact with treated teachers only. In TeachUP, two main possible sources of contamination were identified and addressed: online and off-line contamination.

To avoid **online-contamination**, **two separate online courses** were set up (for all the four courses) within each country, one for the treated and the other for the control group. The two courses were identical, with the only difference being the possibility of getting personalised support that was provided only in the treatment group. Online

interaction was only possible within one's own group course community forum. Online interaction on open platforms or social networks was in principle possible and could not be controlled for.

To minimise the risk of **off-line contamination**, assignment to the treatment or the control group did not occur at the teacher but **at organisation level** (school and ITE, depending on the target population). That means that all teachers belonging to a given school or ITE organisation were entirely assigned to only one of the two groups: a **treatment group**, eligible to receive the personalised support; and a second group, the **controls**, in which none received it. This setup made it difficult for treated and control teachers to interact during the experiment. Of course, treated teachers could in principle communicate with colleagues from other schools, and this could not be controlled for. The allocation to the two groups happened randomly through a lottery-like mechanism. As explained above, this allows interpreting the between-group post-treatment differences as a causal impact of the treatment. To put it otherwise, randomisation ensured that **the two groups were on average identical but for their exposure to the personalised support**. Any difference detected after the program implementation could therefore be attributed to the support.[9]

Table 3.1 shows the distribution of TeachUP teachers in the treatment and the control groups. The size of the two groups in terms of number of teachers are due to the different size of the schools/ITE org. Further details of the randomisation are shown in Appendix D.

Table 3.1 *Randomisation outcome by group*

	Professional Teachers (PTs)			student teachers (STs)		
	control	treated	total	control	treated	total
EU	**541**	**571**	**1112**	**434**	**494**	**928**
TR	**486**	**594**	**1080**	**499**	**471**	**970**
Total	1027	1165	2192	933	965	1898

	schools			ITE		
	control	treated	total	control	treated	total
EU	**198**	**200**	**398**	**33**	**33**	**66**
TR	**57**	**56**	**113**	**194**	**196**	**390**
Total	255	256	511	227	229	456

Beyond making sure that contamination did not alter group comparability, two additional crucial checks had to be performed in order to assess how the randomisation protocol was implemented in reality. The first of these checks was carried out through "**balancing tests**" of the two groups, i.e. a statistical analysis aimed at making sure that treatment and control groups were statistically equivalent along all baseline characteristics. Results reported in Appendix D show that this check was empirically passed. The second set of checks had to do with the issue of **attrition** (or non-response rate because of some participants who started the course but dropped out at some point and did not answer to the follow-up survey) in the post-treatment surveys (see details in section 3.5).

An important feature of the TeachUP instructional design - as illustrated in subsection 2.2 - was that **only a fraction of the teachers in the treatment group received personalised support** (Figure 3.5). The reason for this was the presumption that not necessarily every enrolled teacher would have actually needed support and that personalised support had to be accompanied by a consistent targeting approach. All enrolled teachers were identified as either "in need" or "not in need" of personalised support according to a set of background characteristics collected with the baseline survey or their actual behaviour during the training, as monitored on the platform.

9 While, overall, no issues of contamination were detected, reports from personalised support agents in Turkey suggest that some information from the agents was passed on to the control group.

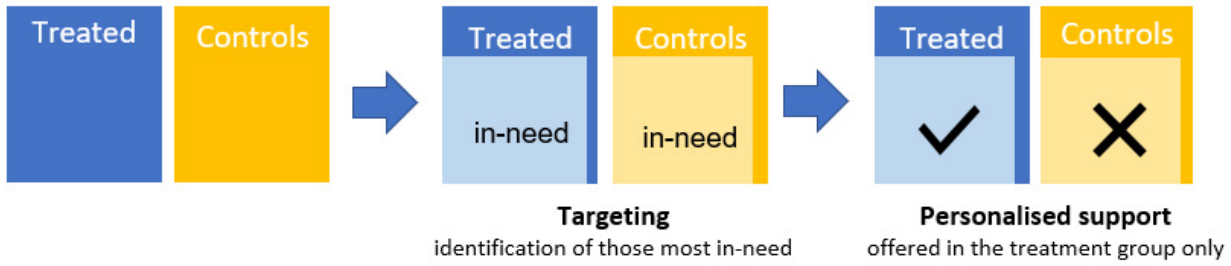


Figure 3.5 TeachUP targeting approach
Only targeted teachers in the treatment group received personalised support

One further aspect should be considered in the analysis: in course 1 the treatment was not delivered in Portugal due to the impossibility of setting up the support model in time for the start of the first course.

3.5. Data collection

The TeachUP evaluation exploited different sources of data that were collected at different critical points of the experiment (Figure 3.6).

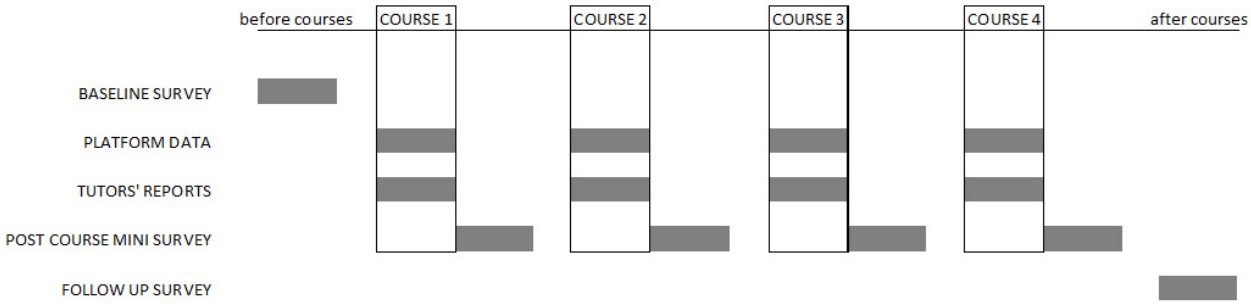


Figure 3.6 Data collection plan

Baseline Survey

A **Baseline Survey** (hereafter also BS) was administered online before the courses started and served as registration to TeachUP. The aim was to collect background information and baseline data on their self-regulated learning experience, views on online learning, digital competencies, teaching beliefs and practices. Two separate questionnaires were drawn, one for professional teachers, the other for student teachers. The questions and their wording were mainly taken and adapted from already existing and validated cross-national surveys (e.g., TALIS, ICILS, PIRLS, TIMMS, Survey of Schools: ICT in Education).[10] Participants were encouraged to register to the entire series of 4 courses, but it was still possible for a participant to start in course 2, 3 or 4 even in absence of participation in the previous TeachUP course(s). Overall, 97% of TeachUP teachers answered all the questions. The English version of the BS is available online on the project's website.[11] It is important to highlight that participants filled in this survey only once, this means that the BS information which was acted upon by personalised agents was collected in a static manner, and thus not updated in real-time as participants progressed through the different courses.

10 All surveys were translated in the countries' official languages. A small pilot took place in each participating country to test and refine the translations. By May 2018, the final version of the BS was uploaded on the platform.

11 http://teachup.eun.org/

Post-course mini surveys

A short survey was run immediately after the end of each course to all enrolled teachers (**post-course mini survey**). The questionnaire (see project's website) was different depending on whether or not the teacher had completed the course. The aim was to assess through very few questions (up to 5 questions) on the one side teachers' satisfaction with the specific course and learning strategies used and, on the other side, the reasons for not participating. Among course completers, response rates were very high (around 90%) and well balanced between treated and control groups, while among course non-completers completion rates were too low (lower than 10%) to make it possible to analyse these data.

Follow-Up Survey

About one month and a half after the end of the fourth course, the last online survey, the **Follow-Up Survey** (hereafter FUS), was administered to all TeachUP teachers (i.e., who had filled in the BS), regardless of their actual participation in the courses (i.e. whether they started/completed, took just the first course or the last).[12] The questionnaire was the same for PTs and STs and contained some of the questions of the BS such as self-regulated learning experience, views on online learning, digital competencies, teaching beliefs and practices in order to estimate the impact of the personalised support that was provided. Additional questions were integrated in order to get additional information on how the personalised support was used by those receiving it.

The survey was kept open longer than initially planned (i.e., until mid-October 2019) as an attempt to reach as many teachers as possible. Moreover, incentives were given to encourage responses.[13] Nonetheless, **only 18% of TeachUP teachers answered the survey** and the proportion of respondents varied substantially across teachers with less or more course participation: **teachers who participated in at least one course showed a 47% response rate** while among those who did not participate in any course the response rate was 7%.[14]

The three main potential drawbacks of these patterns of response in the FUS are: (i) a reduced statistical power due to the **low sample size**; (ii) a **compromised internal validity** of the experiment, due to higher response rates among course completers and to possibly high and different attrition rates between treated and control groups; and (iii) a **reduced external validity** of the results, due to the existence of systematic differences between FUS respondents and the starting sample.

To cope with the first drawback, we were forced to give up the 4-group analysis and **pool together all participants regardless** of their country of belonging and their professional status. To deal with the second aspect, we checked whether **differential attrition** compromised the integrity of the evaluation design by comparing the attrition levels of treated and control groups, and found that the two groups had comparable attrition rates. Third, we checked the **equivalence of the baseline characteristics** in the subsample of teachers completing the FuS with those not completing it. The analysis left out the existence of substantial differences between FUS-completers and FUS-non-completers (see Appendix D for all the details on these statistical checks).

Overall, these checks led us to conclude that, even if heavily underpowered, the analysis could be still conducted by adjusting the impact estimates through multiple regression models. The results are shown in section 7.

12 The administration of the survey followed roughly the same protocol as in the BS, with a massive email invitation to all TeachUP teachers, followed by a reminder to the school heads and ITE organisation directors.

13 These included a survey certificate and online shop vouchers for 150 users.

14 In addition, those completing more courses showed higher response rates. Response rates were highest among those teachers who completed four courses (72%). Significant differences in response rates were also found across countries. More detailed figures are included in appendix D.

Platform-generated data

Beside survey data, the evaluation made use of **platform-generated data** on participants' actions (or inactions) during the course such as indicators of course progression (i.e. start and completion) and markers to identify those most in need (i.e. baseline devised triggers) and therefore eligible for the personalized support (see Section 2). Particularly, these data were used to construct the two main outcomes of our study, i.e. course start rates and course completion rates. These data, being automatically generated, did not suffer from the attrition problems encountered with survey data.

Support agents' reports

Finally, in order to investigate how the personalised support model was actually translated into real actions and interventions, the analysis exploited information provided by **personalised support agents**. These reports were produced by each of them at the end of every intervention following a common grid.

3.6. Outcomes

Based on the TeachUP evaluation questions (section 1), the analysis included two primary outcomes (i.e., course participation and Self-Regulating Learning Online competences).

Course participation

The first outcome of interest (as shown in Figure 1.1) concerns **course participation**. This outcome was measured with three indicators (see Figure 3.7), which were computed relying on platform-generated data (i.e. course start and course completion).

The first indicator (**course start**) was computed among teachers who enrolled in a given course and distinguished those who started the course from those who did not.

The second indicator (**course completion**) was calculated again for teachers who enrolled in a given course and distinguished teachers who completed that course from those who did not.

The third indicator is a different way of looking at **course completion** and was calculated only among teachers who actually started a given course distinguishing teachers who completed that course from teachers who did not.

Hence, three different types of "participation" probability are computed: **starting a course among enrolled** (unconditional estimate), **completing a course among enrolled** (unconditional estimate), and **completing a course conditional on having started it** (Figure 3.7).

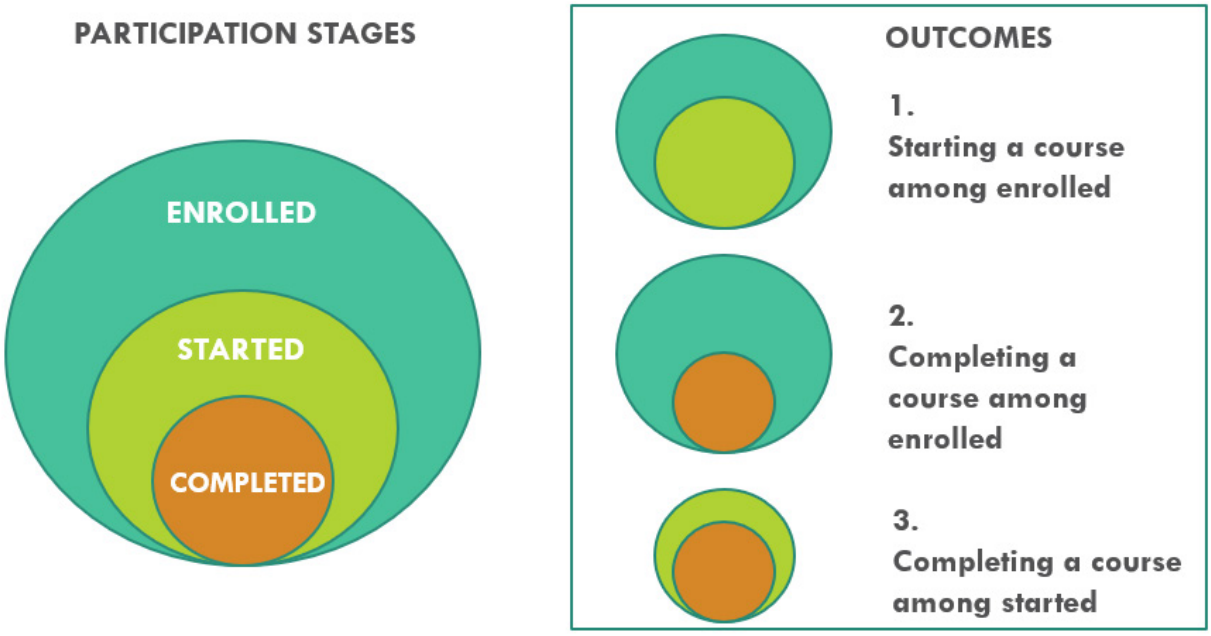


Figure 3.7 *Course participation stages and outcome variables of interest*

Self-regulated Learning Online (SRLO)

The second set of outcomes concerns teachers' **self-regulated learning online (SRLO)**. The very nature of online learning demands control, task management, and motivation in order to successfully complete a set of learning objectives (Van Laer & Elen 2016; Tsai, Shen & Fan 2013; Lynch & Dembo 2004). Various models have been developed to theorize and measure self-regulated learning (SRL) (Zimmerman & Schunk 2011, Panadero et al., 2017). The different models reflect different emphasis or approaches employed to measure and to promote SRL and SRLO (Lee *et al.* 2019). Despite these differences however, a general consensus exists on the importance of SRL development as an essential component for learning (Boekaerts 1999; Zimmerman 2001).

SRLO is a comparatively nascent (Lee et al. 2019) and currently remains an experimental field when it comes to measuring SRL for interventions (Panadero et al. 2016; Lee & Recker 2017; Molenaar et al. 2019). This aspect is even more pronounced when it comes to the actual use of individuals' learning data to intervene and explicitly promote student SRL in online environments (Viberg et al. 2020; Pérez-Álvarez et al. 2018). This being said, several strategies have been identified as effective at helping promote students' SRLO. The following strategies are commonly identified as part of effective learners' strategies (Kizilcec et al. 2017; Milligan & Littlejohn 2015), of which greater detail can be found in the initial project's SRLO landscaping (Triquet, Peeters & Lombaerts 2017):

1. **Goal setting**: Setting of educational goals or sub-goals in order to exert the effort required to achieve those goals (Schunk 2005; Zimmerman 2000)
2. **Strategic planning**: Planning the sequence, timing, and completion of activities directed at learning goals (Zimmerman & Pons 1986)
3. **Self-evaluation**: Setting quality standards and criteria for progress to judge one's own performance (Boud 1995). Activities for monitoring the learning process in relation to defined learning goals (Schunk 2005).
4. **Task strategy**: Organizing, planning, and transforming one's own study time (time management) and tasks (i.e., timing sequencing, pacing, rearrangement of institutional materials) (Effeney & Bahr 2013; Zimmerman & Pons 1986). Activities to improve persistence and effort-regulation in the face of academic challenge (Richardson *et al.* 2012).
5. **Elaboration**: Combining new knowledge with prior knowledge and constructing meaning from learned materials (Niemi *et al.* 2003). Extending or modifying the learning materials to make them more meaningful and memorable (Weinstein *et al.* 2011).

6. **Help seeking**: Asking other people for help, such as the instructor or one's peers, or consulting external help and resources (Pintrich 1999; Richardson *et al.* 2012).

In TeachUP, both general SRLO and course-specific SRLO competences were measured.

General SRLO was measured both in the BS and in the FUS using the six strategy subscales proposed by Kizilcec et al. (2017). The full set of items included in the questionnaire was reduced to three factors via a principal component analysis (see Appendix E). The first one, that we generically called "**task**", includes items related to task strategies, elaboration and self-evaluation strategies, the second one, which we called "**goal**", includes goal setting and strategic planning competences, while the third index focus on "**help**" seeking strategies (Figure 3.6). The six original strategies identified by Kizilcec et al. (2017) were thus reduced to three in this specific context. The individual score for each strategy was computed by averaging factor loadings of corresponding statements. The SRL measure had high reliability for all three factor subscales with Cronbach's of at least 0.90 (see Appendix E).

Figure 3.8 *The six "dimensions" of self-regulated learning online (SRLO)*
Note: In the empirical analyses, yellow and green dimensions were considered as two unique dimensions

The **course-specific SRLO** was collected among TeachUP participants **who completed a course** through the post-course mini surveys. The number of items used to measure SRLO in the mini-surveys was smaller (13 items) as the main goal of these surveys was to capture the degree of course satisfaction. In this case, the factorial analysis of course-specific SRLO allowed for the reduction of ten items into two indices representing "help seeking" and "persistence" competences, respectively. The three remaining items were analysed separately because they could not be reduced to a single index (see Appendix E for details). Additional information on those who did not complete courses was not collected.

In total, SRLO measures were collected at six different points during the project: a general SRLO baseline measure; four voluntary course-specific mini surveys (administered after the end of the course to all those enrolled); and one follow-up measure of general SRLO (Figure 3.9). If participants undertook only the first course, then they would be presented with 3 collection moments: a baseline survey, the mini-survey after course 1, and the Follow-Up Survey.

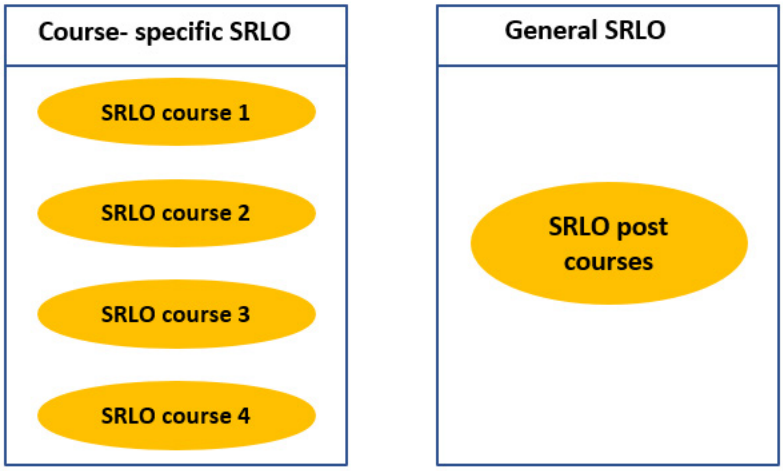


Figure 3.9 *Different measures of self-regulated learning online (SRLO)*

Secondary outcomes

In addition to the main outcomes, the impact evaluation also considered other secondary outcomes such as **attitudes towards online learning**, **teaching beliefs and practices** and **digital competences**. For each of these outcomes, a set of items were included both in the BS and in the FUS in order to test whether the intervention (i.e., the personalised support offered) produced a change. Also, with these secondary outcomes we used the same procedure of data reduction used for SRLO. Appendix E provides a detailed description of the indices found.

3.7. Experiment's timeline

All the activities described in the previous subsections, from sampling to Follow-Up survey delivery, took place between October 2017 and the end of October 2019, as illustrated in Figure 3.10.

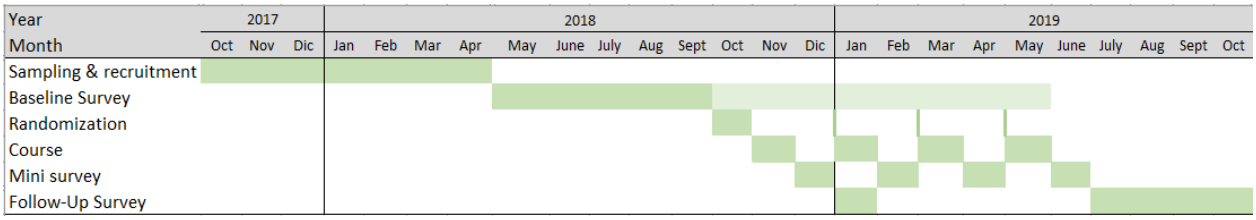


Figure 3.10 *Timeline of the experiment*

4. A PROFILE OF TEACHUP TEACHERS

Before getting to the results of the evaluation, it is worth asking ***who the TeachUP teachers are***. In this section we first provide a profile of these teachers focusing on the most relevant of the observed characteristics (subsection 4.1)[15], we then show different patterns of participation in the four courses offered (subsection 4.2).

4.1. Teacher characteristics at baseline

A total of 4,090 teachers across the ten participating countries signed up to TeachUP by filling in the Baseline Survey. Of them, 3,777 teachers actually enrolled in at least one of the four courses, while the remaining 313 signed up but did not enrol in any course and hence were not considered in the analysis. In what follows, then, we will only refer to those teachers who enrolled in at least one of the four courses (n=3,777, of which 2,026 PTs and 1,751 STs).

Most of the TeachUP teachers were **women** (Figure 4.1). There was some variation across countries, with the highest share of women being found among EU PTs (8.5 out of 10) and the lowest among TR PTs and TR STs (less than 7 out of 10).

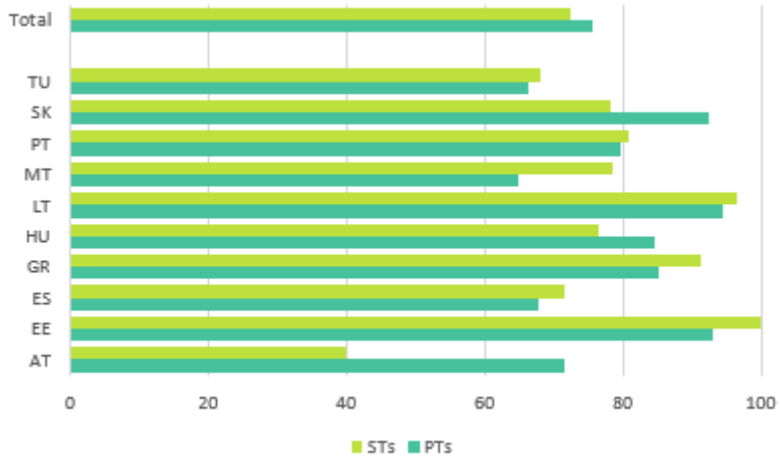


Figure 4.1 *TeachUP student and professional teacher gender distribution, overall and by country*
Most TeachUP teachers were women

Among PTs, the most common **age** group was 30-49 years, with student teachers understandably younger (i.e. the modal age group is below 30 years old). Age composition was not exactly the same across countries: noticeable differences were found between EU countries and Turkey. Turkish ones both in service and in induction were, on average, younger than in the others (Figure 4.2).

15 Additional figures of interest are included in Appendix F.

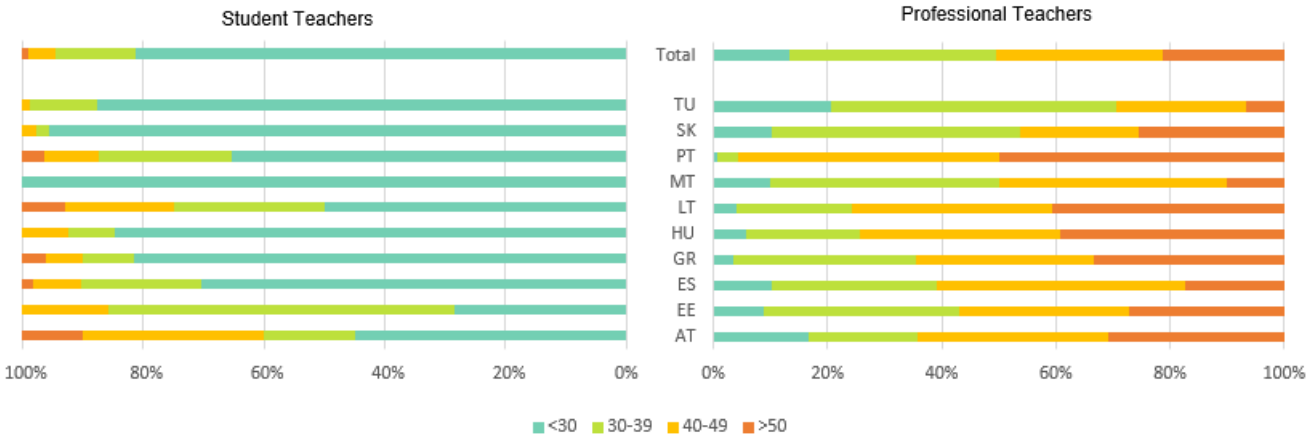


Figure 4.2 *TeachUP student and professional teachers age distribution, overall and by country*
Most student teachers were under 30 years while professional teachers were mainly between 30 and 49 years old

Regarding the subject of specialization, in EU MSs the share of **STEM teachers** was notably larger than in the cohort of Turkish teachers, with negligible differences between professional and student teachers (49-50% vs 39-40%).

The vast majority of participants stated that they had high **digital competences** - i.e., skills in making use of digital resources (Table 4.1). The aspects on which teachers were less self-confident related to social abilities such as participating in a discussion forum or in an online chat session. On average, student teachers and professional teachers showed similar levels of digital competences. Students were, on average, more skilled at judging the reliability of a website, using social media to interact with others, using digital technologies for collaborative work/projects and commenting and behaving appropriately to the situation they were in online. Differences between EU and Turkish teachers were slightly more pronounced, with Turkish teachers typically - yet not always - showing lower digital competences.

Table 4.1 *Percentages of participants who agree with statements concerning their digital competences*

	EU-PT	EU-ST	TR-PT	TR-ST
General abilities				
Conduct an internet search using one or more keywords	92	88	82	82
Judge the reliability of a website	65	73	53	58
Reflect on my online search process	69	70	62	65
Use digital technologies for collaborative work/projects	74	83	67	70
Identify personal needs and select digital tools to solve them	69	67	73	78
Use digital technologies to carry out tasks in a more effective way	76	79	76	82
Seek opportunities to develop my skills to use digital technologies effectively	75	74	67	72
Upload document files	86	88	77	80
Complete multiple choice tests	85	84	81	86
Navigate to a specific point in videos	79	85	74	77
Comment and behave in a way that is appropriate to the situation I find myself in online	76	84	71	77
Decide which information should or should not be shared online	89	89	85	86
Social abilities				
Participate in a discussion forum	61	58	47	44
Participate in an online chat session	61	61	54	57
Use social media to interact with others	69	79	76	80

Almost all professional teachers from EU countries declared that they had used a desktop computer or a laptop or a tablet both at school and outside school in the last 30 days (Table 4.2). This proportion was slightly lower in Turkey and for student teachers who, in general, have only used more ICT devices outside school. Virtually no teachers in EU MSs declared to never use any digital device, while about 2 percent of student and professional teachers in Turkey did.

Table 4.2 *Percentages of participants who used ICT devices in the last 30 days, by group*

	EU-PT	EU-ST	TR-PT	TR-ST
Devices used both at school and outside school	97	83	81	75
Devices used only at school	2	1	6	4
Devices used only outside school	2	16	11	19
No device used either in school or outside the school	0	0	2	2

Almost all TeachUP professional teachers in EU MSs (97%) stated that they are always **connected to the internet** (Table 4.3). This percentage is very high for student teachers in the EU and professional teachers in Turkey as well (95 and 94%, respectively). But among Turkish student teachers the share of "always connected" is lower (82%), partly because they declare to have no access to the internet while at school.

Teachers' perception of their **internet connection quality** at school was rated good (Table 4.3); a little less so at home. The only exception concerns Turkish students who stated that internet connection quality was lower than their domestic one.

Table 4.3 *Access to the internet and quality of the internet connection, by group*

	EU-PT	EU-ST	TR-PT	TR-ST
Access to the internet (%)				
Almost always connected to the internet	97	95	94	82
Quality of internet connection (range 0–10)				
At school	8.1	8.4	7.5	6.5
At home	6.2	6.4	6.6	7.6

Half of participants spent between 1 and 3 hours on **ICT for activities other than work** each day on average (e.g. shopping, socialising, entertainment, booking a hotel, chatting with family and friends) (Table 4.4). The share of participants declaring a high intensity of ICT use (more than 3 hours a day) was found to be highest among students, especially in EU countries (35% on average). On the other hand, the group of teachers declaring the lowest use of ICT (less than 60 minutes a day) is professional teachers from EU countries (30%).

Table 4.4 *Daily use of ICT devices for activities other than work, by group*

	Professional Teachers			Student Teachers		
	Less than 1 hour	1-3 hours	More than 3 hours	Less than 1 hour	1-3 hours	More than 3 hours
AT	23.8	57.1	19.1	15.0	40.0	45.0
EE	27.4	51.0	21.7	0.0	85.7	14.3
ES	31.9	49.3	18.8	13.8	55.8	30.5
GR	29.3	53.6	17.1	11.9	42.4	45.7
HU	35.4	54.9	9.7	15.2	46.7	38.0
LT	15.3	57.6	27.1	7.1	71.4	21.4
MT	30.0	60.0	10.0	7.1	35.7	57.1
PT	42.9	38.5	18.7	22.8	42.1	35.1
SK	33.3	53.9	12.8	15.2	50.0	34.8
TU	23.3	55.5	21.2	17.0	59.6	23.4
Total	26.7	53.4	19.9	15.4	55.4	29.2

In line with OECD (2019) statistics, both student and professional teachers showed low **previous experience in online courses**. About 12% STs and 16% PTs had started at least one course per year in the three years before their participation in the experiment. However, as shown in Figure 4.3, the situation was quite diverse across countries. It is quite evident that some countries stood out: particularly, Austrian and Spanish teachers showed a much higher than average familiarity with online learning, suggesting that online courses for initial teacher training and for teacher continuous professional development are more widespread in these countries (Figure 4.3).

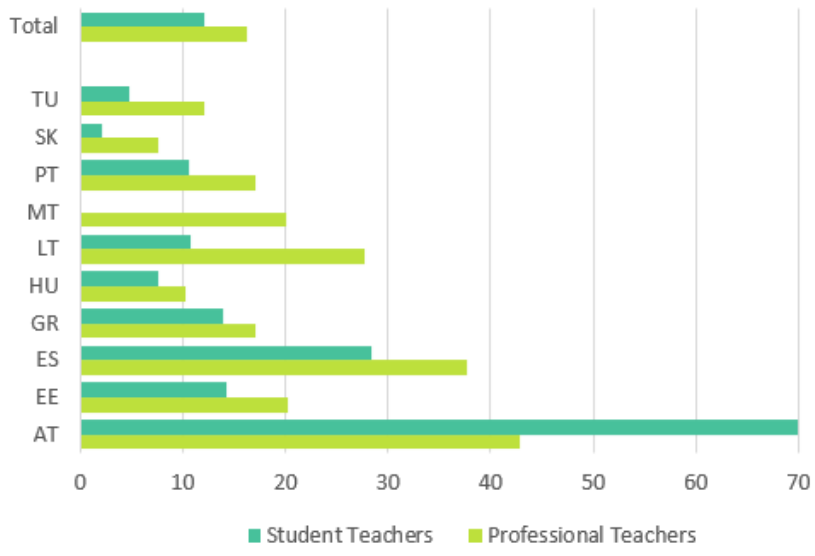


Figure 4.3 *Percentages of teachers who started more than 1 course per year in the last 3 years.*
TeachUP teachers, on average, had very limited past experience in online courses

Overall, at baseline TeachUP PTs and STs show medium levels of **SRLO competences** (between 0.58 and 0.65 on a scale ranging from 0 to 1). Differences between the three SRLO indices (goal setting, task strategies and help-seeking) and across countries were rather limited (Figure 4.4).

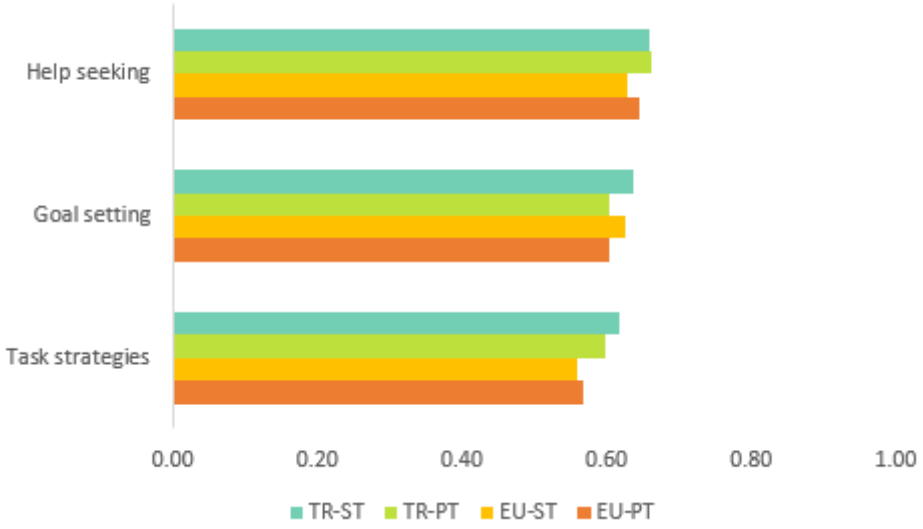


Figure 4.4 *Self-Regulated Learning Online indices (range: 0=low; 1=high)*
TeachUP teachers displayed average levels of SRLO, by group

Teachers' views about online learning and the value they attach to it as a form of training and professional development represent a further aspect of interest in the study. In this regard, **both PTs and STs showed positive views about online learning** (Figure 4.5), suggesting that the teacher workforce generally welcomes investments in online training. For some specific subgroups - i.e. such as PTs in Malta and Sts in Portugal), the average views were exceptionally high.

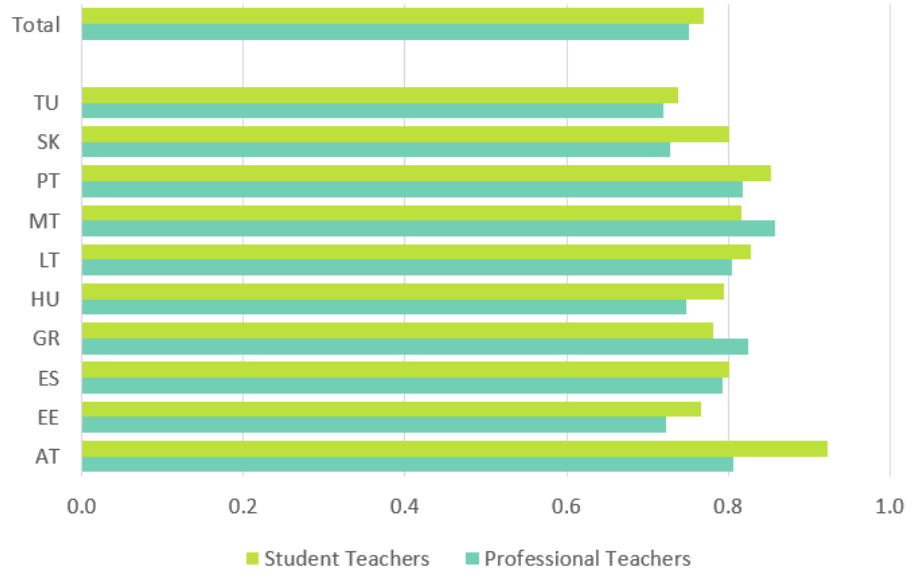


Figure 4.5 *Student and professional teachers' views on online training (range: 0= negative; 1=positive), overall and by country*

Teachers had positive views about online training courses

4.2. Patterns of course participation

TeachUP teachers showed different patterns of participation in the four TeachUP courses. A first indicator that we took into consideration was course **enrolment**. Teachers were free to choose to enroll in one or more courses among those made available. As Figure 4.6 shows, it appears that, apart from a markedly higher preference for course 1 "Formative assessment" among PTs EUs, in general teachers assigned the same preference to the four courses.[16] This was mainly due to the fact that most teachers (66%), in fact, enrolled in all courses.

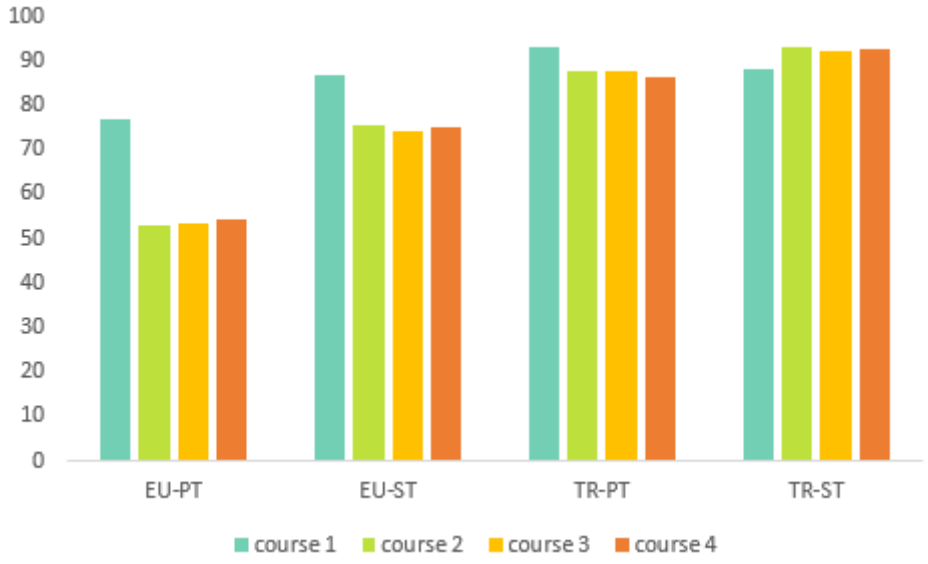


Figure 4.6 *Proportion of enrolled teachers in the four courses, by group*

Note: The figure considers teachers who enrolled in at least one of the four courses (3,777). Users who signed

16 In the BS, the percentage of teachers who stated that they wanted to participate in TeachUP courses was over 80% for each of the four courses.

up to TeachUP but did not register to any of the four courses or decided to unenrol were not considered here and in the following figures and tables.

Enrolling in a course does not imply taking part in it. As it often happens in online courses, individuals enroll but then may not actually start the course for a variety of reasons. Likewise, starting a course does not mean completing it. Figure 4.7 shows the natural flow of enrolment, **start** and **completion** rates observed in TeachUP.[17] Out of ten TeachUP teachers, seven started at least one of the four courses and nearly three completed at least one. Figure 4.7 also shows that the majority of teachers enrolled in all four courses, but when considering those who actually started or completed, the proportion of those attending four courses shrank substantially.

These figures are higher than the estimated median completion rates in scalable online courses such as MOOCs (Jordan 2015). Two possible arguments could explain this fact. First, the key features of the courses that according to the literature favour retention rates (i.e., short duration, high-quality contents, availability of tools for interactions). Second, even if this should be further tested, the specific target population (i.e., teachers in-service and in-training) could in principle be more interested and motivated in training than average online students.

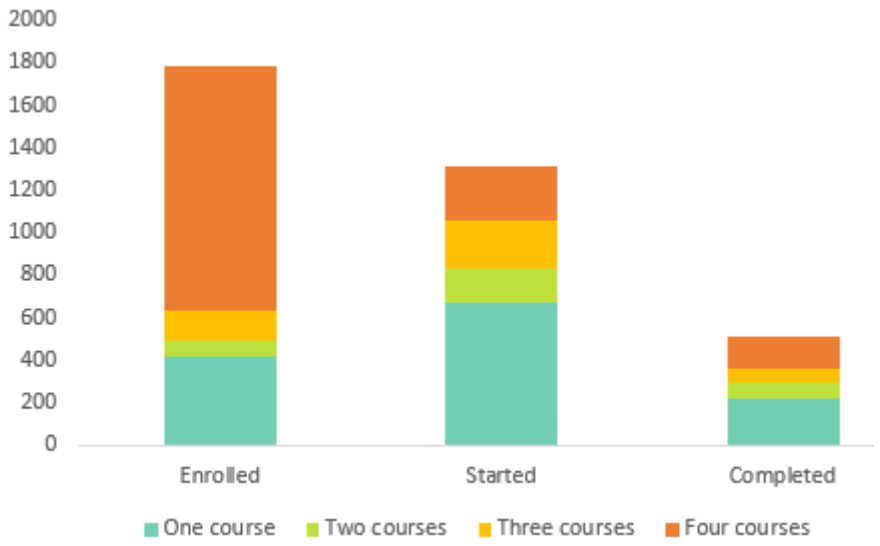


Figure 4.7 *Natural participation patterns in TeachUP*
Out of 10 enrolled teachers, seven started at least one course, and only three completed
Note: Absolute numbers observed in the control group

Figure 4.8 shows the same statistics distinguished by group and course. The **start rates** in course 1 were, depending on the group, between 70% and 50%. They **decreased considerably in subsequent courses**, especially in Turkey and among STs in the EU where they dropped by more than half. For PTs in EU countries, the decrease was less pronounced but still considerable. Completion rates (calculated among enrolled) are, on average, even lower, around 20% in course 1 and slightly lower, but fairly stable - except for STs in Turkey, in subsequent courses. Interestingly, completion rates calculated among starters, showed an opposite pattern in all four groups: i.e., they grew across the four courses. Hence, a kind of positive self-selection took place: the few teachers who started a course after the first one, showed a very high probability of successfully completing it.

17 We focus on teachers belonging to the control group only as their participation was not influenced by the intervention. Hence, we look at the so-called natural (starting and completion) rates.

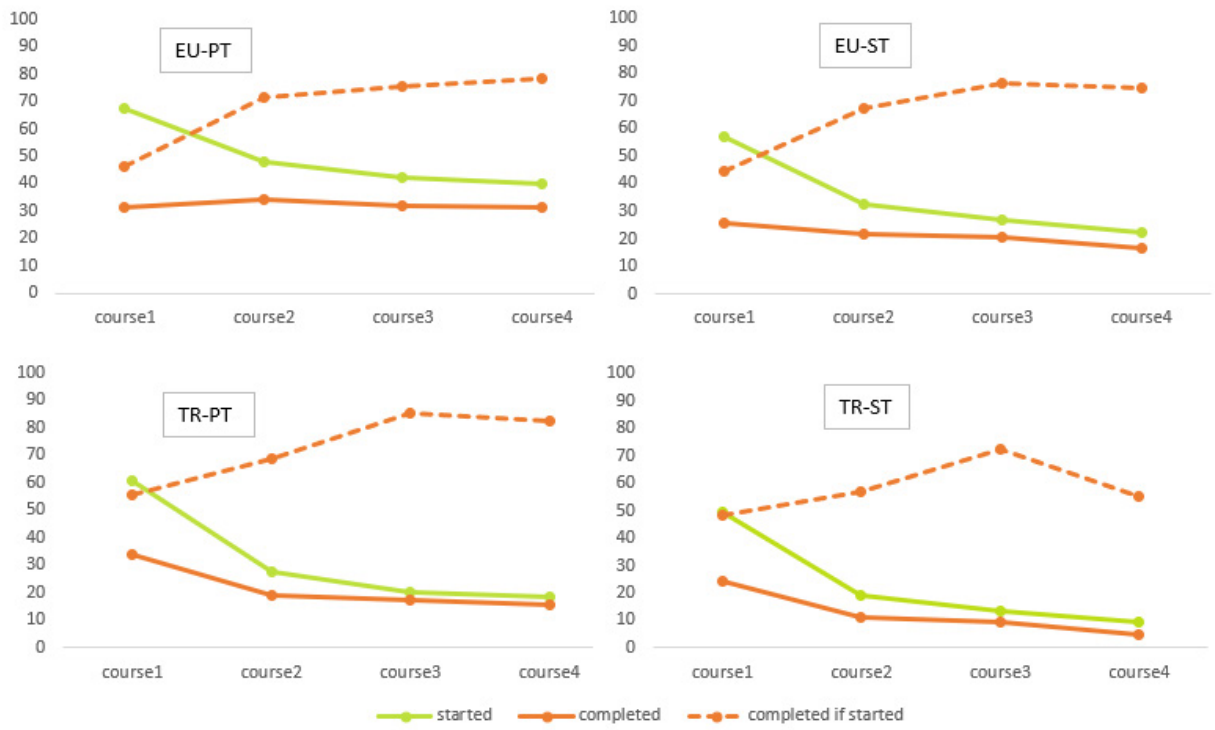


Figure 4.8 *Natural participation patterns, by group*
Start and completion rates among enrolled and completion rates among started, by course
Note: Percentage values computed in the control group

As illustrated in section 3.4, two separate courses were activated in each country and for each of the four courses, leading to a total of 80 separate ***virtual classes***. The number of participants in each of these virtual classes varied as a result of the different rates at which teachers enrolled and started the courses (Table 4.5). In Turkey, the number of participants was substantially higher than the average and median course in the European Member States. In course 1, the number of Turkish participants was 9 times larger than the average number of participants in the EU Member States. In these countries, the average number of participants in each class dropped from 50-55 in course 1 to 19-25 in course 4. Median values for course 1 were 39-40 and 14-23 in course 4. Course 2 and 3 showed intermediate values, yet much closer to course 4.

Table 4.5 *Number of participants per each virtual class, by course and country*

Country	Course 1		Course 2		Course 3		Course 4	
	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment
Austria	11	13	9	10	5	9	5	8
Estonia	39	40	18	27	17	26	12	30
Spain	97	169	54	81	42	63	34	62
Greece	84	69	50	42	48	34	41	42
Hungary	83	33	27	46	21	30	22	22
Lithuania	75	66	34	35	37	24	37	24
Malta	4	19	2	16	1	13	1	15
Portugal	38	62	19	34	13	24	14	23
Slovakia	17	28	4	9	4	6	3	3
Average EU	**50**	**55**	**24**	**33**	**21**	**25**	**19**	**25**
Median EU	**39**	**40**	**19**	**34**	**17**	**24**	**14**	**23**
TU	**462**	**511**	**196**	**206**	**139**	**156**	**115**	**127**

At this point, the question whether there were differences between teachers who completed the courses and those who did not arises. Overall, completers and non-completers do not differ much on the observable baseline

characteristics, i.e. **there are few teacher characteristics determining probability to complete and their predictive power is rather limited**. In other words, with our data, it is difficult to characterize teachers who were more or less successful in TeachUP courses. That said, as shown by Figure 4.9, it is possible to say that among completers there is a slightly higher incidence of women, teachers with higher mastery of the English language (but only in Turkey), teachers teaching STEM; teachers with a good internet connection at school; teachers with some online course experience and some familiarity with professional online learning communities (not for TU STs).

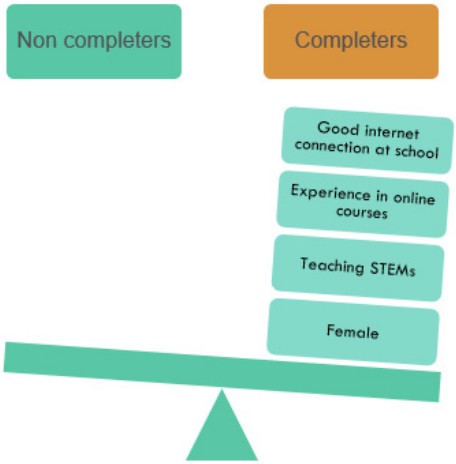


Figure 4.9 *Teachers' characteristics associated with the probability of completing at least one course*
Note: The analysis is restricted to control teachers only, i.e. teachers not receiving the TeachUP online personalised support. Complete figures are reported in Appendix F.

5. THE IMPACT OF PERSONALISED SUPPORT ON COURSE PARTICIPATION

Does online personalised support increase student and professional teachers' participation in online courses? This is the first TeachUP evaluation question, which this section tries to answer.

First, the overall impact estimates are presented and discussed. Second, the prevalence of different effects across groups and types of teachers is investigated through impact models run separately on the different groups. Third, the results of a by-course impact analysis are presented with the aim of gaining a deeper understanding on how personalised support worked as the experimental setting changed.

Box 5.1 *Summary of findings*

Main findings:

- Personalised support **increased course completion rate among enrolled professional teachers in the EU by 10 percentage points**: treatment group teachers in EU had a 42% probability of completing a course vs 32% among controls;
- This result was a combination of some interventions impacting on **start rates** and other interventions impacting on **completion rates**;
- Personalised support did not work for everyone. It did not change course participation of student teachers in the EU nor for professional and student teachers in Turkey.

5.1. Overall impacts on course completion

The impact analysis was performed separately for the four groups identified by teacher status (PTs vs STs) and national context (EU vs Turkey). Within each of these four groups, the estimate of the impact of personalised support was obtained by taking the difference between the average completion rate observed in the treatment group and the same value observed in the control groups. Thanks to randomization, this simple difference between the two averages can be interpreted as the causal effect of the personalised support system.

In reality, in order to take the complex sampling and randomization process and the 4-course setting into due account, the estimation strategy adopted was slightly more complicated than this. More technical details about the estimation strategy behind the numbers that are presented in this section are included in Box 5.2.

Box 5.2 *Analytical approach*

In sub-section 5.1, the outcome variable is the probability of completing among the enrolled. This variable takes value 1 for those who complete a course and value 0 for those who enrolled in that course without completing it;

To account for the fact that teachers could participate in more than one course, in the analysis they were considered as many times as the number of courses they enrolled and the data were analysed by the means of multilevel linear regression models in which observations were "nested" within individuals. This modelling also allowed us to adjust standard errors for clustering.

To improve the statistical precision of the estimates, all models included dummies for the randomization strata.

More technical details are reported in Box 5.3 and in Appendix F.

Starting with professional teachers in EU Member States, the analysis shows that personalised support had a sizable and positive impact among the enrolled. This impact is estimated in 10 percentage points: while teachers in the control group showed a 32% probability of completing a course, for those in the treatment group the same probability resulted to be 42%. Hence, **personalised support boosted completion rate by 10 percentage points among the enrolled**. This result is robust to a series of robustness checks, i.e. different estimation approaches, which lead to the same conclusion (see Box 5.3).

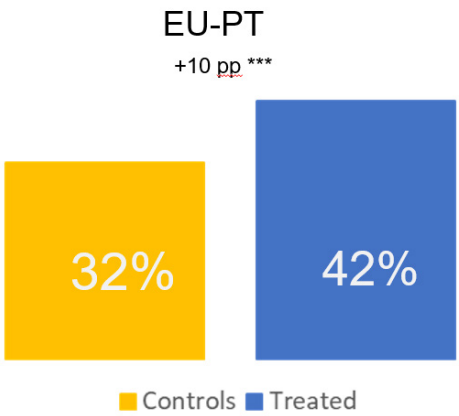


Figure 5.1 *Overall impact on course completion among enrolled professional teachers in EU MSs*
Personalised support increased course completion for professional teachers in EU MSs
Note: Controls' and treated teachers' bars indicate average probabilities of completing a course among enrolled. For teachers who enrolled in more than one course, the average completion rate is considered. The impact estimate is the regression-adjusted difference between treated and control teachers, obtained via a multilevel regression model that allows for correlation of standard errors within individuals' repeated observations. Models control for randomisation blocking variables. Portugal data for course 1 are not considered in this analysis because treatment was not delivered.

The same result was not there in the remaining three groups. For EU STs, TR STs and PTs, the completion rates of treated and control teachers were substantially and statistically the same. Interestingly, the natural completion rates of EU students and Turkish teachers are quite similar, and clearly below EU professional teachers' natural completion rates, while Turkish student teachers' completion rates are even lower.

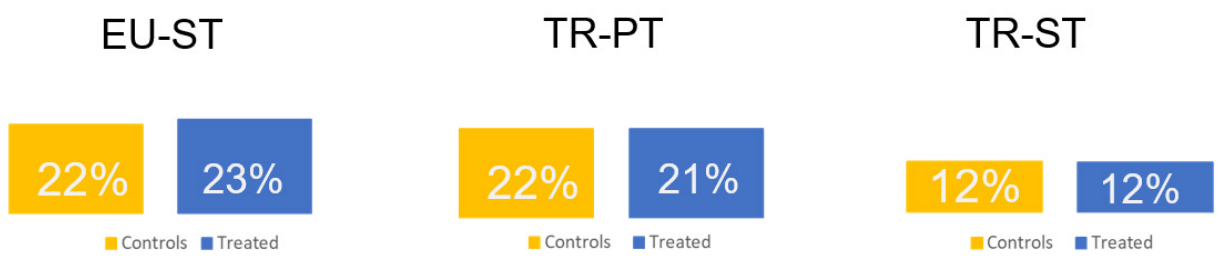


Figure 5.2 *Overall impact on course completion among enrolled, by group*

Personalised support did not increase course completion for the other groups

Note: Controls' and treated teachers' bars indicate average probabilities of completing a course among enrolled. For teachers who enrolled in more than one course, the average completion rate is considered. The impact estimate is the regression-adjusted difference between treated and control teachers, obtained via a multilevel regression model that allows for correlation of standard errors within individuals' repeated observations. Models control for randomisation blocking variables. Portugal data for course 1 are not considered in this analysis because treatment was not delivered.

Box 5.3 *Additional analyses*

To make sure that results reported in Figures 5.4 and 5.5 do not depend on the adopted estimation strategy, a number of other estimates were produced.

First, we replicated all models including a rich set of characteristics collected in the BS to improve precision and found that the results are the same as in the main specification presented above.

Second, we changed the way in which course completion was measured and reached qualitatively the same conclusion:

- Using a continuous variable (i.e., the **number of courses completed**): among EU PTs, treated teachers completed on average 1.06 courses vs .76 among controls (+ .30 courses);
- Using the probability of completing **at least one of the four courses**: among EU PTs, treated teachers had +15 percentage points higher probability of completing at least one course (.48 vs .33).

Third, to test the extent to which the results in the EU group were driven by specific countries, we changed the population on which the effects were estimated by replicating the impact models excluding one country at a time (**leave-one-out validation test**), i.e. the same model was replicated 9 times, each time removing one of the nine EU countries. Results show that for EU PTs, all alternative estimates are statistically indistinguishable from our above-presented estimate and that all are significantly different from zero and ranging from 6.8 to 13.8 percentage points. Also, the null effect found for STs in EU MSs is confirmed by this validation analysis but for two cases out of nine for which impact estimates resulted to be statistically significant, although with different signs, hence not providing a clear indication of a systematic bias in the analysis.

All these additional analyses reinforce the results shown in Figures 5.4 and 5.5 and are included in Appendix G.

A further analysis consisted in replicating the estimation models on the sub-sample of **targeted teachers**, i.e. treated and control teachers that were identified as in need of support based on their profile measured at baseline (as explained in section 3.4). The results on these subs-samples reveal that, for PTs in EU MSs, the impact of personalised support was qualitatively the same and remained statistically insignificant for the three other groups.

The estimates presented so far are average treatment effects. However, it might be the case that some teachers experienced more or less benefits than the average teacher. Such systematic variations in program impacts is called heterogeneity in the effect of the intervention. To investigate this aspect, the sample was split into subgroups defined by some relevant individual characteristics measured at the BS. The bottom line is that there is **no consistent evidence of different effects of personalised support**. More detailed results are summarised in Box 5.4.

Box 5.4 *Heterogeneity of impacts (or subgroup analysis)*

The existence of systematic differences in personalised support impacts across subgroups of STs and PTs was investigated through heterogeneous impact models hypothesizing that the tested intervention could have had different effects according to the pre-treatment level of some characteristics as measured in the baseline survey. To this end, we estimated the treatment effects in the same way as before but within given sub-groups of teachers characterised by low vs high past experience in online courses (having started/completed at least 1 course per year in the last three years) or low vs high frequency of engagement in CPD (for PTs only);

The analysis suggests that, **among PTs, there is no evidence of different effects of personalised support** (from both EU countries and Turkey) depending on the value of the variables listed above.

Some evidence of heterogeneity was found among **STs in EU countries depending on the level of experience** with online courses: personalised support had a positive impact (+13 percentage points, estimated with a model controlling for baseline characteristics) on course completion among those who completed more than 1 online course per year in the past three year, while it had zero effects on those who completed at most one course per year. This result seems to suggest that **having some past experience in online training is a precondition for benefiting from personalised support**.

To interpret this finding, it is important to consider that very few teachers took up the offer of a personalised 1:1 session (see also section 7 and conclusions on this). Hence, what could have made a difference for this particular group of student teachers was probably some other psychological mechanism that was triggered by the personalised support messages. Teachers with some past experience in online courses have higher competences and more positive attitudes towards online training and, thanks to this higher readiness to learn online, they could also be more prone to react to an external support like the one provided in TeachUP (Fixsen *et al*. 2009). This would be in line with research in the behavioural sciences which shows that people primarily focus on hassles or the mental effort required to perform a new action instead of the benefits. Through a process known as output interference the consideration of unwelcome factors interferes with the effective consideration of any subsequent factors (Castelo *et al* 2015). A further discussion about the possible mechanisms explaining the results is proposed in section 9.

5.2. Unfolding the process

In order to gain a deeper understanding of how the above-reported impacts were generated, three main features of the TeachUP experimental settings need to be closely examined: a) the four courses were delivered in sequential order; b) the four courses registered different participation patterns (i.e. different «natural» start and completion rates); c) the personalised support mechanism was partly different across the four courses. This analysis is restricted to EU PTs, as this is the only group on which there is evidence that personalised support had any impact.

The fact that the **four courses were delivered in sequential order** means that teachers gained TeachUP-specific experience throughout the experiment. In courses 2, 3 and 4, over 80% of teachers had completed at least one TeachUP course before; and in both the third and the fourth course, about 70% had completed all previous

courses (Figure 5.3).[18] Moreover, the offer of four subsequent courses could also have reduced the room for personalised support to have an impact. Those who had already participated in a TeachUP course before already knew how to navigate the platform and how the courses were organised. Hence, **it could be hypothesized that personalised support had a higher impact on completion of course 1 (among those who started) and smaller impacts in subsequent courses.**[19]

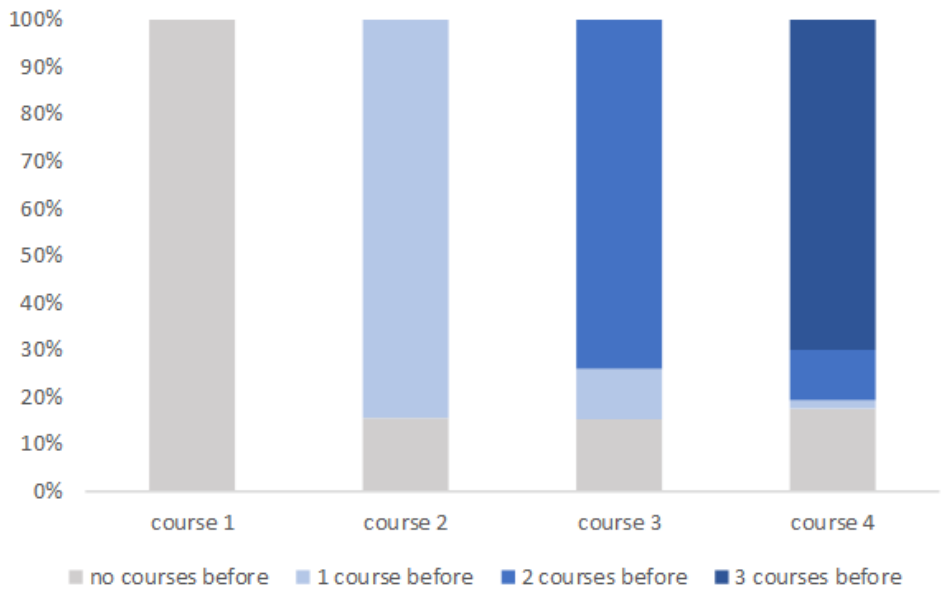


Figure 5.3 *Proportion of teachers who completed TeachUP courses before enrolling in a new course*
Teachers accumulated Teach-up specific experience throughout the four courses

Participation patterns (i.e. different «natural» start and completion rates) were very different across courses. As figure 5.4 shows, the "natural" starting rate (i.e., the proportion of control teachers starting a course among those who enrolled) decreased sharply from course 1 to course 2, and then decreased further, even if slowly, in courses 3 and 4. Overall, the decrease was substantial: 67% started the first course and about 40% started course 4. Conversely, the probability of completing a course among starters increased throughout the courses. It scored about 46% in the first course and boosted up to 78% in the last two courses. Hence, these **teachers showed higher chances of completing a course even if not supported** to do so by the personalised support. This suggests that a **"self-selection" mechanism** was in place. This statement is partly confirmed also by the finding that teachers who enrolled in more than one course showed some slightly higher intrinsic motivation and lower external motivation towards the online training.[20]

These patterns have a potentially important implication for our evaluation, because the higher the natural completion rate (i.e. the completion rate among controls), the lower the room for treatment to have an effect (as it is difficult to

18 The fact that the four courses were offered in series most likely affected participation choices for many reasons: appreciation by participants of the course quality might have induced them to go to the following courses. On the other hand, it might be the case that having already participated in one course could have reduced one's likelihood of participating in another one (e.g., due to constraints in available time to dedicate to extra training).

19 This statement may appear to be in contradiction to the finding that personalised support increased course completion among student teachers in the EU with some past online experience, while it did not have an impact on those with no past experience (see box 5.4). In that case, experience referred to participation in online courses occurred outside TeachUP, while here we refer to TeachUP course-specific experience. It can be assumed that the consequences of these two types of online course experiences are very different in relation to our specific interest.

20 Intrinsic motivation involves engaging in the course because it is personally rewarding (i.e. it is relevant for job, for personal growth and enrichment) while external motivation occurs when one engages in order to get something back or avoid something unpleasant (in this context is when the school head suggested to enroll). Confirming this self-selection mechanism, 90% of those who enrolled in more than one course already planned it at the beginning, as emerging from the baseline survey data.

increase completion rate further if it is already very high). Hence, **it could be hypothesized that personalised support had higher impact on completion of course 1 (among those who start) and smaller impacts on completion in subsequent courses.**

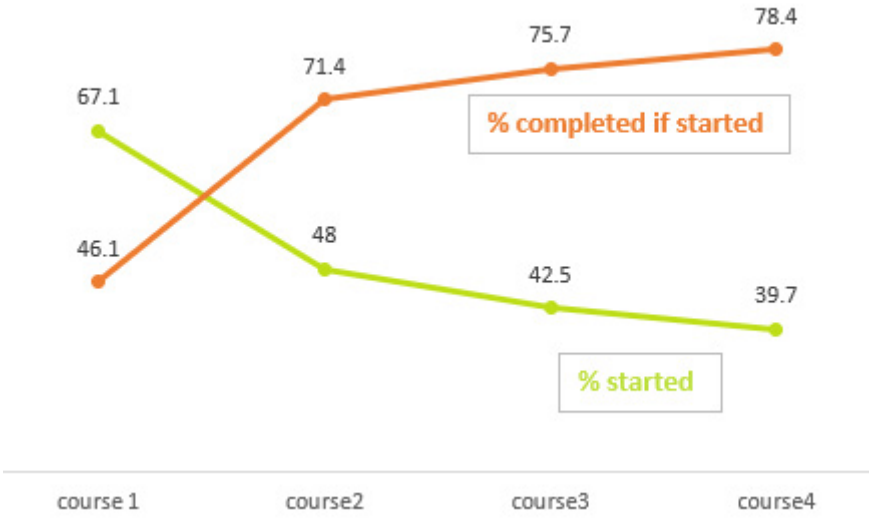


Figure 5.4 Start and completion rates (among starters) in the absence of treatment, EU PTs
As the probability of starting a course goes down, the probability of completing it goes up
Note: Rates estimated for control group teachers

Personalised support offered to teachers was different across the four courses (Figure 5.5). As a consequence of unexpected delays in the setup of the platform, interventions based on participants' behaviour on the course platform were only implemented from course 2 onwards. The most important difference between course 1 and the other courses was that the latter had a prominent intervention affecting teachers' probability to start (i.e., intervention 5 - personalised reminder to start the course among those who enrolled but did not start the course after 5 days), while in course 1 this intervention was not implemented (see also section 6 on the implementation of the different interventions). **Hence, it could be hypothesized that in courses 2, 3, and 4 treated teachers show higher start rates than control teachers.**

Interventions based on:	Course 1	Course 2	Course 3	Course 4	
baseline characteristics	✓	✓	✓	✓*	→ Mostly affecting retention
behaviour on course platform	×	✓	✓	✓	→ Mostly affecting start (Intervention 5)

* In course 4 an additional static intervention was offered

Figure 5.5 Types of interventions implemented in the four courses
Personalised support was full-scale implemented only in courses 2,3 and 4

The consideration of these three features led us to hypothesize that personalised support had positive impacts on start rates in courses 2, 3 and 4 and a positive impact on completion rate (among starters) in course 1. To disentangle the overall impact observed among EU Professional Teachers considering the three features of the TeachUP setting, different impact models were estimated on course start and completion and separately by course. The results are shown in Figure 5.6.[21]

21 Results for the three other groups are not shown here but included in appendix G as they reflect the "zero" effect results already commented on above.

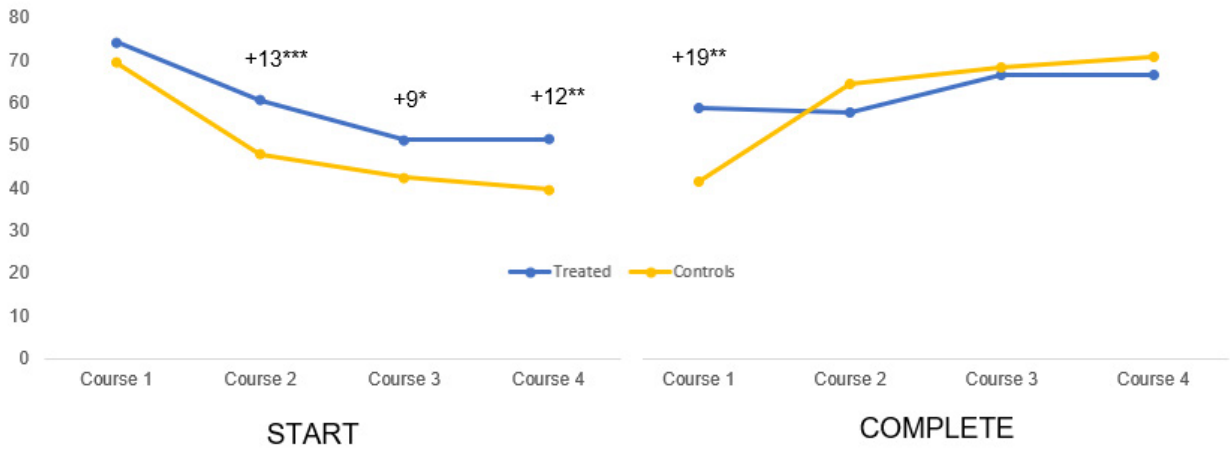


Figure 5.6 *Impacts of personalised support on start and on complete a course, EU PTs*

Personalised support increased the probability to start and to complete the courses

Note: Controls' and treated teachers' line indicate average probabilities of starting or completing a course (among starters). The impact estimates are regression-adjusted differences between treated and control teachers, obtained via linear regression models that control for randomization blocking variables. Standard errors are clustered at the school/organisation level. Portugal data for course 1 are not considered in this analysis because treatment was not delivered.

The yellow line in panel a of Figure 5.6 shows the natural start rate (i.e., the start rate observed in the control group), while the blue line in the same panel shows the start rate in the treatment group. Panel b) follows the same logic but uses "completion" rates instead of start rates.

The results of this analysis confirm the expectations. A positive impact on the probability of starting a course is found only in courses 2, 3 and 4 (left panel), while no difference between treated and non-treated is found in course 1. Intervention 5 (i.e., the reminder to start the course among "slow starters") made a difference for the probability to start. Among teachers who had started courses 2,3 or 4, the number of teachers in the treatment group was larger than the one in the control group and the difference between the two groups was coincident with the number of teachers in the treatment group who received intervention 5 (Figure 5.7).

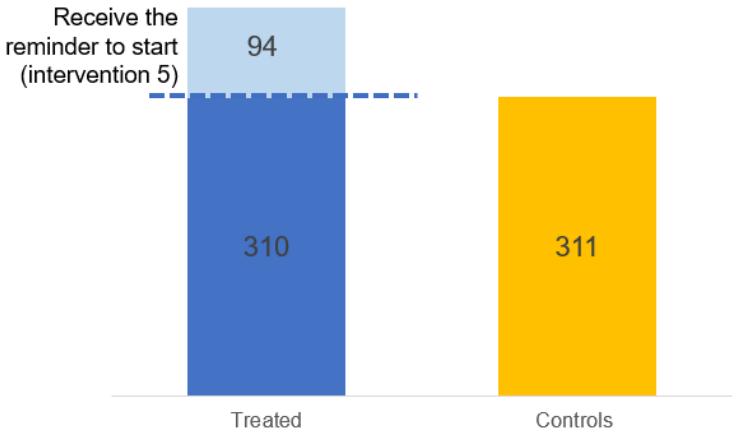


Figure 5.7 *Number of recipients and non-recipients of intervention 5 among teachers who started courses 2,3, and 4, EU PTs*

Treated teachers were more likely to start a course thanks to intervention 5

When looking at completion rates (panel b in Figure 5.6), a large and statistically significant difference between treated and control teachers is only observable in course one (+19pp), while no significant differences are ever found in courses 2,3 and 4.

This latter result is an important one and deserves to be commented further. Personalised support induced "slow starters" to start a course - i.e. it changed the behaviour of teachers who enrolled in a course but would have not started it in absence of personalised support. It is likely that teachers who enrolled but did not start a course faced tighter time schedules or unexpected workloads, but it is also possible that slow starters were teachers lacking motivation or learning planning competences (as they enrolled but then did not start the course activities up to five days of course official start). In line with this hypothesis, baseline information suggests that the likelihood of being a "slow starter" was highest among teachers who, prior to TeachUP, started more online courses than the ones they actually completed. Also, slow starters were those declaring lower educational attainment and lower internet connections at school. All in all, teachers not starting a course in time were likely teachers with a higher-than-average risk of not completing the course. In theory, intervention 5 may have had a negative impact: it may have induced teachers to start a course to then drop out before completing it. This did not happen, as treated and untreated teachers showed the same completion rates in courses 2,3,4 (Figure 5.6 panel b). The reason for this result could be due to the fact that teachers who did not start in time and hence received intervention 5 also had higher likelihood of receiving other personalised support interventions (Figure 5.8): i.e., interventions 1-4 (46 vs 17%) and other interventions triggered by their platform behaviour affecting completion (i.e., interventions 6-8, 35 vs 29%). Given that the former had a neat impact on completion in course 1, it could be plausible to argue that they were effective also in courses 2, 3 and 4.

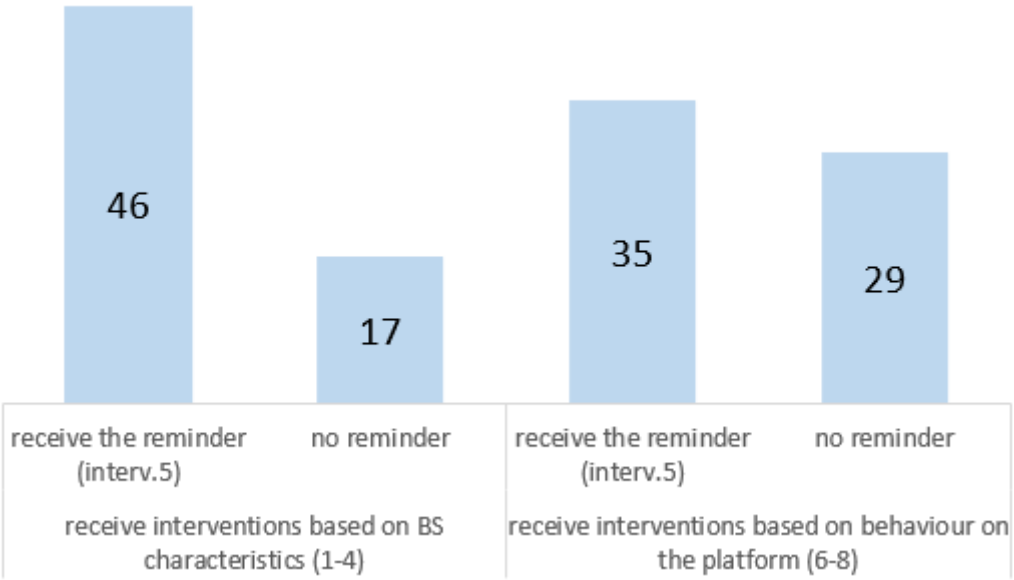


Figure 5.8 *Proportion of recipients and not recipients of intervention 5 who received other interventions.*
Recipients of intervention 5 were also more likely to receive other support interventions
Note: Only teachers in the treatment group who started course 2, 3 and 4 were considered for this analysis.

Overall, these findings lead to the conclusion that personalised support **had a positive impact on completion among the enrolled**. This result was achieved both by helping teachers complete the courses they started and by making teachers actually start the courses they enrolled in. The next section will seek to shed more light into the mechanics of the TeachUP personalised support in the attempt to understand more about the implementation of the personalised support interventions.

6. UNPACKING THE "BLACK BOX"

The TeachUP personalised support as a whole led to an increase in course completion rates for professional teachers in EU MSs but not for the other groups. **But what was personalised support in reality?** Were there differences in the delivery of personalised support across contexts or groups? And how did teachers react to the offered interventions? In the attempt to understand more about how the personalised support protocol was translated into real interventions, this section provides insights into three key implementation aspects. First, it quantifies the number of interventions that were actually activated across the four courses (section 6.1). Second, it explores the extent to which teachers received more than one intervention (section 6.2). Third, it quantifies the occurrences of personalised support 1:1 sessions that actually took place in addition to the personalised messages (section 6.3). Section 6.4 explores the levels of teachers' satisfaction with the courses.

6.1. Quantification of the different interventions

As illustrated in section 2, support interventions were different in *nature* and had different *goals*. Each intervention was activated by a specifically defined trigger and corresponded to specific actions. Two main types of interventions were identified: those triggered by a given set of teachers' characteristics measured as baseline and those triggered by a set of teachers' actions or inactions on the course platforms.

Table 6.1 shows the count of interventions activated during the whole project. Overall, interventions 1-4, which were based on participant profile and offered teachers the same action, and intervention 5 which was based on a participants' inaction - i.e. not starting a course 5 days after launch - made up 91% of all interventions (intervention 5 accounting for 46.5% and interventions 1-4 for 44.5% of all interventions).

A first, important, consideration emerging from these figures is that **the occurrence of enrolling in but not starting in time an online course is very widespread** in the studied sample. This is an important message for future developments in online courses targeted to teachers, as it suggests that a special attention has to be dedicated to the critical passage between enrolment and course start. In this regard, as seen in the previous section, in TeachUP **intervention 5** worked as an effective personalised reminder that succeeded in reducing the gap between enrolment and actual course start.

Interventions based on participant profile and intervention 5 outnumbered all others

Table 6.1 *Number of interventions activated during the whole project*

Intervention	Basis of Triggers	Aim	Occurrences
1-4 personal message including 20 minute 1:1 session	Participant profile	Course Completion	3532
5 Personal message including reminder, encouragement and advice	Behaviour on course platform - not starting a module 5 days after start	Course Start	3692
6-9 Personal message including reminder about deadlines, information on how to find answer, on the work done and support	Behaviour on course platform	Completion & Satisfaction	718

Note: occurrences refer to interventions delivered, not to participants. One participant could receive more than one intervention.

The second consideration is that **a non-negligible fraction of teachers turned out to be identified as "at risk" of not completing the online courses** because of a mix of poor past experience in online learning, poor digital competences or low SRLO. These teachers were offered a personalised message offering advice on how to plan their learning time and self-motivate. Again, this finding is consistent with findings reported in section 5, pointing to the efficacy of intervention 1-4 to increase completion rates.

The distribution of the interventions was found to be similar across the four groups (i.e., EU PTs in EU, STs in EU, PTs in Turkey and STs in Turkey), hence this does not seem to be an explanation for the heterogeneity of impacts (see Appendix G).

6.2. The distribution of interventions across the four courses

A teacher could receive multiple interventions. Figure 6.1 shows the overall breakdown of the intervention mix that teachers received across the four courses (course-specific statistics are shown in Appendix H). It should be remembered that those based on participants' behaviour on the platform were only implemented starting from the second course. This explains why in the first course the proportion of teachers who received support is much lower (38%) than in the other three courses (where it is 92%, 90% and 95% respectively). As shown before, this increase was primarily due to intervention 5 targeted to teachers who had not started the course in time.

Teachers receiving interventions only for their actions/inactions were about 55% in course 2 and 3 and 40% in course 4. Those receiving both types of interventions were one third in course 2 and 3 and a half of course 4 participants. In all the three courses, those receiving only interventions 1-4 were a minority (5% or less).

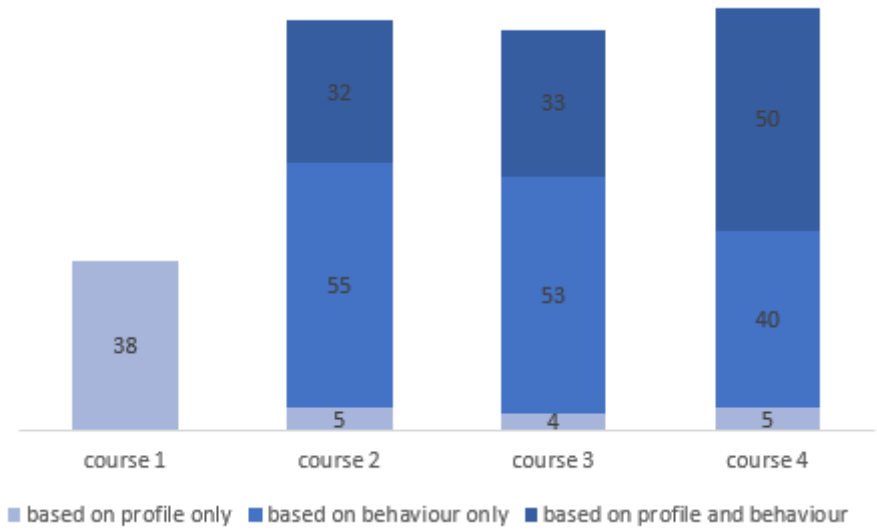


Figure 6.1 *Proportion of teachers in the treatment group who received the different support interventions, by course*

In courses 2, 3 and 4 almost all treated received an intervention

In Turkey, the proportion of teachers who received support based on teacher characteristics measured at baseline was higher than in the EU Member States (Table 6.2). This reflects the different composition of the Turkish teacher sample, who, compared to teachers in the EU, show a higher incidence of factors connected to a lower likelihood of success in online courses. However, as seen in section 5, despite this higher incidence of teachers receiving an offer of support, the latter had no impact on this population.

Table 6.2 Proportion of teachers in the treatment group who received the different support interventions, by group

Basis of trigger by course	EU PTs	EU STs	TR PTs	TR STs	Total
Course 1					
based on profile only	31.3	33.2	44.1	43.0	38.2
Course 2					
based on profile only	9.4	7.7	4.9	1.5	5.4
based on behaviour only	64.5	64.1	55.3	58.6	59.8
based on profile and behaviour	26.1	28.2	39.8	39.9	34.8
Course 3					
based on profile only	9.6	7.0	4.9	0.5	4.9
based on behaviour only	63.5	62.7	54.5	59.1	59.1
based on profile and behaviour	27.0	30.3	40.6	40.4	36.1
Course 4					
based on profile only	13.3	8.5	1.9	1.8	5.4
based on behaviour only	50.0	47.8	42.5	33.9	42.7
based on profile and behaviour	36.7	43.7	55.6	64.3	51.9

6.3. From personalised messages to personalised support sessions

As illustrated in section 6.1, some interventions offered participants the possibility to take part in a 1:1 session. However, the number of 1:1 online sessions that actually took place was very low.

Table 6.3 provides some more detailed figures disaggregated by group and shows that the acceptance rate of the 1:1 session was larger in EU countries (8%) than in TR (4%). The table focuses on the most important of the interventions offering a 1:1 session, i.e., those based on BS characteristics.[22]

Only a minority of teachers took up the 1:1 session offer

Table 6.3 Take-up of 1:1 sessions, by group

GROUP	TREATMENT GROUP	INTERVENTIONS 1-4	ACCEPTANCE OF 1:1 SESSION	%
EU PTs	513	210	17	8.1
EU STs	472	225	18	8.0
TR PTs	567	317	8	2.5
TR STs	440	282	11	3.9
Total	**1,992**	**1,034**	**54**	**5.2**

Interestingly, 79% of all kinds of interactions registered between support agents and participants focused on practical issues linked to the course. This included exchanges about a lack of time to complete the course, technology issues,

22 Also interventions 6 and 9 offered the same possibility. Their inclusion in the analysis would not change the conclusions because of the very limited number of interventions activated.

appreciation or criticism of the course or the offer of support, as well as other practical issues such as deadlines, timelines and live events of the course. Only 21% of exchanges focused on discussions regarding the course topic or its content, aspects linked to self-regulated learning, or other learning strategies and approaches.

The low take-up rates might be partly explained by an imprecise targeting of those most in-need. We came to this conclusion by comparing, among controls (so net of any possible effect of the personalised support), the probability of completing the course among targeted and not targeted. As shown in Figure 6.2, in course 1 they had a very similar probability to complete (similar results also for course 2 and 3). Therefore, we can deduce that targeting was not precise at identifying those at risk of dropping out.[23] Precisely for this reason, in course 4 targeting was slightly revised with the introduction of an additional intervention based on those individual characteristics that were good predictors of the likelihood of dropping out in the previous TeachUP courses.[24] Particularly, exploiting the predictive probability of completing TeachUP courses, we found that motivation, English proficiency, subject of teaching, age, previous experience, internet access, sex, level of education were important element in order to identify those in need. Another important predictor of course completion would have been the previous experience in Teach-up. However, we did not use it to make up the new intervention in course 4 because not all those who started that course had a previous TeachUP experience and our aim was to produce forecast estimates based on previous information that would also apply to first-time course participants. The revised targeting has improved the accuracy in identifying those most in need of help. Unlike what happened in the first three courses, the fourth targeted group showed statistically significantly lower completion rates than the non-targeted (Figure 6.2).

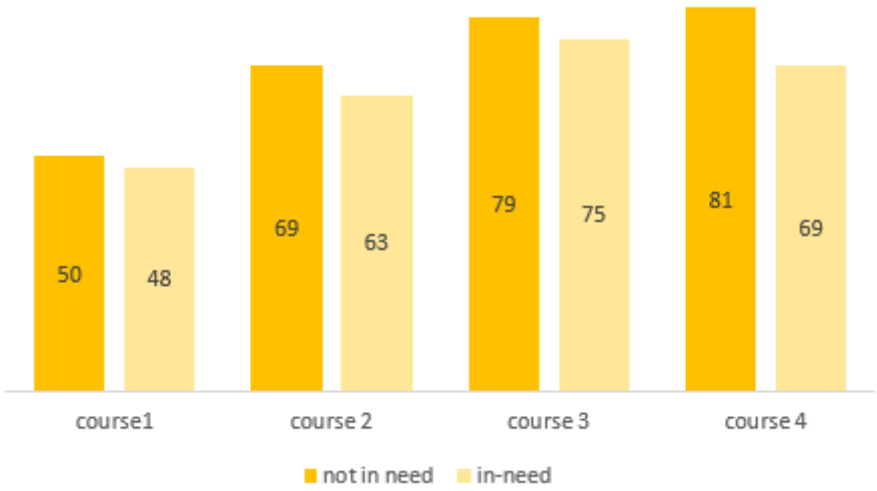


Figure 6.2 *Completion rates of teachers classified as in-need and not in-need among controls*
Targeting based on teacher profiles was more precise in course 4

23 The imprecise nature of the targeting in courses 1-3 is confirmed by the qualitative assessments of the support agents of their exchanges with the participants. Support agents identified 44% of the participants they exchanged with as not in need of support.

24 We estimated (separately for courses 1, 2 and 3) the probability of completing the course using all the baseline features as predictors. Then we compared the distribution of this probability of those who actually completed the course with that of those who dropped out. Theoretically, among the former, this probability should be as close as possible to 1, while among the latter it should be shifted to zero. In practice, however, the two curves were partially overlapped. Through a ROC curve analysis, we were able to identify how much the predictive model was able to distinguish completers from non-completers. Moreover, this tool allowed us to identify a threshold in the predicted probability able to maximize, on the one hand, the group of users who did not complete the course and who, according to their characteristics, had a high risk of doing so and, on the other hand, the group of users who completed despite their expected drop out risk was low. The ROC curve analysis was quite similar in the three courses. To refine targeting in course 4, in the end, we used the predicted probability calculated on course 3 and assigning to a new intervention all enrolled in course 4 with a predicted probability below 0.74.

6.4. Teachers' satisfaction with the courses

Teachers' satisfaction with the courses was analysed exploiting data on the post-course mini surveys for completers and data from the follow-up survey, which was filled in both by completers and non-completers. In both cases, the caveats already made, apply. The mini-surveys data was available for completers only, because non-completers had a very low response rate. Similarly, even if less severe, the problem was there for the follow up survey data, which was filled in by course completers at a higher rate.

Overall the **judgements after the courses were largely positive**: more than 9 out of 10 course completers agree (or strongly agree) with a set of aspects related to **course access, course features,** such as contents, course discussion, activities, communication, and **course perceived impact** (Figure 6.3).

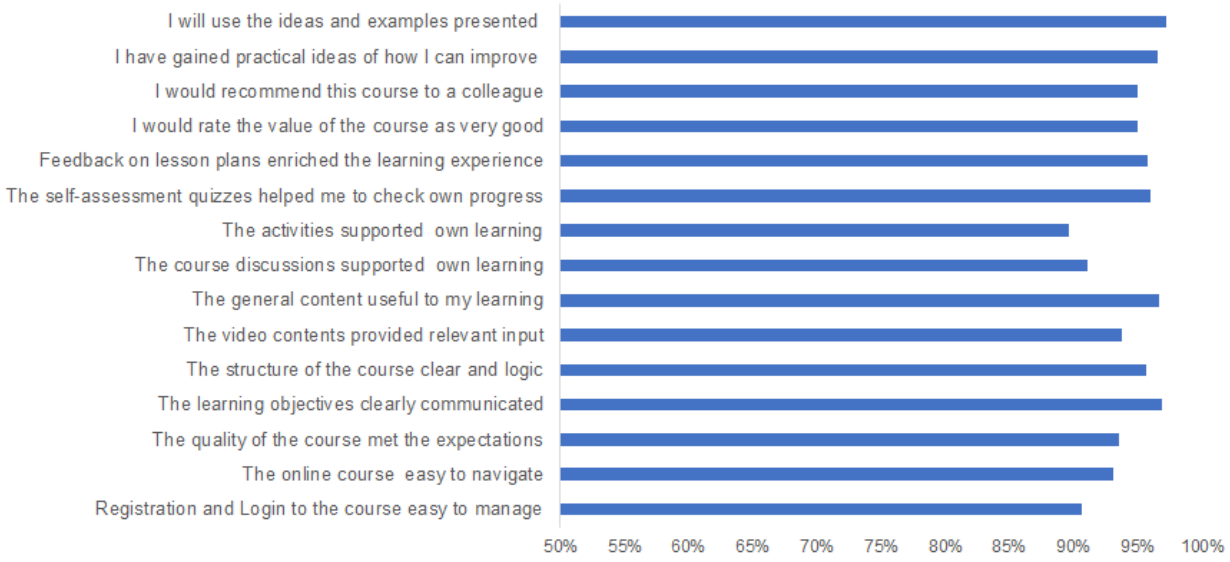


Figure 6.3 *Proportion of course completers who agree or strongly agree to the listed statements*
Teachers were highly satisfied with the courses

Looking at the less positive judgements, it turns out that these mostly concerned the first course and relate to the technical difficulties of access (registration and login) encountered by all teachers at the beginning of the first course.

It has to be recalled that these considerations refer only to those who have concluded courses. However, it is likely that the most critical were those who did not finish the course. On these, the only information comes from FUS.

Here, **among those who did not completed one or more TeachUP courses** (e.g., never even started a course, or started but not completed one or more courses) 38% declared that they faced **time constraints** and 28% experienced **technical issues** (e.g., website malfunctioning, poor connection quality) that **prevented them from successfully engaging with the TeachUP** course(s). Course contents (difficulty or relevance) were not considered matters of concern (overall less than 11% mentioned them).

7. IMPACTS OF PERSONALISED SUPPORT ON SRLO AND ATTITUDES TOWARDS ONLINE TRAINING

The second evaluation question addressed in TeachUP reads: **Does personalised support improve teachers' Self-Regulated Learning Online (SRLO)?** If yes, does it improve them directly or indirectly, i.e. through increased targeted trigger-based tutoring or via greater online course experience?

In order to answer these questions, two different approaches were carried out.

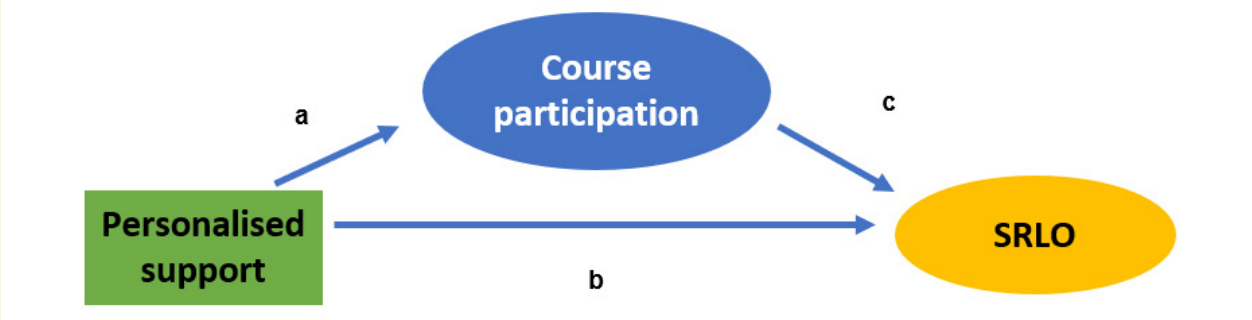
The first type of analysis was aimed at establishing if personalised support had immediate impacts on **course-specific SRLO (using the post-course mini survey)**. To this aim, the analysis investigated the impacts of personalised support on teachers' SRLO among teachers who completed any of the four courses.

The second set of analyses, instead, was aimed at investigating the impact of personalised support on teachers' **general SRLO**, as measured during the four months after the fourth course (using the baseline and the follow-up survey).

This analysis first seeks to establish if, at the end of the experiment, treated teachers displayed higher/lower SRLO than control teachers and, secondly, it provides an answer to the second sub-question outlined above, i.e. if teachers who participated in more TeachUP courses overall also showed higher levels of SRLO.

In addition to SRLO measures, the evaluation also examined if personalised support led to significant changes in the way teachers see or attribute **value to online learning**.

Box 7.1 *Summary of findings*



Main findings:

- Overall, the evidence is not very robust and more research is needed to explore the linkages between personalised support mechanisms and SRLO.
- No short-term impacts of personalised support on course-specific SRLO (a+b+c) were observed.
- No medium-term direct effects on general SRLO (a+b+c) were observed either. However, it would seem that providing personalised support decreased teachers' propensity to seek help from others to solve problems.
- Teachers who completed more courses displayed lower propensity to seek help from others, and higher goal setting strategies (c).

7.1. Course-specific SRLO

As already described in section 3.4, data to investigate the impacts on course-specific SRLO were only available for course completers, as course non-completers responded to the post-course mini-surveys at a very low rate (10%). This data constraint has two important implications.

First, the results in terms of SRLO that we are going to present can be generalised only to the teachers who completed the courses. In other words, we can learn if personalised support changed the ways in how teachers behaved during the courses that they managed to complete but we cannot know if the provided support changed teachers' behaviour in courses that they did not complete. Importantly, teachers who completed a course may have already had diverse and greater inclinations to self-regulate their behaviour in light of achieving their goals.

The second implication deals with the causal interpretation of the results. As shown in section 5, personalised support had an impact on completion (for EU PTs) because it increased teachers' probability of completing the courses. This means that, at least for professional teachers in EU countries, the simple comparison between the treated and the controls in terms of post-treatment SRLO does not easily identify the causal impact of personalised support. This is because, thanks to the treatment, there were treated teachers who completed the courses but would not have completed them in the absence of the support. This implies that when restricting the focus to course completers, treated and control teachers are no longer comparable. To make the two groups as similar as possible, at least in terms of the characteristics measured at baseline, we have included in the models a long list of baseline characteristics. This balance issue between treated and control teachers does not arise in the remaining groups of teachers (i.e., in Turkey and among student teachers in EU-MS) where personalised support had no impact on course completion.

With these two important caveats in mind, let us now turn to the results. Three SRLO dimensions were considered: help seeking; persistence; and time management. While the first two were measured through two indices (whose values ranged from 0 to 10), the third dimension was measured with three separate dummy variables. As no significant differences in the results across the four courses were found, we analyse the four courses together.

Figure 7.1 shows the results of the impact analysis on these three dimensions of course-specific SRLO and finds that personalised support did not significantly change teachers' SRLO in any of the three considered dimensions nor across any of the four groups. A possible reason for the null effect is that, as shown in 6.4, most teachers did not take up the offer of a 1:1 session, which could have, at least in principle, changed more deeply the ways teachers behaved on the course. Another possible reason, refers to the nature of the data collected: having information only on those who succeeded in completing the course implies that the data refer to teachers that, in some way, already possessed the SRLO needed to complete the course and this makes it more difficult for the provided support to improve it further.

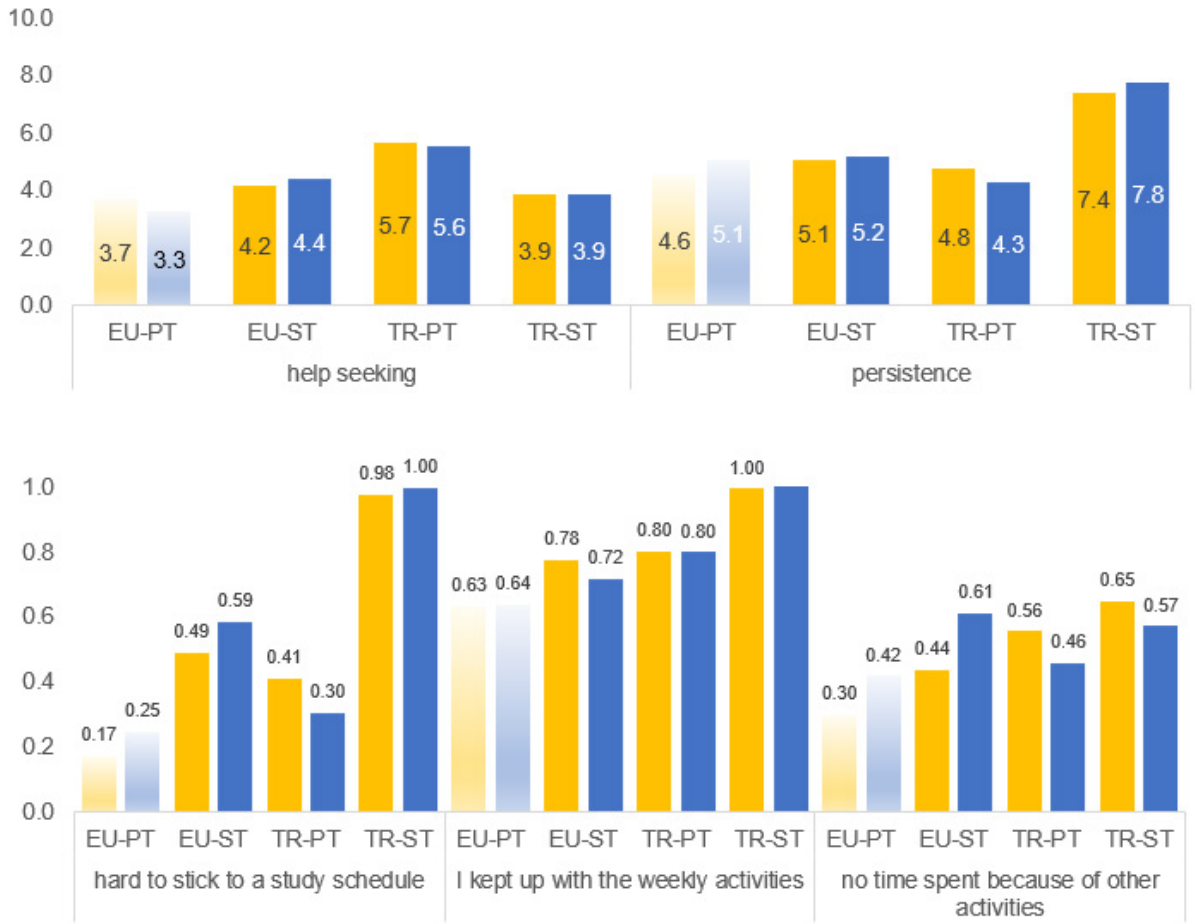


Figure 7.1 *Impacts on help-seeking, persistence and time management, by group*

Personalised support had no impact on course-specific SRLO among course completers

Note: the bars referring to EU-PTs are shaded because the treated-control differences cannot be interpreted as causal impacts of personalised support to the same level of reliability as the effects estimated for the other groups. Among EU-PTs who completed a course, treated and controls are not comparable because personalised support increased the probability of completing a course of teachers who, in the absence of the treatment, would have not completed it.

In addition to the impact estimates, the data show some interesting insights into SRLO levels and variations across groups.

Regarding help seeking, the four groups are roughly comparable and show relatively low help seeking values (i.e. low propensity to ask help to others), with PTs in TR showing slightly higher levels.

Regarding persistence, the group that stands out positively is clearly Turkish student teachers, who perceive themselves as more persistent than all other groups.

When turning the attention to the three time management indicators, globally, it looks like all involved teachers had limited difficulties in keeping up with the tasks required by the course. Students in TR show the highest values on all three indicators but these results are not consistent between each other.

7.2. General SRLO

To measure teachers' general SRLO, we used data from the Follow-Up survey. Beyond measuring general SRLO of teachers instead of course-specific SRLO, another important distinction is that when studying general SRLO we consider all teachers and student teachers who enrolled in TeachUP regardless of whether or not they have completed a course.

Unfortunately, because of the high attrition rates (i.e., only 18% of teachers responded to the follow-up survey, as illustrated in section 3.5), impact estimates are possible only for the overall sample, i.e. not distinguishing between the four different teacher groups as done in the previous analyses.

The analysis of general SRLO involved three indices: goal setting and strategic planning; task strategies, elaboration and self-evaluation; and help seeking. Figure 7.2 shows the SRLO average levels of treated and control teachers. Overall, as shown in Figure 7.2, no statistically significant differences between treated and control teachers were found across any of the three SRLO dimensions.

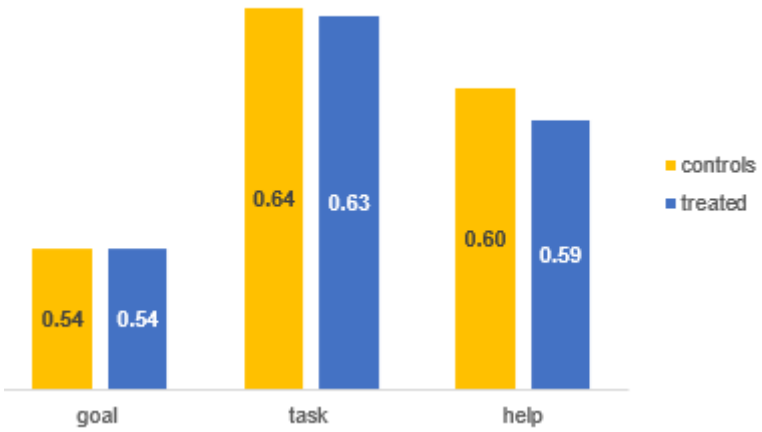


Figure 7.2 *Average impacts on goal setting; task strategies and evaluation; and help seeking indices*
Personalised support did not change teachers' general SRLO
Note: The three indices range between 0 and 1.

Box 7.2 *A deeper look into the findings points to small effects on help seeking*

When changing the metrics of the impact estimates shown in Figure 7.3, i.e. converting them in **"effect size" estimates**, we confirm that the treated/control differences are not only statistically insignificant but also substantially small. Effect size is an alternative way to look at the same thing allowing a comparison of the size of effects across groups and experiments. It is calculated by dividing the impact estimate obtained above by the standard deviation of the dependent variable in the sample models. In educational research, effect sizes larger than ±0.25 are considered substantially relevant (What Works Clearinghouse 2014). In our case, they are estimated at 0.03 for "goal", 0.06 for "task" and -0.13 for "help seeking". Hence, **with the partial exemption of help seeking, this additional analysis confirms the null effects of support on SRLO**.

To check the extent to which these results are affected by small **sample size**, we performed a simulation analysis of the results. We simulated two alternative scenarios: a) that the pool of respondents was doubled (ca 1,400) and b) an 100% response rate (3,777 respondents). **In these two simulated scenarios, only the impact on "help seeking" would be significant**, reinforcing the conclusion that the treatment had no meaningful impact on the remaining dimensions.

Based on Figure 7.2 and Box 7.2, it is confirmed that personalised support did not change teachers SRLO and only weakly modified help seeking. To deepen these results, we replicated the analyses by splitting the sample according to the pre-treatment levels of help-seeking as measured in the Baseline Survey. The idea was that of exploring heterogeneity of the impacts and testing if personalised support had higher/lower impacts on teachers who joined TeachUP with higher/lower levels of help-seeking.[25] Based on this analysis, a statistically significant and negative impact of the treatment on help-seeking was found for those who already had high levels of help seeking before the launch of the courses (Figure 7.3).

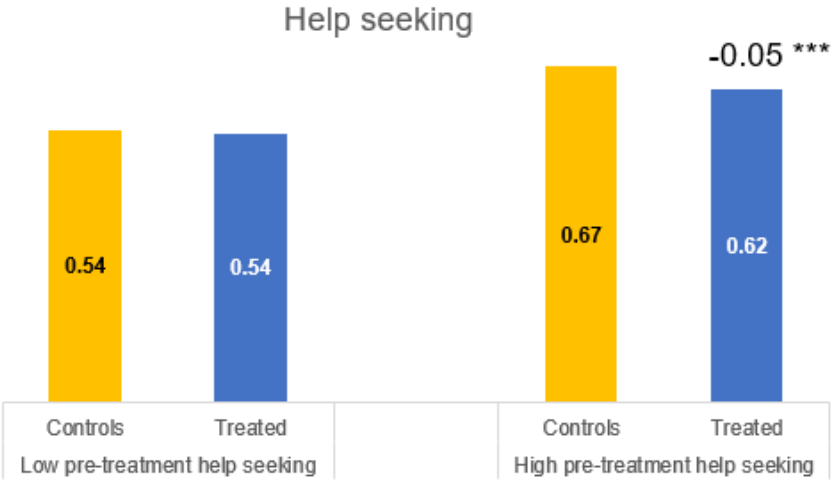


Figure 7.3 *Heterogeneous impact of personalised support on help seeking*
Help seeking was significantly reduced among those who showed a high score before joining TeachUP

Overall, we found evidence that personalised support reduced help-seeking among teachers, i.e. it reduced their perceived need to ask others for help to cope with encountered difficulties.

How to interpret this finding? To answer this question, it is necessary to go back to the specific items used in the definition of the help seeking. The three items used to measure help seeking (also used in Kizilcec et al 2017) were:

i) when I do not understand something, I ask others for help;
ii) I try to identify others whom I could ask for help if necessary: and
iii) I ask others for more information when I need it.

So, in what sense could personalised support have reduced help seeking? In line with other research (Kizilec ***et al.*** 2017), two possible interpretations could be advanced. First, thanks to the received personalised support, treated teachers may have felt more confident in their ability to succeed in the course and solve their 'problems' by themselves. In other words, the offered personalised support may have given them a boost of self-confidence and self-efficacy. This would have reduced their perceived need of help. A second possible explanation, instead, is that treated teachers were aware of the existence of an ***ad hoc*** support provided by the personalised support agents and hence felt less inclined to ask generic and undefined "others" for help: they knew whom to ask for qualified help in case they needed it.

Furthermore, according to some literature, the use and interpretation of help seeking indices require some methodological ***caveats***. According to Kizilec and colleagues (2017), high level of help-seeking does not seem to be related with positive learning outcomes. The authors found that help seeking was a strong negative predictor of goal attainment, while e.g. goal setting was a strong positive predictor. Aleven et. al. show (2006) showed that "help abuse" correlates negatively with learning. Help abuse is defined here as situations where the students misused the help facilities provided by a cognitive tutor or used them unnecessarily.

25 For setting goals and strategic plans competences, we found no difference between treated and controls regardless of the pre-treatment levels.

Lastly, this effect may also be a result of the short nature and time-span of courses which in general demand further attention as to their impacts and potential for SRLO development understanding and promotion respectively.

7.3. The link between course participation and SRLO

Now, let us assess the relationship between participation in TeachUP courses and SRLO. Do teachers who participate in more courses also show higher ability to learn online?

To answer this question, we adopted a different methodological approach from the one applied to estimate the impacts of personalised support. Contrary to personalised support, course participation was not randomised, hence a simple comparison of teachers completing and teachers not completing a course would not yield a reliable estimate of the impact of course participation on teachers' SRLO: as shown in section 4.2 (Figure 4.9), completers and non-completers are different. Among other things, the former have more online course experience prior to joining TeachUP. Instead of simply comparing the average SRLO of those completing more courses with the average SRLO of those completing fewer courses, the impact was estimated adopting a "difference-in-differences" approach, which also exploited information on SRLO coming from the BS and allowed for a double comparison, i.e. a difference of the differences (Figure 7.4).

In the following figures, what is important is not really the level of the two lines (i.e. the SRLO scores measured before and after the intervention) but whether the two lines remain parallel or diverge as the number of completed courses increases. Only in the second case, we can say that TeachUP courses have had an effect on SRLO. This effect corresponds to the distance between the lines that is the difference between the level of SRLO at a given number of courses completed.

Let's start with the task strategy dimension. Here we found no differences according to the number of courses completed. Turning to the second dimension (goal setting and strategic planning), there is a slight effect among those who completed all the TeachUP courses. As defining goals is specifically asked to TeachUP course participants, it is likely that **those who completed four of these courses have acquired the capacity to setting appropriate goals and formulating appropriate strategic plans**. This result would be in line with prior studies finding a positive association between familiarity and past experience in MOOCs with self-efficacy in MOOCs (Littlejohn et al 2016).

Where we see a clear effect even for those who have completed only one course is the help seeking dimension. Particularly, **the higher the number of courses completed, the lower the perceived need to seek others' help to cope with difficulties**, so the higher the level of autonomy.

Figure 7.4 Variation of SRLO before (BS) and after (FUS) courses according to the number of courses completed
The higher the number of courses completed, the lower the perceived need to seek others' help and the higher teachers' goal and strategic planning competences

As recalled many times throughout this section, the study of SRLO faced some methodological issues, which are summarised in Box 7.3.

Box 7.3 *Methodological issues in the analysis of SRLO*

The strength of the evidence collected in respect to the SRLO dimensions is weakened by some issues:

- Given that the course completion SRLO is self-reported by teachers, this information may be affected, as many psychometric indices are, by measurement error and cultural bias;
- Short-term impacts on course-specific SRLO could be estimated on course completers only;
- Medium-term impacts on general SRLO were greatly affected by the low response rates (particularly among teachers who did not complete any course) in the Follow-up Survey;
- The link between course participation and SRLO could not be studied exploiting the randomisation, as teachers were randomly assigned to personalised support not to courses.

7.4. Impact on other secondary outcomes

In addition to course completion and SRLO, the impact evaluation also considered other secondary outcomes, namely **attitudes towards online learning**, **teaching beliefs and practices** and **digital competences**. For each of them, a set of items were included both in the BS and in the FUS to test whether personalised support induced any change. As explained in Section 3, the full set of items used to measure these characteristics was reduced to some indices via a principal component analysis.[26]

To analyse the impact, we employed the same procedure used with the SRLO: first, by trying to determine whether, at the end of the experiment, the treated teachers had higher or lower levels than the control teachers and, second, by trying to determine whether the teachers who participated in more than one TeachUP course showed higher levels.

Concerning attitudes towards online learning, we did not find any impact of personalised support. Moreover, comparing the indices measured before and after the courses according to the number of TeachUP courses completed we found no evidence of an association between what teachers think about online courses and the TeachUP participation.

We came to the same conclusions with regards to the other two characteristics. We can assume that it is unlikely that personalised support could change attitudes, teaching practices and digital competences in such a short time frame.

26 This procedure allowed the reduction of 15 items of digital competences into 2 components, the 12 items exploring teaching practices into two components and the 5 items of online views into one component. The individual score for each component is an index computed by averaging the factor loadings of corresponding statements.

8. MODES OF LEARNING ASSESSMENT

In the previous pages, we have seen how the TeachUP policy experimentation provided an answer regarding what specific support measures could work to increase success in the delivery of online training for teachers.

Another challenge in scalable online courses is the **assessment of the work of course participants**. As the number of participants increases, the usual assessment via external evaluators becomes more difficult and costly. For this reason, peer assessment provides a popular alternative to expert assessment in online learning. But, **is peer assessment a viable approach to assess learning achievements in scalable online courses**?

This section investigates the reliability of peer assessment by trying to answer to the following three questions:

1. Are assessment scores given by peers in online teacher training contexts reliable, i.e. do they accurately represent the quality of the work and are they consistent among different scorers?
2. Is the feedback provided by peers as part of peer assessment in online teacher training contexts useful and constructive?
3. How do teachers perceive the process of peer assessment in online teacher training contexts?

In particular, subsection 8.1 describes how peer assessment was organised in Teachup. Subsection 8.2 tries to shed more light on the similarities and differences between external evaluators' and peers' assessment while subsection 8.3 aims at understanding more about how differently peers scored the same lesson plan. Finally, subsection 8.4 looks at how both assessments were judged by course participants (i.e., professional and student teachers). For a complete description on the purpose of peer assessment in teacher training and on how peer assessment was organised in Teachup, please see "*TeachUP report: Expert vs. Peer assessment*".

Box 8.1 *Summary of findings*

Results

- Both peers and expert evaluators assessed the quality of teachers' work (i.e., the "lesson plan" that was the final course work) as very high.
- However, peers scores were systematically higher than experts'. This difference is statistically significant but small.
- Peers' feedback on teachers' plans was typically less detailed, less constructive and slightly more positive than feedback provided by experts.
- The assessments provided by peers on the same teacher's plan were generally consistent, even if there was some variability.
- Teachers agreed equally with the assessment received from colleagues as well as the experts; moreover 8 of 10 teachers thought that both assessments were fair and useful.

8.1. Peer and expert assessment in TeachUP courses

As part of the Teach-UP third online course on Collaborative Learning and to qualify for the course badge and certificate, participants had to **submit their own lesson plan for assessment.** The purpose of the lesson plan was to implement and consolidate what was learned during the course in a final product that would allow for easy transfer of learning to the everyday practice in the classroom.

Participants themselves were also put in the role of assessors, as they also had to **provide feedback to three different lesson plans of other course participants**. The peer assessment on the TeachUP courses was not

anonymous in order to encourage course participants to see the process as an open formative exchange between two professionals rather than a final summative evaluation of one's work.

The assessment was based on a which consisted of **8 assessment criteria** linked to specific areas addressed during the course as well as general good practices for a lesson plan: 1) Classroom culture for collaboration; 2) Fostering student agency; 3) Effective elements of collaboration; 4) Assessment of collaborative learning; 5) Tools for collaborative learning; 6) Alignment with learning objectives; 7) Diversity of activities; 8) Balance between individual and group work.

Each criterion was scored 1-4 with 4 representing the highest score (from 1 "the lesson plan requires a lot of work in that area" to 4 "the lesson plan is excellent in that area"). The rubric included a description and an example of what was expected for a certain score. Based on this rubric, the reviewer completed an assessment rubric in spider web format (or, bulls-eye) allowing a clear visual display of the scoring of the 8 criteria from the rubric (see Appendix J for a full description of rubric areas and descriptors). This rubric also included a box for providing more qualitative feedback in text-form.[27]

Among course 3 completers, a **15% sample** in both treated and control groups were randomly selected. **Overall, 106 teachers** (66 in-service and 40 students) **received an assessment also by an external evaluator[28]** with expertise in teachers' new competences and a practical hands-on understanding of how to implement collaborative learning in the classroom. The external expert used the **same rubrics and instructions** as previous course participants did to assess the lesson plans of their peers. Specifically, they completed the same bull's eye diagram using assessment rubrics on the same 8 aspects of the lesson plan. Together with the lesson plans, experts also received the peer assessments of those lesson plans.

This way, **a randomly selected group of participants received** not only three assessments from his/her own peers but **also an external expert assessment,** therefore allowing us to check whether peer assessment is reliable and consistent with expert assessments as well as across peer assessors.

8.2. Similarities and differences between peer and expert assessment

To compare numerical scores, we computed the average score of the eight assessment categories for each single assessment, and then the average of scores that a single lesson plan received from several peers. Then, we compared the overall average scores from all peer and all expert assessments.

As shown in Figure 8.1, lesson plans were considered very good both by peers and by external evaluators. However, **peers, on average, gave higher scores than external evaluators** (3.6 versus 3.2). This difference is not very big, but still statistically significant. Moreover, this was more pronounced for student teachers than professional teachers (0.3 and 0.5 points respectively).

27 Assessors were asked to explain briefly any areas of weakness identified in the lesson plan and provide suggestions on how to improve the lesson plan.

28 The experts were identified and selected by the TeachUP partners in each country according to the common selection criteria. Moreover, they were centrally contracted and paid for their tasks to assess the randomly selected lesson plans. The Slovak external evaluator did not send the assessment of the one Slovak lesson plan and no Maltese expert was found to conduct the assessment.

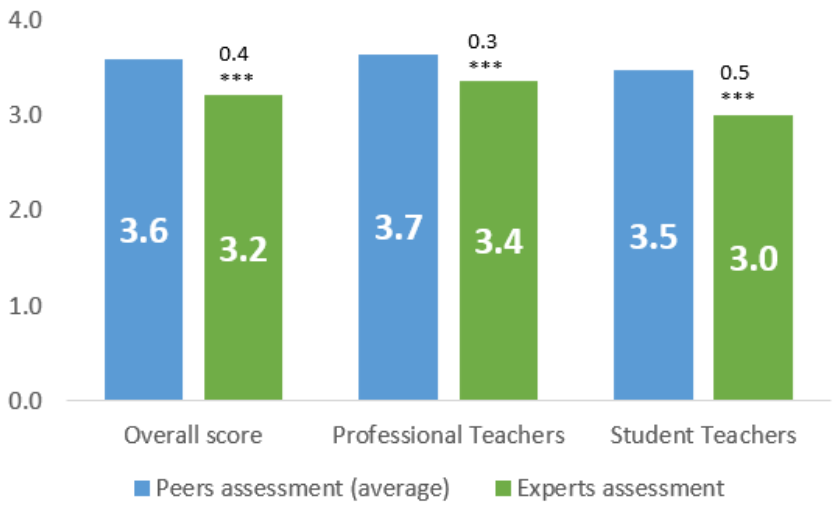


Figure 8.1 *Peers average score and external evaluators score, overall and by teacher type*
Peer scores were higher than those given by external evaluators
Note: The overall score is the mean of 8 categories scores ranging from 1 "the lesson plan requires a lot of work" to 4"the lesson plan is excellent "

When controlling for teacher characteristics (i.e. removing, through a multilevel regression model, the effect of all observed and unobserved individual teachers' characteristics by doing "within-teacher" comparisons of evaluators and peers), the conclusion is the same: expert assessment is lower than peer assessment (mean is 3.27 and 3.53 respectively). The overall finding that peers provided slightly higher scores than experts seems to confirm previous research findings (Kilic & Cakan 2007).

Differences between peers and expert scores were there in all of the eight score components (Figure 8.2). All scores provided by peers were, on average, higher than 3.5 while experts' scores range from 2.9 to 3.4. Since the range of variation of the scores is 0-4, this means that **score variability was slightly larger among experts than among peers.**[29] Differences between peers and experts were more pronounced and statistically significant when the object of evaluation concerned the focus of the assessment, effectiveness of designed activities, the tools (digital and non-digital) used in the lesson, the alignment of the lesson plan with all of its learning objectives and the balance between individual and group work (Figure 8.2).

29 This is shown by the larger range between minimum and maximum score as well as by other possible indices (e.g., in the overall sample, the "coefficient of variation" -the ratio of standard deviation over the mean- is 0.16 for peers and 0.22 for external evaluators).

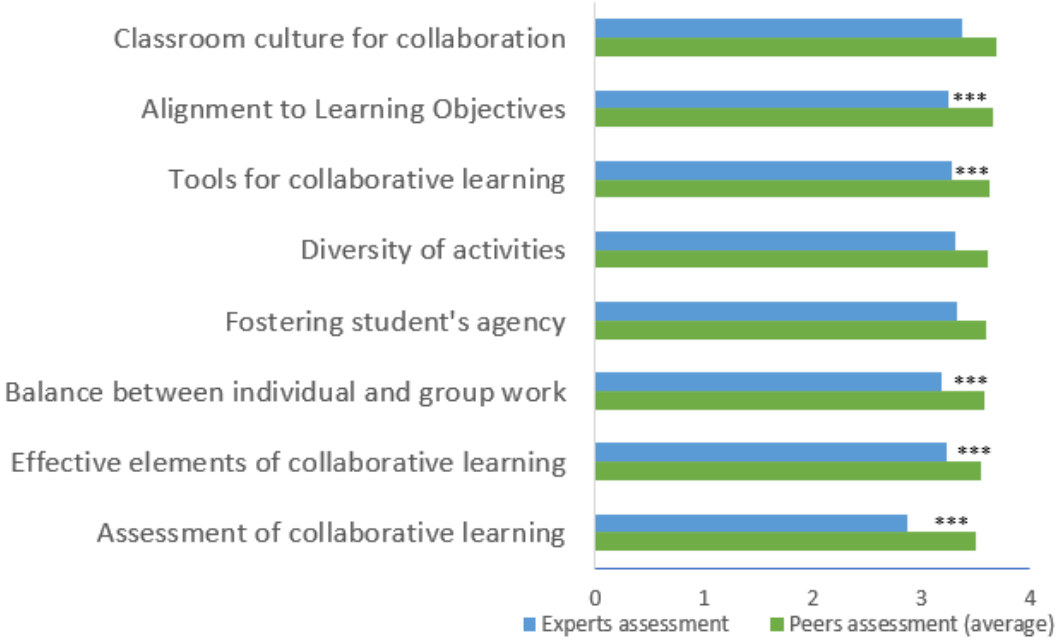


Figure 8.2 *Overall peers average score and expert score, by category*
Peers gave higher scores than experts
Note: stars represent significant differences at 95% confidence level.

It is not possible to provide a conclusive explanation as to why scores by peers were slightly higher. Below five possible explanations are provided, several of which might be at play, based on feedback from course participants via the survey and interviews with experts:

- A first plausible explanation deals with the **feedback culture**: if you are not used to assessing others' work you are probably less comfortable to provide lower scores. And this might have happened especially in those countries where this practice is less common.
- Another possible explanation is related to the **online environment** in the sense that giving someone you do not know feedback in written form online might make you less at ease with giving critical comments.
- A third reason concerns the **understanding of the concepts**: experts might have a better understanding of what is real mastery of collaborative learning, so what peers think is already excellent, based on their experience, is only good for experts.
- It could be that **familiarity in giving and receiving feedback** on the work of colleagues both in general and in particular online may have played a role. Indeed, the assessment categories that describe more complex concepts, such as assessment of collaborative learning, with which teachers were perhaps less familiar, show bigger differences. This might be the case since experts might be more able or at ease to identify weaknesses for such areas.
- Finally, it cannot be excluded that for some course participants **providing feedback with a good score required less explanation** and was therefore likely to be produced quickly. This may have been an additional incentive to provide a higher score. We found some clues in favour of this interpretation by comparing the overall score of peers with the length of their feedback. Indeed, very high scores from peers were more likely to be coupled with very short feedback.[30]

Figure 8.3 shows, for each possible score given by the experts to each category, the distribution of the peers' score. For instance, given an expert score equal to 1, peer average scores ranged from 1, hence exactly like experts' score, to 2.9 (blue rectangle).[31] When experts scored 2 or 3, peers' scores were systematically higher in all areas even if discrepancies decreased (note the areas of the rectangles getting smaller). The expert/peer difference

30 Peer who gave very short feedback had an overall score of 3.8 while those who gave very long feedback scored 3.6. The difference is small and still non-statistically significant also because of the low sample size.

31 It's still worth remembering that scores so low are quite few.

decreases as experts give a score equal to 4 (yellow rectangle) but at the maximum score, it is likely that variability is constrained by a "ceiling effect". To sum up, expert/peer difference decreases as expert scores increase (i.e. the height of the rectangles in the Figure is reduced).

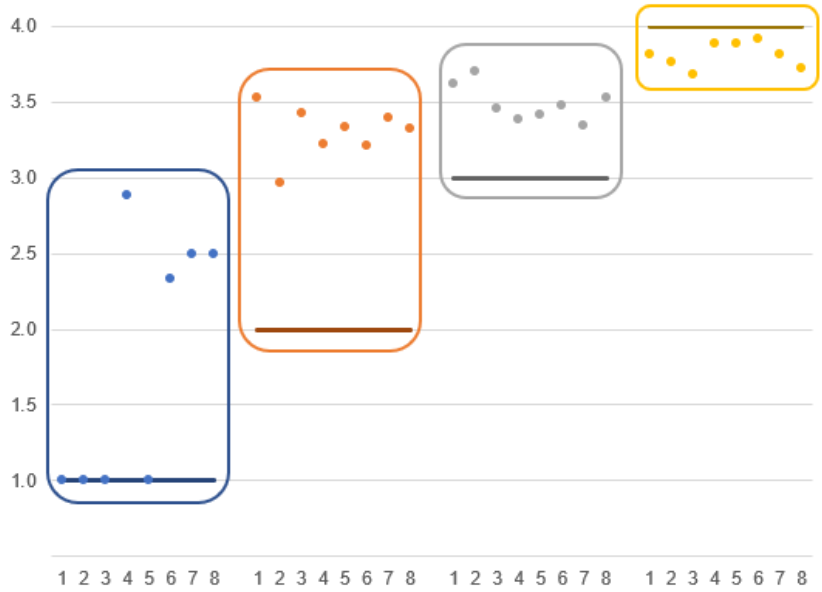


Figure 8.3 *Expert scores (line) and overall peers average score (dots) by the 8 categories*
The expert/peer difference in scores decreases as expert scores increase
Note: categories of assessment are: (1) Effective use of Formative Assessment Techniques; (2) Inbuilt Flexibility; (3) Feedback; (4) Goal Setting; (5) Balance between individual and group work; (6) Tools; (7) Diversity of activities; (8) Alignment to Learning Objectives.

In addition to the eight scores, peers and external evaluators were also asked to give qualitative feedback on teachers' lesson plans. This judgment was then recorded into three indicators: the length of the feedback, the tone of the feedback and the presence of constructive comments. The general assumption was that a feedback that had a certain length, a neutral tone and was constructive (including concrete suggestions for improvement) was likely to be perceived as valid and useful by the person receiving it.

Overall, **peers' and experts' qualitative feedback was quite similar**. **Peers' feedback was overall slightly more positive in tone** (Figure 8.4). However, the difference was not big. This overall slightly more positive tone of qualitative feedback by peers seems to be in line with their slightly higher overall scores.

Expert feedback overall also **included more concrete suggestions** for improvement (Figure 8.4). The biggest difference between peers and expert qualitative feedback was, however, their length. **Experts' qualitative feedback was overall longer** than the one provided by peers (Figure 8.4). There might be several reasons for this difference. Experts might have felt more at ease to provide more detailed feedback, as they overall had a higher level of expertise on the subject matter itself and more experience in assessing others. Also, they might have felt more motivated or obliged to provide more detailed feedback, since they were contracted for the task.

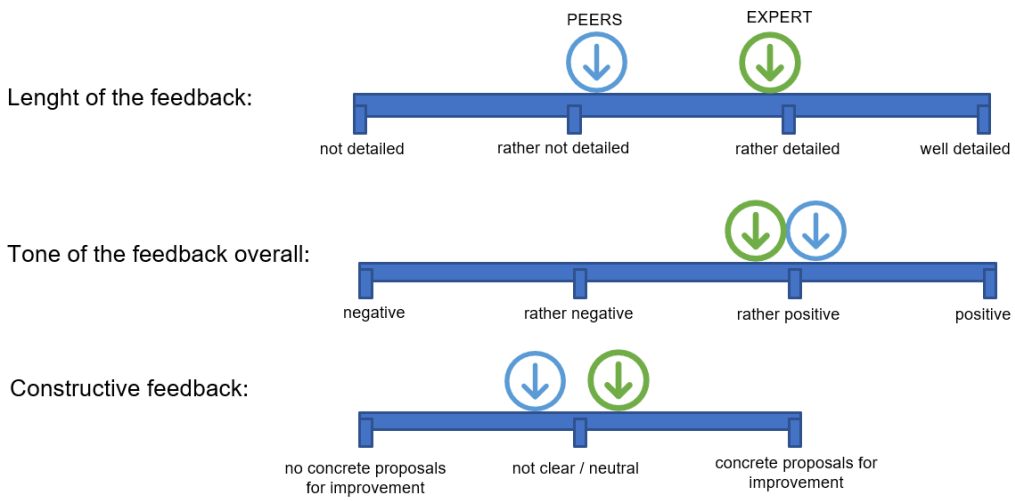


Figure 8.4 *Characteristics of peers' and experts' qualitative feedback on teachers' lesson plans*

8.3. Variability in peer assessment scores

As we have seen, peers gave an average score of 3.6. This average score was calculated in two steps: (a) we first calculated the average of scores that a single teacher plan received from peers; and then (b) we computed the average of (a) across all teachers who received at least one peer score.

However, as said, each teacher received up to 3 scores from peers. So, how much variability in peer scores is there for the same teacher plan (first step above)? In other words, did **different peers score to the same teacher plan in a similar way?**[32]

To answer this question, we decomposed the overall peer score variability in two parts: the differences in mean scores between teachers and the differences between peer scores for the same teacher. The results indicate that the overall peer score variability comes more from teacher plan differences (70%) than for different grading from peers (30%). In other words, **peer assessment varies more between teachers (evaluated)** - which is good - **than between peers (evaluators)**. However, a non-negligible part (30%) of the overall variance in peer assessment score is due to different scoring by peers, suggesting that there is room for interpretation in how a given assignment is evaluated.

8.4. Teachers' opinions on assessment

Another question about assessment that TeachUP wanted to answer is how both assessments are judged by teachers and student teachers. To this aim, a **short survey** (translated in each national language) was sent to the 106 teachers whose lesson plan had been randomly selected to be assessed by an external evaluator.[33] Teachers were asked how they valued the different assessment types and their learning in the online course in general.

In general, teachers expressed a positive opinion of the two evaluations (Figure 8.5): more than eight out of ten confirm that the **assessment was fair and useful**. However, teachers and student teachers perceived the evaluation provided by experts as a bit less fair: 96% of lesson plan authors perceived the peer assessment as fair, compared to only 79% for experts. From the results of this survey, it is not possible to say the reason for this difference

32 Scores do not need to be exactly the same, but they should be within a reasonable bandwidth to claim inter-rater reliability.

33 Out of the 106 selected teachers who completed course 3, 72 answered the survey on the assessment. Therefore, it must be stressed that the results cannot be generalised to the entire TeachUP sample.

but it is plausible that the slightly lower scores and less positive feedback of expert assessment in comparison of peers assessment might have played a role.

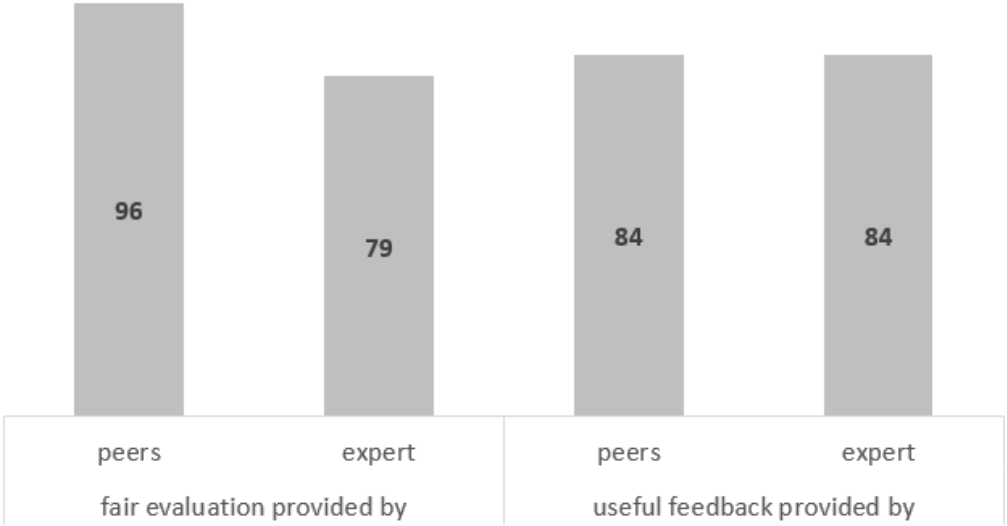


Figure 8.5 Percentage of agreement with statements concerning the assessment by peers and external experts in the course
The largest majority of teachers agreed with the assessment they received

Assessment carried out by peers or external evaluators is judged as quite consistent with the self-assessment. On a scale from 1 "totally disagree" to 6 "totally agree" the average score given by teachers was 4.9 for the assessment made by experts and 4.9 for the assessment made by peers. It is therefore not surprising that the majority of lesson plan authors did **not** express **a clear preference** for which assessment form they would find more useful for their learning in future courses (77%).[34]

However, when asked about the usefulness of the individual assessment activities for their learning, 72% found expert assessment very useful and 65% found very useful assessing the lesson plan of their peers, while only 51% found the assessment of their peers very useful (Figure 8.6). The process of assessing a peer is found to be more useful than the process of being assessed by a peer. And that the overall usefulness of peer assessment is made up of both elements.

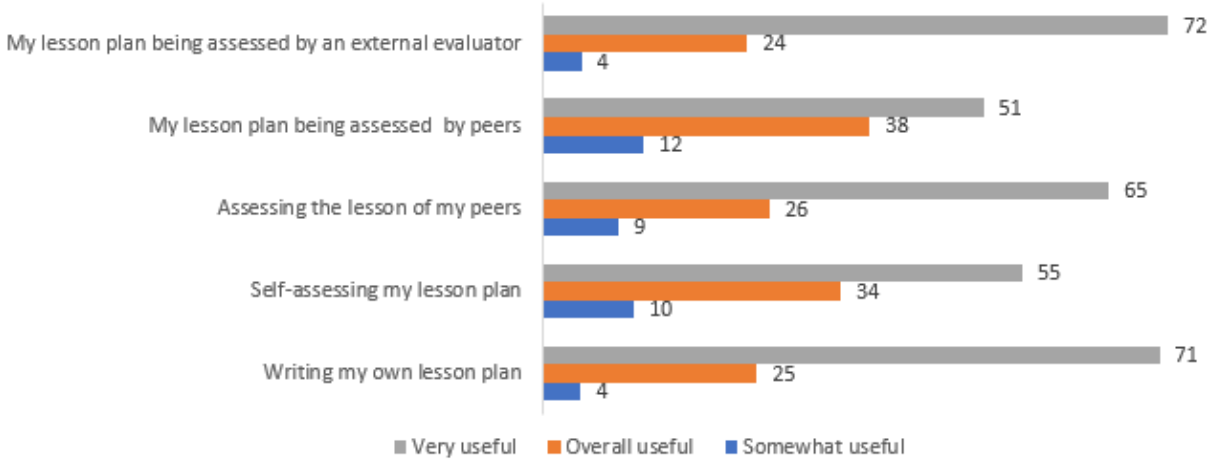


Figure 8.6 Percentage of teachers finding the received overall assessment as useful for their learning
The majority of the individual assessment activities were perceived by all lesson plan authors as overall useful or very useful

34 12% found expert assessment more useful; and 11% found peer assessment more useful.

9. MAIN LESSONS AND POLICY IMPLICATIONS

9.1. What we have learned

Summary of findings

The TeachUP online personalised support proved to be an effective approach to increase course participation among **professional teachers in EU Member States** (Figure 9.1).

Teachers receiving the TeachUP personalised support showed **10 percentage points higher probability of course completion** than control teachers, who received no form of support. Beyond being highly statistically significant, this impact is also substantial as it raised completion rates from 32 to 42%.

The evaluation also provided evidence of an effect of course participation on **teachers' SRLO** as measured via a survey in the months after the last course. The more courses a teacher completed within the TeachUP policy experimentation, the higher his/her ability in setting her learning goals and the lower her propensity to seek others' help to solve problems encountered during the online course. Hence, by increasing course participation, **personalised support could also have indirect effects on teachers' SRLO**, helping them become more independent learners.

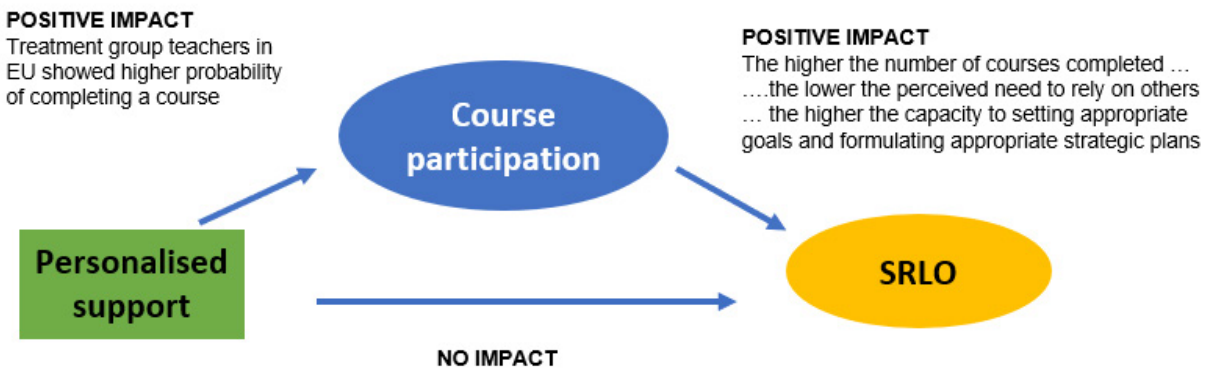


Figure 9.1 Summary of the TeachUP evaluation's findings

While, as just illustrated, the tested intervention proved to be effective for professional teachers in EU Member states, the same conclusion was not reached for the other teacher groups studied in TeachUP. Among **student teachers in the EU**, personalised support had an impact on completion rates **only among those students who joined TeachUP with some past experience in online learning** (i.e., teachers who completed at least 1 course per year in the past three years). Also, personalised support intervention had **no impact on professional and student teachers in Turkey**, who saw no improvement in course completion rates thanks to the intervention.

How should we interpret these results? Why did the TeachUP personalised work for some teachers and why did it not for others?

Understanding why it worked for some

The TeachUP personalised support was delivered through a set of personalised messages targeted to teachers, who were identified as 'in need' of support to either actually start or successfully complete the tasks requested by the courses. Neat experimental evidence for professional teachers in EU countries shows that **the overall impact on course completion** was a result of both **support interventions making people complete** the courses they started and **interventions making teachers actually start the courses** they enrolled in.

But **through which specific mechanisms did personalised support induce teachers to start or complete the courses at higher rates**? Because offers of concrete help (i.e., the 1:1 online sessions) were taken up by a very small fraction of teachers, others had to be the mechanisms in place. **What made a difference was not just the concrete support provided, but the message itself**. Different groups of teachers could perceive and react to the provided stimulus in different ways (Shulman and Shulman 2004). The messages triggered teachers' engagement with the course because of a mix of psychological reactions to those messages. Even though the data do not allow testing all possible mechanisms, listing them is a worthwhile effort to better understand what the potential channels linking personalised support and participation in online courses are.

Among these mechanisms, one could hypothesize: a) prompting; b) external observation; c) guidance. A prompting effect can be hypothesized if the personalized messages were perceived as encouragement to start/complete a course. The fact of receiving a message could be perceived by itself as a **prompt**, in some cases even a simple reminder, to keep up with the course. Furthermore, personalised messages could have altered teachers' behaviours - i.e., induced them to start or complete the course - by making them feel **observed**. In other words, the simple fact of knowing that someone was watching over them could have induced teachers to behave differently than they would have otherwise. Finally, teachers may have benefited from the **guidance** and the suggestions included in the messages (e.g., to set aside specific times in their calendar for following the course, to consult certain resources to help with the work required, or where and when to submit their work). This guidance and suggestions were tailored on teachers' needs, which were explicitly stated in those messages (i.e., difficulties faced by teachers and the "estimated" likelihood for them to fail). This could also have triggered a positive reaction in teachers in terms of additional effort put on the course.

As said, our data make it impossible to disentangle the possible mechanisms, but for sure it can be concluded that **what made a difference was receiving one (or more) personalised messages containing a concrete offer of support and not the concrete support in itself**. At the same time, **we cannot dismissively conclude that simple e-mail messages would have led to the same results**. The messages sent in TeachUP were highly **personalised and contained a concrete offer of support**. It is possible that this promise was precisely what made teachers engage more with the courses.

...and why it did not for others

Why personalised support did not work equally well for the other three groups (i.e., student teachers in the EU and student and professional teachers in Turkey) is another crucial point of discussion. A number of factors could be responsible for this result, but the evaluation does not allow reaching any conclusive answers.

A first set of explanations may have to do with the **different characteristics and needs** of teachers.

When comparing **professional and student teachers in the EU**, the latter showed systematically lower natural completion rates (22% vs 32%). This would suggest that there was ample room for personalised support to have an impact. However, this did not happen, or it happened only among those students who had some past experience in online courses, i.e. students with higher readiness to engage in online training. These teachers most likely did not need any specific guidance or concrete support but benefited only from the psychological effect of prompting messages. Qualitative evidence collected in the process (i.e. during the Country Dialogue Labs) also

suggests that the "observation" mechanism could be weak in this population because students are less sensitive to the 'social pressure/monitoring' aspect, and that they might have perceived the possible relationship with the agent as quite hierarchical, and therefore did not want it. More search is needed to investigate this hypothesis further.

When comparing **professional teachers in EU Member States and in Turkey**, many contextual and individual differences become apparent. On average, teachers in Turkey are younger, less frequently female, have lower English proficiency, lower education levels, lower past experience in online courses, overall lower quality internet access and less positive views about online courses. Overall, such a profile should predict lower completion rates and higher need of personalised support, hence a higher room for the treatment to have an impact. But this did not happen.

A possible reason why this did not happen can be traced back to some differences in the **patterns of course participation** in TeachUP. When compared to professional teachers in the EU, Turkish professional teachers showed overall lower natural start rates and higher natural completion rates among those who started. This could have reduced the room for personalised support to have an impact on completion (it is difficult to raise completion rates when they are already very high). However, this would not explain why personalised support had no effect on start rates and neither would it explain why no effect was found for student teachers in Turkey, who did not show the same course participation patterns as Turkish professional teachers.

A possible factor explaining the null effect of personalised support in Turkey may then be related to **how the courses were delivered**. Notably, the course moderator was particularly active and experienced and might have made additional support less relevant by establishing a vibrant and supportive course community where participants supported each other. This could have been facilitated by the higher **number of participants** in Turkey, which was between 4 and 9 times larger, depending on the course, than in European Member States. Furthermore, reports from personalised support agents in Turkey suggest that some **contamination** between treated and control groups occurred with information from the agents and details about their role passed to the control group.

9.2. Assessing the scalability of the TeachUP support model

The TeachUP personalised support proved to be effective for professional teachers in the EU, but the scalability of this "model" needs to be closely assessed. There are at least four crucial questions that a policy maker wanting to adopt this model may ask:

1. Can the results be generalized to the entire teacher population or are they valid only for teachers sampled for this particular study?
2. Will the results apply to my country as well?
3. Will the results apply also to other types of online courses?
4. How much would it cost to increase course completion rates?

Can the results be generalized to all teachers?

Do the results apply to the entire teacher population? Would the tested support work to the same extent if it was extended to the entire teacher population?

TeachUP teachers were not randomly picked from the population of teachers in the different countries. Within a representative sample of schools (and ITE organisations), teachers were free to either accept or decline the invite to take part in the project. Hence, while the geographical representativeness was ensured by design, it was still possible that teachers who decided to take up the invitation to participate were different from those who declined the invitation.

The extent to which the TeachUP teacher differed from the average teacher is crucial for the 'external validity' of the TeachUP experiment, i.e. to the extent to which the results can be "transported" to the entirety of teachers.

Thanks to the availability of external microdata from OECD's *Teaching and Learning International Survey* (*TALIS*) 2018 (2020), we could compare the baseline characteristics of TeachUP professional teachers in the EU with the characteristics of representative samples of professional teachers in the same countries. Unfortunately, a similar comparison could not be performed for student teachers as such data do not exist. The results of the TeachUP and TALIS comparison suggest that, **based on the observable characteristics included in both datasets, TeachUP teachers are similar to the entire population of teachers in Europe**.[35] TeachUP teachers show the same age profiles, the same educational attainment and similar professional experience and weekly workload, but they were more likely to be women than in the overall teacher population as estimated with TALIS data. Differences about participation in CPD activities were found only for Turkey: TeachUP teachers have less experience than in TALIS. Additional details are available in Appendix I.

Are results expected to be the same in all countries?

The entire TeachUP policy experimentation was designed to provide overall impact estimates of the personalised support and not country-specific ones. To estimate county-specific impacts, a much larger sample of teachers would have been needed. Such a condition was only possibly for Turkey, thanks to the large sample size obtained for this country.

However, some validation tests (based on the iterative exclusion of one of the participating countries, see appendix G) show that results are not too sensitive to the exclusion of countries from the sample (estimates are always statistically significant and ranging from 6.8 to 13.8 percentage points). Hence, while we cannot conclude that personalised support worked in all countries, **we can exclude that the overall positive impact was driven by some particular countries**.

Would personalised support work also in other types of online courses?

Online courses can vary substantially regarding the instructional design used, size, length, platform used, etc. It is unlikely that the dynamics created by the personalised support mechanism and which most likely resulted in the observed impact would be substantially different on a course with a different length or platform used. For example, the dynamics of feeling observed, being prompted, or receiving useful guidance are not particularly dependent on the length of a course.

It is however possible that the size and/or instructional design of a course can affect these dynamics in different ways. Small online courses with a strong instructor presence who can afford to provide 1:1 support in regular intervals would most likely make the personalised support mechanism less effective, given that the instructor already creates dynamics of feeling observed, being prompted, and receiving useful guidance. Larger online courses where instructor presence is less possible but which succeed in creating a very active and supportive course community might also reduce the impact of the personalised support mechanism as the community can more effectively cater to the needs of those targeted through the personalised support mechanism. Such community might also reduce the threshold for participants in need to become active and engaged in the course.

Accordingly, the personalised support mechanism is likely to be most effective in scenarios where instructor presence is not available and at the same time there is not (yet) an active and supportive course community in place.

35 Greece did not participate in TALIS 2018; hence it was not included in this comparison.

How much would it cost?

When resources are limited, it is important to learn about an intervention's **cost-effectiveness**, and not only its effectiveness, before deciding whether to scale it up or not. The concept of cost-effectiveness draws the attention to the costs that were sustained to achieve a given impact. In the case of TeachUP, personalised support was found effective for professional teachers in EU Member States. But how much did it cost to induce one teacher more to complete a course? And how generalizable to other contexts would this cost-effectiveness estimate be?

The calculation of the TeachUP support's cost-effectiveness is obtained, as illustrated in Box 9.1, and led to estimate the cost of one unit of impact at 131 euros. This means that, **on average and across the 9 EU countries, 131 euros were spent to induce one additional teacher to complete a course thanks to the intervention**.

To assess whether it is high or low, this value should be compared with estimates coming from alternative programs, which - to our knowledge - unfortunately are not available. Moreover, it is important to recognise that this estimation is based on the entire approach used in the TeachUP experiment and not just those interventions and processes that are responsible for the impact observed. The estimation also incorporates time spent on actions that are unlikely to be implemented in a non-experimental context, such as a full day workshop (including preparation and follow-up) for support agents introducing them to the experiment and approach.

Moreover, it should also be acknowledged that costs varied widely across countries and the figure presented is an average estimate. Differences in costs between countries did not only stem from average wage costs of support agents but were also based on the number of participants in the courses and the time support agents took to implement the interventions. For example, in Austria very few participants joined and completed the courses which meant that the fixed costs of the mechanism (training, recruitment, planning, etc.) was part of the reason why Austria had the highest costs per participant (€186.6), while in countries with higher numbers like Greece costs per participant were much lower (€33.9), and in some other countries costs were even lower, i.e. in Lithuania (€11.03). Hence, considering that TeachUP courses were characterized by relatively low numbers of participants (ranging between 19 and 55 in the European countries) and assuming that personalised support would work to the same extent also in the presence of more participants, the cost-effectiveness of the tested personalised support could improve in courses with higher numbers of teachers.

Box 9.1 *Estimating cost-effectiveness of personalised support for professional teachers in the EU*

Once it is established that an intervention has had an impact, its cost-effectiveness – defined as the cost needed to produce a one-unit impact – depends on three factors; a) the so-called 'deadweight' that, in our case, is captured by the natural course completion rate (i.e. the completion rate observed among the controls) and expresses the percentage of teachers who would have completed a course even in the absence of personalised support. Hence – if the personalised support's impact size and costs are held fixed – the higher the deadweight, the lower the personalised support's cost-effectiveness; b) the size of the impact; (c) per-person program costs.

The first cost-effectiveness indicator is the simple ratio between the deadweight and the impact. The value of this index expresses the number of supported teachers that would have completed the course event in the absence of personalised support for every teacher that completed a course thanks to the personalised support. It is a measure of "waste" or inefficiency: the lower this number, the better the cost-effectiveness. Overall, as shown in table 5.1, about three teachers (3.23) needed to be supported in vain to make one additional teacher complete a course.

The second cost-effectiveness indicator adds personalised support costs into the calculation and is obtained by simply multiplying the estimated per-person cost by the previous cost-effectiveness indicator. Based on a weighted average estimate on the cost of personalised support for each enrolled teacher (approximately 29 euros; cost estimates for course 4, see Appendix H for details), it is estimated that cost to produce one unit of impact (i.e., one additional teacher completing a course) amounted to €130.8 (=30.9+99.9 euros). However, it should also be acknowledged that, as shown in appendix H, costs varied widely across countries and the one presented is an average estimate, which is subject to variability.

Table 9.1 *Personalised support's cost-effectiveness for EU PTs*

Controls (deadweight) (a)	0.32
Impact (b)	0.10
Cost-effectiveness I (c=a/b)	3.23
Estimated per-person cost (d)	30.9 euros
Cost-effectiveness II (e=c*d)	99.9 euros

9.3. Implications for ITE and CPD

ITE and CPD providers are key to address the increasingly complex challenges teachers face in their classrooms. ITE providers need to adequately prepare new teachers with the skills and competences that will allow them to continually develop their practice so that they have the ability to positively and proactively tackle the challenges they face during their teaching careers. CPD providers need to offer concrete and timely support to those dealing with the complexities of classroom practice on a daily basis usually entirely by themselves.

Effective face-to-face training will and should remain the bedrock of how ITE and CPD providers support initial and in-service teachers. The social nature of learning and the kind of experiences and bonds between professionals that

can be developed in face-to-face settings, especially at school-level, are irreplaceable (OECD, 2010). However, offering only face-to-face training formats is not sufficient, as indicated by the barriers teachers report to accessing such formats, and ITE and CPD providers need to also consider more versatile and flexible formats that are built more around the needs of teachers.

CPD Providers

The TeachUP results have shown that scalable online courses like MOOCs have the potential to engage and effectively support large numbers of teachers with their professional development. CPD providers should therefore consider developing and running their own MOOCs or consider how to reuse existing MOOCs provided at national and international level.

CPD providers are also well placed to offer personalised support alongside a MOOC, either by means of the mechanism tested in TeachUP or through a face-to-face personalised support offer that runs alongside and in tandem with a MOOC. While the TeachUP experimentation did not investigate such a scenario, it is likely that a blended approach would benefit even more those specifically targeted in TeachUP, i.e. those with low levels of online learning experience, self-regulated learning competence, and digital competence.

Considering the substantial resources required for the production of a new MOOC, it might therefore be efficient and effective for CPD providers to consider offering a personalised support mechanism alongside existing MOOCs – provided those MOOCs exist on competences/topics at stake - rather than developing new MOOCs from scratch themselves. This would allow CPD providers to focus more resources on the individual support they can offer to those teachers who are most in need of support, rather than spending resources on developing a course themselves.

When actively engaged in autonomous MOOC production, CPD providers should engage or at the very least establish close collaborations with actors involved in MOOC production, so as to ensure the relevance and adequacy of MOOC instructional designs. Either way, using MOOCs would require the development of new capacities of teacher trainers so as to ensure that instructional design, content production, facilitation and support on MOOCs follows high-quality and evidence-based models.

ITE Providers

The TeachUP results are less clear in regard to what they can tell us about the suitability of scalable online courses for student teachers. Accordingly, ITE providers will need to investigate further to understand what the conditions are for MOOCs to play a role in initial teacher education. First and foremost, understanding why student teachers seemed to be less engaged by the type of MOOCs offered in TeachUP could offer valuable insights for ITE providers in how to shape any potential online offer for student teachers. For example, investigating questions such as: would student teachers engage more in courses targeted only at them, how integrated and aligned with their usual programme of study do MOOCs have to be, and what role does the timing of running a MOOC in the annual calendar play for student teachers, would provide ITE providers with a clearer basis from which to develop their own MOOCs or integrate existing MOOCs into their study programmes.

Furthermore, ITE providers would benefit from investigating further why a personalised support offer along the lines of that implemented in TeachUP had limited impact, again allowing them to shape any online offer for student teachers more effectively. In this regard, investigating alternative approaches to personalised support, potentially more embedded in existing ITE structures such as a school placement or induction phase, could be of interest. For example, could student or newly-qualified teachers use MOOCs from national or international providers during their school placement or induction phase with personalised support offered by school or ITE-based mentors? Taking a MOOC together with their mentor could not only support student teachers and newly qualified teachers in benefitting from the MOOC itself, it could also offer important experience of and a framework for ongoing discussions about their professional practice with colleagues, thereby familiarising them to the idea of becoming a reflective teaching professional who is responsible for their own professional development, and equipped with the skills to do so.

Apart from such further investigations, the fact that the TeachUP results highlight potential for MOOCs as a tool for teacher professional development, means that ITE providers should play a role in accustoming student teachers to the idea of engaging in self-directed online professional development courses and communities, thereby making them more likely and capable to benefit from MOOCs as they enter the profession.

9.4. Further steps for research

Within the TeachUP policy experimentation, a number of challenges were encountered and not all of them could be answered.

Regarding personalised support mechanisms to improve retention in online courses, further investigation in the future would particularly need to explore the potential of human-chatbots integration and the use of predictive analytics to improve personalisation. Also, future studies should attempt to shed light on the mechanisms that lay behind an effective personalised support as well as investigate the effectiveness of different personalised support mechanisms depending on the contextual conditions and user background characteristics.

In regard to SRLO, such competence is recognised as very important for ensuring success in online courses, however its measurement is still difficult and more research is needed (Rovers et al 2019).

Regarding assessment, more research is needed to further strengthen the purpose of peer assessment being a learning experience for course participants and, in particular, to further improve peer assessment mechanisms, in view of possible accreditation of online courses. It would also be of interest to compare peer and expert feedback in terms of their effectiveness for short-and longer-term learning gains.

REFERENCES

Aleven, V., Mclaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. International Journal of Artificial Intelligence in Education, 16(2), 101-128.

Bawa, Papia. (2016). "Retention in Online Courses: Exploring Issues and Solutions--A Literature Review." *SAGE Open* 6(1).

Bettinger, Eric, and Rachel Baker. (2011). "The Effects of Student Coaching in College: An Evaluation of a Randomized Experiment in Student Mentoring. NBER Working Paper No. 16881." *National Bureau of Economic Research*.

Bettinger, Eric, Jing Liu, and Susanna Loeb. (2016). "Connections Matter: How Interactive Peers Affect Students in Online College Courses." *Journal of Policy Analysis and Management* 35(4):932–954.

Boud, David. (2013). *Enhancing Learning Through Self-Assessment*. Routledge.

Briggs, Amanda, and Shayne Spaulding. (2018). "Three Ways Technology Can Help Nontraditional Students Succeed in Online Coursework." *Urban Institute*. Retrieved April 23, 2020 (https://www.urban.org/urban-wire/three-ways-technology-can-help-nontraditional-students-succeed-online-coursework).

Boekaerts, M. (1999). Self-regulated learning: Where we are today. International journal of educational research, 31(6), 445-457.

Carr, Sarah. (2000). "As Distance Education Comes of Age, the Challenge Is Keeping the Students." *Chronicle of Higher Education* 46(23).

Castaño-Muñoz, J., Kalz, M., Kreijns, K., & Punie, Y. (2018). Who is taking MOOCs for teachers' professional development on the use of ICT? A cross-sectional study from Spain. Technology, Pedagogy and Education, 27(5), 607–624. https://doi.org/10.1080/1475939X.2018.1528997

Castelo, N., Hardy, E., House, J., Mazar, N., Tsai, C., & Zhao, M. (2015). Moving citizens online: Using salience & message framing to motivate behavior change. Behavioral Science & Policy, 1(2), 57–68. https://doi.org/10.1353/bsp.2015.0018

Castles, J. (2004). "Persistence and the Adult Learner: Factors Affecting Persistence in Open University Students." *Active Learning in Higher Education* 5.

Cheung, L. L. W., and A. C. N. Kan. (2002). "Evaluation of Factors Related to Student Performance in a Distance-Learning Business Communication Course." *Journal of Education for Business* 77.

Clay, M. N., S. Rowland, and A. Packard. (2009). "Improving Undergraduate Online Retention through Gated Advisement and Redundant Communication." *Journal of College Student Retention: Research, Theory and Practice* 10.

Dupin-Bryant, P. (2004). "Pre-Entry Variables Related to Retention in Online Distance Education." *American Journal of Distance Education* 18.

Effeney, Gerard, Annemaree Carroll, and Nan Bahr. (2013). "Self-Regulated Learning: Key Strategies and Their Sources in a Sample of Adolescent Males." *Australian Journal of Educational & Developmental Psychology* 13.

Emanuel, Ezekiel J. (2013). "Online Education: MOOCs Taken by Educated Few." *Nature* 503(7476):342.

European Commission. Education, Audiovisual, and Culture Executive Agency. (2015). "The Teaching Profession in Europe: Practices, Perceptions, and Policies." *Eurydice Report*.

European Schoolnet, and University of Liege, Psychology and Education. (2013). "Survey of Schools: ICT in Education. Benchmarking Access, Use And Attitudes To Technology In Europe's Schools." Retrieved May 4, 2020 (http://essie.eun.org/homepage).

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. Review of educational research, 70(3), 287-322.

Fike, David S., and Renea Fike. (2008). "Predictors of First-Year Student Retention in the Community College:" *Community College Review*.

Fixsen, D. L., , Blase, K. A., Horner, R., and Sugai, G. (2009). "Scaling-up Brief 3: Readiness for Change | FPG Child Development Institute." Retrieved May 18, 2020 (https://fpg.unc.edu/node/6556).

Fyle, C. (2013). Teacher Education MOOCs for Developing World Contexts : Issues and Design Considerations. Sixth International Conference of MIT's Learning International Networks Consortium (LINC). Retrieved from http://linc.mit.edu/linc2013/proceedings/Session3/Session3Fyle.pdf

Goel, A. K., & Polepeddi, L. (2019). Jill Watson: A Virtual Teaching Assistant for Online Education. In C. Dede, J. Richards, & B. Saxberg (Eds.), Learning Engineering for Online Education: Theoretical Contexts and Design-Based Examples. New York: Routledge.

Ho, A.D.; Chuang, I.; Coleman, C.; Whitehill, J.; Northcutt, C.; Williams, J. J.; Hansen, J.; Lopez, G.; Peterson, R. (2015). HarvardX and MITx: Two years of massive open online courses. (10), 1–37. https://doi.org/10.2139/ssrn.2586847

Jordan, Katy. (2014). "Initial Trends in Enrolment and Completion of Massive Open Online Courses." *International Review of Research in Open and Distributed Learning* 15(1):133–160.

Jordan, Katy. (2015). "Massive Open Online Course Completion Rates Revisited: Assessment, Length and Attrition." *International Review of Research in Open and Distributed Learning* 16(3):341–358.

Kai, Shimin, Juan Miguel L. Andres, Luc Paquette, Ryan S. Baker, Kati Molnar, Harriet Watkins, and Michael Moore. (2017). *Predicting Student Retention from Behavior in an Online Orientation Course*. International Educational Data Mining Society.

Kember, D. (1995). *Open Learning Courses for Adults: A Model of Student Progress*. Englewood Cliffs, NJ: Educational Technology Publications.

Kilic, Gulsen Bagci, and Mehtap Cakan. (2007). "Peer Assessment of Elementary Science Teaching Skills." *Journal of Science Teacher Education* 18(1):91–107.

Kizilcec, René F., Mar Pérez-Sanagustín, and Jorge J. Maldonado. (2017). "Self-Regulated Learning Strategies Predict Learner Behavior and Goal Attainment in Massive Open Online Courses." *Computers & Education* 104:18–33.

Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., ... & Tingley, D. (2020). Scaling up behavioral science interventions in online education. *Proceedings of the National Academy of Sciences*.

Koutsodimou, K., & Jimoyiannis, A. (2015). MOOCs for teacher professional development: investigating views and perceptions of the participants. MOOCS FOR TEACHER PROFESSIONAL DEVELOPMENT: INVESTIGATING VIEWS AND PERCEPTIONS OF THE Konstantina Koutsodimou , Athanassios Jimoyiannis. (November).

Laurillard, D. (2016). The educational problem that MOOCs could solve: Professional development for teachers of disadvantaged students. Research in Learning Technology, 24(1), 29369. https://doi.org/10.3402/rlt.v24.29369

Lee, Daeyeoul, Sunnie Lee Watson, and William R. Watson. (2019). "Systematic Literature Review on Self-Regulated Learning in Massive Open Online Courses." *Australasian Journal of Educational Technology* 35(1).

Lee, Ji-Eun, and Mimi Recker. (2017). "Measuring Students' Use of Self-Regulated Learning Strategies from Learning Management System Data: An Evidence-Centered Design Approach." *National Science Foundation Project No. SMA1338487*.

Lee, Youngju, and Jaeho Choi. (2011). "A Review of Online Course Dropout Research: Implications for Practice and Future Research." *Educational Technology Research and Development* 59(5):593–618.

Levy, Yair. (2007). "Comparing Dropouts and Persistence in E-Learning Courses." *Computers & Education* 48(2):185–204.

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. The Internet and Higher Education, 29, 40-48.

Lynch, R., & Dembo, M. (2004). The relationship between self-regulation and online learning in a blended learning context. The International Review of Research in Open and Distributed Learning, 5(2).

Milligan, Colin; Littlejohn, Allison (2015): MOOC Design Recommendations. Figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.1420557.v1

Misra, P. K. (2018). MOOCs for Teacher Professional Development: Reflections and Suggested Actions. Open Praxis, 10(1), 67. https://doi.org/10.5944/openpraxis.10.1.780

Molenaar, I., Horvers, A., Baker, R.S. (2019). What can moment-by-moment learning curves tell about students' self-regulated learning? *Learning and Instruction.*

Muirhead, B. n.d. "Encouraging Interaction in Online Classes." *International Journal of Instructional Technology and Distance Learning* 1(6).

Niemi, H., Nevgi, A., & Virtanen, P. (2003). Towards self-regulation in web-based learning. Journal of Educational Media, 28(1), 49-71.

OECD. (2009). Creating Effective Teaching and Learning Environments: First Results from TALIS. Paris: OECD.

OECD. (2014). TALIS 2013 results: an international perspective on teaching and learning. In TALIS. https://doi.org/10.1787/9789264196261-en

OECD (2019), TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners, TALIS, OECD Publishing, Paris. https://doi.org/10.1787/1d0bc92a-en

OECD (2020) TALIS 2018 Results (Volume II): Teachers and School Leaders as Valued Professionals, TALIS, OECD Publishing, Paris. https://doi.org/10.1787/19cf08df-en

Osborn, V. (2001). "Identifying At-Risk Students in Videoconferencing and Web-Based Distance Education." *American Journal of Distance Education* 15.

Packham, G., P.Jones, C. Miller, and B. Thomas. (2004). "E-Learning and Retention: Key Factors Influencing Student Withdrawal." *Education Training* 46.

Panadero, Ernesto. (2017). "A Review of Self-Regulated Learning: Six Models and Four Directions for Research." *Frontiers in Psychology* 8.

Panadero, E., Klug, J. & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand, Scandinavian Journal of Educational Research, 60:6, 723-735, DOI: 10.1080/00313831.2015.1066436

Parsons, S. A., Hutchison, A. C., Hall, L. A., Parsons, A. W., Ives, S. T., & Leggett, A. B. (2019a). U.S. teachers' perceptions of online professional development. Teaching and Teacher Education, 82, 33–42. https://doi.org/10.1016/j.tate.2019.03.006

Pierrakeas, C., M. Xenos, C. Panagiotakopoulos, and D. Vergidis. (2004). "A Comparative Study of Dropout Rates and Causes for Two Different Distance Education Courses." *International Review of Research in Open and Distance Learning* 5.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. International Journal of Educational Research, 31(6), 459-470.

Poellhuber, B., M. Chomienne, and T. Karsenti. (2008). "The Effect of Peer Collaboration and Collaborative Learning on Self-Efficacy and Persistence in a Learner-Paced Continuous Intake Model." *Journal of Distance Education* 22.

Powell, C. G., & Bodur, Y. (2019). Teachers' perceptions of an online professional development experience: Implications for a design and implementation framework. Teaching and Teacher Education. https://doi.org/10.1016/j.tate.2018.09.004Pérez-Álvarez, R., Maldonado-Mahauad, J., & Pérez-Sanagustín, M. 2018. Design of a Tool to Support Self-Regulated Learning Strategies in MOOCs. J. UCS, 24(8), 1090--1109

Prestridge, Sarah, and Jo Tondeur. (2015). "Exploring Elements That Support Teachers Engagement in Online Professional Development." *Education Sciences* 5(3):199–219.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and metaanalysis, Psychological Bulletin, 138(2), 353.

Rovers, S.F.E., Clarebout, G., Savelberg, H.H.C.M. et al. (2019) Granularity matters: comparing different ways of measuring self-regulated learning. Metacognition Learning 14, 1–19. https://doi.org/10.1007/s11409-019-09188-6

Sadler, P.M., & Good, E. (2006). The Impact of Self- and Peer-Grading on Student Learning. Educational Assessment, 11, 1 - 31.

Shulman, S and Shulman, J (2004) How and What Teachers Learn: a shifting perspective in Journal of Curriculum Studies Vol. 36, No. 2, pp 257-271

Scott, K. (2016). Usability Heuristics for Bots. Retrieved May 19, 2020, from Chatbots Magazine website: https://chatbotsmagazine.com/usability-heuristics-for-bots-7075132d2c92

Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. Educational Psychologist, 40(2), 85-94

Tello, S. F. 2007. "An Analysis of Student Persistence in Online Education." *International Journal of Information and Communication Technology Education* 3.

Triquet, K., Peeters, J., & Lombaerts, K. (2017) Self-Regulated Learning Online: Benefits, Empirical Foundations, Multi-level and Multi-modal Promotion, and the Evaluation thereof for Teacher Professional Development, Contributing SRL Part of TeachUP Deliverable D1.1

Tsai, C. W., Shen, P. D., & Fan, Y. T. (2013). Research trends in self-regulated learning research in online learning environments: A review of studies published in selected journals from 2003 to 2012. British Journal of Educational Technology, 44(5), E107-E110.

Van Laer, S., & Elen, J. (2016). Learners' Self-Regulatory Behaviour Profiles in Blended Learning Environments. In Proceedings of the 7th Biennial Meeting of the EARLI Special Interest Group 16 Metacognition (pp. 117-118). Radboud University Nijmegen.

Viberg, O., Khalil, M., & Baars, M. (2020, March). Self-regulated learning and learning analytics in online learning environments: a review of empirical research. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (pp. 524-533).

Yuan, L., & Powell, S. (2013). MOOCs and open education: Implications for higher education (JISC CETIS white paper). Retrieved from http://publications.cetis.ac.uk/2013/667

Wastiau, Patricia, Roger Blamire, Caroline Kearney, Valerie Quittre, Eva Van de Gaer, and Christian Monseur. (2013). "The Use of ICT in Education: A Survey of Schools in Europe." *European Journal of Education* 48(1):11–27.

Weinstein, C. E., Acee, T. W., & Jung, J. (2011). Self-regulation and learning strategies. New Directions for Teaching and Learning, 2011(126), 45-53.

Willging, P. A., and Johnson S. D. (2009). "Factors That Influence Students' Decision to Dropout of Online Courses." *Journal of Asynchronous Learning Networks* 13(3): 115–27.

What Works Clearinghouse (2014). Procedures and Standards Handbook Version 3.0. U.S. Department of Education. Washington, DC: Institute of Education Sciences.

Zimmerman, Barry J. (2000). "Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective." Pp. 13–39 in *Handbook of Self-Regulation*, edited by M. Boekaerts, P. R. Pintrich, and M. Zeidner. San Diego: Academic Press.

Zimmerman, B. J. (2001) Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), Self-regulated learning and academic achievement: Theoretical perspectives (2nd ed., pp. 1-38). New York: Lawrence Erlbaum Associates.

Zimmerman, Barry J., and Manuel Martinez Pons. (1986). "Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies." *American Educational Research Journal* 23(4):614–28.

Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance. In B. J. Zimmerman and D. H. Schunk (Eds.), Handbook of self-regulation of learning and performance (pp. 1-12). New York: Routledge.

APPENDIX A

REVIEW OF THE LITERATURE ON ONLINE COURSE RETENTION

This appendix summarizes the most relevant research evidence on online course retention. While there is no great consensus on how to define and measure retention (or, dropout, completion, success) in online courses, there is a good level of agreement in identifying **retention as one of the main challenges in online courses** (Lee and Choi 2011; Jordan 2014; 2015; Kizilcec et al. 2020). Typically, scalable and open online courses show higher levels of dropout than smaller and closed online courses.

While many studies address the issue of effective online professional development for teachers (Fyle 2013; Prestridge and Tondeur 2015; Koutsodimou & Jimoyiannis, 2015; Laurillard, 2016; Castaño-Muñoz et al., 2018; Misra, 2018; Parsons et al. 2019; Powell & Bodur, 2019), there is very limited evidence on teachers' retention in online courses. Most available studies on retention in online courses concern online programs delivered by post-secondary education institutions or massive online courses (MOOCS). Thus, the available evidence refers to contexts, settings and populations that may not be easily comparable to teachers and teacher training, even if many participants of MOOCs, regardless of the MOOC topic, are often teachers (Ho et al. 2015).

The lower completion rates in online vs traditional courses are a well-known challenge in the literature on distance education in post-secondary education (Carr 2000; Fike & Fike 2008; Lee & Choi 2011). Evidence of low retention rates is available also in literature studying MOOCs. One of the most recent and broad studies in the field is the one by Jordan (2014, 2015). The author examines 221 University MOOCs and concludes that completion rates (defined as the percentage of enrolled students who completed the course) are extremely variable across the analyzed courses and vary between 0.7% and 52.1%, with a median value of 12.6%.

Factors affecting course retention and completion

There has been a lot of research aimed at identifying the factors leading to low retention in online courses. This research mostly regards the case of post-secondary distance education. According to Lee and Choi (2011), these factors can be grouped into three main categories: *Student* factors; *Environmental* factors; and *Course/Program* factors.

Individual factors

There is non-conclusive evidence of the association between **participants' demographic characteristics** and online course retention. Regarding gender, some studies find that gender plays no role (Tello 2007), while others find that females tend to dropout less frequently than males even if no clear explanation for this pattern is found (Packam *et al.* 2004; Willging & Johnson 2009). Mixed evidence exists also with regards to the age of participants, with some studies reporting no association and others finding that older students show higher dropout rates (Lee & Choi, 2011).

Typically, it is found that the **educational background** of participants matters (Lee & Choi, 2011). Students with higher levels of education (Emanuel 2013) as well as students who are at an advanced stage of their training program have been found to be less likely to drop out (Levy 2007). Similarly, students with higher past academic achievement (Castels 2004) tend to show higher completion rates. Importantly, **past experience** (i.e. having

already participated in online courses) is typically found to be a strong predictor of success in online courses (Cheung and Kang 2002; Dupin-Bryant 2004).

There are also many studies pointing to the importance of **psychological attributes** and skills of the participants. Some studies suggest that the lack of motivation may be particularly important in online courses (Osborn 2011) because of the self-directed and self-learned nature of these courses compared to face-to-face ones (Bawa 2016). This brings in the issue of participants' **skills**. Several studies show that students who possess higher self-segulated learning competences such as time-management skills and resilience show higher attainments in online courses (Lee & Choi 2011; Bawa 2016; Kizilcec *et al.* 2017). Finally, digital and internet skills are typically found to correlate positively with success in online courses (Osborn 2001; Dupin-Bryant 2004; Yuan & Powell 2013).

Environmental aspects

Persistence and retention in online courses are determined by the extent to which course participants are capable of integrating study demands with external factors such as insufficient time, unexpected events and distractions (Kember 1995). Teachers who typically have **full-time jobs** and **important workloads** may struggle to find the needed time and need to take online courses at home (Pierrakeas *et al.* 2004). Moreover, as they often take these courses at home, they may be constrained by **family and social commitments** (Lee & Choi 2011).

Course factors and design features

According to some studies on MOOCs, there is an overall trend of improvement, as more recent courses show higher completion rates than less recent ones (Jordan 2015). This latter result may reflect a trend of continuous improvement in course quality and design. Indeed, how courses are designed is unanimously found to be important for increasing completion rates (Lee & Choi 2011; Jordan, 2015).

Among the online course factors that are found to be positively related to persistence rates, there is the degree to which courses allow for interactions (Bettinger et al. 2016). The lower frequency and lower quality of **student–teacher interactions** in online vs face-to-face courses could lead to dropout (Bawa 2016). In online courses, learners may "communicate with their instructors more to get help with a problem and less to take actual guidance to facilitate their learning". The existence of multiple communication options in online setups helps, but they risk not being used "as extensively as they should be, simply because the usage is largely dependent on the learners' own initiatives". Hence, online courses should be set up in a way that teacher-student interactions are facilitated. **Peer-to-peer interactions in virtual classrooms** are found to be positively linked with retention in online college courses (Tello 2007). This holds true particularly for students who are relatively less likely to be engaged in online discussion, which greatly benefit from increased exposure to more interactive peers (Bettinger et al 2016)

Additional course features that are found to be positively linked with completion rates are the **length of a course** and the modes of assessment. Ceteris paribus, it seems that the longer a course, the lower its completion rates (Jordan 2015). Courses using **self-assessment** as the only method of assessment have higher completion rates (Jordan 2015)

Promising approaches to improve retention in online courses

Robust evaluation studies on what works to increase retention in online courses are limited. However, it is possible to identify some promising interventions that could improve online course retention.

The first area to be considered is **advising and guidance support**. Students enrolled in online courses, especially in scalable online courses with large numbers of participants, have limited chances of having 1:1 interactions with

advisers and, especially the most at-risk among them, may benefit from ad hoc supports (Briggs and Spaulding 2018).[36] There is evidence of the effectiveness of technology-mediated coaching services. A randomised controlled trial conducted among college students found that students who received coaching service for assistance with goal setting, time management, and other support were more likely to attend college than students who had not received it (Bettinger and Baker 2011). There is also correlational evidence suggesting that **orientation programs** that introduce students to the demands of the online classes could be beneficial to students' successful completion of online courses (Clay et al. 2009; Bawa 2016).

A second suggestion coming from the literature is **leveraging on predictive data analytics to provide learners with personalized and equitable support** (Briggs and Spaulding 2018). Data-driven technology (artificial intelligence, machine learning, and predictive analytics) can allow tailoring support to individual needs and ultimately increase course retention (e.g., by. improving advising, course scheduling, personalised instruction). Kai ***et al.*** (2017) apply machine learning techniques to study the likelihood of registering in an online program at the University of Arkansas among students who participated in an online orientation course. The authors find that the degree of action and interaction during the online orientation course correlate positively with enrolment in the university program.

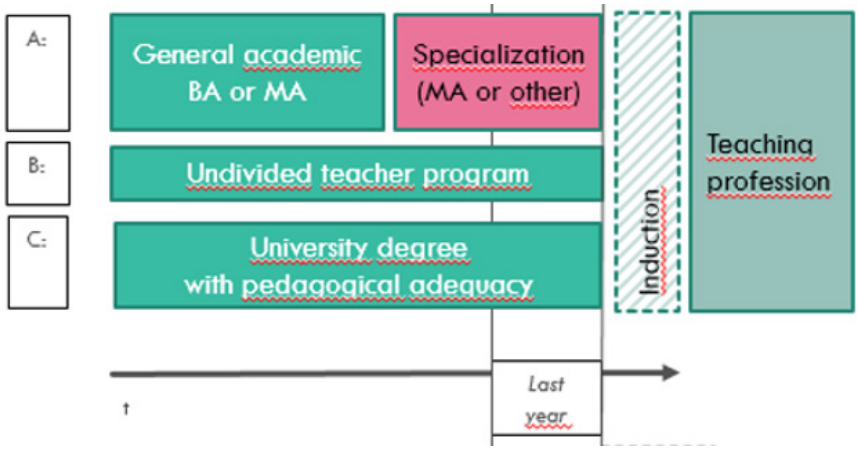
Ensuring high levels of interaction in the courses is a third factor expected to increase retention. To improve interaction, it is important that online courses are designed in such a way to foster effective dialogue, transparency and more social interaction between peers and students-teachers (Bawa 2016). Also, **classes should be structured to favor collaborative learning** (Poelhuber ***et al.*** 2008). Muirhead (2004) recommends that instructors develop strategies that will enhance their guidance for the students, such as creating a timeline for feedback and having a specific feedback rubric (Bawa 2016).

Finally, **enhancing instructors training:** teaching in a real class is different from teaching in a virtual one, teachers should be specifically trained to do it and should be incentivized to practice beforehand (Bawa 2016).

36 https://www.urban.org/urban-wire/three-ways-technology-can-help-nontraditional-students-succeed-online-coursework

APPENDIX B
ITE ORGANISATIONS

An ***ad hoc* internal online survey** has been carried out in order to gather comparable information concerning the organisation of initial teacher training in the countries participating in the TeachUP policy experimentation. Such information was necessary to wrap up a common protocol to sample student teachers.



Type A: Austria, Portugal, Slovakia, Spain, Turkey
Type A +B: Estonia, Hungary, Lithuania
Type A+B+C: Greece
Sources: survey on NCs
Figure B1 *Initial teacher education (ITE) organisations in the different countries*

The most common way to become a teacher is to get a mandatory teacher specialization lasting one, two or three years after a general academic degree (type A). However, in some countries you can also have an undivided teacher program that lasts longer and prepares directly for teaching profession (type B).

In addition to these two there is another way to become a teacher that is extremely common in Greece: from some not pedagogical university (such as Math) you may become a secondary teacher without any additional ITE (type C). The main difference between type B and type C is that in the latter there are not only those interested in becoming teachers but also other students that are not the target of TeachUP.

Table B1 *Organisation of ITE and adopted solution for sampling*

COUNTRY	TYPE	SHORT DESCRIPTION OF THE ITE ORGANISATION	CRITERIA USED FOR ITE ORGANISATIONS (LEVEL-1 SAMPLING UNITS)
AUSTRIA	A	After completing a bachelor degree, future teachers need to attend 1-year of ITE that is provided by Colleges	Last year of colleges providing ITE
ESTONIA	A+B	Future teachers must complete a pedagogical master degree, without having to attend any additional ITE course	Last year of pedagogical master degrees

COUNTRY	TYPE	SHORT DESCRIPTION OF THE ITE ORGANISATION	CRITERIA USED FOR ITE ORGANISATIONS (LEVEL-1 SAMPLING UNITS)
GREECE	A+B+C	Future teachers of ISCED-2 level have two options: 1) obtaining a bachelor degree in a university department with pedagogical courses (e.g., Mathematics, Physics, Literature etc.) without having to attend any additional ITE course; 2) completing a bachelor university degree without pedagogical adequacy plus: i) either attending additional ITE provided by ASPETE schools or university (1 year) ii) or completing a postgraduate university degree in the field of educational science (2 years)	Last year of University departments of Early Childhood Education and Primary Education
	B	Future teachers of ISCED-1 level need a bachelor degree of the University departments of Early Childhood Education and Primary Education	
HUNGARY	A+B	Future teachers have two options: 1) completing a generalistic Bachelor degree, + a teacher MSC/MA degree; 2) completing an "undivided" teacher program (lasting 5 or 6 years). In both cases, without having to attend any additional ITE course	Last year of: 1) teacher MSC/MA 2) undivided teacher program
LITHUANIA	A+B	Future teachers have two options: 1) *integrated model* - trainee teachers follow a professional route from the start and get both bachelor degree and teacher qualification without any additional courses; 2) *consecutive model* - after completing any bachelor degree, completing a degree for teacher qualification provided by universities (lasting 12 months)	Last year of 1) universities offering the integrated model 2) universities providing teacher qualification degrees *Note*: both tracks are available within the same university
MALTA	B	After completing a 4-year Bachelor Degree in Education at the University of Malta, future teachers go through a 2-year induction programme by the Ministry of Education (Quality Assurance Department)	Last year of Bachelor degree in education (University of Malta)
PORTUGAL	A	Future teachers have to complete a master teaching degree (2 years)	Last year of the master degree in teaching (university or some other higher education institutions)
SLOVAKIA	A	Future teachers have to complete a 2-year university master degree for secondary education in pedagogical and teachers faculties)	Last year if pedagogical and teachers' faculties
SPAIN	A	Future teachers have to go through three phases: Phase 1: 1-year ITE course (Master's Degree in Teaching in Secondary Schools) delivered by Universities Phase 2: Public Exam Phase 3: Induction (one year)	Last year of Master's degree in teaching Note: On request of the Spanish national partner, the experimentation will involve future teachers in academic year 2016/2017 that will follow the MOOC in the induction period

COUNTRY	TYPE	SHORT DESCRIPTION OF THE ITE ORGANISATION	CRITERIA USED FOR ITE ORGANISATIONS (LEVEL-1 SAMPLING UNITS)
TURKEY	A	Future teachers' training process develops in three phases: Phase 1: ITE courses delivered by the universities (14 weeks, from mid-September till the end of December) Phase 2: Public Exam Phase 3: Induction (one year) organised at provincial level	81 Provincial Directorates of National Education managing student teachers' induction year. *Note*: On request of the Turkish national partner, the experimentation will The focus on teachers in induction (Phase 3)

APPENDIX C
SAMPLE

Response rates resulted to be very heterogeneous across groups and countries. Among schools, on average 16 out of 100 invited schools have at least one enrolled teacher (61% in Turkey and from 4% to 100% in EU countries). Among ITE organisations the acceptance rates were, on average, even lower (3.4%) mainly due to non-response or explicit refusals in Turkey (3%). In the other countries the rates were higher, ranging from 10% in Estonia to 82% in Hungary (Malta had only one ITE so the rate is less informative).

It is worth mentioning that in Turkey the sampling design differs from the one implemented in the other countries because of the peculiarity of the pre-service teachers' population. Student teachers involved in TeachUP were, indeed, teachers attending the induction period exactly in a secondary school. To avoid that in-service and pre-service teachers belonging to the same schools could be part of the sample, we divided randomly the 81 Turkish provinces in two groups. Among the first group of 43 provinces we drew a sample of schools and invited only professional teachers while among the second group of 38 provinces we drew a sample of schools and invited only pre-service teachers. As the average number of teachers in induction in each school was below 5 it was necessary to contact many more schools than in other countries in order to reach the target.

Another peculiarity in the sampling design was Austria. In that case, as the numbers of volunteers in sample were very low despite the oversampling, on request of the Austrian national partner we added to the random sample also a 'convenience sample' of teachers involving all teachers who had already proven to be interested in Teach-up topics in previous projects regardless if they were teaching at a school that was in the sample or not.

Table C1 *Summary of the sampling process concerning schools*

COUNTRY	TARGET PROFESSIONAL TEACHERS	TARGET SCHOOLS	INVITED SCHOOLS	TEACHUP SCHOOLS	ACCEPTANCE RATE (%)
AT	300	76	497	29	5.8
EE	150	45	183	78	42.6
ES	500	66	282	23	8.2
GR	200	78	313	72	23.0
HU	300	111	721	84	11.7
LT	200	64	254	48	18.9
MT	50	10	10	10	100.0
PT	300	37	51	29	56.9
SK	200	78	613	25	4.1
TU	550	185	185	113	61.1
Total	2,750	741	3109	511	16.4

Table C2 *Summary of the sampling process concerning ITE organisations*

COUNTRY	TARGET STUDENT TEACHERS	TARGET ITE	INVITED ITE	TEACHUP ITE	ACCEPTANCE RATE %
AT	300	14	14	3	21.4
EE	150	2	19	2	10.5
ES	500	12	19	13	68.4
GR	200	11	11	8	72.7
HU	300	11	11	9	81.8
LT	200	11	11	6	54.5
MT	25	1	1	1	100.0
PT	300	24	24	16	66.7
SK	200	10	10	8	80.0
TU	550	641	13,151	390	3.0
Total	2,725	737	13271	456	3.4

APPENDIX D

RANDOMISATION AND EXPERIMENT'S INTEGRITY CHECKS

Randomization

To reduce treatment-control contamination all TeachUP teachers belonging to a treated school/ITE organisation were treated. On the other hand, all TeachUP teachers belonging to a control school/ITE organisation were controls. The process of allocation of the experimental units via random assignment to the two groups was firstly implemented before the launch of the first online course. Then, In order to increase the number of participants, as project subscription remained open throughout the experimental period, newly registered PTs or STs were randomized right before the start of a new course. This happened very rarely (see figure below)

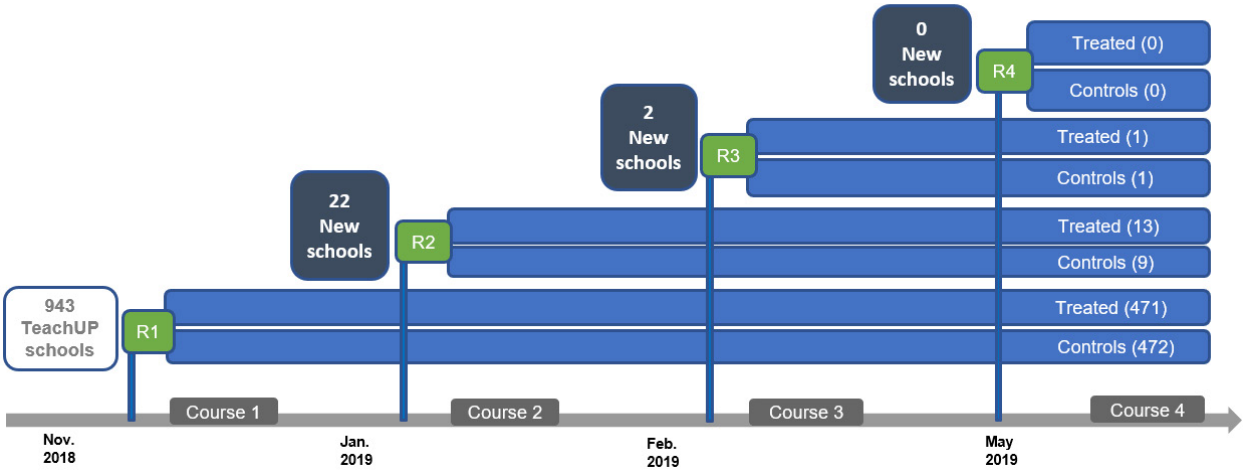


Figure D1 *Randomisation process*

The results of randomisation in terms of schools/ITE organisations are shown below.

Table C1 summarizes the result of the randomisation and shows the number of treated and control schools in each country. By definition, the size of the two groups in terms of number of schools is, overall, the same. Differences within countries are marginal and due to rounding procedures.

Table D1 *The randomisation of schools and ITE*

COUNTRIES	TEACHUP SCHOOLS			TEACHUP ITE		
	TOTAL	CONTROLS	TREATED	TOTAL	CONTROLS	TREATED
AT	29	15	14	3	1	2
EE	78	39	39	2	1	1
ES	23	12	11	13	8	5
GR	72	36	36	8	4	4
HU	84	42	42	9	4	5
LT	48	24	24	6	3	3

COUNTRIES	TEACHUP SCHOOLS			TEACHUP ITE		
	TOTAL	CONTROLS	TREATED	TOTAL	CONTROLS	TREATED
MT	10	5	5	1	0	1
PT	29	13	16	16	8	8
SK	25	12	13	8	4	4
TU	113	57	56	390	194	196
Total	**511**	**255**	**256**	**456**	**227**	**229**

As the size of the schools and ITE organisations did not appear in the algorithm used to perform the randomisation, the relative size in terms of teachers of treatment and control groups is different across countries (Table C2). In some countries, the treatment group is larger, while in others it is smaller in number. It means that in the former group, schools/ITE organisations assigned to the treatment group have more participants than those assigned to the control group while in the latter it is the other way around. These unbalances are often negligible and do not affect the impact evaluation analysis.

Table D2 *Number of Professional Teachers and Student Teachers belonging to treated and control groups*

COUNTRIES	PROFESSIONAL TEACHERS			STUDENT TEACHERS		
	TOTAL	CONTROLS	TREATED	TOTAL	CONTROLS	TREATED
AT	42	24	18	19	14	5
EE	180	93	87	7	2	5
ES	75	29	46	446	225	221
GR	153	65	88	157	111	46
HU	197	109	88	105	22	83
LT	191	116	75	34	6	28
MT	23	8	15	30	0	30
PT	203	78	125	73	40	33
SK	46	17	29	57	14	43
TU	1,078	484	594	970	499	471
Total	**2,188**	**1023**	**1165**	**1898**	**933**	**965**

Equivalence tests

Two crucial checks were performed in order to assess the quality of the randomisation: the balancing tests and the analysis of the attrition rates.

In order to estimate the statistical significance of differences between the two groups in terms of number of student and professional teachers, we regress the average number of teachers enrolled in Teach-up by school/ITE on a dummy variable indicating the treatment status. The OLS models also include dummies for the blocking variables used in the randomisation procedure, that is the geographical strata used. Table C3 shows the results of this analysis. It

reports separately for PTs and STs: (i) the mean size of schools/ITE in treatment and control group; (ii) the ●coefficient of the regression model described, expressing the estimated mean difference between the two groups; and (iii) to assess the statistical significance of the mean differences we also show the corresponding p-values.

Overall, the differences are small and not statistically significant. Hence, treated and control schools/ITE have not only a similar profile in terms of geographical distribution within the country (achieved by design) but also in terms of school/ITE size and teachers' average propensity of participating in Teach-up. These conclusions apply also to within-country analyses.

Table D3 *Balancing test, schools/ITE size*

VARIABLES	CONTROLS	TREATED	β	P-VAL
Schools: Number of TeachUP teachers (mean)	3.97	4.46	0.559	0.322
ITE: Number of Teach-UP teachers (mean)	4.09	4.03	-0.195	0.777

**** p<0.01; ** p<0.05; * p<0.1*

The following tables contain additional balancing tests aimed at assessing the comparability of treated and control teachers with respect to a number of variables collected with the Benchmark Survey such as teachers' gender, age, and subject taught and the value of the main outcomes at the pre-treatment stage.

Table D4 *Balancing test on enrolled teachers, by group*

	EU PT	EU ST	TR PT	TR ST
Other teacher	-0.061		-0.083	
	(0.055)		(0.067)	
Teacher	0.000		0.000	
	(.)		(.)	
Student		-0.073		-0.262
		(0.109)		(0.162)
Teacher in induction		-0.040		-0.166
		(0.123)		(0.134)
Other student		-0.038		-0.243*
		(0.137)		(0.144)
Female	-0.055	0.028	0.022	-0.032
	(0.046)	(0.032)	(0.039)	(0.039)
30 years old or less	0.029		0.016	
	(0.111)		(0.114)	
30-39 years old	0.055		0.043	
	(0.073)		(0.089)	
40-49 years old	0.044		0.029	
	(0.050)		(0.059)	

	EU PT	EU ST	TR PT	TR ST
More than 50 years old	0.000		0.000	
	(.)		(.)	
English proficiency	0.002	0.010	0.003	0.023***
	(0.007)	(0.008)	(0.005)	(0.007)
At least a master degree	0.005	-0.009	0.028	-0.161
	(0.038)	(0.047)	(0.052)	(0.135)
Teach at Primary education	-0.020	0.101	0.146	-0.054
	(0.095)	(0.094)	(0.150)	(0.118)
Teach at Lower secondary education	0.016	-0.053	0.153	-0.214**
	(0.064)	(0.049)	(0.108)	(0.090)
Teach at Upper secondary education	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Teach Humanities (arts, social, language, reading)	0.031	0.024	-0.099*	-0.059
	(0.046)	(0.043)	(0.057)	(0.050)
Teach Science (math, science, technology)	0.064	-0.014	-0.117**	0.048
	(0.046)	(0.043)	(0.057)	(0.052)
Teach Other subject	0.001	0.002	-0.066	0.043
	(0.047)	(0.036)	(0.073)	(0.056)
Availability of devices (ICT)	-0.018	-0.004	0.007	-0.025**
	(0.011)	(0.009)	(0.009)	(0.011)
Have access to the internet	0.012	0.043	-0.068	-0.039
	(0.083)	(0.099)	(0.086)	(0.052)
Quality of Internet in school	-0.022**	-0.001	-0.000	0.000
	(0.010)	(0.008)	(0.009)	(0.008)
Quality of Internet at home	0.015	-0.009	0.000	-0.001
	(0.011)	(0.011)	(0.009)	(0.008)
Daily use of computer: Less than 60 min	-0.010	-0.049	-0.001	-0.022
	(0.047)	(0.055)	(0.047)	(0.061)
Daily use of computer: 1-3 hours	-0.030	-0.105***	-0.008	0.005
	(0.044)	(0.037)	(0.046)	(0.043)
Daily use of computer: More than 3 hours	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Digital competencies general	0.290	-0.109	-0.152	-0.205
	(0.189)	(0.142)	(0.135)	(0.183)
Digital competencies social	0.034	0.009	-0.333***	-0.032
	(0.125)	(0.089)	(0.110)	(0.142)
Online courses started in the past three years	-0.144	0.067	-0.232**	-0.085
	(0.097)	(0.091)	(0.100)	(0.140)

	EU PT	EU ST	TR PT	TR ST
Online courses completed in the past three years	0.131	-0.101	0.238**	0.114
	(0.103)	(0.100)	(0.101)	(0.151)
Learning experience: goal, strategy	-0.210	0.117	-0.124	0.312*
	(0.133)	(0.089)	(0.153)	(0.189)
Learning experience: help seeking	0.148	0.153	0.177	0.079
	(0.108)	(0.098)	(0.109)	(0.141)
Online courses views	0.038	-0.046	0.100	0.038
	(0.142)	(0.133)	(0.124)	(0.128)
Expected formal recognitions of TeachUP certificate	-0.043	0.032	0.007	-0.044
	(0.037)	(0.028)	(0.029)	(0.051)
Teaching practices: social	0.002	0.187	-0.108	0.028
	(0.151)	(0.133)	(0.134)	(0.177)
Teaching practices: content	-0.111	0.089	-0.000	0.070
	(0.110)	(0.110)	(0.095)	(0.135)
Start of teaching career (year)	-0.001		-0.006	
	(0.003)		(0.004)	
Teaching load: Less than 25 hours per week	-0.066		-0.036	
	(0.044)		(0.043)	
Teaching load: Between 26 and 35 hours per week	0.027		-0.031	
	(0.039)		(0.038)	
Teaching load: More than 36 hours per week	0.000		0.000	
	(.)		(.)	
CPD: At most once a year	0.036		-0.078	
	(0.048)		(0.057)	
CPD: Every six months	0.022		-0.108	
	(0.044)		(0.067)	
CPD: At least every three months	0.000		0.000	
	(.)		(.)	
Number of actions for encouragement received	-0.034**		0.012	
	(0.016)		(0.016)	
Online communities: No, I have never done it	0.070		0.034	
	(0.055)		(0.059)	
Online communities: Yes, I have done it but very rarely	-0.049		0.035	
	(0.049)		(0.042)	
Online communities: Yes, I have done it occasionally	-0.025		0.051	
	(0.047)		(0.034)	
Online communities: Yes, I have done it very often	0.000		0.000	
	(.)		(.)	

	EU PT	EU ST	TR PT	TR ST
_cons	3.394	0.385	13.567*	0.792***
	(5.826)	(0.254)	(8.145)	(0.261)
F	1.379	0.963	1.647	1.541
p	0.075	0.529	0.026	0.043
N	976.000	788.000	983.000	890.000

Attrition

The second set of checks had to do with the issue of attrition (or non response rate) in the post-treatment surveys. Attrition is the process by which some sample units, after taking part in the first survey (the BS in our case), drop out and do not take part in subsequent waves (the mini-survey or the FuS in our case). In this section, we check whether attrition compromised the integrity of the evaluation design. We do that by: i) comparing the attrition levels by group; and ii) by running the equivalence analyses of the previous section on the subsample of teachers completing the FuS.

Regarding the **mini survey**, as mentioned above, there were two questionnaires depending on whether or not the participant had completed the course. Among course completers, response rates were very high and well balanced between treated and controls. On the contrary, among course non-completers, response rates were less than 10% making the survey unusable for this group.

Table D5 *Response rates in the mini post course surveys*

	COURSE 1		COURSE 2		COURSE 3		COURSE 4	
	CONTROLS	TREATED	CONTROLS	TREATED	CONTROLS	TREATED	CONTROLS	TREATED
COURSE COMPLETERS								
EU PT	89%	82%	88%	92%	86%	94%	89%	91%
EU ST	84%	91%	86%	92%	87%	90%	94%	90%
TR PT	88%	91%	92%	90%	88%	93%	92%	83%
TR ST	90%	91%	81%	80%	86%	91%	91%	89%
COURSE NON-COMPLETERS								
Overall	7%	6%	9%	9%	11%	9%	6%	7%

Concerning the **FollowUP Survey** we registered low response rates (18.8%). We used the Baseline Survey to compare respondents and not respondents and we found some differences but not dramatically large. Respondents tend to be more experienced in online training and, in general, more involved in training; have better views on online courses; be more mathematics/science teacher than humanities. In measuring the causal effect, we account for these differences.

Table D6 *Number of respondents to the FUS and response rates*

TYPE	NO FUS	FUS	TOTAL	OVERALL %	TREATED %	CONTROLS %
EU Professional	640	364	1,004	36.3	37.4	35.0
EU Student Teacher	689	149	838	17.8	15.7	20.5

TYPE	NO FUS	FUS	TOTAL	OVERALL %	TREATED %	CONTROLS %
TR Professional	886	136	1,022	13.3	12.5	14.3
TR Student	851	62	913	6.8	6.1	7.4
EU PTs without ES and SK	550	346	896	38.6	39.8	37.4
EU STs without ES and SK	280	104	384	27.1	24.1	30.2
Total	**3,066**	**711**	**3,777***	**18.8**	**18.3**	**19.4**

**About 315 users signed up to TeachUP but did not register to any of the four courses nor completed the FUS, hence are not considered here*

Table D7 *Number of respondents to the FUS and response rates by country*

COUNTRY	FUS NON RESP	FUS RESP.	TOTAL	OVERALL %	T %	C %
AT	27	35	62	56.5	60.9	53.9
EE	95	70	165	42.4	43.0	41.9
ES	420	57	477	11.9	12.8	10.7
GR	167	125	292	42.8	47.2	39.5
HU	213	56	269	20.8	22.9	18.6
LT	123	82	205	40.0	35.8	43.6
MT	27	21	48	43.8	38.1*	83.3*
PT	178	61	239	25.5	26.1	24.7
SK	79	6	85	7.1	6.9	7.4*
TU	1,737	198	1,935	10.2	9.7	10.8
Total	**3,066**	**711**	**3,777**	**18.8**	**18.3**	**19.4**

**Less than 10 units*
Note: T= Treated C= controls

Table D8 *FUS Respondents by number of courses completed*

	NUMBER OF COURSES COMPLETED					
	NONE	1	2	3	4	TOTAL
%	7.3	27.5	39.0	57.1	72.4	18.8
N	195	127	64	84	241	711

Table D9 *Baseline characteristics of PT FUS respondents and non respondents (means and t-test on the difference)*

	EU -PT			TR-PT		
	NO FUS MEAN	FUS MEAN	P	NO FUS MEAN	FUS MEAN	P
Teacher	0.92	0.89	0.039	0.92	0.90	0.449
Other teacher	0.08	0.11	0.054	0.08	0.10	0.449
Female	0.85	0.85	0.942	0.65	0.73	0.063
30 years old or less	0.07	0.03	0.006	0.21	0.17	0.269
30-39 years old	0.25	0.20	0.105	0.50	0.52	0.610
40-49 years old	0.34	0.38	0.207	0.22	0.26	0.351
More than 50 years old	0.34	0.39	0.138	0.07	0.05	0.430
English proficiency	5.29	5.49	0.285	3.74	4.18	0.105
At least a master degree	0.60	0.57	0.461	0.10	0.12	0.551
Teach at Primary education	0.14	0.22	0.001	0.03	0.03	0.986
Teach at Lower secondary education	0.48	0.47	0.784	0.95	0.95	0.994
Teach at Upper secondary education	0.38	0.31	0.026	0.02	0.02	0.970
Teach Humanities (arts, social, language, reading)	0.65	0.69	0.199	0.51	0.49	0.590
Teach Science (math, science, technology)	0.48	0.52	0.153	0.38	0.49	0.009
Teach Other subject (vocational, gym, religion)	0.17	0.19	0.510	0.13	0.06	0.018
Availability of devices (ICT)	5.36	5.37	0.929	5.34	5.50	0.333
Have access to the internet	0.97	0.97	0.996	0.94	0.98	0.051
Quality of Internet in school	6.60	6.84	0.094	5.25	5.10	0.550
Quality of Internet at home	8.46	8.43	0.754	7.75	8.26	0.019
Daily use of a computer: Less than 60 min	0.30	0.30	0.941	0.24	0.18	0.146
Daily use of a computer: 1/ 3 hours	0.51	0.53	0.545	0.55	0.57	0.637
Daily use of a computer: More than 3 hours	0.19	0.17	0.489	0.21	0.24	0.354
Digital competencies general	0.67	0.68	0.147	0.65	0.69	0.010
Digital competencies social	0.66	0.67	0.181	0.63	0.62	0.409
Online courses started in the past 3 years	0.16	0.28	0.000	0.12	0.14	0.419
Online courses completed in the past 3 years	0.14	0.25	0.000	0.11	0.15	0.148
Learning experience: goal, strategy	0.60	0.61	0.102	0.60	0.61	0.960
Learning experience: help seeking	0.64	0.65	0.965	0.66	0.69	0.014
Online courses views	0.77	0.82	0.000	0.71	0.78	0.000
Expected formal recognitions of TeachUP	0.63	0.55	0.017	0.77	0.81	0.324
Teaching practices: social	0.64	0.64	0.950	0.68	0.71	0.012
Teaching practices: content	0.54	0.51	0.007	0.65	0.65	0.922

	EU -PT			TR-PT		
	NO FUS	FUS		NO FUS	FUS	
	MEAN	MEAN	P	MEAN	MEAN	P
Start of teaching career (year)	1999	1996	0.001	2005	2005	0.778
Teaching load: Less than 25 hours per week	0.20	0.17	0.279	0.31	0.26	0.250
Teaching load: 26/35 hours per week	0.25	0.25	0.937	0.46	0.46	0.959
Teaching load: More than 36 hours per week	0.55	0.58	0.430	0.24	0.29	0.198
CPD: At most once a year	0.43	0.36	0.023	0.65	0.58	0.126
CPD: Every six months	0.29	0.27	0.399	0.23	0.29	0.161
CPD: At least every three months	0.27	0.37	0.001	0.12	0.13	0.676
Nr of actions for encouragement received	1.55	1.60	0.459	1.08	1.13	0.546
Online communities: never	0.19	0.17	0.337	0.13	0.10	0.222
Online communities: very rarely	0.29	0.23	0.081	0.23	0.26	0.545
Online communities: occasionally	0.31	0.37	0.035	0.39	0.38	0.797
Online communities: very often	0.22	0.23	0.715	0.24	0.27	0.520
Observations	640	364	1004	886	136	1022

Table D10 *Baseline characteristics of ST FUS respondents and non respondents (means and t-test on the difference)*

	EU-ST			TR-ST		
	NO FUS	FUS		NO FUS	FUS	
	MEAN	MEAN	P	MEAN	MEAN	P
Student	0.81	0.66	0.000	0.06	0.02	0.131
Teacher in induction	0.13	0.17	0.160	0.78	0.79	0.887
Other student	0.03	0.07	0.004	0.12	0.13	0.875
Female	0.77	0.79	0.519	0.68	0.65	0.527
English proficiency	6.53	7.14	0.001	3.60	4.40	0.027
At least a master degree	0.37	0.31	0.148	0.02	0.03	0.348
Primary education	0.23	0.34	0.005	0.07	0.08	0.775
Lower secondary education	0.23	0.20	0.434	0.89	0.87	0.645
Upper secondary education	0.54	0.46	0.072	0.04	0.05	0.715
Humanities (arts, social, language, reading)	0.63	0.62	0.802	0.44	0.61	0.010
Science (math, science, technology)	0.47	0.63	0.000	0.40	0.37	0.632
Other subject (vocational, gym, religion)	0.26	0.35	0.025	0.23	0.11	0.035
Availability of devices (ICT)	5.23	5.17	0.655	4.73	4.66	0.756
Have access to the internet	0.95	0.93	0.297	0.82	0.79	0.559
Quality of Internet in school	6.46	6.74	0.187	4.08	3.36	0.067
Quality of Internet at home	8.42	8.20	0.113	6.75	6.66	0.785

	EU-ST			TR-ST		
	NO FUS	FUS		NO FUS	FUS	
	MEAN	MEAN	P	MEAN	MEAN	P
Daily use of a computer: Less than 60 min	0.13	0.16	0.488	0.16	0.26	0.055
Daily use of a computer: 1/ 3 hours	0.50	0.55	0.297	0.60	0.48	0.063
Daily use of a computer: More than 3 hours	0.37	0.30	0.112	0.23	0.26	0.649
Digital competencies general	0.67	0.67	0.917	0.67	0.67	0.633
Digital competencies social	0.67	0.69	0.242	0.63	0.65	0.426
Online courses started in the past 3 years	0.19	0.26	0.041	0.04	0.13	0.002
Online courses completed in the past 3 years	0.16	0.22	0.097	0.03	0.11	0.001
Learning experience: goal, strategy	0.63	0.61	0.272	0.64	0.63	0.609
Learning experience: help seeking	0.63	0.61	0.108	0.66	0.64	0.287
Online courses views	0.80	0.81	0.214	0.74	0.76	0.298
Expected formal recognitions of TeachUP	0.51	0.62	0.020	0.84	0.81	0.492
Teaching practices: social	0.70	0.70	0.819	0.72	0.70	0.276
Teaching practices: content	0.54	0.54	0.910	0.66	0.67	0.768
Observations	689	149	838	851	62	913

APPENDIX E
MEASURES

The battery in the questionnaire had 24 statements about SRL strategies on how characteristic they were for them on a labeled 5-point scale (from 1 to 5)[37]: goal setting strategies (4 statements), strategic planning (4), self-evaluation (3), task strategies (6), elaboration (3), and help seeking (4). To compute an individual SRL score we used factorial analysis.[38] The determination of the number of factors to extract was guided by the theory, but also informed by running the analysis extracting different numbers of factors and seeing which number of factors yielded the most interpretable results.

Table E1 *Self-evaluated SRLO: Rotated factor loadings (pattern matrix)*

ITEMS		FACTOR1 GOAL	FACTOR2 TASK	FACTOR3 HELP
1	I set personal standards for performance in my learning.	0.655	0.314	-0.249
2	I set short-term (daily or weekly) goals as well as long-term goals (for the whole course).	0.688	0.368	-0.355
3	I set goals to help me manage studying time for my learning.	0.715	0.361	-0.373
4	I set realistic deadlines for learning.	0.712	0.288	-0.281
5	I ask myself questions about what I am to study before I begin to learn.	0.702	0.113	-0.027
6	I think of alternative ways to solve a problem and choose the best one.	0.740	0.079	0.038
7	When planning my learning, I use and adapt strategies that have worked in the past.	0.744	0.098	-0.078
8	I organise my study time to accomplish my goals to the best of my ability.	0.762	0.183	-0.177
9	I try to translate new information into my own words.	0.709	0.078	0.081
10	I ask myself how what I am learning is related to what I already know.	0.750	0.032	0.187
11	I change strategies when I do not make progress while learning.	0.758	0.017	0.122
12	When I study for a course, I make notes to help me organize my thoughts.	0.643	-0.079	0.043
13	I create my own examples to make information more meaningful.	0.738	-0.018	0.217
14	I read beyond the core course materials to improve my understanding.	0.700	-0.046	0.252
15	When I am learning, I try to relate new information I find to what I already know.	0.778	-0.054	0.222
16	When I am learning, I combine different sources of information (for example: people, web sites, printed material).	0.755	-0.076	0.196
17	I try to apply my previous experience when learning.	0.760	-0.056	0.191
18	I know how well I have learned once I have finished a task.	0.746	-0.042	0.157
19	I ask myself if there were other ways to do things after I finish learning.	0.720	-0.071	0.173
20	I think about what I have learned after I finish.	0.748	-0.118	0.133

37 Respondents indicated how they typically behave when participating in online learning experiences. Multiple-choice options for each item were: Not at all true for me (1); Sometimes true for me (2); Quite true for me (3); True for me (4); Very true for me (5).

38 This mathematical procedure seeks underlying unobservable (latent) variables, called factors, that are reflected in the observed variables.

ITEMS		FACTOR1 GOAL	FACTOR2 TASK	FACTOR3 HELP
21	When I do not understand something, I ask others for help.	0.665	-0.554	-0.262
22	I try to identify others whom I can ask for help if necessary	0.675	-0.522	-0.258
23	I ask others for more information when I need it.	0.665	-0.517	-0.268
24	Even if I am having trouble learning, I prefer to do the work on my own.	0.331	0.332	0.372
		Alpha 0.90	**Alpha 0.93**	**Alpha 0.90**

Table E2 *Self-evaluated digital competence (Rotated factor loadings (pattern matrix)*

ITEMS		INTERNET ABILITIES I	INTERNET ABILITIES II
1	Conduct an internet search using one or more keywords	0.669	-0.128
2	Judge the reliability of a website	0.669	0.123
3	Reflect on my online search process	0.733	0.134
4	Participate in a discussion forum	0.639	0.603
5	Participate in an online chat session	0.661	0.598
6	Use social media to interact with others	0.633	0.283
7	Use digital technologies for collaborative work/projects	0.765	-0.066
8	Identify personal needs and select digital tools to solve them	0.771	-0.163
9	Use digital technologies to carry out tasks in a more effective way	0.792	-0.284
10	Seek opportunities to develop my skills to use digital technologies effectively	0.753	-0.219
11	Upload document files	0.773	-0.220
12	Complete multiple choice tests	0.752	-0.183
13	Navigate to a specific point in videos	0.771	-0.114
14	Comment and behave in a way that is appropriate to the situation I find myself in online	0.762	0.068
15	Decide which information should or should not be shared online	0.725	-0.200
		Alpha 0.79	**Alpha 0.93**

Table E3 *Views on online courses. Rotated factor loadings (pattern matrix)*

ITEMS		ONLINE VIEWS
1	Online courses are an effective way of improving my competences and increasing my knowledge	0.861
2	Online courses are a very enjoyable experience	0.835
3	Online courses are a great learning opportunity because they are accessible anywhere and anytime	0.885
4	Online courses are a great learning opportunity because they allow exchanges with people from many different countries	0.834

ITEMS		ONLINE VIEWS
5	I will definitely take part in some online courses in the future	0.862
		Alpha
		0.93

Table E4 *Self-evaluated teaching practices. Rotated factor loadings (pattern matrix)*

ITEMS		TEACHING PRACTICES I	TEACHING PRACTICES II
1	I believe that expanding on students' ideas is an effective way to build my curriculum.	0.601	-0.361
2	An essential part of my teacher role is supporting a student's family when problems are interfering with a student's learning.	0.699	-0.272
3	I involve students in evaluating their own work and setting their own goals.	0.659	-0.437
4	I make it a priority in my classroom to give students time to work together when I am not directing them.	0.670	-0.371
5	I invite parents to volunteer in or visit my classroom almost any time.	0.668	-0.134
6	I prefer to assess students informally through observations and conferences.	0.688	-0.220
7	I often create thematic units based on the students' interests and ideas.	0.699	-0.284
8	I like to make curriculum choices for students because they can't know what they need to learn.	0.559	0.260
9	I base student grades primarily on homework, quizzes, and tests.	0.440	0.596
10	To be sure that I teach students all necessary content and skills, I follow a textbook or workbook.	0.587	0.561
11	I teach subjects separately, although I am aware of the overlap of content and skills.	0.619	0.502
12	I find that textbooks and other published materials are the best sources for creating my curriculum.	0.575	0.583
		Alpha	**Alpha**
		0.84	**0.81**

APPENDIX F

DESCRIPTIVE FIGURES AND PARTICIPANT PROFILES

Table F1 *Digital competences, by country (% of PT and ST who declared they are able to)*

	COUNTRY									
	AT	EE	ES	GR	HU	LT	MT	PT	SK	TU
GENERAL ABILITIES										
Conduct an internet search using one or more keywords	94	93	89	89	93	88	94	90	86	82
Judge the reliability of a website	84	69	71	70	62	51	75	77	60	56
Reflect on my online search process	81	63	71	69	68	62	77	81	58	63
Use digital technologies for collaborative work/projects	84	71	84	78	74	69	71	90	69	69
Identify personal needs and select digital tools to solve them	73	65	69	64	65	70	56	82	52	75
Use digital technologies to carry out tasks in a more effective way	77	64	79	79	75	75	75	89	71	79
Seek opportunities to develop my skills to use digital technologies effectively	84	68	78	77	67	67	69	87	58	69
Upload document files	89	88	92	79	86	80	92	94	76	79
Complete multiple-choice tests	87	87	83	82	92	86	79	85	79	83
Navigate to a specific point in videos	87	80	88	81	87	71	81	85	55	75
Comment and behave in a way that is appropriate to the situation I find myself in online	84	82	87	79	77	72	83	82	61	74
Decide which information should or should not be shared online	87	89	89	90	88	84	94	94	79	86
SOCIAL ABILITIES										
Participate in a discussion forum	73	68	58	58	58	59	63	68	39	46
Participate in an online chat session	74	70	56	65	60	61	56	62	46	56
Use social media to interact with others	79	82	72	77	72	68	88	70	69	78

Table F.2 *Quality of internet connection (0 "low - 10 "high")*

COUNTRY	PROFESSIONAL TEACHERS		STUDENT TEACHERS	
	AT SCHOOL	AT HOME	AT SCHOOL	AT HOME
AT	7.4	5.1	6.7	7.2
EE	7.1	6.6	7.7	3.9
ES	8.3	5.8	8.5	6.2
GR	8.7	6.7	8.5	7.0
HU	8.5	6.4	8.4	6.5
LT	7.4	4.8	7.4	6.4
MT	8.6	7.1	9.7	7.2

COUNTRY	PROFESSIONAL TEACHERS		STUDENT TEACHERS	
	AT SCHOOL	AT HOME	AT SCHOOL	AT HOME
PT	8.8	7.0	8.3	6.3
SK	8.1	4.8	7.9	6.2
TU	7.5	6.6	6.5	7.6
Total	**7.8**	**6.4**	**7.4**	**7.0**

Table F.3 *Percentages of participants who used ICT devices in the last 30 days*

	PROFESSIONAL TEACHERS				STUDENT TEACHERS			
COUNTRY	BOTH AT SCHOOL AND OUTSIDE SCHOOL	ONLY AT SCHOOL	ONLY OUTSIDE SCHOOL	NO DEVICE USED	BOTH AT SCHOOL AND OUTSIDE SCHOOL	ONLY AT SCHOOL	ONLY OUTSIDE SCHOOL	NO DEVICE USED
AT	97.6	0.0	2.4	0.0	95.0	0.0	5.0	0.0
EE	97.5	0.6	1.9	0.0	85.7	0.0	14.3	0.0
ES	100.0	0.0	0.0	0.0	85.8	1.7	12.0	0.5
GR	94.3	2.1	2.8	0.7	70.2	0.0	28.5	1.3
HU	98.3	0.6	1.1	0.0	78.5	2.2	19.4	0.0
LT	92.7	5.1	2.3	0.0	75.0	7.1	17.9	0.0
MT	90.0	0.0	10.0	0.0	96.4	0.0	3.6	0.0
PT	98.9	1.1	0.0	0.0	91.2	0.0	8.8	0.0
SK	100.0	0.0	0.0	0.0	84.8	0.0	15.2	0.0
TU	80.7	6.1	11.3	2.0	75.1	3.9	19.3	1.6
Total	**88.6**	**3.9**	**6.5**	**1.0**	**78.8**	**2.7**	**17.5**	**1.1**

Table F.4 *SRLO indices (0= low, 1=high) by country*

COUNTRY	TASK STRATEGIES		GOAL SETTING		HELP SEEKING	
	PT	ST	PT	ST	PT	ST
AT	0.59	0.60	0.57	0.55	0.60	0.61
EE	0.54	0.47	0.60	0.56	0.62	0.68
ES	0.55	0.56	0.63	0.63	0.65	0.64
GR	0.57	0.57	0.63	0.63	0.67	0.60
HU	0.54	0.54	0.62	0.62	0.63	0.60
LT	0.59	0.63	0.58	0.61	0.65	0.60
MT	0.57	0.50	0.67	0.66	0.58	0.71
PT	0.58	0.57	0.60	0.63	0.68	0.68
SK	0.59	0.55	0.56	0.62	0.58	0.61
TU	0.60	0.62	0.60	0.64	0.66	0.66

COUNTRY	TASK STRATEGIES		GOAL SETTING		HELP SEEKING	
	PT	ST	PT	ST	PT	ST
Total	**0.58**	**0.59**	**0.60**	**0.63**	**0.65**	**0.65**

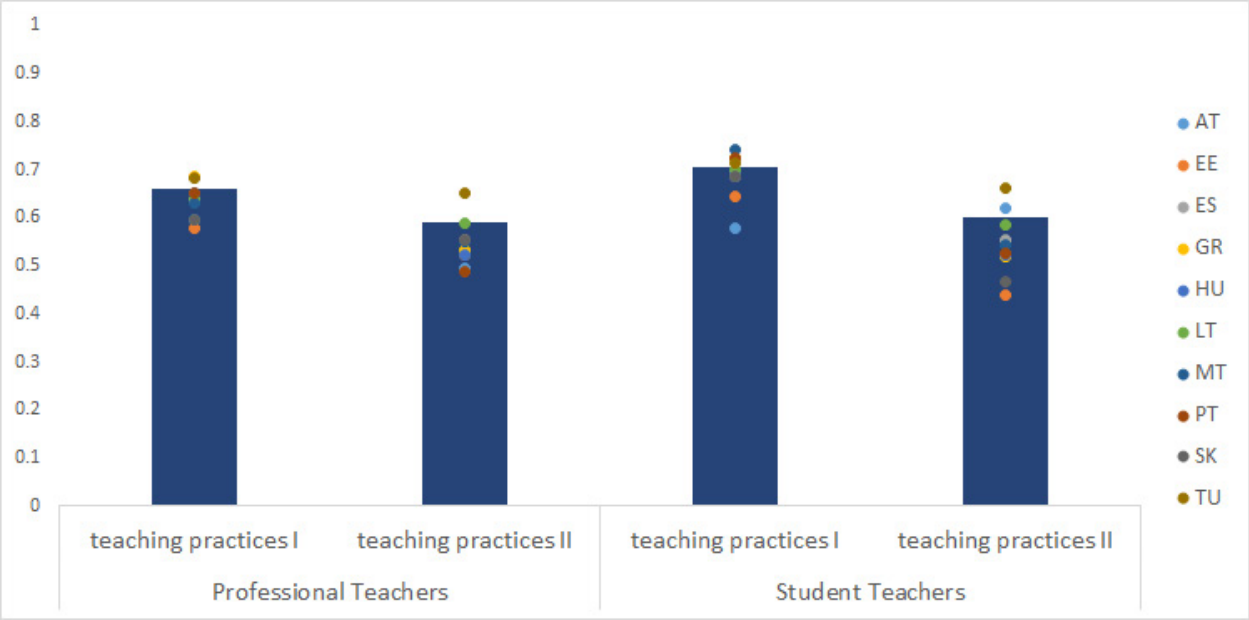


Figure F1 *Self-assessed teaching practices among PTs and STs*
Note:Two dimensions: (range 0-1)

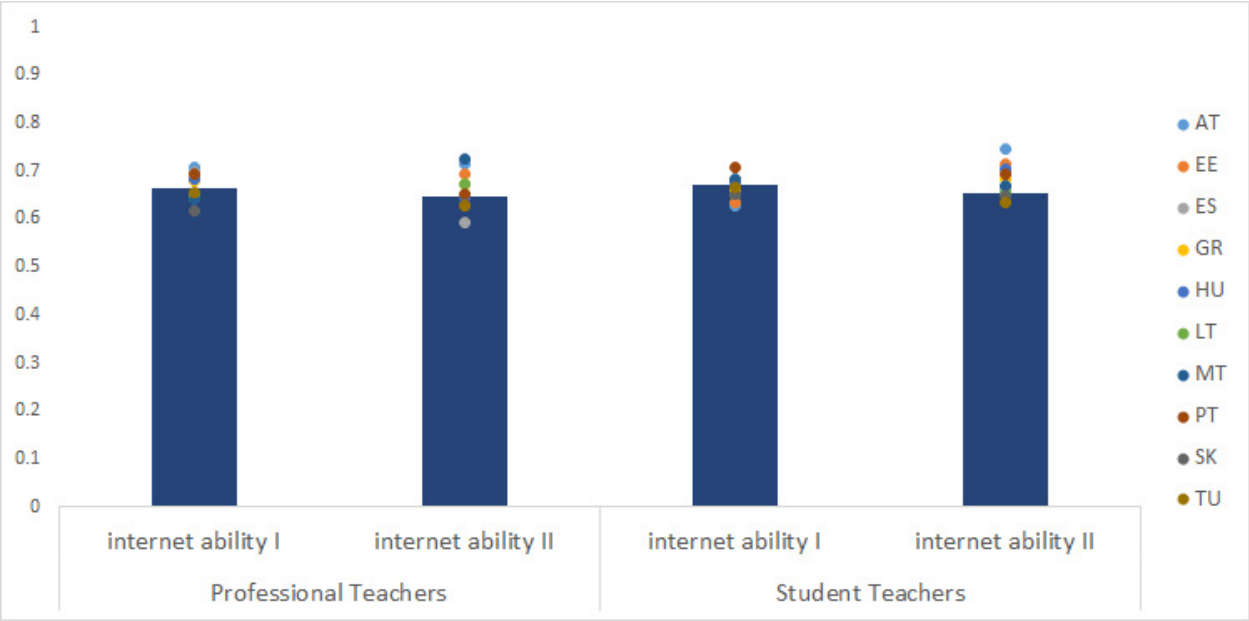


Figure F2 *Self-assessed internet competences*
Note:Two dimensions: (range 0-1)

Table F5 *Number of teachers enrolled by course in each country (absolute values)*

COUNTRY	COURSE 1	COURSE 2	COURSE 3	COURSE 4
AT	40	33	42	29
EE	118	94	99	96
ES	427	358	356	364
GR	246	207	216	223
HU	189	152	138	153
LT	181	88	91	94
MT	45	37	38	38
PT	194	136	126	127
SK	63	60	53	52
TU	1,757	1,748	1,737	1,729
Total	**3,260**	**2,913**	**2,896**	**2,905**

Table F6 *Number of teachers started by course (absolute values)*

COUNTRY	COURSE 1	COURSE 2	COURSE 3	COURSE 4
AT	24	19	14	13
EE	79	45	43	42
ES	266	135	105	96
GR	153	92	82	83
HU	116	73	51	44
LT	141	69	61	61
MT	23	18	14	16
PT	100	53	37	37
SK	45	13	10	6
TU	973	402	295	242
Total	**1,920**	**919**	**712**	**640**

Table F7 *Number of teachers completed by course (absolute values)*

COUNTRY	COURSE 1	COURSE 2	COURSE 3	COURSE 4
AT	14	7	9	8
EE	49	42	39	31
ES	130	93	79	70
GR	84	66	67	64
HU	51	37	33	27
LT	86	51	49	51

COUNTRY	COURSE 1	COURSE 2	COURSE 3	COURSE 4
MT	14	14	14	16
PT	44	28	25	25
SK	11	8	5	1
TU	464	258	223	182
Total	947	604	543	475

Table F8 *Characteristics of completers and non-completers, by ST and PT (percentages)*

CHARACTERISTICS	EU-PT		EU-ST		TR-PT		TR-ST	
	NC	C	NC	C	NC	C	NC	C
Female	83.6	87.5	74.4	84.5	64.5	70.7	68.1	68.3
Aged 40+	71.1	71.9	11	8.6	29.7	29.2	1.1	1.4
High English proficiency	48.8	47.7	75.5	76.8	26.5	33.2	19.7	29.8
Teaches maths, science, technology	47.7	51.8	49.3	50.6	37.3	43.6	38.0	46.6
Good internet quality at school	53.1	58.3	55.7	58.4	34.8	41.9	24.3	22.2
Some experience in online courses	13.8	25.1	16.9	19.0	10.7	12.1	2.6	7.2
Experience in online communities for teachers	53.1	58.2	51.6	44.4	63.4	63.8	34.8	75.0

NC= non-completers; C=completers

APPENDIX G

DETAILED RESULTS ON COURSE COMPLETION

Impacts on Professional Teachers in EU MSs

Table G1 *Multilevel model impact estimates on course completion among enrolled, EU-PTs*

	FULL SAMPLE		BS-TARGETED	
	(1)	(2)	(3)	(4)
Control mean	.324		.256	
ITT	.100	0.093	.117	.085
	(.028)	(.028)	(.039)	(.039)
Blocking variables	Y	Y	Y	Y
PT-specific controls	N	Y	N	Y
N observations	2,244	2,186	785	770
N individuals	943	917	414	408

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G2 *Course-specific impacts on course start among enrolled, EU-PTs*

	COURSE 1		COURSE 2		COURSE 3		COURSE 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITT	0.047	0.043	0.127***	0.141***	0.088*	0.089*	0.118**	0.108**
	(0.041)	(0.043)	(0.047)	(0.049)	(0.051)	(0.049)	(0.050)	(0.050)
Blocking variables	Y	Y	Y	Y	Y	Y	Y	Y
PT-specific controls	N	Y	N	Y	N	Y	N	Y
N	628	611	533	519	536	523	547	533

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G3 *Course-specific impacts on course completion among started, EU-PTs*

	Course 1		Course 2		Course 3		Course 4	
	(1)	(3)	(4)	(6)	(7)	(9)	(10)	(12)
ITT	0.190***	0.145***	-0.073	-0.074	-0.018	-0.010	-0.048	-0.053
	(0.050)	(0.051)	(0.069)	(0.071)	(0.067)	(0.064)	(0.067)	(0.076)

	Course 1		Course 2		Course 3		Course 4	
	(1)	(3)	(4)	(6)	(7)	(9)	(10)	(12)
Blocking variables	Y	Y	Y	Y	Y	Y	Y	Y
PT-specific controls	N	Y	N	Y	N	Y	N	Y
N	448	436	282	270	239	232	241	235

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G4 *Impacts on the number of courses started and on the likelihood of starting at least one course*

	NUMBER OF COURSES STARTED		STARTED AT LEAST ONE COURSE	
	(1)	(2)	(3)	(4)
ITT	0.016	0.026	0.013	0.016
	(0.094)	(0.091)	(0.033)	(0.033)
Blocking variables	Y	Y	Y	Y
PT-specific controls	N	Y	N	Y
N	822	797	822	797

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*
Portugal is excluded from this analysis

Table G5 *Impacts on the number of courses completed and on the likelihood of completing at least one course*

	NUMBER OF COURSES COMPLETED		COMPLETED AT LEAST ONE COURSE	
	(1)	(2)	(3)	(4)
ITT	0.301***	0.296***	0.145***	0.133***
	(0.112)	(0.113)	(0.041)	(0.041)
Blocking variables	Y	Y	Y	Y
PT-specific controls	N	Y	N	Y
N	822	797	822	797

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*
Portugal is excluded from this analysis

Impacts on Student Teachers in EU MSs

Table G6 *Multilevel model impact estimates on course completion among enrolled, EU-STs*

	FULL SAMPLE		BS-TARGETED	
	(1)	(2)	(3)	(4)
Control mean	.215		.199	
ITT	.013	.023	.015	.008
	(.022)	(.023)	(.031)	(.032)
Blocking variables	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y
N observations	2,565	2,423	859	929
N individuals	824	776	396	383

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G7 *Course-specific impacts on course start among enrolled, EU-STs*

	COURSE 1		COURSE 2		COURSE 3		COURSE 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITT	0.001	-0.030	0.101**	0.103**	0.075	0.077	0.089**	0.095**
	(0.054)	(0.052)	(0.044)	(0.047)	(0.048)	(0.047)	(0.037)	(0.043)
Blocking variables	Y	Y	Y	Y	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y	N	Y	N	Y
N	681	643	632	600	623	587	629	593

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G8 *Course-specific impacts on course completion among started, EU-STs*

	COURSE 1		COURSE 2		COURSE 3		COURSE 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITT	0.060	0.011	-0.094	-0.160*	-0.056	-0.102	-0.223**	-0.247***
	(0.061)	(0.059)	(0.084)	(0.094)	(0.087)	(0.101)	(0.084)	(0.085)
Blocking variables	Y	Y	Y	Y	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y	N	Y	N	Y

	COURSE 1		COURSE 2		COURSE 3		COURSE 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
N	423	398	235	225	178	169	157	148

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G9 *Impacts on the number of courses started and on the likelihood of starting at least one course, EU-STs*

	NUMBER OF COURSES STARTED		STARTED AT LEAST ONE COURSE	
	(1)	(2)	(3)	(4)
ITT	0.069	0.103	0.029	0.033
	(0.165)	(0.186)	(0.063)	(0.063)
Blocking variables	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y
N	781	734	838	788

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*
Portugal is excluded from this analysis

Table G10 *Impacts on the number of courses completed and on the likelihood of completing at least one course, EU-STs*

	NUMBER OF COURSES COMPLETED		COMPLETED AT LEAST ONE COURSE	
	(1)	(2)	(3)	(4)
ITT	0.065	0.032	0.070	0.050
	(0.142)	(0.138)	(0.045)	(0.043)
Blocking variables	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y
N	781	734	838	788

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*
Portugal is excluded from this analysis

Impacts on Professional Teachers in Turkey

Table G11 *Multilevel model impact estimates on course completion among enrolled, TR-PTs*

	FULL SAMPLE		BS-TARGETED	
	(1)	(2)	(3)	(4)
Control mean	.215		.184	
ITT	<-.001	-.002	<.001	-0.11
	(.024)	(.024)	(.032)	(.032)
Blocking variables	Y	Y	Y	Y
PT-specific controls	N	Y	N	Y
N observations	3,630	3,489	1,660	1,628
N individuals	1,022	983	550	542

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G12 *Course-specific impacts on course start among enrolled, TR-PTs*

	COURSE 1		COURSE 2		COURSE 3		COURSE 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITT	-0.024	-0.027	0.042	0.022	0.061	0.051	0.059	0.059
	(0.069)	(0.063)	(0.058)	(0.059)	(0.059)	(0.058)	(0.057)	(0.057)
Blocking variables	Y	Y	Y	Y	Y	Y	Y	Y
PT-specific controls	N	Y	N	Y	N	Y	N	Y
N	953	918	898	863	895	859	884	849

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G13 *Course-specific impacts on course completion among started, TR-PTs*

	COURSE 1		COURSE 2		COURSE 3		COURSE 4	
	(1)	(3)	(4)	(6)	(7)	(9)	(10)	(12)
ITT	-0.045	-0.078	0.091	0.144	-0.063	-0.100	0.032	0.037
	(0.086)	(0.087)	(0.092)	(0.103)	(0.089)	(0.112)	(0.089)	(0.107)
Blocking variables	Y	Y	Y	Y	Y	Y	Y	Y
PT-specific controls	N	Y	N	Y	N	Y	N	Y
N	576	553	244	230	182	172	165	154

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G14 *Impacts on the number of courses started and on the likelihood of starting at least one course, TR-PTs*

	NUMBER OF COURSES STARTED		STARTED AT LEAST ONE COURSE	
	(1)	(2)	(3)	(4)
ITT	0.086	-0.123	0.036	0.003
	(0.149)	(0.137)	(0.055)	(0.048)
Blocking variables	Y	Y	Y	Y
PT-specific controls	N	Y	N	Y
N	1,022	983	1,022	983

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G15 *Impacts on the number of courses completed and on the likelihood of completing at least one course, TR-PTs*

	NUMBER OF COURSES COMPLETED		COMPLETED AT LEAST ONE COURSE	
	(1)	(2)	(3)	(4)
ITT	0.014	0.011	-0.049	-0.056
	(0.193)	(0.186)	(0.064)	(0.062)
Blocking variables	Y	Y	Y	Y
PT-specific controls	N	Y	N	Y
N	1,022	983	1,022	983

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Impacts on Student Teachers in Turkey

Table G16 *Multilevel model impact estimates on course completion among enrolled, TR-STs*

	Full sample		BS-targeted	
	(1)	(2)	(3)	(4)
Control mean	.123		.118	
ITT	-.002	-.006	-.026	-.019
	(.017)	(.017)	(.020)	(.020)
Blocking variables	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y
N observations	3,341	3,265	1,589	1,574
N individuals	913	890	589	584

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G17 *Course-specific impacts on course start among enrolled, TR-STs*

	COURSE 1		COURSE 2		COURSE 3		COURSE 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITT	-0.011	-0.027	-0.010	0.022	-0.000	0.051	-0.001	0.059
	(0.049)	(0.063)	(0.036)	(0.059)	(0.033)	(0.058)	(0.026)	(0.057)
Blocking variables	Y	Y	Y	Y	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y	N	Y	N	Y
N	804	918	850	863	842	859	845	849

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G18 *Course-specific impacts on course completion among started, TR-STs*

	COURSE 1		COURSE 2		COURSE 3		COURSE 4	
	(1)	(3)	(4)	(6)	(7)	(9)	(10)	(12)
ITT	-0.013	-0.016	0.065	0.065	-0.109	-0.122	0.183*	-0.058
	(0.064)	(0.063)	(0.086)	(0.102)	(0.106)	(0.119)	(0.103)	(0.169)
Blocking variables	Y	Y	Y	Y	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y	N	Y	N	Y
N	399	392	158	154	113	110	77	74

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

Table G19 *Impacts on the number of courses started and on the likelihood of starting at least one course, TR-STs*

	NUMBER OF COURSES STARTED		STARTED AT LEAST ONE COURSE	
	(1)	(2)	(3)	(4)
ITT	-0.068	-0.023	-0.023	-0.015
	(0.111)	(0.094)	(0.032)	(0.030)
Blocking variables	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y
N	913	890	913	890

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

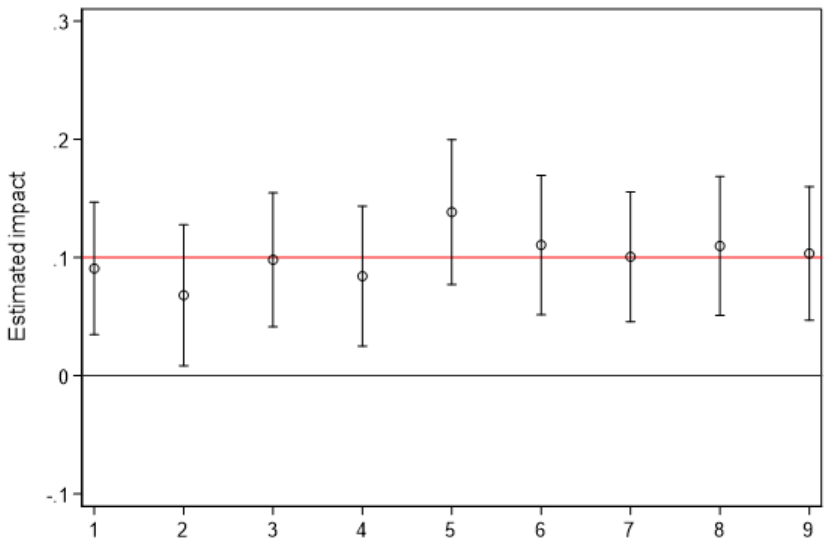
Table G20 *Impacts on the number of courses completed and on the likelihood of completing at least one course, TR-STs*

	NUMBER OF COURSES COMPLETED		COMPLETED AT LEAST ONE COURSE	
	(1)	(2)	(3)	(4)
ITT	-0.013	-0.030	-0.017	-0.026
	(0.094)	(0.090)	(0.041)	(0.041)
Blocking variables	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y
N	913	890	913	890

Standard errors in parentheses
** p<0.10, ** p<0.05, *** p<0.01*

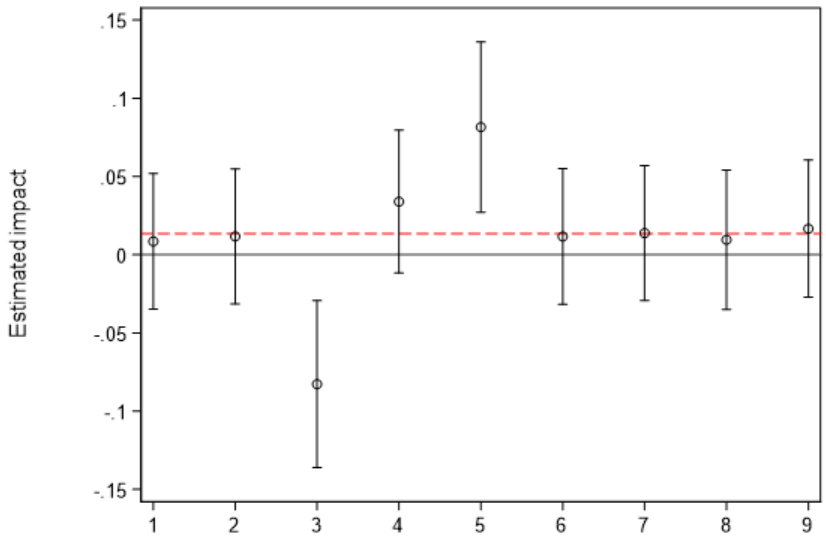
Leave-one-out validation tests

Figure G21 *Leave-one-out estimates of impacts for EU PTs*



Note: the red horizontal line is the estimated impact on the entire EU PT sample (Figure 5.4)

Figure G22 *Leave-one-out estimates of impacts for EU STs*



Note: the red horizontal line is the estimated impact on the entire EU ST sample (Figure 5.5)

APPENDIX H

FIGURES ABOUT PERSONALISED SUPPORT

Table H1 *Distribution of interventions activated during the whole project by participants and course*

		INTERVENTIONS					
#	TRIGGER	ACTION	COURSE 1	COURSE 2	COURSE 3	COURSE 4	TOTAL
1	User indicates low levels in at least 2 of the following indicators: - belief about effectiveness at online learning, - expectations of likelihood to take online courses in the future, - Education level lower than Masters	In course 1-3 the agent contacted the user with a personalised message that included an offer for a video call, template and guidance for succeeding in online learning using a learning plan. In course 4 the agent contacted the user with a general message offering support of any kind at any point during the course should the user like to reach out.	170	158	157	156	641
2	User indicates low competence in at least 2 of the following self-regulated learning components: - Goal setting - Task strategies/time management - Help seeking - Self-evaluation - Elaboration	In course 1-3 the agent contacted the user with a personalised message that included general guidance on the importance of SRLO, an example of how to develop the SRLO components, as well as an offer for a video call to discuss specific SRLO strategies. In course 4 the agent contacted the user with a general message offering support of any kind at any point during the course should the user like to reach out.	147	130	131	131	539
3	User has low experience of online courses or reports low ability and confidence with online learning technologies	In course 1-3 the agent contacted the user with a personalised message that included an offer for a video call to "walk-through" the course interface and tools or to identify a set of questions they have about the course. The agent would then provide answers to these questions. In course 4 the agent contacted the user with a general message offering support of any kind at any point during the course should the user like to reach out.	549	497	490	492	2028
4*	Weighted mix of low motivation, low English proficiency, subject of teaching, age, previous experience, internet access, gender, and level of education	Support agent contacted the user with a general message offering support of any kind at any point during the course should the user like to reach out.				274	274

		INTERVENTIONS					
#	TRIGGER	ACTION	COURSE 1	COURSE 2	COURSE 3	COURSE 4	TOTAL
5	User has not started 5 days after module launch.	Support agent contacted the user with a personalised message reminding of the importance to stay on track and encouraging the user to keep going as well as offering advice on how to plan their learning time and self-motivate.		1160	1227	274	2661
6	User has not submitted their work for the peer assessment activity 2 days prior to the deadline	Support agent contacted the user with a personalised message about the deadline, offering tips on how to complete the work and where to find support, including the possibility for a video call prior to the deadline, in order to answer any final questions.		225	111	117	453
7	A user has made two or more support requests via the contact form of the course platform within a period of 1 week and has never visited the course FAQ page.	Support agent contacted the user with a personalised message highlighting ways how users can find answers to their questions, including finding peer support but in particular highlighting the FAQ page and the support section of the forum.		5	7	0	12
8	User indicates dissatisfaction/confusion with the feedback provided by peers in the peer assessment activity	Support agent provided feedback to the work submitted.		5	0	3	8
9	User has visited less than 70% of module sections 1 week after module launch	Support agent sent a personal message offering support to help user benefit more from the course content, including a possibility to book a 1:1 session to discuss how to use the content.		36	103	106	245

Table H2 *Average per-person overall cost of personalised support per country and country weight*

COUNTRY	COST PER PARTICIPANT	COUNTRY WEIGHT
AT	186.58	5.08
EE	25.79	17.51
ES	12.34	6.64
GR	33.92	18.54
HU	14.34	15.42
LT	11.03	17.34
MT	31.62	2.41

COUNTRY	COST PER PARTICIPANT	COUNTRY WEIGHT
PT	26.06	12.43
SK	36.49	4.63

Note: calculations based on course 4. Country weight given by sample size in impact model

Weighted average per-person cost: 29.35249

See TeachUP report 'Implementing Personalised Support in Scalable Online Courses' for a detailed analysis on costs.

APPENDIX I

TALIS VS TEACHUP COMPARISON

Table I.1 *EU countries: TeachUP vs TALIS*

In bold significant and relevant differences between teach-up and Talis sample

CHARACTERISTICS	EU TEACHUP MEAN	EU TALIS MEAN	P
Female	0.850	0.693	0.000
Age below 30	0.060	0.056	0.229
Age between30-39	0.215	0.199	0.103
Age between 40-49	0.363	0.336	0.001
Age 50+	0.362	0.409	0.000
Master	0.604	0.603	0.941
Experiences as a teacher in total	20.24	18.61	0.029
Cpd (I version)	0.921	0.922	0.146
Cpd (II version)	0.921	0.853	0.000
Incentives	0.636	0.714	0.000
Hours worked per week	7.229	7.141	0.007
Observations	862	30017	30879

Note: Greece excluded from the sample because not part of the TALIS sample.

Table I.2 *Turkey: TeachUP vs TALIS*

CHARACTERISTICS	TR TEACHUP MEAN	TR TALIS MEAN	P
Female	0.663	0.558	0.000
Age below 30	0.206	0.259	0.849
Age between30-39	0.498	0.481	0.131
Age between 40-49	0.228	0.198	0.353
Age 50+	0.068	0.063	0.109
Master	0.104	0.070	0.003
Experiences as a teacher in total	12.534	10.878	0.436
Cpd (I version)	0.791	0.928	0.000
Cpd (II version)	0.791	0.905	0.000
Incentives	0.776	0.721	0.000
Hours worked per week	6.275	6.258	0.430
Observations	1022	3952	4974

Table I.3 *TALIS: EU countries vs Turkey*

CHARACTERISTICS	EU TALIS MEAN	TR TALIS MEAN	P
female	0.693	0.558	0.000
Age below 30	0.056	0.259	0.000
Age between30-39	0.199	0.481	0.000
Age between 40-49	0.336	0.198	0.000
Age 50+	0.409	0.063	0.000
Master	0.603	0.070	0.000
Experiences as a teacher in total	18.612	10.878	0.000
Cpd (I version)	0.922	0.928	0.124
Cpd (II version)	0.853	0.905	0.000
Incentives	0.714	0.721	0.000
Hours worked per week	7.141	6.258	0.000
Observations	30017	3952	33969

Overall very different confirming what we already saw using only teachup data

Table I.4 *Other TALIS questions non-directly comparable with the TeachUP Baseline Survey*

	EU	TR
WHEN YOU BEGAN WORK AT THIS SCHOOL, WERE THE FOLLOWING PROVISIONS PART OF YOUR INDUCTION?		
Online courses/seminars	58.6	26.1
Online activities (e.g. virtual communities)	34.6	24.2
WERE ANY OF THE TOPICS LISTED BELOW INCLUDED IN YOUR PROFESSIONAL DEVELOPMENT ACTIVITIES DURING THE LAST 12 MONTHS		
Student assessment practices	65.7	55.3
ICT skills for teaching	60.8	62.8
Teaching cross	48.0	51.2
Analysis and use of student assessments	47.0	41.6
THINKING OF THE PROFESSIONAL DEVELOPMENT ACTIVITY THAT HAD THE GREATEST POSITIVE IMPACT ON YOUR TEACHING DURING THE LAST 12 MONTHS, DID IT HAVE ANY OF THE FOLLOWING CHARACTERISTICS?		
It provided opportunities for active learning	76.6	80.9
It provided opportunities for collaborative learning	74.7	75.7
It provided opportunities to practise/apply new ideas and knowledge in my classroom	82.0	86.8
It focused on innovation in my teaching.	76.8	71.4

	EU	TR
FOR EACH OF THE AREAS LISTED BELOW, PLEASE INDICATE THE EXTENT TO WHICH YOU CURRENTLY NEED PROFESSIONAL DEVELOPMENT [HIGH +MODERATE LEVEL]		
ICT skills for teaching	33.4	60.9
Teaching cross curricular skills (e.g. creativity, critical thinking, problem solving)	28.7	56.3
HOW STRONGLY DO YOU AGREE OR DISAGREE THAT THE FOLLOWING PRESENT BARRIERS TO YOUR PARTICIPATION IN PROFESSIONAL DEVELOPMENT? [STRONGLY AGREE]		
Professional development is too expensive.	8.0	9.4
There is a lack of employer support.	14.2	10.7
Professional development conflicts with my work schedule.	13.8	17.8
I do not have time because of family responsibilities.	9.1	14.8
There is no relevant professional development offered.	13.7	13.4
There are no incentives for participating in	24.1	28.7

APPENDIX J

PEER AND EXPERT ASSESSMENT

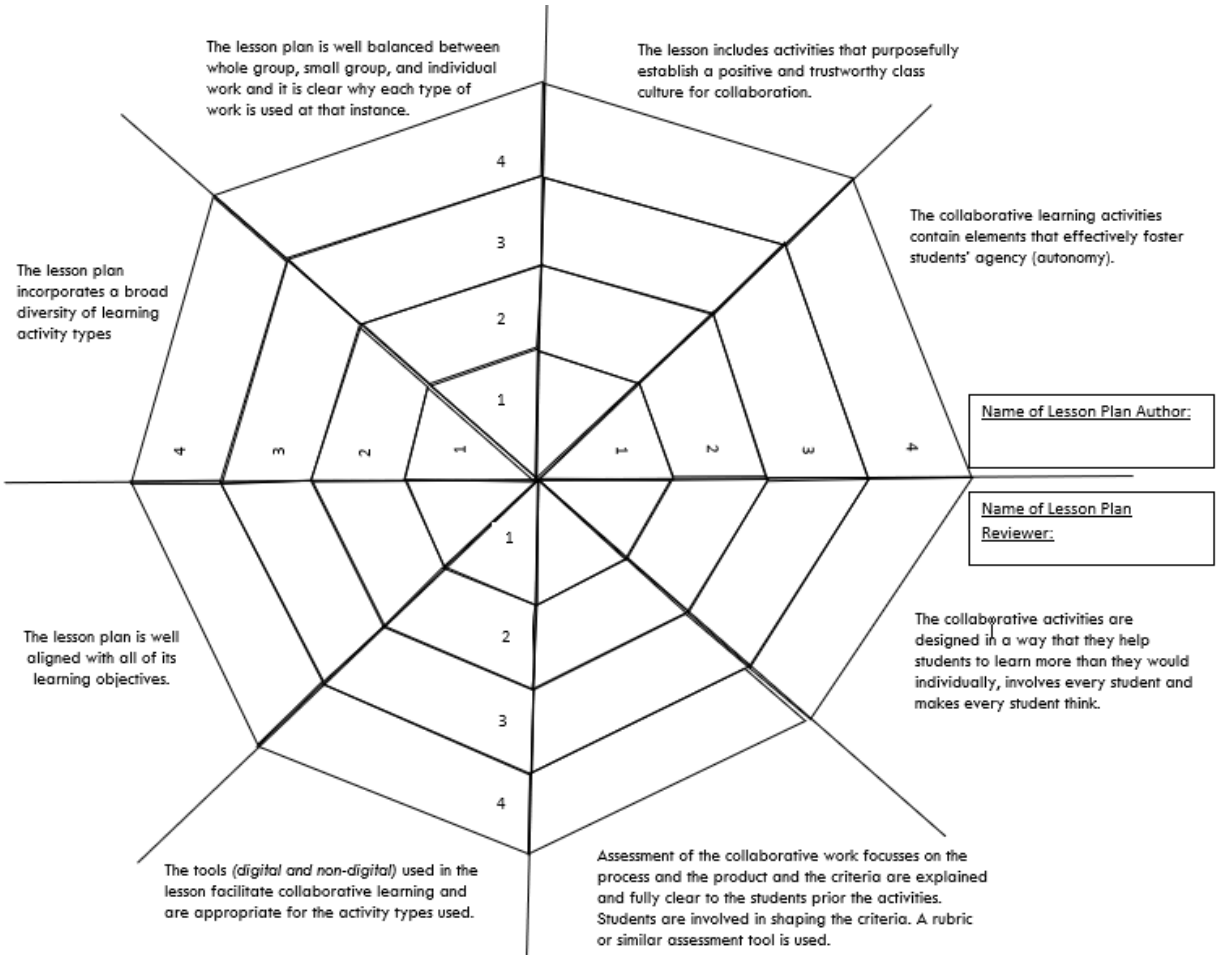


Fig. J.2 *Rubric areas and descriptors*

AREA	DESCRIPTORS			
	DESCRIPTOR LEVEL 4	DESCRIPTOR LEVEL 3	DESCRIPTOR LEVEL 2	DESCRIPTOR LEVEL 1
Classroom culture for collaboration	The lesson includes activities that purposefully establish a positive and trustworthy class culture for collaboration, for example through teambuilding activities and thoughtful division of students into groups.	The lesson includes a few activities that establish a positive and trustworthy class culture for collaboration.	The lesson includes no activities that establish a positive and trustworthy class culture for collaboration.	The lesson includes activities which will reduce trust and positive atmosphere in the class.

AREA	DESCRIPTORS			
	DESCRIPTOR LEVEL 4	DESCRIPTOR LEVEL 3	DESCRIPTOR LEVEL 2	DESCRIPTOR LEVEL 1
Fostering student's agency	The collaborative learning activities contain elements that effectively foster students' agency (autonomy).	The collaborative learning activities contain some elements designed to foster students' agency, but these do not fully support more student autonomy.	The collaborative learning activities contain very few elements designed to foster students' agency (autonomy).	The collaborative learning activities do not contain any elements designed to foster students' agency (autonomy).
Effective elements of collaborative learning	The collaborative activities are designed in a way that helps students to learn more than they would individually, involves every student and makes every student think.	The collaborative activities are designed in a way that helps students to learn somewhat more than they would individually, it involves most of the students and evokes most students to think.	The collaborative activities are designed in a way that students learn slightly more than they would individually, it involves only part of the students and does not really evoke students to think.	The collaborative activities are designed in a way that students learn the same as they would if they worked individually, it involves only part of the students and does not evoke students to think.
Assessment of collaborative learning	Assessment of the collaborative work focusses on the process and the product and the criteria are explained and fully clear to the students prior the activities. Students are involved in shaping the criteria. A rubric or similar assessment tool is used.	Assessment of the collaborative work focusses on the process and the product. A rubric or similar assessment tool is used.	Assessment of the collaborative work focusses only on the product. A rubric or similar assessment tool is used.	Assessment of the collaborative work focusses only on the product and there is no information provided about the criteria used for the assessment.
Tools for collaborative learning (digital and non-digital)	The tools used in the lesson facilitate collaborative learning and are appropriate for the activity types used. For example, the eTwinning portal is used to get students to collaborate in a project-based context with students from another class.	The tools used in the lesson facilitate collaborative learning but they are not always fully appropriate for the activity types used. For example, a Google document is used to collect shared notes but activities offer little opportunity for students to actually take notes. Or a common sheet is distributed to the group to fill in but there is no guidance on how to work on the task collaboratively.	The tools used in the lesson do not facilitate collaborative learning and they are not fully appropriate for the activity types used.	There are hardly any, or no digital or non-digital tools used in the lesson.

AREA	DESCRIPTORS			
	DESCRIPTOR LEVEL 4	DESCRIPTOR LEVEL 3	DESCRIPTOR LEVEL 2	DESCRIPTOR LEVEL 1
Alignment to Learning Objectives	The lesson plan is well aligned with all of its learning objectives.	The lesson plan is partially aligned with its learning objectives, the majority of learning objectives are reflected in the activities.	The lesson plan is mostly unaligned with its learning objectives, there are only one or two learning objectives which are reflected in the activities.	The lesson plan is not at all aligned with its learning objectives.
Diversity of activities	The lesson plan incorporates a broad diversity of learning activity types.	The lesson plan incorporates some diversity of learning activity types.	The lesson plan incorporates little diversity of learning activity types.	The lesson plan does not incorporate diversity of learning activity types, there is mainly one type of activity.
Balance between individual and group work	The lesson plan is well balanced between whole group, small group, and individual work and it is clear why each type of work is used at that instance.	The lesson plan is mostly balanced between whole group, small group, and individual work. It is not always clear why each type of work is used at that instance.	The lesson plan is not very balanced between whole group, small group, and individual work. It is not clear why each type of work is used at that instance.	The lesson plan includes only one type of work throughout the lesson and it is not clear why this type of work is used.

Table J.1 *Number of subjects receiving assessment by peers by country and teacher status*

COUNTRY	PROFESSIONAL TEACHERS				STUDENT TEACHERS				
	# PEERS				# PEERS				
	1	2	3	TOTAL	1	2	3	TOTAL	TOTAL
AT		1	1	2	1			1	3
EE		5	1	6					6
ES			2	2		3	8	11	13
GR		1	3	4		1	4	5	9
HU		2	3	5	1		1	2	7
LT		4	9	13					13
MT			1	1		1		1	2
PT	1		6	7					7
SK						1		1	1
TR		6	20	26	2	4	13	19	45
Total	**1**	**19**	**46**	**66**	**4**	**10**	**26**	**40**	**106**
%	**1.5**	**28.8**	**69.7**	**100.0**	**10.0**	**25.0**	**65.0**	**100.0**	

Table J.2 *Average score by categories*

	PEERS ASSESSMENT (AVERAGE)				EXPERTS ASSESSMENT			
	MEAN	SE(MEAN)	P50	N	MEAN	SE(MEAN)	P50	N
1. Effective use of Formative Assessment Techniques	3.69	0.06	4	82	3.37	0.07	4	104
2. Inbuilt Flexibility	3.59	0.07	4	83	3.33	0.08	4	105
3. Feedback	3.55	0.07	4	81	3.23	0.08	3	103
4. Goal Setting	3.50	0.09	4	81	2.87	0.10	3	105
5. Balance between individual and group work	3.62	0.07	4	81	3.27	0.08	3	104
6. Tools	3.65	0.08	4	80	3.24	0.08	3	105
7. Diversity of activities	3.61	0.06	4	83	3.31	0.08	4	105
8. Alignment to Learning Objectives	3.58	0.07	4	82	3.18	0.09	3	105
Overall score	3.59	0.06	3.8	83	3.22	0.07	3.5	105
Overall score PTs only	3.65	0.07	3.8	55	3.36	0.08	3.6	65
Overall score STs only	3.48	0.12	3.6	28	3.00	0.12	3.1	40

Note: The overall score is the mean of the scores of the 8 categories. 41 peers (almost all from Turkey) did not score any of the 8 categories. For 23 teachers (of which 21 from TR) they were no other peers so it's impossible to compute an overall score. Moreover, 1 expert (from TR) did not score any of the 8 categories. In red the minimum score, in blue the maximum score.

PROJECT COORDINATOR

European Schoolnet

www.europeanschoolnet.org
Belgium

RESEARCH ORGANIZATION

FONDAZIONE BRUNO KESSLER
IRVAPP
RESEARCH INSTITUTE FOR THE EVALUATION OF PUBLIC POLICIES

irvapp.fbk.eu
Italy

PARTNERS

BMB Bundesministerium für Bildung
irvapp.fbk.eu
Austria

HITSA Information Technology Foundation for Education
hitsa.ee
Estonia

CTI "DIOPHANTUS"
cti.gr
Greece

EDUCATIONAL AUTHORITY
oktatas.hu
Hungary

VYTAUTO DIDŽIOJO UNIVERSITETAS MCMXXII
leu.lt
Lithuania

NACIONALINĖ ŠVIETIMO AGENTŪRA
nsa.smm.lt
Lithuania

MINISTRY FOR EDUCATION AND EMPLOYMENT
education.gov.mt
Malta

CFAE centro-oeste
cfaecentro-oeste.pt
Portugal

REPÚBLICA PORTUGUESA | EDUCAÇÃO
direção-geral educação
dge.mec.pt
Portugal

Universidade do Minho
uminho.pt
Portugal

mpc METODICKO-PEDAGOGICKÉ CENTRUM
mpc-edu.sk
Slovakia

UNIVERSITAS SOLIENSIS
uniza.sk
Slovakia

GOBIERNO DE ESPAÑA | MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL
intef INSTITUTO NACIONAL DE TECNOLOGÍAS EDUCATIVAS Y DE FORMACIÓN DEL PROFESORADO
educalab.es
Spain

yegitek.meb.gov.tr
Turkey

gtc SCOTLAND
gtcs.org.uk
United Kingdom

TEACHUP

www teachup.eun.org | #TeachUP_eu |
groups/TeachUP.experimentation | @ teachup@eun.org

Erasmus+

The TeachUP project is a European Policy Experimentation co-funded by the European Commission via the Erasmus+ programme. This brochure reflects the views only of the authors and it does not represent the opinion of the European Commission, and the European Commission is not responsible or liable.