Econometric Analysis

Deliverable 2.3

Davide Azzolini, Sonia Marzadro, Enrico Rettore

FBK-IRVAPP

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

Content

1 INTRODUCTION

This report illustrates the data that were collected and the statistical analyses that were performed to carry out the TeachUP experimental evaluation. All the activities described took place between October 2017 and the end of October 2019.

Section 1 provides a detailed description of the entire data collection plan and illustrates the questionnaires. Section 2 describes the outcome variables used in the evaluation. Section 3 describes the statistical checks that were performed to assess the integrity of the randomisation, by looking both at group equivalence at baseline as well as at overall and differential attrition in the FollowUp survey. Finally, section 4, illustrates the econometric approach used to gauge the impact of personalised support on course participation and completion and SRLO. Section 4 also describes all the statistical analyses that were performed to test the robustness of the impact estimates.

2 DATA COLLECTION

Different sources of data were collected at different critical points of the experiment (Figure 1).

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

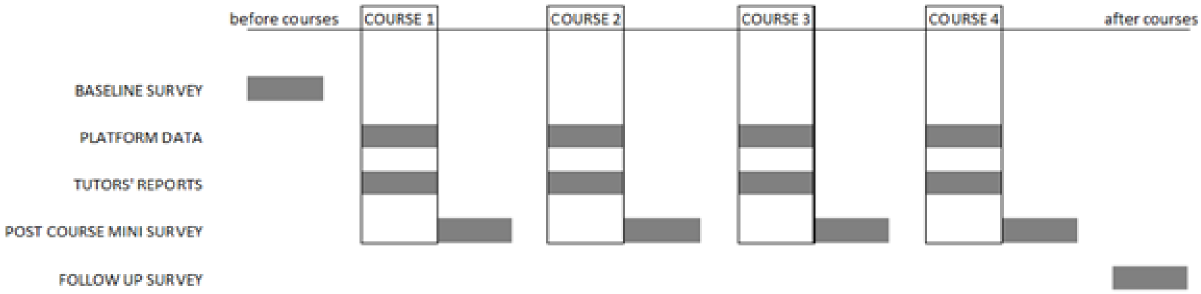


Figure 1 Data collection plan

Baseline Survey

A **Baseline Survey** (hereafter also BS) was administered online before the courses started and served as registration for 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). 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. 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 (http://teachup.eun.org/). 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.

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

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 still be conducted by adjusting the impact estimates through multiple regression models. The results are shown in section 7.

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. 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**

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

support agents. These reports were produced by each of them at the end of every intervention following a common grid.

3 MEASURES

The analysis included two primary outcomes (i.e., course participation and self- regulating learning online) and three secondary outcomes (such as attitudes towards online learning, teaching beliefs and practices and digital competences).

3.1 COURSE PARTICIPATION

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

1. 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.
2. 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.
3. 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.

Three different probabilities 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.**

3.2 SELF-REGULATED LEARNING ONLINE (SRLO)

The second set of outcomes concerns teachers' **self-regulated learning online** (SRLO). The 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, a general consensus exists on the importance of SRL development as an essential component for learning (Boekaerts 1999; Zimmerman 2001).

Measuring SRLO remains a very work-in-progress area in research (Panadero et al. 2016; Lee & Recker 2017; Molenaar et al. 2019). 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

General SRLO was measured both in the BS and in the FUS using the six strategy subscales proposed by Kizilcec et al. (2017). 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): 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 factor analysis (Table 1). 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.

The full set of items included in the questionnaire was reduced to three factors via a principal component analysis. 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. 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 α of at least 0.90.

Table 1 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

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

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
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

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

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

Course-specific SRLO

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 factor 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. Additional information on those who did not complete courses was not collected.

3.3 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 (see tables 2, 3, and 4)

Table 2 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

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

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 3 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
5	I will definitely take part in some online courses in the future	0.862
		Alpha 0.93

Table 4 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

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

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 0.84	Alpha 0.81

4 INTEGRITY CHECKS

We carried out two types of checks to assess how the randomisation protocol was actually implemented. First, by performing a series of "balancing tests" of the two groups we estimated the statistical significance of differences between the treatment and control group in terms of number PTs and STs.

The second set of checks addressed the issue of attrition, or non-response rate. In our case, this occurred when participants started the course and took the first survey (the BS) then, at some point, dropped out and did not answer the follow-up surveys (the mini-survey or the FUS). We checked whether attrition compromised the integrity of the evaluation design by comparing the attrition levels by group and then running equivalence analyses with the subgroup of teachers who completed the FUS.

4.1 GROUP EQUIVALENCE

In order to estimate the statistical significance of differences between the two groups in terms of number of student and professional teachers, we regressED the average number of teachers enrolled in TeachUP by school/ITE organisation on a dummy variable indicating the treatment status. The OLS models also include dummies for the blocking variables used in the randomisation procedure, i.e. the geographical strata used. Table 5 shows the results of this analysis. It reports separately for PTs and STs: (i) the mean size of schools/ITE organisations in treatment and control group; (ii) the β-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.

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

Overall, the differences are small and not statistically significant (Table 5). Hence, treated and control schools/ITE organisations have not only a similar profile in terms of geographical distribution within the country (achieved by design) but also in terms of school/ITE organisation size and teachers' average propensity for participating in TeachUP. These conclusions apply also to within-country analyses.

Table 5 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 TeachUP teachers (mean)	4.09	4.03	-0.195	0.777

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

The following table (Table 6) contains 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 6 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	

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

	(0.073)		(0.089)	
40-49 years old	0.044		0.029	
	(0.050)		(0.059)	
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)
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	

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

	(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	
	(.)		(.)	
_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

4.2 ATTRITION (SURVEYS)

We checked whether attrition compromised the integrity of the evaluation design by comparing the attrition levels by group and then running equivalence analyses with the subgroup of teachers who completed the FUS.

With regards to the mini survey, there were actually two questionnaires depending on whether or not the participant had completed the course. Among course completers, response rates were very high (80-94%) and well balanced between treated and controls. On the contrary, response rates among course non-completers were below 11%, making the results de facto worthless (Table 7).

Table 7 Response rates in the mini post course surveys

	course 1		course 2		course 3		course 4	
	controls	treate d	controls	treated	controls	treated	controls	treated

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

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%

With regards to the FUS, we registered low response rates (18.8%) (Table 8). We used the Baseline Survey to compare respondents to non-respondents and found some modest differences. Respondents tend to be more experienced in online training and, in general, more involved in training; have better views on online courses; teach mathematics/science rather than humanities. In measuring the impact of personalized support, we accounted for these differences.

Table 8 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**
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

5 IMPACT ESTIMATIONS

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

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

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 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. Hence two main kinds of estimation models were run:

- Overall estimates on completion rates among enrolled;
- By-course estimates
- Robustness checks

5.2 OVERALL IMPACT ESTIMATES

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 and did not complete it.

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.

To improve the statistical precision of the estimates, all models included dummies for the randomization strata. Portugal data for course 1 are not considered in this analysis because treatment was not delivered.

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 in and the data were analysed by means of multilevel linear regression models in which observations were "nested" within individuals. This modelling also allowed us to adjust standard errors for clustering.

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 was 42% (Table 9). Hence, **personalised support boosted completion rate by 10 percentage points among the enrolled**. This result is robust to the inclusion of a long list of baseline characteristics.

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 (Table 10, 11 and 12).

Table 9 Multilevel model impact estimates on course completion among enrolled, EU-PTs

	Full sample		BS-targeted	
	(1)	(2)	(3)	(4)

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

	(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 10 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 11 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 12 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)

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

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

5.3 BY-COURSE IMPACT ANALYSIS

Impact analysis by course looked at two distinct outcomes: the probability of starting a course among enrolled and the probability of completing a course among those who started one.

The analyses were replicated, as above, with models including only randomization blocking variables or also baseline characteristics in addition to them.

The analytical model used was a linear probability model.

The results are shown in Tables 13-20.

Table 13 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 (0.041)	0.043 (0.043)	0.127*** (0.047)	0.141*** (0.049)	0.088* (0.051)	0.089* (0.049)	0.118** (0.050)	0.108** (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 14 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*** (0.050)	0.145*** (0.051)	-0.073 (0.069)	-0.074 (0.071)	-0.018 (0.067)	-0.010 (0.064)	-0.048 (0.067)	-0.053 (0.076)
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 15 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 16 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
N	423	398	235	225	178	169	157	148

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

Table 17 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 18 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

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

	(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 19 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 20 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

5.4 ROBUSTNESS CHECKS

To make sure the results were robust, we replicated – as shown above - 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 conclusions (Tables 21-28):

- Using the probability of completing at least one of the four courses: among EU PTs, treated teachers had +14.5 percentage points higher probability of completing at least one course (.48 vs .33).
- 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);

Table 21 Impacts on the number of courses started and on the likelihood of starting at least one course, EU-PTs

	Number of courses started		Started at least one course	
	(1)	(2)	(3)	(4)
ITT	0.016 (0.094)	0.026 (0.091)	0.013 (0.033)	0.016 (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 22 Impacts on the number of courses completed and on the likelihood of completing at least one course, EU-PTs

	Number of courses completed		Completed at least one course	
	(1)	(2)	(3)	(4)
ITT	0.301*** (0.112)	0.296*** (0.113)	0.145*** (0.041)	0.133*** (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

Table 23 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 (0.165)	0.103 (0.186)	0.029 (0.063)	0.033 (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

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

Portugal is excluded from this analysis

Table 24 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

Table 25 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 26 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

Table 27 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 28 Impacts on the number of courses completed and on the likelihood of completing at least one course, TR-STs

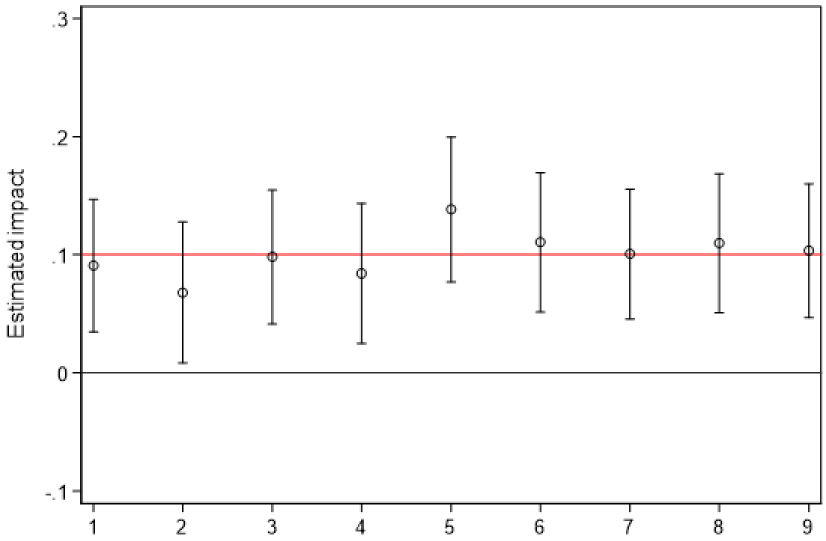
	Number of courses completed		Completed at least one course	
	(1)	(2)	(3)	(4)
ITT	-0.013	-0.030	-0.017	-0.026
	(0.094)	(0.090)	(0.041)	(0.041)
Blocking variables	Y	Y	Y	Y
ST-specific controls	N	Y	N	Y
N	913	890	913	890

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

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, Figures 2 and 3), 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. 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.

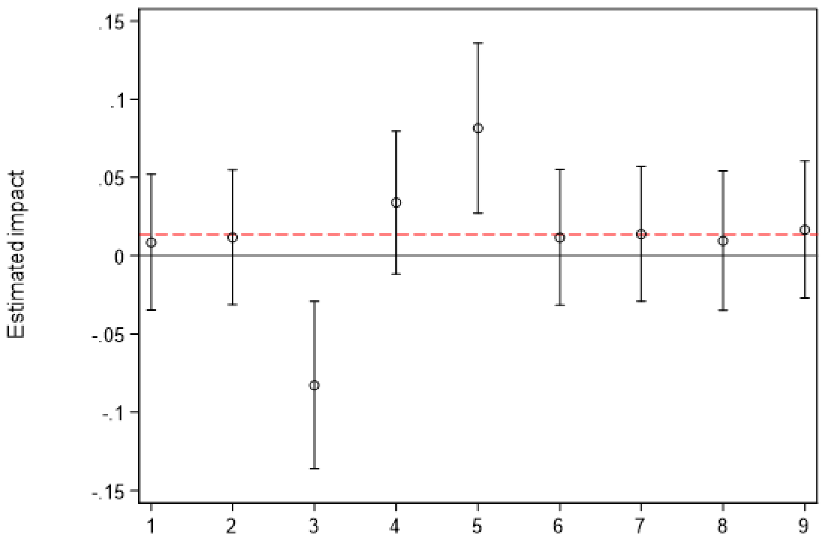
A further analysis consisted in replicating the estimation models on the sub-sample of targeted teachers, i.e. treated and control teachers that were identified as in need of support based on their profile measured at baseline. The results on these sub-samples reveal that, for PTs in EU MSs, the impact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

Figure 2 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 3 Leave-one-out estimates of impacts for EU STs



Note: the red horizontal line is the estimated impact on the entire EU ST sample

REFERENCES

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

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

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.

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.

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.

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

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

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

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

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

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

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

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.

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

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.

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