MENTEP
Evaluation Report

Results of the field trials: The impact of the technology-enhanced teaching self-assessment tool (TET-SAT)

May 2018

MENTEP
MENTORING TECHNOLOGY-ENHANCED PEDAGOGY

European Schoolnet

Erasmus+

Authors:
Giovanni Abbiati, FBK - IRVAPP
Davide Azzolini, FBK - IRVAPP
Daniela Piazzalunga, FBK - IRVAPP
Enrico Rettore, FBK - IRVAPP
Antonio Schizzerotto, FBK - IRVAPP

Editors:
Anja Balanskat, European Schoolnet
Roger Blamire, European Schoolnet
Katja Engelhardt, European Schoolnet
Patricia Wastiau, European Schoolnet

Please cite this publication as:
Abbiati, G.; Azzolini, D.; Piazzalunga, D.; Rettore, E. and Schizzerotto, A. (2018).
MENTEP Evaluation Report, Results of the field trials: The impact of the technology- enhanced self-assessment tool (TET-SAT) - European Schoolnet, FBK- IRVAPP, Brussels.

Contents

1. Introduction

This deliverable provides an extensive and technical account of the counterfactual impact evaluation carried out in the framework of the MENTEP policy experimentation.

The report is structured as follows. Section 2 introduces the MENTEP evaluation question, describes the main features of the tested intervention and the outcome measures employed in the study. In addition, section 2 illustrates and discusses the hypothesised links between the different components of the intervention and the outcomes. Section 3 is entirely dedicated to the description of the implementation of the experimental protocol with regard to the sampling and recruitment operations, data collection and randomisation. Section 4 describes the results of the statistical tests of the integrity of the experimental design, i.e. group equivalence and attrition. Section 5 is a statistical description of the teachers involved in the experimentation as well as an analysis of the actual take-up of the intervention and the reaction of teachers towards it. Section 6 contains the results of the experiment. Section 7 concludes by discussing the results and drawing policy implications arising from them.

MENTEP National Workshop, Lithuania

2. The evaluation question and the theory of change of the intervention

Research indicates that most teachers in Europe have positive opinions about the impact of ICT on students and learn about ICT in their own time, but many teachers lack digital competence for teaching and have limited use of ICT in their classrooms (European Commission, 2013).

The MENTEP policy experimentation builds upon a rich literature in pedagogical and education science literature (Andrade, et al. 2009; Hattie, J.A.C. 2003; Hattie, 2008; Tondeur, J., et al. 2013) investigating whether and how the practice of self-assessment can enhance that digital competence. Exploring this approach further, in the MENTEP project an innovative self-assessment tool (TET-SAT – Technology-Enhanced Teaching Self-Assessment Tool) to help lower-secondary education teachers improve their technology-enhanced teaching competences was co-designed and tested.

The MENTEP evaluation starts with a simple core question: *"Does the Technology-Enhanced Teaching Self-Assessment Tool (TET-SAT) have an impact on teachers' Technology-Enhanced Teaching (TET) competences?"* As with all impact evaluation questions, this question identifies two necessary components: an outcome (TET competences) and a treatment (the TET-SAT). In the next two sections, the two components are illustrated, and in section 2.3 the links between the two are discussed.

2.1 The outcome

TET competence can be defined as "proficiency in using ICT in teaching, applying pedagogical and didactic judgment, and being aware of its implications for learning" (Laurillard, D. 2012). In other words, it is a multi-dimensional competence which includes a variety of different types of competences, behaviours and attitudes.

Within the MENTEP policy experimentation, the following indicators are used to measure teacher TET competences:

- Self-reported TET competences;
- Views on ICT in teaching and learning;
- Use of ICT in the classroom;
- Use of ICT devices and software;
- Collaboration using ICT;
- Participation in ICT-related training.

Data for these indicators is collected from teachers taking part in the MENTEP project via the online Benchmark and Follow-up Surveys which use internationally validated questions and scales. More details on the measurement approach employed are provided in section 3.2. Because the Follow-up Survey was run shortly after the implementation of the experiment, only the first two indicators are taken into account in estimating the impact of the TET-SAT (see also section 2.3 on this). Information on the other indicators, collected in the Benchmark Survey, was used to elaborate a descriptive profile of the teachers involved in the various stages of the experiment.

2.2 The treatment

The TET-SAT was co-designed by scientific and national experts from the countries participating in MENTEP. It was inspired by existing tools in use (e.g. Teacher Mentor, Norway) and EU tools and frameworks (e.g. UNESCO, DIGCOMP) and covered four areas: digital pedagogy, digital content use and production, digital communication and collaboration, digital citizenship (Figure 1).

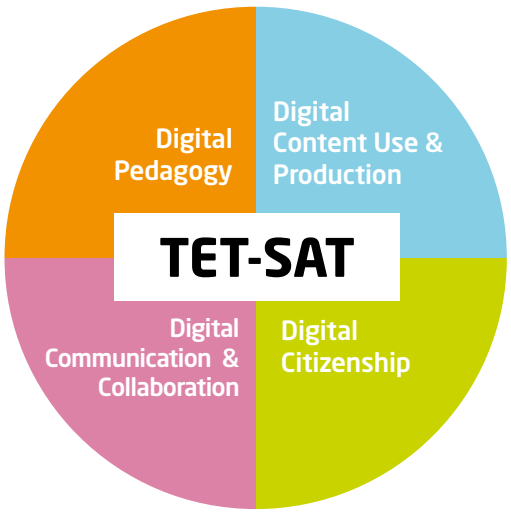


Figure 1 The four areas of the TET-SAT

Within each of the four areas a rich set of descriptive items is presented to users with the aim of covering the relevant sub-dimensions. Each item comprises five statements describing related practical pedagogical situations. Teachers using the TET-SAT are asked to read the five statements, reflect on their actual teaching practice and select the one that most closely matches their own pedagogical behaviour.

In order to associate a competence score to teachers, the five scenarios are ranked from the lowest (1) to the highest competence level (5). However, to stimulate self-reflection and discourage a response set and social desirability bias, the scenarios' ranking is not disclosed - i.e. the five scenarios are presented in random order. Figure 2 shows an illustrative example concerning the design of learning activities. More details on the topics covered within the four areas are provided in D.2.2 Blue- print (in English of TET-SAT).

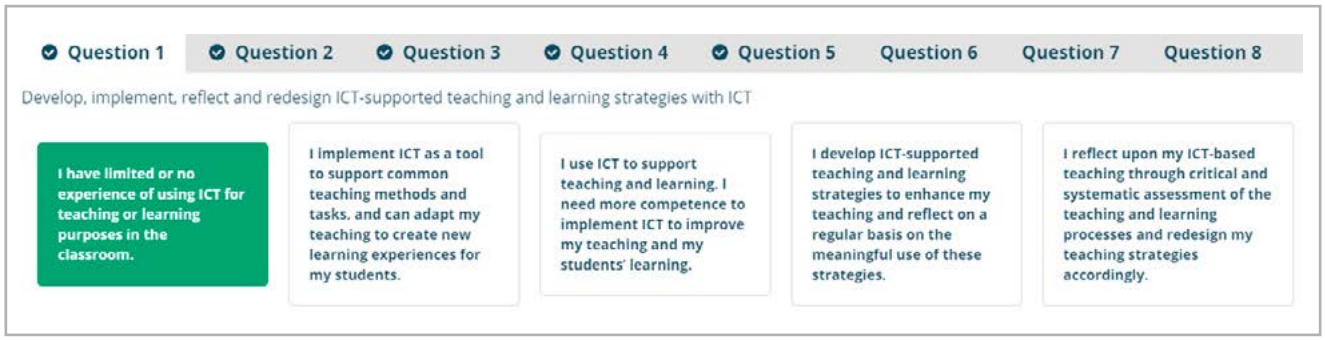


Figure 2 Example of a TET-SAT item

Beyond triggering teachers' self-reflection, the TET-SAT provides teachers with two additional inputs: a feedback score and free access to specially created eco-systems of training resources in their countries and at European level (e.g. European Schoolnet Academy courses). The feedback score is computed as a sum of the responses chosen to the different items in the TET-SAT. Alongside their overall score and the score in each of the four areas, users can also compare themselves with others from their own country as well as from other participating countries. Hence, teachers not only receive a personal score but also learn how they perform relative to peers. Both the score and the link to national and European eco-systems of training resources are revealed on completion of the TET-SAT.

2.3 Hypothesised channels linking the treatment to the outcome

In principle, the three TET-SAT components just described (self-reflection, feedback score and access to training resources) could have an impact on teachers' TET competences. Self-reflection could trigger teachers' critical appraisal of their pedagogical practice and increase their awareness of the potential benefits offered by information and communication technology (ICT) to teaching and learning. The feedback score could also convey important information to teachers. Here, it is important to consider that the information provided by the score could vary depending on the prior beliefs of teachers concerning their own competences. For some, the feedback score could simply confirm what they already know about their

own competences: this is the case, for example, of teachers thinking that they possess a good level of competence and receiving a 'good score'. Yet, for some other teachers the information revealed by the feedback score could be more than just a confirmation of what they already know. There may be teachers who under- or over-estimate their competence levels. For example, some may think that they are very well able to use digital resources in teaching but then find out - thanks to the feedback score - that they in fact have still substantial progress to make. For those teachers, the feedback score could represent an 'information shock' (Gonzalez, 2017)[1]. Third, free access to specific training resources could increase teachers' knowledge of and competence in the adoption and use of new technologies.

The timing dimension has to be considered as well when discussing the possible impacts of the TET-SAT (Figure 3). It can be hypothesised that the first two TET-SAT components have an instant effect on those TET competences related to attitudes and beliefs. The third component (giving teachers free access to training resources) can hardly exert an immediate effect, because of the time needed from participating in training and applying in the classroom the knowledge and skills acquired. On the other hand it could have some impact in the medium term because of the new knowledge acquisition through the courses. Given the fact that research on this specific topic is still in its infancy, it is difficult to form more elaborated and precise research hypotheses. One of the main contributions of the current study could be precisely that of providing an experimental empirical basis on which future studies could build to elaborate more solid research hypotheses. Considering that a) teachers only had a few months to make use of the tool and b) the short time from exposure to the treatment and the Follow-up Survey, the evaluation focuses on the two outcomes that can realistically be changed in the short term (i.e. self-reported TET competences and views on ICT in teaching and learning) leaving the other four dimensions to be studied in the future.

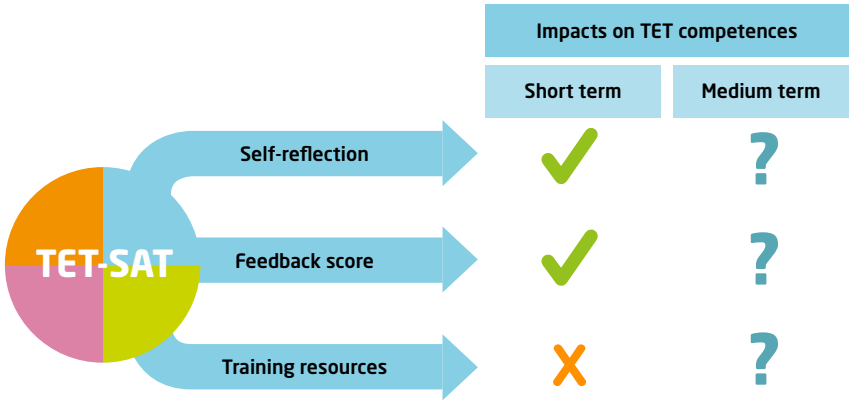


Figure 3 The expected channels linking the TET-SAT to TET competences

1 More on the feedback score can be found in sections 5.3 and section 6.3.

3. The experimental design

In what follows, we briefly summarise the main points of the MENTEP evaluation design and the approach employed for the estimation of the effects of the TET-SAT[2].

The evaluation design uses a randomised controlled trial protocol. Once a group of teachers from a list of invited schools has agreed to participate in the project, a random subset of them is encouraged to use the TET-SAT through a series of encouragement e-mails. The effectiveness of the TET-SAT is measured by comparing the group of teachers encouraged to use the TET-SAT and the group of non-encouraged ones on a given set of outcomes collected after the implementation of the intervention. The random selection of the units of analysis (in our case the teachers) eliminates the "selection bias" affecting most observational research designs and enables credible causal estimates of the effectiveness of a policy or other types of intervention. The underlying condition is that the random selection generates two (or more) groups that are, on average, statistically comparable. This means that the group of teachers not exposed to the intervention (i.e. does not receive the encouragement letters) can be considered as a good approximation of what would have happened to the treatment group of teachers if they had not received the encouragement letters (this is the so-called counterfactual situation). To claim that the difference in outcomes between the treatment group and its approximated counterfactual is a credible estimation of the causal effect of an intervention, some conditions concerning the 'integrity' of the experimental design must be empirically verified (section 4).

The section is organised as follows. Subsection 3.1 explains the sampling procedure. The ambition of the MENTEP research design to combine a strong internal validity given by the counterfactual approach (i.e.: the extent to which a difference in outcomes observed between groups can be genuinely attributed to the treatment, that is the TET-SAT) with external validity at the country level

2 The contents of this section have already been extensively described in the previous technical deliverables (3.1 Technical Report on the sampling; 3.2 Technical Report on the experimentation protocol; 3.3 Technical Report on the econometric analysis; 4.2 Data collection and management protocol).

(i.e. whether the findings are likely to be true of all eligible schools in partner countries, even if they did not take part in the project). Subsection 3.2 describes the data collection plan, which was entirely based on ad hoc surveys. Administrative and pre-existing data sources were in fact used only during the sampling phase. Finally, subsection 3.3 deals with the randomisation process. The analyses on the integrity of the experiment are in section 4.

3.1 Sampling

The sampling of the schools to be invited to take part in the project was implemented with the aim of providing MENTEP with external validity, i.e. making it possible to generalise the results of the analyses to the reference population of schools. In the context of the MENTEP project, that is all eligible publicly funded ISCED-2 schools in the participating countries[3].

The reason for the inclusion of eligibility criteria in the choice of the sample is twofold. First, teachers should not face high barriers to the use of ICT in their teaching; second, the sample of schools must be comparable across countries. Eligible schools had to a) have a student/PC(tablet) ratio no greater than 5:1; b) have internet access in the classrooms of at least 10mbps (ideally); c) have a good degree of communication facilities (i.e. virtual learning environments or a learning platform, or, at least, professional email addresses used by teachers). Some countries, as detailed in deliverable 3.1 Technical Report on the sampling, needed ad hoc adjustments given the lack of data to assess their eligibility. While in most countries these criteria were easily met by all the schools, in Italy, Lithuania, Spain and Greece only a fraction of all the schools could be included (60%, 59%, 32% and 17% respectively).

Once the complete list of eligible schools was defined, a stratified random sample was drawn (the strata defined generally on a geographical basis). Each school in a given stratum has a probability of being selected proportional to its size in terms of students. The quota of schools drawn in each stratum depends on the stratum share of the ISCED-2 student population. In each country, the starting aim of this procedure was to reach 50 participating schools.

An oversampling list was provided to National Coordinators (NCs), in order to deal with school refusals.

3 A detailed account of the sampling procedure and results can be found in the deliverable 3.1 Technical Report on the sampling.

Table 1 summarises the output of the sampling process.

Country	Target sample	Invitations	Refusals	MENTEP Schools
Cyprus	50	69	35	**34**
Czech Republic	50	165	113	**52**
Estonia	50	73	41	**32**
Finland	49	103	51	**52**
France*	-	-	-	**-**
Greece	50	100	50	**50**
Italy	50	101	51	**50**
Lithuania	50	91	41	**50**
Portugal	50	85	35	**50**
Slovenia	50	59	9	**50**
Spain	49	201	152	**49**
Total	**498**	**1047**	**578**	**469**

Note: In France, the sampling process was carried out under the supervision of the ministry of education. Teachers, and not schools, were directly contacted for participation from a sample carried out by the DEPP (Evaluation, Foresight and Performance Department of the ministry). The DEPP sample was representative of the French teaching population, based on the MENTEP criteria.

Table 1. Number of target sample schools, invited schools, schools refusing to participate and schools recruited (MENTEP schools), by country

In order to have sufficient statistical power to estimate the impact of the TET-SAT on teacher attitudes and behaviour, a minimum threshold of 1,000 teachers per country was set. However, as table 2 shows, this number was not reached in all countries. To achieve this threshold, all teachers with contact information were invited to take part in the project by filling in the Benchmark Survey.[4] The last column of the table reports the number of teachers who completed it, hence actively taking part in the project (for this reason we labelled them 'MENTEP teachers'). In total, the project had the active participation of 7,391 teachers.

Country	MENTEP Schools	Teachers with contact information	MENTEP teachers
Cyprus*	34	590	422
Czech Republic*	52	813	523
Estonia	32	705	314
Finland	52	1,969	847
France	-	-	712
Greece*	50	393	326
Italy	50	1,752	745
Lithuania	50	1,519	812
Portugal	50	2,797	798
Slovenia	50	931	858
Spain	49	2,246	1,034
Total	**469**	**14,649**	**7,391**

Note: In these countries, most schools sent only the list of teachers giving consent to handle personal information

4 An intermediate random sampling of teachers to be invited was set at this stage; however, at the end, all the teachers with contact information have been invited to take part to the project. See the Technical report on sampling (deliverable 3.1 Technical Report on the sampling) for additional details.

Table 2. Total number of teachers with available contact information and number of sampled teachers, by country

3.2 Data collection plan

The MENTEP evaluation exploited three sources of data:

- A Benchmark Survey (BS hereafter), administered online before the treatment. The aim of the BS was twofold: to collect background information on teachers and baseline data on their TET competences and attitudes. Background information was collected with two aims: i) to assess the heterogeneity of the TET impact across teachers and ii) to implement an alternative estimation solution in case the main strategy, based on the comparison between encouraged and non-encouraged teachers, failed.
- A Follow-up Survey (FuS hereafter), administered online after the treatment to the teachers who completed the BS. The aim of the FuS was to gather information on the main outcomes (views and self-reported TET-competence) in order to estimate the impact of the TET-SAT. Some questions were added in order to get additional information on how the TET-SAT was received by those using it.
- Finally, thanks to the fact that all interactions teachers had in the TET-SAT were stored in the platform's registers, we could also use this source of information to better understand which teachers used the tools and how they used it.

The surveys were built in accordance with the logical framework of the project and the relevant literature about ICT in teaching and learning. The following areas were selected to be covered in the BS: socio-demographic information; professional experience; familiarity with ICT; TET use in the classroom, attitudes towards ICT in education, self-reported competence in the TET [5]. The questions and their wording were chosen mainly from already existing validated cross-national surveys (e.g. TALIS, ICILS, PIRLS, TIMMS, Survey of Schools: ICT in Education).

The first drafts of the survey integrated comments from the partners and were translated in partners' official languages. A small pilot took place in each participating country to test and refine the translations.

By late September 2016, the final version of the BS was ready and uploaded to the platform. The FuS final draft was ready by April 2017. (For the final version of the surveys please see appendices 1 and 2 of the deliverable 4.2 "Data collection and analysis protocol".)

5 Self-assessed ability questions were designed to measure this construct in order to differentiate as much as possible the survey from the tool, the latter containing items being designed to stimulate teachers.

The BS was scheduled for the beginning of the 2016/2017 school year (September-October 2016). However, since the final lists of teachers in many countries were only completed in November the beginning of the BS data collection period had to be rescheduled. The invitation of teachers in the MENTEP schools to take part in the BS was sent via e-mail by the National Coordinators (NCs) by mid-December 2016. The BS was closed in the second week of February, once all the teachers (both sample and oversample) were invited to take part. (A detailed description of the procedure and the protocols can be found in the Sampling Protocol and in the Guidelines for the field trials documents.)

Next, starting from March 2017, encouraged teachers were invited to use the tool, also in this case via e-mail, by NCs. Finally, the timing of the FuS was chosen to take into account differences in the school calendar in partner countries. For this reason, the FuS was launched between the end of the first week of May and the following week. The administration of the survey followed roughly the same protocol as in the BS, with a massive email invitation to the teachers (in this case: only those answering the BS) closely followed by a reminder to the school heads.

Figure 4 visually represents the timeline of the data collection plan.

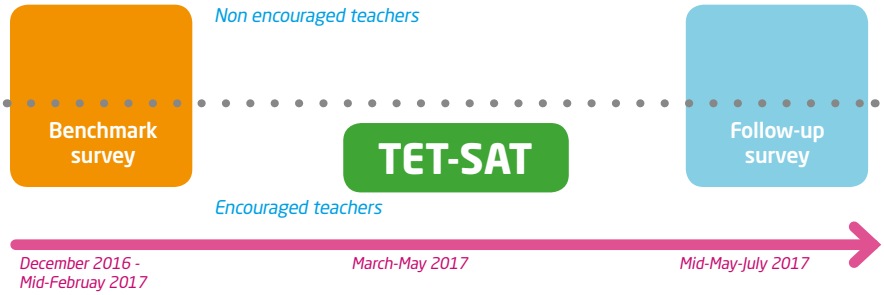


Figure 4 Data collection plan

3.3 The randomisation process

By randomisation we mean the process of allocation of the experimental units (in this project: schools and teachers) via random assignment to the two (or more) groups of units. The aim is the creation of statistically equivalent groups at the benchmark stage in order to infer the effectiveness of an intervention by comparing their post-treatment outcomes.

In the context of MENTEP, this procedure was organised in two stages, to take into account the particularity of the school context and of the design, which sees, on the one hand, teachers in schools and, on the other, the creation of a sub-group of peers who will not receive encouragement even if they work in encouraged schools.

Initially the randomisation plan was, as explained below, rather straightforward: (i) MENTEP schools are randomly assigned to the encouraged and non-encouraged group; and (ii) random teachers within encouraged schools are encouraged while others in the same schools are not encouraged to use the TET-SAT (the peers). The latter group is needed in order to assess the existence and strength of "peer effects" (or, spillover effects) of the treatment – i.e. detecting any possible change in the outcomes of the peers induced by the fact of having or not having colleagues who are "treated". Because of some particular national contexts, we were forced to invite all teachers in encouraged schools to use the TET-SAT because of the low number of participating teachers per school. This has immediate implications for the estimation of treatment effects: in fact, while the overall impact is measurable in all countries, peer effects can be detected only where the second step of the randomisation has been implemented following the original protocol.[6]

To accommodate national specificities three slightly different designs were therefore put in place for the second step of the randomisation (see table 3); the first step, that is the randomisation of schools, was implemented in the same way in all countries.

In countries in group A (labelled "simple randomisation design") the randomisation of teachers within encouraged schools turned out to be impossible given the low number of teachers. These countries are Estonia, Cyprus, Greece, France and Czech Republic.

In the remaining countries we followed the original randomisation plan: MENTEP teachers in encouraged schools are randomly assigned to two groups: encouraged and non-encouraged teachers (the so-called "peers"). The proportion of encouraged teachers in encouraged schools is set at 60%. In order to reinforce the intention to analyse peer effects, non-MENTEP teachers – i.e. teachers not taking part into the Benchmark Survey – in encouraged schools are also randomly encouraged. Because not all countries agreed to send these additional encouragements, this 'reinforced design' was implemented only in Finland, Lithuania and Italy. In these three countries, three quotas (40%, 60% and 80%) of non-MENTEP teachers were randomly encouraged. We will refer to this group of countries (LT, FI, IT) as "Reinforced peer effects design" group (design B2). The group of countries with no reinforcement of the peer effect design (ES, PT, SI) will be referred to as "Peer effects design" group (design B1).

6 The full report on the randomisation process can be found in the deliverable 3.3 Technical report on the econometric analysis.

Randomisation design	Randomisation level		
	MENTEP Schools	MENTEP Teachers in MENTEP Schools	Non-MENTEP Teachers in MENTEP schools
A - Simple randomised design	Yes	No	No
B1 - Peer effects design	Yes	Yes	No
B2 - Reinforced peer effects design	Yes	Yes	Yes

Table 3. The MENTEP randomisation design

Table 4 summarises the number of treatment and control units (schools), while Table 5 reports the number of teachers included in each group.

Country	Encouraged	Non-Encouraged	Total
Overall	369	366	**735**
CY	17	17	**34**
CZ	26	26	**52**
EE	16	16	**32**
ES	25	24	**49**
FI	26	26	**52**
FR	133	132	**265**
GR	25	25	**50**
IT	26	25	**51**
LT	25	25	**50**
PT	25	25	**50**
SI	25	25	**50**

Table 4. Number of encouraged and non-encouraged MENTEP schools, overall and by country

Country	Randomisation design	Non-encouraged schools	Encouraged schools		
		Non-encouraged MENTEP teachers	Encouraged MENTEP teachers	Non-encouraged MENTEP teachers	Encouraged non-MENTEP teachers
Overall	-	3,641	2,750	1,000	885
CY	A	212	210	-	-
CZ	A	273	250	-	-
EE	A	179	135	-	-
ES	B1	486	340	208	-
FI	B2	380	291	176	325
FR	A	356	356	-	-
GR	A	174	152	-	-
IT	B2	338	256	151	353
LT	B2	403	255	154	207
PT	B1	395	246	157	-
SI	B1	445	259	154	-

Table 5. The distribution of teachers across school types and encouragement status, overall and by country

4. Integrity of the experiment

This section summarises the analyses of the internal validity of the experiment, by testing the statistical equivalence of the groups generated by randomisation at the Benchmark Survey and their attrition at the Follow-up Survey.

The analyses presented in this chapter show that the groups are statistically equivalent on a large set of observable characteristics. Hence, there is a basis for extending conclusions to non-observable characteristics influencing the outcomes. We reach the same conclusion when we look at attrition rates.

In this chapter we will briefly go through internal validity (§ 4.1) and attrition analyses (§ 4.2). The full report on the integrity of the experiment is contained in deliverable 3.3 Technical report on the econometric analysis.

4.1 Group equivalence

Group equivalence was assessed by comparing characteristics of the two groups of schools (encouraged and non-encouraged) and of the three groups of teachers (encouraged, peers and non-encouraged) by means of linear probability models in which we regressed a dummy variable indexing the treatment status of the unit on the set of teacher characteristics. In non-technical words, we checked that the probability of being assigned to the treatment or control group was not affected by any available pre-treatment characteristic of the sample units. As concerns teachers, we separately performed equivalence tests comparing (i) non-encouraged teachers in non-encouraged schools (NET) with encouraged teachers (ET) in all countries; (ii) encouraged teachers (ET) with non-encouraged teachers in encouraged schools (NET-SE) only in countries with randomisation designs B1 and B2. Such a comparison allows us to consider the differences in randomisation designs across countries.

Both models on schools and teachers control for sampling strata (see D3.1 Technical Report on the sampling); standard errors of models on teachers are clustered at school level.

The variables used for the comparison are the following:

- Schools (administrative archives and results of the BS survey)
  - » Number of ISCED-2 teachers for which we possess contact information
  - » Number of students in ISCED-2 grades
  - » Percentage of teachers completing the BS survey
- Teachers (administrative archives and BS survey data)
  - » Gender
  - » Age
  - » Subject taught
  - » Previous ICT training
  - » Collaboration in ICT-related matters index[7]
  - » Views on ICT index
  - » Self-reported TET competence index
  - » Use of educational ICT tools index
  - » Use of basic ICT tools index

Table 6 shows the results of the analysis on schools for the overall sample. For each Y variable we report: (i) the average value in encouraged and non-encouraged schools; (ii) the coefficient on the treatment status expressing the estimated mean difference between the two groups of schools; and (iii) the corresponding p-value.[8] No statistical or substantial difference was detected between groups in any of the variables. Schools whose geographical distribution is by design comparable can be considered statistically equivalent also in terms of size and willingness of their teachers to participate in MENTEP. There is no way of replicating the same test on other school-level characteristics because the available administrative data are not homogeneous across countries.

Variables	School Group		Coeff.	p-value
	Non encouraged	Encouraged		
BS respondents (%)	66.2	66.2	-0.038	0.983
Number of ISCED-2 teachers	20.1	19.6	-0.500	0.631
Number of students	238.5	245.7	6.890	0.507

Table 6. School balancing test, school characteristics

7 A detailed description of the construction of the indexes can be found in the deliverable 4.2 Data Analysis Protocol

8 The p-value indicates the statistical strength of the association (the higher the value, the lower the statistical significance). At conventional levels, a p-value lower than 0.05 indicates a statistically significant coefficient.

Table 7 shows the results of the balancing tests for teachers. For each variable, we first report the average value for each group (columns 2 to 4). The other columns show the coefficients of two distinct linear probability models regressing the treatment status on the whole list of pre-treatment characteristics. More precisely, columns 5 and 6 contain the values of the estimated regression coefficients and their corresponding p-values for the comparison of NET versus ET. The same analysis is replicated for the comparison of NET versus NET-ES (columns 7 and 8). As apparent from the table, there are only small differences between groups, even when statistically significant (one case). Hence, we can conclude that the randomised groups are comparable.[9]

Variables (1)		Value			NET vs ET		NET vs NET-ES	
		(2)	(3)	(4)	(5)	(6)	(7)	(8)
		NET	ET	NET-ES	Coef.	p-val	Coef,	p-val
Gender (%)	Female	74.6	73.7	77.8	-0.004	0.733	0.027	0.100
Age (%)	<30	3.2	1.2	1.2	-0.002	0.686	-0.005	0.346
	30-49	57.7	57.6	55.9	-0.009	0.531	0.005	0.814
	49<	39.1	39.2	42.9	0.011	0.446	0.001	0.993
Subject (%)	Humanities	43.3	41.9	41.9	-0.012	0.274	0.002	0.891
	Science	31.1	31.8	29.0	0.004	0.684	-0.001	0.972
	Other	25.6	26.3	29.2	0.008	0.449	0.002	0.907
ICT training (%)	None	8.5	10.1	7.4	0.016**	0.034	-0.007	0.516
Positive views towards ICT	std. factor	-0.01	-0.00	0.046	0.017	0.541	0.011	0.786
Use of ICT in lessons	std. factor	0.005	-0.00	0.019	-0.002	0.922	-0.045	0.254
Collabora-tion with colleagues about TET	sum	1.86	1.80	1.78	-0.078**	0.029	-0.047	0.385
Basic ICT tools used in classroom	std. factor	0.011	-0.05	0.091	-0.033	0.235	0.014	0.737
Educational ICT tools used in the classroom	std. factor	0.016	-0.00	-0.01	-0.035	0.188	-0.027	0.501
Self-reported TET competence	std. factor	0.011	-0.02	0.040	-0.012	0.653	-0.032	0.415

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

Table 7. Teachers balancing test 4.2 Overall and differential attrition

9 Comparison of groups by country can be found in the deliverable 3.3 Technical report on the econometric analysis.

4.2 Overall and differential attrition

Attrition is the process by which some sample units, after taking part in the first wave of the survey (the BS in our case), drop out and do not take part in subsequent waves (the FuS in our case). In this section, we check whether attrition compromised the integrity of the evaluation design. We do that by comparing the attrition levels by group and by re-running the equivalence analyses of the previous section on the subsample of teachers completing the FuS.[10]

The first element is the overall attrition level. As seen in table 8, the response rate was quite high (75.6%), especially if we take into account that the survey was administered online. As concerns the differential attrition, i.e. the difference in attrition level by group, we see that peers and non-encouraged teachers have higher response rates than encouraged teachers (82%, 78.6% and 69.7% respectively).

School type	Teacher group	N of teachers			Response rate
		Benchmark	Follow-up	(FuS-BS)	
Overall					
Non-encouraged	Non-encouraged	3,641	2,861	-780	78.60%
Encouraged	Non-encouraged	1,000	820	-180	82.00%
Encouraged	Encouraged	2,750	1,917	-833	69.70%
Total		**7,391**	**5,598**	**-1,793**	**75.70%**
Countries with no peer effect estimation (rand. design A)					
Non-encouraged	Non-encouraged	1,194	900	-294	75.40%
Encouraged	Encouraged	1,103	757	-346	68.60%
Total		**2,297**	**1,657**	**-640**	**72.10%**
Countries with peer effect estimation (rand. design B1, B2)					
Non-encouraged	Non-encouraged	2,447	1,961	-486	80.10%
Encouraged	Non-encouraged	1,000	820	-180	82.00%
Encouraged	Encouraged	1,647	1,160	-487	70.40%
Total		**5,094**	**3,941**	**1,153**	**77.40%**

Table 8. Response rates by randomisation design and randomisation group

10 Also in this case, detailed by country figures are available in the deliverable 3.3 Technical report on the econometric analysis.

The level of differential attrition can be considered as moderate (What Works Clearinghouse, 2014), but it nevertheless calls for a deeper investigation of the sample on which the impact of the intervention will be estimated. For this reason, the equivalence test between groups was run again on FuS respondents following the same procedure as the analyses presented above. The three groups are still comparable, as no systematic difference between groups was detected.

Variables		Value			NET vs ET		NET vs NET-ES	
		NET	ET	NET-ES	β	p-val	β	p-val
Gender (%)	Female	74.8	74.3	76.7	-0.003	0.807	0.024	0.193
Age (%)	<30	3.0	2.7	2.0	-0.005	0.282	-0.006	0.348
	30-49	58.1	59.0	55.9	0.003	0.849	0.006	0.816
	49<	38.9	38.3	42.1	0.0023	0.889	<0.01	0.998
Subject (%)	Humanities	42.6	41.4	42.3	-0.013	0.318	0.010	0.602
	Science	32.5	33.3	30.4	0.005	0.687	-0.007	0.672
	Other	25.0	25.3	27.3	0.008	0.533	-0.002	0.882
ICT training (%)	None	9.1	9.1	8.4	-0.002	0.796	0.014	0.222
Positive views towards ICT	*std. factor*	0.00	0.03	0.06	0.038	0.229	0.050	0.274
Use of ICT in lessons	*std. factor*	0.02	0.04	0.04	0.016	0.598	-0.017	0.700
Collaboration with col-leagues about TET	*std. factor*	1.87	1.85	1.79	-0.040	0.334	-0.024	0.664
Basic ICT tools used in classroom	*std. factor*	0.03	-0.03	0.12	-0.031	0.315	0.023	0.625
Educational ICT tools used in the class-room	*std. factor*	0.03	0.03	-0.01	-0.013	0.679	-0.002	0.968
Self-reported TET competence	*std. factor*	0.06	0.03	0.08	-0.009	0.782	-0.028	0.508

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

Table 9. Balancing test for the analytical sample (teachers answering to the FuS)

5. Process analysis

5.1 Who are the participating teachers? Some figures

As seen in sections 3 and 4, a total of 469 schools and 7,391 teachers were recruited across the 11 MENTEP countries. Out of the total number of teachers who became "MENTEP teachers" by filling in the Benchmark Survey, 5,598 also responded to the Follow-up Survey and hence became part of the analytical sample used to produce the TET-SAT's impact estimates (see section 6).

Beyond this quantitative description of the teacher sample, it is important to shed light on some of its qualitative characteristics. In what follows (tables 10, 11, 12 and 13), we focus on those that are theoretically more relevant in regard to the MENTEP evaluation question, while we include additional figures of interest in Appendix A.

The overall picture coming from the BS data is that MENTEP teachers have a good familiarity with ICT, a very high perception of their competences to exploit digital resources in teaching and very positive views concerning the benefits of new technology for both teaching and learning. For instance, all of them have some device at home and about 50% spend more than one hour a day using ICT at home (Table 11).

Individual characteristics	Average
Females (%)	75.1
Age groups (%)	
Less than 40	25
Between 40 and 50	35.8
More than 50	39.2
Subject taught (%)	
Humanities	42.2
Scientific fields	32.3
Other subjects	25.5
Weekly working hours	
Teaching	18.0
Preparing lessons	8.2
Admin. Duties	4.3
Other	3.3
Total	**33.8**
Years of experience	20.7
Share of teachers with a permanent contract	12
Total	**5,598**

Table 10. Sociodemographic and professional background of MENTEP teachers

Individual characteristics	Average
Age when first used a PC (%)	
9 years/younger	6.9
10-19 years	37
20-29 years	33.2
30-39 years	16.7
40 years/older	6.2
ICT time at home (%)	
0-60 min/day	49.8
1-3 h/day	38.7
3+ h/day	11.5
Availability of ICT devices at home (%)	
tower PC	54.5
portable PC	88
Tablet	58.8
internet connection	93.8
cell-phone internet	83.3
Printer	74.3
ebook reader	15.1
Total	**5,598**

Table 11. MENTEP teachers' familiarity with ICT

Moreover, at the time of the BS, MENTEP teachers consider that they possess very high TET competence: 90% that they could stimulate students to use ICT in a critical manner, and 71% that they can redesign ICT apps in view of specific educational settings (Table 12).

I am able to...	%
Stimulate students to use ICT in a critical manner	90
Support students in searching information by means of ICT	95
Support students to communicate with ICT in a safe, responsible and effective way	90
(Re)design ICT applications in view of a specific educational setting	71
Select ICT applications effectively in creating a learning environment	77
Total	**5,598**

Note: The values refer to the cumulative relative frequency of teachers answering slightly agree, agree and totally agree.

Table 12. MENTEP teachers' self-reported TET competences in the BS

Finally, teachers exhibit positive views about ICT in teaching and learning - e.g. 94% think that using ICT at school enables students to access better sources of information and 76% that it helps students develop greater interest in learning (Table 13). At the same time, teachers are still positive, but to a lesser extent, that ICT can help children develop self-regulation in learning as well as enhance their academic performance.

Using ICT at school...	%
Enables students to access better sources of information	94
Helps students to consolidate and process information more effectively	84
Helps students learn to collaborate with other students	75
Enables students to communicate more effectively with others	64
Helps students develop greater interest in learning	76
Helps students work at a level appropriate to their learning skills	76
Helps students develop skills in planning and self-regulation of their work	65
Improves academic performance of students	60
Total	**5,598**

Table 13. MENTEP teachers' opinions on the usefulness of ICT in teaching and learning

5.2 Treatment compliance and reactions of the teachers to the tool

As in all experiments, not all units assigned to the treatment group actually made use of the intervention offered, i.e. not every teacher to whom the set of encouragement emails was sent actually decided to try out the TET-SAT and complete it. As shown in Table 14, the proportion of encouraged teachers deciding to start the TET-SAT among those receiving the encouragement letter is 33.8 percent and of those who started and completed it is 26.7 percent.[11] This fraction, as shown in the Table, varies considerably across participating countries. The figures can (and should) be read from two distinct perspectives.

From a methodological point of view (i.e. in regard to the experimental design), the encouragement worked fully, allowing us to carry out the planned econometric analyses described in section 6: a non-negligible fraction of teachers reacted to the set of encouragement emails while none of the non-encouraged teachers used the TET-SAT. This is a case of so-called "one-side non-compliance", i.e. there are only "no-shows" (1,820 teachers did not use the TET-SAT, even if encouraged to do so) and zero "crossovers" (non-encouraged teachers using the TET-SAT).

11 Among the teachers who made use (and completed) the TET-SAT, the fraction of those who used it more than once during the experimental period happened to be low (8%).

From a substantive point of view, these numbers can be seen as an estimate of the 'natural' take-up rate of such an intervention. Policy makers considering scaling-up the TET-SAT can see from these data that a light-touch invitation sent to the entire teacher population might result in a take-up rate ranging between one-fourth and one-third. Yet, as the table shows, the TET-SAT take-up rate shows impressive variability across countries (ranging from a minimum of 16% to a maximum of 61%, when considering the proportion of teachers starting using the tool). This variability is possibly accounted for by a mixture of contextual factors (e.g. a culture of self-assessment in national professional development frameworks, or the diffusion of new technologies in schools) and MENTEP-specific organisational aspects (e.g. actual implementation of the experimental protocol).

Country	Total number of encouraged teachers	Encouraged teachers who started the TET-SAT		Encouraged teachers who started and completed the TET-SAT	
		N	%	N	%
CY	210	112	53.3	98	46.7
CZ	250	42	16.8	32	12.8
EE	135	36	26.7	25	18.5
ES	340	126	37.1	103	30.3
FI	291	48	16.5	35	12.0
FR	356	116	32.6	83	23.3
GR	152	77	50.7	68	44.7
IT	256	62	24.2	45	17.6
LT	255	89	34.9	69	27.1
PT	246	64	26.0	42	17.1
SI	259	158	61.0	134	51.7
Total	**2750**	**930**	**33.8**	**734**	**26.7**

Table 14. TET-SAT take-up rates

Concerning the reasons for not using the TET-SAT (Table 15), teachers reported as the main one that they were not aware of it. This is an unexpected finding, considering that a set of emails was successfully sent to all encouraged teachers. According to some National Coordinators, teachers in some countries may have confused the emails concerning the survey with the ones to invite them to use the TET-SAT. It is not clear the extent to which this explanation holds across countries. However, this finding calls for more attention in future implementations of such a design to carefully monitoring communication with the experimental units. The second - and less unexpected - reason mentioned by teachers is lack of time. Based on available anecdotal evidence as well as on the MENTEP figures (Table 10), 90% of teacher working time is dedicated to teaching, preparing lessons and carrying out administrative duties, leaving little room for training.

Main reason	Overall
Unaware of it	32%
Time constraints	30%
Not interested in self-assessment	10%
Already competent	4%
Could not access	4%
Do not use ICT	3%
Not interested in training	1%
Other	15%
Number of respondents	1,091

Table 15. The reasons why encouraged teachers did not use the TET-SAT as reported at the Follow-up Survey

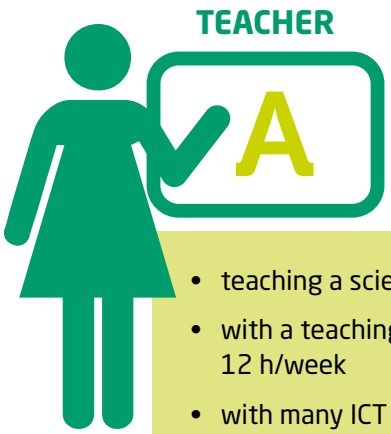
It goes without saying that teachers using the TET-SAT is a self-selected subset of the encouraged teachers and that, for this reason, the main characteristics of these two groups of teachers are, on average, not comparable. As shown in Table 16, some teacher characteristics are prominently associated with the use of the tool. In other words, some teacher characteristics (i.e. teaching a science subject, working more hours in non-teaching tasks, using ICT at home and, most importantly, having positive views on ICT in teaching and learning and rating levels of TET competences as high) are found to be predictive of the probability of accepting the invitation to use the TET-SAT.

Variable	Sign of the effect on the probability of using the TET-SAT
Scientific Subjects	+
High teaching hours	-
High non-teaching hours	+
Use of ICT devices at home	+
Positive views on ICT in teaching	+
ICT collaboration with other teachers	+
Use of ICT mainstream applications (e.g. Office)	+
Training in ICT	+
self-reported TET competence	+

Note: The associations reported in the table are obtained via a linear regression probability model. Only the statistically significant associations (p-value <0.05) are reported. The coefficients associated to all other characteristics included in the model (e.g., sex, age, years of professional experience) are not statistically significant at the conventional levels, and hence not included in the table.

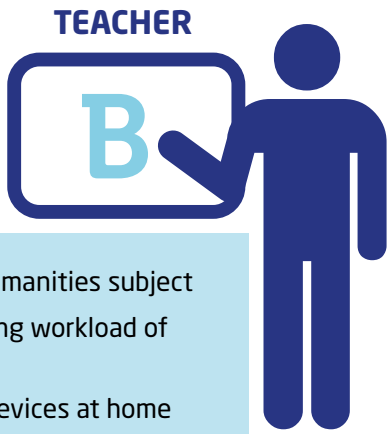
Table 16. Teachers' characteristics significantly associated with the use of the TET-SAT

On the basis of these data, it is possible to sketch two 'ideal' teacher profiles:



- teaching a scientific subject
- with a teaching workload of 12 h/week
- with many ICT devices at home
- with positive views on ICT
- who is highly collaborative
- who followed training in ICT
- who has a high level of self-assessed ICT ability

has a probability of using the TET-SAT of 50%.



- teaching a humanities subject
- with a teaching workload of 18 h/week
- with no ICT devices at home
- with negative views on ICT
- who is less collaborative
- who did no training in ICT for teaching
- who has a low level of self-assessed ICT ability

has a probability of using the TET-SAT of 15%.

Finally, among those teachers who completed the TET-SAT, the level of satisfaction with the self-assessment experience and its usefulness, as well as the satisfaction with the technical features of the platform, were quite high (Tables 17 and 18). 64% of teachers stated that the TET-SAT helped them assess their competence; 69% of them believed that the feedback page was useful to assess their competences; 60% stated that they would recommend the tool, and so on. Hence, those teachers using the TET-SAT were, overall, quite satisfied with it.

Item	Agree
TET-SAT helped me to assess my competence	64%
TET-SAT helped me to re-think use of ICT in teaching	62%
TET-SAT took too much time	38%
TET-SAT was boring	30%
TET-SAT was useful	63%
TET-SAT was easy	75%
Feedback page useful to assess my competences	69%
Resources useful to improve teaching	51%
Overall satisfied	63%
Self-comparison useful	55%
I would recommend tool	60%
Inspired to try new practices in my teaching	54%
I prefer to use self-assessment tool to other methods of ass.	57%
Number of respondents	818

Table 17. Teachers' opinions on the usefulness of the TET-SAT

Features	Overall rate (min 1; max 10)
Graphic interface	7.4
Online access	8.1
Organization	7.6
Feedback page	7.6
User friendliness	7.7
Language easy to understand	7.8
Contents overall	7.4
Contents pedagogy	7.4
Contents content use and production	7.4
Contents communication and collaboration	7.3
Contents citizenship	7.2
National training resources	7.3
International training resources	7.2
Number of respondents	615

Table 18. Teachers' satisfaction with the main technical features of the TET-SAT

5.3 The TET-SAT feedback score

Let us now open up the TET-SAT tool and shed light on one of its key components: the feedback score given to teachers on completion of the questionnaire (Figure 1). As argued above (section 2.2), the feedback score can be understood as a piece of "objective" information that teachers receive about their TET-competences by relating the score to a specific competence levels underpinned by objective criteria. In principle, receiving this information can alter the way teachers perceive their own competences. However, the information contained in the score has a different value for teachers depending on the extent to which that score is in line with what they thought about themselves before taking the TET-SAT. Although for some teachers the feedback score would just be a confirmation of what they already thought (i.e. when the feedback score matches the pre-treatment self-assessment), for other teachers the feedback could represent an information update. The latter can be a negative one (when teachers receive a competence score which is lower than what they thought) or a positive one (when the feedback 'tells' teachers that they are more competent than they thought). In this section, we first show some descriptive statistics on the scores received by teachers (Table 19) and then investigate the extent to which teachers may have been differently affected by the feedback score.

Country	Overall score TET-SAT	Digital pedagogy	Digital content use and production	Digital comm. and collabor.	Digital citizenship	N
CY	50.9	49.5	51.0	45.2	53.5	98
CZ	50.0	48.0	50.9	41.9	56.3	32
EE	60.0	62.2	55.5	58.2	59.9	25
ES	52.0	55.2	47.6	47.6	51.3	103
FI	55.2	57.8	53.3	48.6	56.0	35
FR	53.0	55.0	51.7	45.8	54.9	83
GR	51.4	50.8	52.6	45.7	52.0	68
IT	53.1	52.5	49.9	47.4	57.9	45
LT	58.0	60.7	48.5	55.5	56.0	69
PT	57.0	59.6	52.3	53.9	57.6	42
SI	52.0	50.6	49.7	45.4	58.3	134
Overall (mean)	53.2	53.8	50.5	47.8	55.3	734
Overall (median)	51.0	53.3	47.5	45	55	734

Note: The score ranges from 25 to 95.

Table 19. The TET-SAT feedback score by country

Figure 5 shows a scatter plot of teachers' self-reported TET competences collected in the BS (vertical axis) and teachers TET-SAT feedback score (horizontal axis). The red vertical and horizontal lines represent the average values of the two variables. It is important to stress the purely descriptive purpose of this figure. By no means do we claim that the two measures are directly comparable. Teachers do not directly compare the two measures as only the feedback score is known to them in numeric form, while the level of self-reported competence is computed by pooling the items of the scale used in the BS.

Let us simplify the analysis by considering only the location of these teachers above or below the average. A possible interpretation of this figure is that for most teachers (about 68% - **blue dots**) the feedback score confirmed their self-reported level of TET competences. Only a small fraction of teachers (8% - **red dots**) received a positive update, while a sizeable proportion of teachers (28% - **green dots**) "discovered", thanks to the TET-SAT, that their level of TET-competences is lower than they thought.

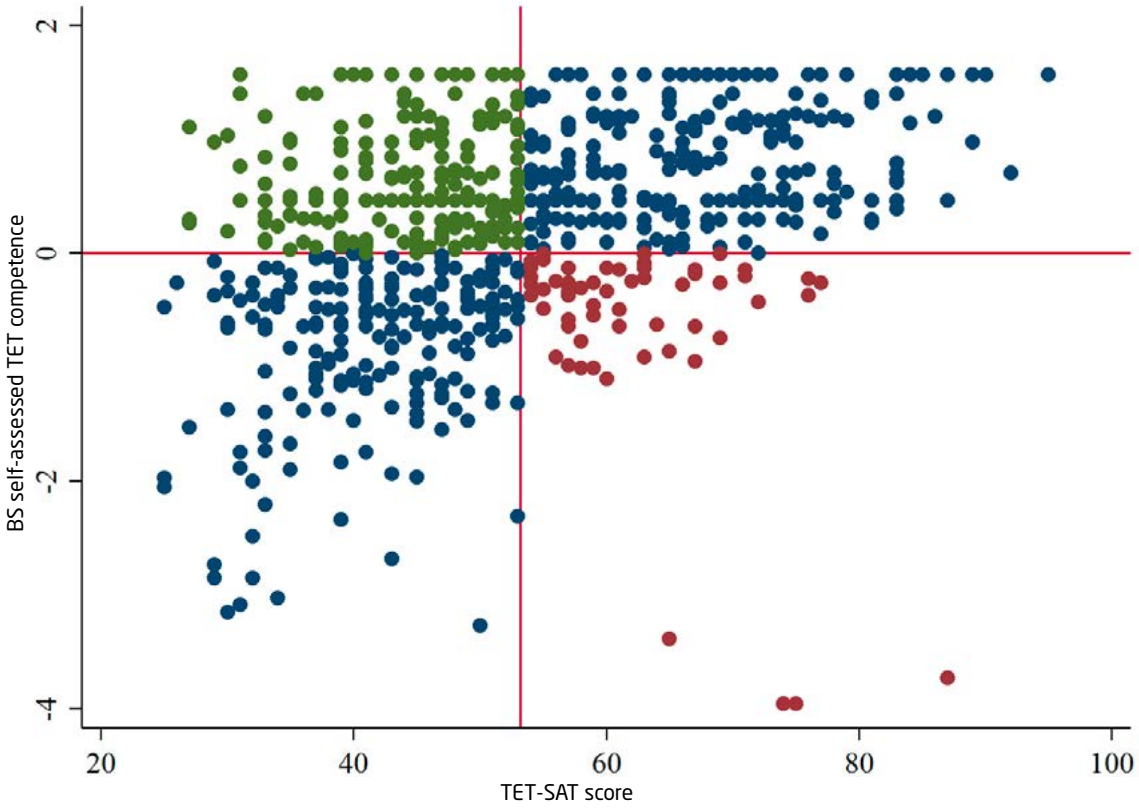


Figure 5. The "information value" of the TET-SAT feedback score

We shall elaborate on this hypothesis in a formal way in the section on the estimation of the causal effect of the TET-SAT.

6. Treatment effects

In this section, we present the estimate of the short-term impact of the TET-SAT on self-reported TET-competence and on views on ICT in teaching.

On the other outcomes (i.e. frequency and modalities of use of digital resources or further training on ICT), the TET-SAT did not have any significant effect, and this is explained by the fact that we are only observing short-term effects, as explained in section 2. The section is organised as follows. Sub-sections 6.1 and 6.2 present the result of the main effect and the peer effect respectively. In section 6.3, we present additional evidence on the link between the feedback score and outcomes.

6.1 Main effects of the treatment on self-reported TET competences and on views on ICT in teaching

The main effects of the TET-SAT are obtained by comparing test and control teachers. Several causal parameters may arise from this, but we are mainly interested in estimating the effect that the TET-SAT had, on average, on the teachers actually using it. In experimental designs with one-sided partial compliance (see section 5), this parameter is known as "Average Treatment Effect on the Treated" (ATT hereafter). The idea behind this parameter is to "correct" the comparison among test and control teachers by accounting for the fact that only a fraction of the test teachers used the TET-SAT.

More precisely, we assess the impact of the TET-SAT by estimating separately its impact on self-reported TET competence and on views on ICT in education. In order to retrieve causal estimates of the TET-SAT we rely on instrumental variable models in which the actual participation of each teacher in the TET-SAT (a binary variable taking value 1 if the teachers logged into the TET-SAT) is the core explanatory variable and the receipt of the randomised set of encouragement emails is used as an exogenous instrument. The idea behind this procedure is that

the difference in outcomes between test and control groups is boosted by the proportion of subjects (i.e. teachers) actually treated among those assigned to the treatment.

We control for stratification blocks[12] to take into account the design of the sampling and of the randomisation. Moreover, we include as a control variable the pre-treatment value of the outcome (collected at the BS) to improve the accuracy of our estimates. Finally, standard errors are clustered at school level.

An important limitation of our analysis concerns the impossibility to break down the impact at country level. The number of FuS respondents is large enough to provide an estimate of the overall impact of the TET-SAT but not to allow us to estimate country-specific impacts with any confidence. We can, nevertheless, run heterogeneity analyses breaking down the sample according to other important variables.

Table 20 reports the effects of the TET-SAT on the outcomes on the overall sample. The results show that the use of the TET-SAT leads teachers to revise more critically their competence in TET and their opinions about ICT in education. Both results, significant at 5% and at 1% respectively, come from a standardised scale and hence can be read as effect sizes. Teachers using the tool, in other words, self-evaluated themselves, on average, 0.13 standard deviations less than control group teachers and revised their opinions downward by 0.35 standard deviations.

If we use instead the original scale of the variables we find that the effect on self-reported TET competence is equal to -0.12 on a scale ranging from 1 to 6, with the average score for control group teachers as large as 4.7. This means that the effect is rather small in absolute terms, and that test teachers still have a very high opinion of themselves in rating their TET competences, even if a bit lower than that of control group teachers.

As concerns views on the use of ICT in teaching, the average value for control group teachers is as large as 3, on a scale ranging from 1 to 4 (1 means total disagreement with positive statements about ICT and 4 total agreement). The estimated effect on the original scale is as large as -0.16, this means that the respondents share, by and large, positive views about ICT in education and that the downward revision induced by the treatment does not alter respondents' generally positive feelings about ICT.

12 For most countries, the strata identify geographical administrative aggregations using (or, in some specific cases, adapting) the Nomenclature of territorial units for statistics (NUTS). For small countries, we either decided not to create strata (Cyprus) or used the type of school program instead of geographical aggregations (Lithuania). For more details, see deliverable 3.1 Technical report on the Sampling.

	Average value for controls	Effect Size	S.E.	N
Self-reported TET competence				
Treatment	0.02	-0.138**	(0.0611)	4,690
Positive views on ICT				
Treatment	0.02	-0.351***	(0.0607)	4,771

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

Table 20. Average Treatment effect on the Treated (ATT) of the TET-SAT on self-reported TET competence and views on ICT in teaching, overall sample. Standard errors clustered at the school level

In tables 21 and 22, we study the heterogeneity of the causal effect along some dimensions. We start by looking at the difference across genders and age groups. The effect on TET competence is stronger among females than among males (-0.159 vs -0.067), even if in both cases the the coefficient is negative and the confidence intervals of the two effects strongly overlap. On the other hand, no difference is detected among teachers of different age groups. The effect on views is highly similar for males and females, while it is stronger among older teachers. Note that also in this case the sign of the effect points to the same direction for both groups and that the confidence intervals of the estimates strongly overlap.

	Average value for controls	Effect size	S.E.	N
Self-reported TET-competences				
Males	0.12	-0.067	(0.104)	1,192
Females	-0.01	-0.159**	(0.075)	3,498
Teachers <40 years	-0.01	-0.172	(0.117)	1,198
Teachers >40 years	0.03	-0.133*	(0.070)	3,491
Positive views on ICT				
Males	0.08	-0.399***	(0.098)	1,208
Females	0.01	-0.356***	(0.072)	3,551
Teachers <40 years	-0.16	-0.270***	(0.100)	1,216
Teachers >40 years	0.09	-0.407***	(0.071)	3,542

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

Table 21. Average Treatment effect on the Treated (ATT) of the TET-SAT on self-reported TET competence and views on ICT in teaching, by subgroup. Standard errors clustered at the school level.

In Table 22 we break down the estimates by the pre-treatment level of the outcome variables. The highest impact of the TET-SAT is found for teachers who started

with higher levels of the two tested outcomes. In other words, the TET-SAT had an impact at the top end of the self-reported competence and views distribution.

Stratification	TET-SAT impacts
Self-reported TET competences in BS	Self-reported TET-competences
High competences	-0.143**
Low competences	-0.092
Views on ICT in BS	Positive views on ICT
High positive views	-0.441***
Low positive views	-0.231***

Note: High and low levels are determined using the median values of the two variables.
*** p<0.01; ** p<0.05; * p<0.1

Table 22. ATT estimates stratified by pre-treatment self-reported TET competence and views on ICT

Finally, in Table 23 we break down the estimates by level of the take-up rate in the country. We showed in section 5.2 that the take-up rate of the treatment among teachers invited to use the TET-SAT is highly variable across countries. Cross-country cultural differences with respect to the use of ICT in teaching might explain this heterogeneity, among other reasons. These in turn might even be a factor explaining heterogeneity in the causal effect of the TET-SAT. As in the previous cases, the sign of the impact is uniformly negative, but in high take-up rate countries the impact of the TET-SAT is larger on both outcomes.

	Effect size	S.E.	N
Self-reported TET competence			
Low take-up	-.093	.103	2,300
High take-up	-.151**	.076	2,390
Positive views on ICT			
Low take-up	-.302***	.103	2,353
High take-up	-.376***	.075	2,418

Note: High/low means above/below the average take-up rate.
*** p<0.01; ** p<0.05; * p<0.1

Table 23. ATT estimates stratified by country TET-SAT take-up rate

6.2 Peer effects of the treatment on self-reported TET competence and views on ICT in teaching

One of the innovative features of the MENTEP research design is the possibility to estimate peer effects of the intervention by comparing a random subset of teachers teaching in encouraged schools that did not receive the set of encouragement

emails (NET-ES) with teachers from non-encouraged schools (NET). The estimation has important policy implications. The type of tool promoted is designed to appeal particularly to teachers who access and use it on voluntary basis, given the strong emphasis on self-assessment and voluntary access to further training. This leaves out teachers that are less interested in this kind of initiative. The detection of a peer effect of the TET-SAT could signal the existence of processes of spillover occurring within the school and hence be informative of the extent to which teachers collaborate or, at least, exchange information at school.

Note that the peer effect can be estimated only in a subset of the MENTEP countries, namely countries where we could implement randomisation designs B1 or B2 (see section 4 for details): Spain, Lithuania, Italy, Finland, Portugal and Slovenia. Based on the evidence set out in this section, there is very little sign of a peer effect, if any.

The results of this analysis reveal that there is no apparent peer effect, either on self-reported TET-competences or on views about ICT (Tables 24 and 25). Table 24 shows the ATT estimates [13] on all countries and on the sub-group of B1-design countries are similar, proving that there is no substantial sample selection when focusing only on B1 countries. As shown in Table 24, no statistically significant peer effects are detected: the coefficient related to views is as sizeable as the main ITT effect, but it is only marginally significant.

	All countries	Only B1 design countries*	
	ATT	ATT	Peer effects
Positive views on ICT	-0.351***	-0.127*	0.061*
Self-reported TET-competences	-0.138**	-0.116	-0.026

*B1 design countries are: Slovenia, Lithuania, Italy, Spain, Portugal, Finland.
*** p<0.01; ** p<0.05; * p<0.1

Table 24. Peer effects in B1-design countries

Table 25 presents a similar exercise on B2-design countries, i.e. countries in which the proportion of encouraged teachers within a school was allowed to vary by randomly encouraging non-MENTEP teachers in the school. Table 25 first shows that the ATT estimates in B2-design countries are comparable to those obtained in the full sample.

13 The ITT estimate is given by the simple comparison of the randomised groups (i.e. test and control groups) without taking into account the existence of partial compliance (i.e. no-shows).

A similar picture as the one depicted in Table 24 emerges: no statistically strong evidence in support of the existence of peer effects of the TET-SAT is found.

	All countries	Only B2 design countries*	
	ATT	ATT	Peer effects
Positive views on ICT	-0.351***	-0.34**	0.14*
Self-reported TET-competences	-0.138**	-0.14	-0.04

*B2 design countries are Italy, Lithuania and Finland.
*** p<0.01; ** p<0.05; * p<0.1

Table 25. Peer effects in B2-design countries

To conclude, there exists very weak – if any – evidence of peer effects, leading us to conclude that treated teachers' changed views and perceptions of their own competences do not affect their "untreated" colleagues.

6.3 Interpreting the main effects: the role of the feedback score

In this section, we provide collateral evidence on the likely role of the feedback score in the process determining the impact of using the TET-SAT documented in the previous section. Here we focus on the 672 teachers who filled in the Benchmark and the Follow-up Survey and completed the TET-SAT, i.e. teachers receiving the feedback score from the system, an impartial assessment of their level of competence in using ICT for teaching. The evidence we provide below is that the feedback score had an impact on self-reported competence (if not on opinions on using ICT in teaching) reported by teachers in the FuS.

The evidence we provide must be interpreted with care since it is not experimental evidence as provided in the previous sections on the main and peer effects. We proceed along the logic of the so-called Difference-in-Differences. With reference to a specific outcome – either self-reported competence or opinions on using ICT in teaching - we have available two measurements, one before the experimental period and one after it.

The assumption on which we rely to derive an estimate of the impact on the feedback score on the outcome is that in the absence of the feedback score the variation of the outcome between the two time periods would have been on average the same for all the teachers using the TET-SAT, whatever their starting level of self-reported competence and of opinions on using ICT in teaching. Clearly, there is no way to check against the data the validity of this assumption precisely because all the teachers using the TET-SAT did receive the feedback

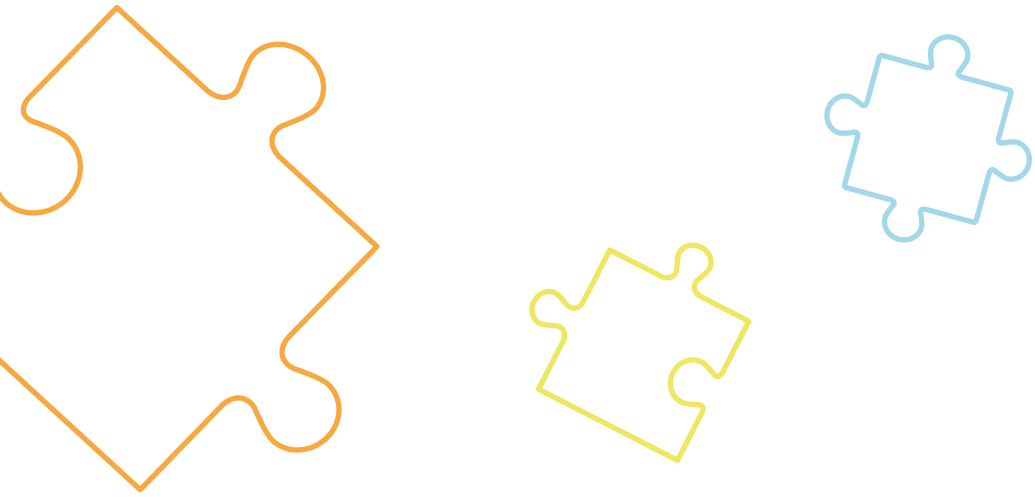
score. Nonetheless, over the short lapse of time between the BS and the FuS it seems unlikely that teachers experienced anything but the TET-SAT relevant to the outcomes we are considering. In this sense our assumption is credible, we believe.

Relying on the Diff-in-Diff's assumption, in order to measure the impact of the feedback score on the outcome it suffices to run the regression of the variation of the outcome in the lapse of time between the BS and the FuS on the feedback score. Results of this regression are in Table 26, separately for the two outcomes. There is a clear and far from negligible effect of the feedback score on the self-reported ICT competence: one standard error increase of the feedback score induces 0.1 standard error increase in the self-reported ICT competence. On the other hand, the effect of the feedback score on opinions on using ICT in teaching is statistically zero.

	Self-reported TET competence	Opinions on using ICT in teaching
TET-SAT feedback score	0.0969**	-0.0383
	(0.0488)	(0.0346)
N	672	

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

Table 26. Diff-in-Diff's estimate of the impact of the feedback score on self-reported TET competence and on opinions on using ICT in teaching (heteroscedasticity robust standard error; all variables are standardised).[14]



14 Heteroscedasticity refers to the situation in which the variability of a variable changes depending on the values taken by a second variable.

7 Conclusions and policy implications

Our concluding remarks on the MENTEP policy experimentation start by stressing the overall success of its implementation. MENTEP is, to the best of our knowledge, the largest randomised controlled trial, in terms of participating countries, ever carried out on the topic of teacher training. Its success provides a lesson for all policy makers by proving that running robust policy experimentations on salient policy topics is not only recommended but also practically feasible. From the initial design of the experimental protocol to the definition of its application in the different contexts, the MENTEP experience provides a wealth of lessons that can be used by policy makers to promote evaluation capacity building across Europe.

Concerning the research findings, several points need to be stressed:

First, the low take-up rate of the TET-SAT (between one third and one fourth of teachers) alerts us to the "natural" take-up rate of such a policy and its variability across national contexts. In other words, it indicates the extent to which teachers would adopt such a programme if the policy were to be taken to scale. At the same time, this result draws the attention of policy makers on the need to consider carefully the issue of how to raise the interest and cooperation of teachers to improve their participation rate (e.g. face-to-face or virtual meetings, phone calls).

Second, and related to the previous point, teachers who made use of the tool are a self-selected sub-population of teachers. Younger teachers, teaching in scientific subjects and having a familiarity with ICT are those who reacted more positively to the offer of using the TET-SAT.

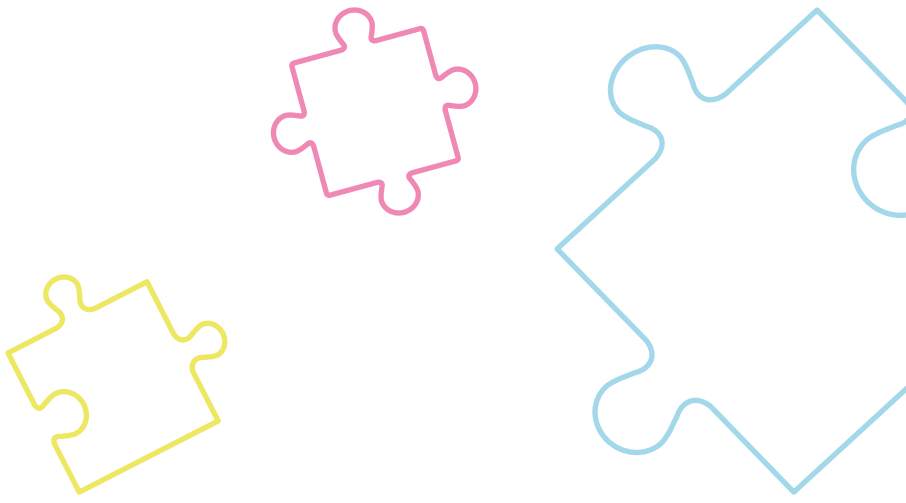
Third, teachers who used the TET-SAT showed a high level of satisfaction with it both with respect to technical features of the tool and with respect to its perceived usefulness.

Fourth, due to time constraints on the project, the evaluation research has focused on short-term impacts only: opinions on the use of ICT in teaching and learning and self-reported TET competences. The main experimental results show that the TET-SAT made teachers using it revise slightly downwards their beliefs about their own competence in using ICT for teaching and their views on whether ICT is useful in teaching and learning. We showed that these effects are larger among the subset of teachers who had higher starting levels of both self-reported competences and views, hence the TET-SAT had an impact on the top-end tail of the distribution.

Fifth, a possible explanation we propose for this downward revision of self-reported competence is that it is driven by the feedback score given to teachers by the system on completing the TET-SAT, an "objective" assessment of their TET-SAT competences.

Sixth, we conclude that the downward revision of perceived TET-ability and views is a sign of the fact that the TET-SAT triggered self-reflection on respondents' practices, helping them to adjust in a more informed way their prior assumptions about ICT in education and their own TET competence.

Future research, exploiting a second Follow-up Survey – not initially planned – and implemented in addition in Spring 2018, could yield some further insights on the persistence of these effects and on the existence of effects on other types of TET outcomes, e.g. actual behaviours and use of ICT in teaching or participation in further training on ICT. The usability of the data of this second Follow-up Survey is still under the careful scrutiny of the research team, because the attrition rate proved to be very high (85%) and a number of statistical checks have to be performed before analysing them.



References

Andrade, H. Valtcheva, A. (2009). Promoting learning and achievement through self-assessment, Theory into practice, 48:1, 12-19.

Bloom, H.S. (2008). The core analytics of randomized experiments for social research. *The Sage handbook of social research methods*, 115-133.

European Commission (2013). Survey of schools: ICT in education. Benchmarking access, use and attitudes to technology in European schools. doi: 10.2759/94499.

Gonzalez, N. (2017). How Learning About One's Ability Affects Educational Investments: Evidence from the Advanced Placement Program. Working Paper 52. Oakland, CA: Mathematica Policy Research, February 2017.

Hattie, J.A.C. (2003). Teachers make a difference: What is the research evidence? Paper presented at the Building Teacher Quality: What does the research tell us? ACER Research, Conference, Melbourne, Australia. Retrieved from: *http://research.acer.edu.au/research_conference_2003/4/*

Hattie, J. (2008), Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement, Routledge, London/New York.

Laurillard, D. (2012). Teaching as a Design Science, Routledge, New-York/Abigdon.

Tondeur, J., Pareja Roblin, N., van Braak, J., Fisser, P. & Voogt, J. (2013). Technological pedagogical content knowledge in teacher education: in search of a new curriculum. Educational Studies, 39, 239-243.

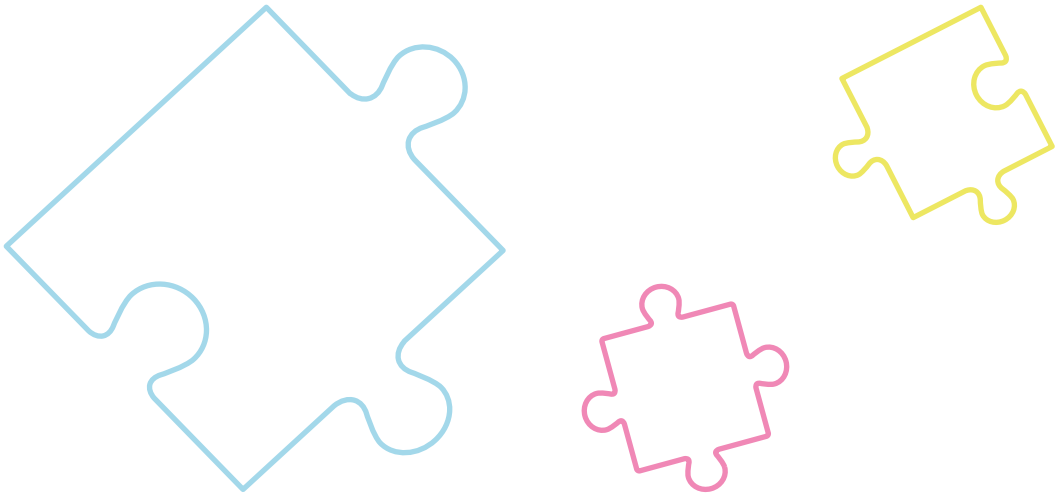
What Works Clearinghouse (2014). Procedures and standards handbook (Version 3.0). Washington, DC: US Department of Education.

Appendix A

Table A1 - ICT-related behaviour of MENTEP teachers in BS

Behaviours	%
Use of ICT in some lessons, in more lessons, in every lesson	
Presenting information through direct class instruction	95
Providing remedial or enrichment support to individual students or small groups of student	77
Enabling student-led whole-class discussions and presentations	74
Assessing students' learning through written tests	56
Providing feedback to students	76
Reinforcing learning of skills through repetition of examples	85
Supporting collaboration among students	75
Mediating communication between students and experts or external mentors	31
Enabling students to collaborate with other students (within or outside school)	56
Collaborating with parents or guardians in supporting students' learning	56
Supporting inquiry learning	76
Assigning written task/ exercises / homework to students	77
Facilitating / supporting individual or collaborative oral presentation by students	82
Communicating with students out of the classroom	61
Collaboration using ICT	
I work together with other teachers on improving the use of ICT in classroom teaching	56
I collaborate with colleagues to develop ICT based lessons based on the curriculum	42
I observe how other teachers use ICT in teaching	48
I work with other teachers on cross-curricula projects involving ICT	38
No collaboration with ICT	16
How often do you use the following tools? At least in some lessons	
Educational software	50
Tutorial software	53
Digital learning games	59
Word/Power Point	92
Spreadsheets	53
Multimedia production tools	52
Data logging and monitoring tools	15
Simulations and modelling software	19
Social media	28
Communication software	72
Computer-based information resources	93
Interactive whiteboard	52
Graphing or drawing software	32
E-portfolios	17
Mobile devices	55
Learning management systems	33

Has participated in any training courses in the past three school years	
Introductory courses	23
Advanced courses on applications	17
Advanced courses on internet use	15
Equipment-specific training	41
Courses on the pedagogical use of ICT	42
Subject-specific training on learning applications	20
Course on multimedia	13
Participation in online communities	20
ICT training provided by school staff	41
Personal learning about ICT	57
Other	23
None of the above	8
Total	**5,598**



The success of the MENTEP policy experimentation summarised in this report was only possible thanks to the active participation and commitment of all MENTEP partners, the high level support of the National Coordinators, who coordinated the field trials in their country, the teachers participating in the experimentation and the valuable input of the scientific committee members.

MENTEP scientific Committee:
Janet Looney, European Institute of Education and Social Policy -EIESP
Diana Laurillard, University College London, Knowledge Lab
Marco Caliendo, University of Potsdam

European Schoolnet, BE
www.eun.org

CNDP – National Centre for Pedagogical Documentation, FR
www.reseau-canope.fr

CPI – Cyprus Pedagogical Institute, CY
www.pi.ac.cy/pi/index.php?lang=en

CTI – Computer Technology Institute & Press "Diophantus", EL
www.cti.gr

FBK-IRVAPP – The Research Institute for the Evaluation of Public Policies, IT
http://irvapp.fbk.eu

DZS – The Centre for International Cooperation in Education, CZ
www.dzs.cz

Finnish National Agency for Education, FI
www.oph.fi

DGE – Directorate-General for Education, PT
http://dge.mec.pt

HITSA – Information technology Foundation for education, EE
www.hitsa.ee

INDIRE – National Institute for Documentary Research and Innovation in Education), IT
www.indire.it

INTEF – Ministry of Education, Culture and Sport, ES
http://educalab.es/intef

Ministry of Education and Culture, CY
www.moec.gov.cy/en/index.html

Senter for IKT i utdanningen – Norwegian Directorate for Education and Training
https://iktsenteret.no

STIL – Agency for IT and Learning, DK
www.stil.dk

UPC – Education Development Centre, LT
www.upc.smm.lt

ZRSS – National Education Institute, SI
www.zrss.si

About MENTEP

MENTEP (Mentoring Technology-Enhanced Pedagogy) was a major European Research project to boost teachers' competence and confidence to use Information and Communications Technologies (ICT) in the classroom. The project, which run from March 2015 to May 2018, investigated the potential of an online Self-Assessment Tool (SAT) to empower teachers to progress in their Technology-Enhanced Teaching (TET) competence at their own pace.National public authorities and 11,000 teachers in 11 partner countries participated in the project: Cyprus, Czech Republic, Estonia, Finland, France, Greece, Italy, Lithuania, Portugal, Slovenia and Spain. European Schoolnet, a network of 30 ministries of Education in Europe coordinated the project and FBK-IRVAPP, the research institute for the evaluation of public policies in Italy, was responsible for the quantitative evaluation of the project.Participating Teachers completed two short surveys, one at the beginning and one at the end of the school year 2016/2017 about their use of ICT in school. All information provided by teachers was treated anonymously and privacy rights are fully respected. At the end of the project, MENTEP teachers received a certification of participation.

Find out more **http://mentep.eun.org/**

Follow MENTEP

http://mentep.eun.org

#MENTEP

Erasmus+

The MENTEP project is a European Policy Experimentation funded by the European Commission via the Erasmus+ programme. This publication 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 for any use that may be made of the information contained therein.