E-assessment of student-teachers' competence as new teachers

Admiraal, W.; Janssen, T.; Huizenga, J.; Kranenburg, F.; Taconis, R.; Corda, A.

Publication date
2014
Document Version
Final published version
Published in
TOJET

E-ASSESSMENT OF STUDENT-TEACHERS' COMPETENCE AS NEW TEACHERS

Wilfried Admiraal
Leiden University, Leiden, the Netherlands
w.f.admiraal@iclon.leidenuniv.nl

Tanja Janssen
University of Amsterdam, Amsterdam, the Netherlands
(t.m.janssen@uva.nl, j.c. Huizenga@uva)

Jantina Huizenga
University of Amsterdam, Amsterdam, the Netherlands
(t.m.janssen@uva.nl, j.c. Huizenga@uva)

Frans Kranenburg
Utrecht University, Utrecht, the Netherlands
f.a.n.kranenburg@uu.nl

Ruurd Taconis
Eindhoven Technical University, Eindhoven, the Netherlands
r.taconis@tue.nl

Alessandra Corda
Amsterdam University of Applied Sciences, Amsterdam, the Netherlands
a.corda@hva.nl

ABSTRACT
In teacher education programmes, text-based portfolios are generally used to assess student-teachers' competence as new teachers. However, striking discrepancies are known to exist between the competencies reflected in a written portfolio and the competencies observed in actual classroom practice. Multiple assessments should be used to provide a more valid assessment of student-teachers' competence as new teachers. Technology can support this kind of multiple and flexible ways of assessment. In a Research & Development project, four types of e-assessments were designed, implemented and evaluated in 27 interventions in 13 post-graduated teacher education programs in the Netherlands. Teacher educators reported positive outcomes of the interventions in terms of new procedures, materials and tools. No significant effects were found of the implementation of the four types of e-assessments on the evaluation by either teacher educators or student-teachers. A possible explanation for this absence of effects might be teething problems of the interventions implemented.

INTRODUCTION
Assessment and evaluation are increasingly important in all educational sectors. In teacher education programs, text-based self-evaluations are generally used to assess student-teachers' competence as new teachers (Fox, White, & Kidd, 2011; Winsor, Butt, & Reeves, 1999). However, this kind of written self-evaluation does not give valid evidence of teacher competencies that are typically used to guide the curriculum of teacher education programs. Consequently, observation of student-teachers' performance are increasingly used for assessment, such as class observations, teaching materials and tests. Simultaneously, assessment is used for both formative and summative purposes: assessments are not only used to measure student-teachers' competencies, but also to feed back student-teachers which competencies they already possess, in what phase of development they are and how they can acquire teacher competencies. Technology can support this kind of multiple and flexible ways of assessment. The objective of this paper is to provide insight into how multiple e-assessments of student-teachers' competence as new teachers can be designed in an efficient and effective way.

Student-teachers' Competence as New Teachers
In 2005, in response to national and international calls for improved teacher education and greater educational accountability, the Dutch Ministry of Education decided to develop a standard for all teachers in secondary education. Subsequently, a standard was developed resembling the Professional Standards for Teachers in England (http://www.tda.gov.uk/), the National Professional Standard for Teachers in Australia (see http://www.nsw.gov.au/), and the Professional Teaching Standards in the United States (see http://www.nbpts.org/). The Dutch Teacher Standard includes pedagogical, interpersonal, organizational, methodological, relational (colleagues, community), and reflective competencies (see the Association for the Professional Quality of Teachers, http://www.lerarenweb.nl/). The first four competencies (i.e., pedagogical,

interpersonal, organizational, and methodological competencies) can be assessed on the basis of teacher performance in the classroom. While the relational competencies that pertain to colleagues and the community are important, student-teachers usually gain only limited experience with these competencies during their training. All six competencies refer to the professional role of the teacher in three types of situations: working with students, working with colleagues, and working in the school. The seventh competence is reflection, which is seen as important for a teacher's ongoing personal and professional development (Day, 1993; Hatton & Smith, 1995; Korthagen, 1992). All of the seven competencies of the Dutch standard are described according to rubrics of key knowledge, skills and attitudes that teachers must have at various levels. Teacher education programs typically use the competencies outlined in the national standard to guide their curriculum design and assessment. The problem, of course, is how to assess the competencies and thereby demonstrate that teachers meet the required standards.

Assessment of Student-teacher Competence

In the 1980s, written teaching portfolios were introduced into teacher education to stimulate student-teachers to think more carefully about their teaching practices and subject matter (see, for example, Bartell, Kayne, & Morin, 1998; Darling-Hammond & Snyder, 2000; Fox et al., 2011; Winsor et al., 1999; Woodward & Nanlohy, 200a, 2004b). Portfolios are argued to be suited not only for learning purposes but also for assessment purposes as they represent: "a way to define, display, and store evidence of a teacher's knowledge and skills that is based on multiple sources of evidence collected over time in authentic settings" (p. 58) [10]. Student teachers can include, for instance, the following in assessment portfolios: their ideas regarding teaching, summaries of relevant theories, samples of lesson plans, observational notes on their teaching, and reflections upon their teaching practices. While such documents cover a wide range of knowledge and competence, striking discrepancies are known to exist between the competencies reflected in a written portfolio and the competencies observed in actual classroom practice. That is, student-teachers can sometimes present excellent written portfolios while their teaching performance is evaluated by school and university supervisors as rather weak (cf., Darling-Hammond & Snyder, 2001) and (cf., Burroughs, 2001; Uhlenbeck, 2002).

When Delandshere and Arens (2003) analyzed the written portfolios submitted to three teacher education programs in the USA, they encountered major problems with the evidence submitted for assessment purposes. Most of the written portfolios consisted of meta-data (e.g., statements of beliefs, lesson plans, mentor observations, reflections on teaching experiences). In other words, the data was removed from actual practice and thus indirect; the portfolios showed the student teachers' views on classroom events and their beliefs about teaching. As Delandshere and Arens point out, however, the assessment of teaching performance requires direct evidence and thus data on the teacher's actual work in the classroom.

In contrast to such indirect sources of data, video recording allows direct teaching evidence to be included in an assessment portfolio. The use of video recordings allows direct evidence of teaching to be included in a narrative. Compared to written or oral accounts, video narratives are likely to provide information on a wider variety of teacher competencies and more specific information on the contexts in which the competencies are demonstrated. This rich picture of teacher competencies and practices obtained in specific contexts can be assumed not only to provide highly valid information but also can be used for analytic and varied reflection.

There is much empirical work on the use of video for learning, mostly in teacher education (e.g., Bower, Cavanagh, Moloney, & Dao, 2011; Rosaen, Lundeberg, Cooper, Fritzen, & Terpstra, 2009) and in professional development programs with (experienced) teachers (Borko, Jacobs, Eiteljorg, & Pittman, 2008; Rich & Hannafin, 2009). For example, in their evaluation study of the use of video in web-based computer-mediated communication in teacher education, Lee and Wu (2006) found that student-teachers reflect more thoroughly on their teaching, pinpointing the areas of required improvement better, compared to situations in which student-teachers had to rely on their recall of their practices only. Likewise, these authors showed that student-teachers were also willing to share their experiences with and learn from their peers. Moreover, the authors found that – compared to micro-teaching sessions in which student-teachers had to rely on their recall only - peer feedback became more concrete and associated with specific points in the video clips. This feedback was also appreciated more by student-teachers. Finally, watching, analyzing and reflecting upon the video-taped practices of *others* enabled the student-teachers to learn from good teaching models and guard against bad ones. Experiences with how the use of video clips can be further integrated into the professional development of teachers confirm these findings (e.g., Video Clubs in Sherin & Van Es, 2009).

However, due to the lack of empirical studies on video portfolios with teachers or student-teachers for assessment purposes, it is still unclear if the inclusion of direct evidence about the functioning of student-teachers in the classroom facilitates a valid assessment of student-teachers' competence.

e-Assessment of Student-teachers' Competence

The licensing and certification of teachers today is performance-based and thus recognizes teaching as a highly complex, highly contextual, and highly personal activity cf., Darling-Hammond & Snyder, 2000; Moss et al., 2004; Schutz & Moss, 2004). In teacher education programs, performance-based assessment is often supplemented with other information from portfolios, which can include lesson plans, reflections, feedback from students, and feedback from supervisors, superiors and colleagues (Wolf & Dietz, 1998). A portfolio should show not only that the student-teacher knows and understands theory but also that the student-teacher can act in accordance with theory and detect discrepancies between what is taught in theory and what occurs in actual practice.

This complex combination of teacher competencies asks for multiple assessment procedures in teacher education. Technology might support these new, complex ways of assessment. Recent years have been characterized by extensive growth in the use of technology in education, such as virtual learning environments, simulation software, virtual experiments, visualization of complex models as well as tools which enables students and teachers to communicate and collaborate through email, electronic forums, and instant-messaging systems. However, the use of technology in assessment procedure (i.e., e-assessment) is an under-researched area. e-Assessments convey practical benefits such as accessibility of practices, flexibility in updating information, and incorporating multimedia resources (Fill & Ottewil, 2006), in addition to efficiency for both teacher educators and student-teachers. As teaching has been recognized as a highly complex, highly contextual, and highly personal activity, e-assessments might be helpful in order to assess student-teachers' competence as new teachers in an efficient and effective way.

Problem of this Study

The problem of the present study was how multiple e-assessments of student-teachers' competence as new teachers could be designed in such a way that these could be carried out in an efficient and effective way and provide a valid assessment of student-teachers' competence as new teachers. Research questions were:

1. How do interventions on e-assessment affect the use and evaluation of these e-assessments by teacher educators?
2. How do interventions on e-assessment affect the evaluation of these e-assessments by student teachers?
3. How do teacher educators perceive the implementation of the interventions on e-assessment?

METHODS

Research Context

Teacher preparation includes certification at three levels: primary education, lower secondary education (pre-vocational secondary education and the three lower grades of senior general secondary education and pre-university education) and all levels of secondary education. The latter programs are mainly based in research universities and the former two programs are mainly organized by universities of applied sciences.

The context of this study is the post-graduate teaching education program in the Netherlands. Students who graduate are licensed to teach at all levels of secondary education in the Netherlands. Teacher preparation for certification to teach at all levels of secondary education usually takes a one-year full time (or two-years 50% part-time) master program as a follow-up of a master program in a particular school subject (e.g. mathematics or a foreign language). This means that teachers who are licensed to teach at all levels of secondary education have two Masters: one in a school subject or related domain and one in teaching this school subject. The curricula of these teacher education programs exist of 50% courses at the teacher education institution and 50% teaching in school. The common goal of these master programs is to connect theory and practice of teaching in secondary education.

In a Dutch national Research & Development project, Non satis scire (funded by the SURF foundation, http://www.surf.nl/), teacher educators and master students of teacher education programs of all 13 Dutch research universities participated. Teacher educators collaboratively design, implement, and evaluate both formative and summative assessments of student-teachers' competence as new teachers. Four e-assessment types have been addressed: 1) knowledge tests on learning and instruction, 2) providing feedback on students' plans for research on teaching practice, 3) providing feedback on students' web-based video clips of teaching practice and 4) digital self-assessments of student-teachers' reflection.

Design of the Study

In a multiple-case study research design, 27 interventions were carried out, spread over 13 teacher education programs and the four forms of e-assessment (see Table 1). In order to answer research questions 1 and 2, for each type of e-assessment teacher educators and students from the experimental condition (programs that carried

out the particular type of e-assessment) were compared with teacher educators and students from the control condition (i.e., programs that were not part of the experimental conditions). In order to answer research question 3, a multiple case study design was used (Yin, 2014) using multiple data sets about each of the programs.

Table 1. Overview of the design

Intervention	Participating TE programs	
	Experimental condition	Control condition
1. Knowledge tests	4	9
2. Feedback on students' research plans	9	4
3. Feedback on students' video clips	11	2
4. Digital self-assessment	4	9

Data and Procedures

Data were collected of 115 teacher educators and 644 master students from 13 universities. A digital pre-test and post-test questionnaire was administered to teacher educators to evaluate the four interventions on two aspects: 1) the extent to which different forms of e-assessments were used and 2) the extent to which these forms were valued. A similar pre- and post-test questionnaire was administered (on paper) to students from the 13 universities. In addition, observations of work meetings and evaluation reports were used to map teacher-trainers' experiences with the various forms of e-assessment. Finally, all educational materials (study guides, readers, tests, video clips, student reflections, research plan, feedback forms and completed assessment rubrics) were collected and analyzed to support or contradict interpretations from the questionnaire data and work meetings.

Questionnaire for Teacher Educators. In addition to their gender, age, teaching experience and teaching position, teacher educators were asked to evaluate the use of 1) a corpus of shared items of a knowledge test on learning and instruction; 2) digital knowledge tests; 3) peer feedback on research plans; 4) peer assessment on research plans; 5) digital rubrics to support the assessment of research plans; 6) video recording of student-teachers' practices and 7) self-evaluations.

First, we asked teacher educators to indicate the variety of their use of the assessment types. The frequency of use was measured by 2 to 5 yes/no items, with items like, "Did you use the digital corpus of knowledge items?" (Shared test items), "Did students provide written feedback on their research plans?"(Peer feedback) or "Did you provide feedback on the basis of students' video clips of their teaching practice?" (Video).

Second, the evaluation of each of the assessment types was measured using a series of 4 to 7 similar Likert-type scale statements, with *1*= completely disagree to *5*= completely agree. Example items are "The use of digital tests has a positive effect on the time that is needed to feed back the test results (Digital knowledge test), "Peer feedback has a positive effect on the time teachers spend on providing feedback" (Peer feedback), or "The use of web-based video clips of students' teaching practice has a positive effect on students' insight into their own teaching competence" (Video).

In Table 2, the descriptive statistics are presented for the frequency of use and for the evaluation of each of the assessment types. Of the 115 teacher educators, 60 completed both the pre-test and the post-test. The reliability of the seven evaluation scales met our norm of 0.70, for the first scale with only 4 items after using the Spearman-Brown correction for test length.

Table 2. Descriptive statistics teacher-educator questionnaire

	Frequency scale*	Evaluation scale	Cronbach's α	Exp cond N	Contr cond N
Shared test items**	0 – 3	1-5	.58	26	34
Dig. knowl. tests	0 – 2	1-5	.72	26	34
Peer feedback	0 – 5	1-5	.74	52	8
Peer assessment	0 – 3	1-5	.77	52	8
Rubrics	0 – 4	1-5	.82	52	8
Video	0 – 5	1-5	.77	52	8
Self-assessment	0 – 3	1-5	.78	13	37

* 0 = assessment instrument is not used; 2/5 = instrument is used in various ways

** this scale included only 4 items

Questionnaire for Students. In addition to their university, gender and age, students were asked to report their evaluation of 1) digital knowledge tests; 2) peer feedback on research plans; 3) peer assessment on research plans; 4) digital rubrics to support the assessment of research plans; 5) video recording of student-teachers' practices and 6) self-evaluations.

The items of this part of the student questionnaire were similar to those in the teacher questionnaire. For each of the e-assessments types, a series of 4 or 5 statements were used to measure students' evaluation. These statements were answered on a Likert-type scale, with *1*= completely disagree to *5*= completely agree. Example items are "I receive feedback about my test results more timely in the case of a digital test compared to a paper-and-pencil test" (Digital knowledge test), "I can learn a lot from provide providing peer feedback on research proposals" (Peer feedback), or "Supervision using a web-based video clips of my teaching practice is better than supervision on the basis of life observation of my supervisor" (Video).

In Table 3, the descriptive statistics are presented for the evaluation of each of the seven assessment types. The reliability of five evaluation scales met our norm of 0.70. The first scale was excluded from the analyses as the reliability appeared to be low. As shown in Table 2, the distribution of participants in both conditions is strongly skewed, which lowers the chance to find any significant differences between both conditions.

Work Meetings and Evaluation Reports. During the project period of two years two or three teacher educators per teacher education program that participated in the four types of e-assessment interventions attended three work meetings and completed evaluation reports which were used as input for these meetings. The information from the meetings and reports was summarized.

Table 3. Descriptive statistics student questionnaire

	Evaluation scale	Cronbachs α	Exp cond. *N*	Control cond. *N*
Dig. knowl tests*	1-5	--	--	--
Peer feedback	1-5	.79	131	5
Peer assessment	1-5	.76	126	5
Rubrics	1-5	.84	130	5
Video	1-5	.78	109	25
Self-assessment	1-5	.78	5	125

* this scale is excluded because the reliability was too low

Analyses

A mix-method analysis procedure was used. For the questionnaire data, repeated measures analyses were used to examine possible differences in evaluation before and after the interventions. In these analyses, each intervention condition was compared with the three other forms of e-assessment (which form the control condition). The qualitative data in the written protocols of the work meetings and evaluation reports were combined into a thick description (Geertz, 1973) of each of the 27 interventions indicating teacher educators' self-reported experiences with the particular form of e-assessment.

RESULTS

Use and Evaluation by Teacher Educators

The results of the repeated measures analyses of variance for teacher educators are summarized in Table 4 (frequency of use) and Table 5 (evaluation).

Table 4. Results for teacher educators: frequency of use of assessment procedure (means and standard deviations between brackets)

	Experimental condition		Control condition	
	Pre-test	Post-test	Pre-test	Post-test
Shared test items	1.6 (1.4)	1.4 (1.4)	0.8 (1.2)	1.1 (1.4)
Dig. knowl. tests	0.2 (0.5)	0.3 (0.6)	0.1 (0.2)	0.1 (0.2)
Peer feedback	2.2 (1.8)	2.3 (1.8)	0.1 (0.4)	0.6 (1.2)
Peer assessesment	0.4 (0.9)	0.4 (1.0)	0.0 (0.0)	0.0 (0.0)
Rubrics	2.2 (1.7)	2.2 (1.7)	0.1 (0.4)	1.0 (1.9)
Video	1.8 (1.5)	2.0 (1.5)	0.5 (0.5)	1.3 (1.6)
Self-assessment	0.8 (0.4)	0.8 (0.4)	0.8 (0.8)	0.9 (0.7)

Table 5. Results for teacher educators: evaluation of assessment procedure (means and standard deviations between brackets)

	Experimental condition		Control condition	
	Pre-test	Post-test	Pre-test	Post-test
Shared test items	3.6 (0.6)	3.3 (0.6)	3.5 (0.5)	3.2 (0.5)
Dig. knowl. test	3.2 (0.3)	3.1 (0.7)	3.1 (0.6)	3.0 (0.5)
Peer feedback	3.6 (0.5)	3.4 (0.5)	3.8 (0.3)	3.5 (0.5)
Peer assessment	3.2 (0.6)	3.2 (0.4)	3.7 (0.4)	3.5 (0.5)
Rubrics	3.5 (0.5)	3.5 (0.6)	3.9 (0.1)	4.0 (0.3)
Video	3.2 (0.6)	3.2 (0.6)	3.1 (0.4)	3.1 (0.6)
Self-assessment	3.6 (0.4)	3.6 (0.6)	3.4 (0.5)	3.4 (0.5)

Note. Scale is *1* =totally disagree, *5* =totally agree that the particular e-assessment has a beneficial effect

The analyses did not show a significant increase in teacher educators' use of the particular assessment procedure, compared to the control condition (consisting of programs that did not use the particular e-assessment form). As shown in Table 4, teacher educators in the intervention condition did generally differ in their use of the particular assessment form from the control condition, but these differences already existed a priori (with all Fs< 1.71 and all ps>.20). It appears that teacher educators apparently decided to participate in the interventions that included the assessment form they already used in their regular practice. A marginal trend was found for the use of a digital knowledge test ($F(1,58)= 3.50$; $p= 0.06$) indicating that teacher educators in the experimental condition tended to increase their use of a digital knowledge test after the intervention, compared to teacher educators from the control condition.

In Table 5, the results are summarized for the evaluation of the e-assessment types by teacher educators. Again, no differences were found between the experimental and control conditions, indicating that teacher educators from the intervention condition generally did not evaluate the e-assessment forms differently, compared to the other teacher educators (with all Fs <0.25 and all ps >.62). Finally, no significant correlations were found between the use of the assessment types by teacher educators and their evaluations of the particular form of e-assessment (with all rs < .25).

Evaluation by Student-teachers

In Table 6, the results of the repeated measures analyses on the data of the master students are summarized. No significant differences were found between students from the experimental and control condition on the evaluation of the e-assessment types (all Fs < 1.85 and all ps >18). A marginal trend was found for the evaluation of peer feedback ($F(1,134)= 3.35$; $p= 0.07$) indicating that students in the experimental condition generally tended to report a negative evaluation of peer feedback after the intervention, compared to students from the control condition. Generally, students from the experimental condition tended to show lower evaluation scores after the intervention with respect to all types of assessment, compared to the pre-test and compared to students from the control condition. It should be noted that the distribution of numbers of students in the experimental and in the control conditions is strongly skewed. In order to decrease this skewedness, students' practice of the particular e-assessment (yes/no) was used to define the experimental en control condition. Although this increased the number of students in the control condition (i.e. students who were part of an intervention, but did not practice the particular assessment), similar results were found as shown in Table 6.

Table 6. Results for master students: evaluation of assessment procedures (means and standard deviations between brackets)

	Experimental condition		Control condition	
	Pre-test	Post-test	Pre-test	Post-test
Peer feedback	3.5 (0.5)	3.3 (0.6)	3.6 (0.3)	3.9 (0.6)
Peer assessment	3.4 (0.6)	3.2 (0.7)	3.3 (0.9)	3.6 (0.7)
Rubrics	3.6 (0.7)	3.4 (0.8)	3.5 (1.1)	3.8 (0.6)
Video	4.0 (0.5)	3.8 (0.7)	3.9 (0.4)	3.7 (0.7)
Self-assessment	3.8 (0.2)	3.6 (1.2)	3.5 (0.6)	3.8 (0.6)

Note. Scale is *1* =totally disagree, *5* =totally agree that instrument has a beneficial effect

Teacher-educators' Perceptions of the e-Assessment Interventions

In Table 7, the results of the qualitative analyses of the work meetings and evaluation reports of the teacher educators are summarized. These analyses show the particularities of using the four forms of assessments. One of the results from the analysis of the educational materials was that teacher educators used the assessments in a formative way, instead of or in addition to summative assessments. This result aligns with observations from

Admiraal, Van Duin, Hoeksma, and Van de Kamp (2011) that teacher educators strongly prefer the role of mentor or coach, guiding students during their learning process, instead of the role of assessor, which includes judging the quality of students' competence. Moreover, many educational and procedural outcomes can be distinguished such as the setup of a digital repository of test items, quality improvement of knowledge tests, and procedures and rubrics for peer feedback on research plans and for feedback and assessment of web-based video of teaching practices.

DISCUSSION AND CONCLUSION

Assessment procedures and criteria were developed and evaluated for testing student-teachers' knowledge of teaching, for assessing a written research proposal using peer feedback, peer assessment and rubrics, for judging video clips of teaching practices and student-teachers' self-evaluations. Although teacher educators reported positive outcomes of the interventions in terms of e-assessment procedures and tools (research question 3), no significant effects were found of the implementation and the evaluation of these procedures and tools (research question 1).

Teacher educators did use a particular type of assessment significantly more in the experimental condition than in the control condition, but these differences already existed a priori. So, it seems that teacher educators participated more in the type of assessment they already used before the intervention started. Student-teachers showed a less positive evaluation of the assessment type after the intervention than at the beginning and compared to the students in the control condition, although differences were not significant (research question 2). It might be that most interventions in the teacher education programs involved in this study were in a so-called experimental phase, showing teething problems in the implementation of the assessment procedures, materials and tools. This would explain why teacher educators are quite positive about the educational outcomes of the study reporting new procedures, materials and tools that were absent before.

Table 7. Results from the qualitative analyses of the work meetings and evaluation reports

Shared tests and test items	Sharing the knowledge tests - used in the various training institutes - was evaluated positively by all participants. Participants reported that they reflected more on good ways of testing and how to improve test items
Digital knowledge tests	Participants indicated that they wished to experiment further with digital testing. Digital testing appeared to be especially advantageous for larger training institutes. However, within these institutes organizational hindrances (i.e. lack of large enough computer rooms) were also reported.
Peer feedback	One participant reported that the developed peer feedback procedure had helped to diminish the workload of teacher-trainers in evaluating research plans written by students. Two other participants indicated that the procedure had a beneficial effect on students' study progress. All participants agreed that peer feedback had an added value for the assessment of research plans.
Peer assessment	Participants agreed that (summative) peer assessment of students' research plans was not feasible, because of the extra workload for students and teacher-trainers. Participants also doubted the quality of students' assessments.
Rubrics	Participants agreed that using rubrics for peer feedback helped to make the assessment criteria more transparent for students and teacher-trainers, and helped to improve the quality of the feedback.
Video	Three findings were reported, on which participants agreed: - Much attention needs to be paid to the technological and organizational aspects before video can be adequately used as an instrument to assess students' classroom practices. - According to participants video cannot replace live observation of classroom practice; rather, video is seen as complementary. Usually, video is used for formative and not for summative assessment. - Discussions of video recordings and feedback on classroom practice should take place in a safe environment (teacher-student, or in small groups)
Self-assessments	According to participants students need help to be able to reflect on their classroom practice and competencies as new teachers. (Digital) self-assessment instruments can be used, but need to be properly "framed" in the curriculum.

Limitations

As this project was carried out as a Research & Development project aimed at the implementation of e-assessments in teacher education, some limitations of the research design should be mentioned here. Firstly,

there might be a bias of self-selection. Teacher education institutes chose to implement two to three interventions with e-assessment in their programs, which means that all teacher educators and students of a particular program participated in the experimental condition that was connected to the particular e-assessment form of their institute. So, the self-selection was on the program level instead of the individual level, and therefore we think that potential confounding effects are quite minimal. Secondly, due to this self-selection of teacher education programs, the distribution of participants in the experimental and control condition was highly skewed, except for the self-assessment intervention. This considerably decreased the power of our analyses and might therefore explain why no significant differences were found between participants of the experimental and control conditions. Thirdly, self-reports of implementations and evaluations were used instead of registration measures such as observation or performance tests. Teacher educators could have under- or over-estimated their use of a particular e-assessment form, although no differences were found in their evaluation of the e-assessment forms. It might be that teacher educators over-estimated their implementation of e-assessment forms as most of them knew they were part of a R&D project that had the aim of stimulating the use of particular e-assessment forms.

Implications for Teacher Education

In the next years, the procedures and criteria that were designed, implemented and evaluated in the current project should be re-designed and re-tested in order to be used as input for curriculum changes in teacher training programs. As we mentioned earlier, teething problems might have explained why the interventions were not evaluated positively. Some interventions were not fully developed at the time of the evaluations and in some programs the infrastructure did not fully support the interventions (absence of a web-video server or no large computer rooms to administer the digital tests). Recent research on the technical infrastructure of teacher education program in the Netherlands (Admiraal, Lockhorst, Smit, & Weijers, 2013) showed a quite conventional picture: basic technology such as computers, WiFi, electronic whiteboards, virtual learning environments and presentation software was available, but not commonly used, and more advanced or innovative technology was less available. So, future pedagogical interventions in the domain of e-assessment in teacher education should concur with a supportive technological infrastructure.

REFERENCES

Admiraal, W., Hoeksma, Kamp, M.-T. van de, & Duin, G. van. (2011): Assessment of teacher competence using video portfolios: Reliability, construct validity and consequential validity. *Teaching and Teacher Education, 27*, 1019-1028. doi:10.1016/j.tate.2011.04.002.

Admiraal, W., Lockhorst, D., Smit, B., & Weijers, S. (2013). The Integrative Model of Behavior Prediction to explain technology use in post-graduate teacher education programs in the Netherlands. *International Journal of Higher Education, 4*(2), 172-178, doi:10.5430/ijhe.

Bartell, C., Kayne, C., & Morin, J. A. (1998). Teaching portfolios in teacher education. *Teacher Education Quarterly, 25*, 23-32.

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education, 24*, 417–436. doi:10.1016/j.tate.2006.11.012.

Bower, M., Cavanagh, M., Moloney, R., & Dao, M. M. (2011). Developing communication competence using an online video reflection system: Pre-service teachers' experiences. *Asia-Pacific Journal of Teacher Education, 39*, 311-326. doi: 10.1080/1359866X.2011.614685.

Burroughs, R. (2001). Composing standards and composing teachers. *Journal of Teacher Education, 52*, 223-232. doi: 10.1177/0022487101052003005.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education, 16*, 523-545.

Day, C. (1999). Reflection: A necessary but not sufficient condition for professional development. *British Educational Research Journal, 19*, 83-93. doi:10.1080/0141192930190107.

Delandshere, G., & Arens, S. A. (2003). Examining the quality of evidence in pre-service teacher portfolios. *Journal of Teacher Education, 54*, 57-73. doi: 10.1177/0022487102238658.

Fill, K., & Ottewil, R. (2006). Sink or swim: Taking advantage of developments in video streaming. *Innovations in Education and Teaching International, 43*, 397–408.

Fox, R. K., White, C. S., & Kidd, Js, K. (2011). Program portfolios: Documenting teachers' growth in reflection-based inquiry. *Teachers and Teaching: Theory and Practice, 17*, 149-167. doi: 10.1080/13540602.2011.538506.

Geertz, C. (1973). *The interpretation of cultures.* New York: Basic Books.

Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education, 11*, 33-49.

Korthagen, F. A. J. (1992). Techniques for stimulating reflection in teacher education seminars. *Teaching and Teacher Education, 18*, 265-274.

Lee, G. C., & Wu, C. (2006). Enhancing the teaching experience of pre-service teachers through the use of video in web-based computer-mediated communication (CMC). *Innovations in Education and Teaching International, 43*, 369–380. doi: 10.1080/14703290600973836.

Moss, P. A., Sutherland, L.-M., Haniford, L., Miller, R., Johnson, D., Geist, P. K., et al. (2004). Interrogating the generalizability of portfolio assessments of beginning teachers: A qualitative study. *Education Policy Analysis Archives, 12*(32).

Rich, P. J., & Hannafin, M. (2009). Video annotation tools: Technologies to scaffold, structure, and transform teacher reflection. *Journal of Teacher Education, 60*, 52-67. doi: 10.1177/0022487108328486.

Rosaen, C. L., Lundeberg, M., Cooper, M., Fritzen, A., & Terpstra, M. (2009). Noticing noticing: How does investigation of video records change how teachers reflect on their experiences? *Journal of Teacher Education, 59*, 347-360. doi: 10.1177/0022487108322128.

Sherin, M. G., Van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*, 20-37. doi: 10.1177/0022487108328155.

Schutz, A. M., & Moss, P. A. (2004). Reasonable decisions in portfolio assessment: Evaluating complex evidence of teaching. *Education Policy Analysis Archives, 12*(33).

Uhlenbeck, A. (2002). The development of an assessment procedure for beginning teachers of English as a foreign language. Dissertation. Leiden, the Netherlands: University of Leiden.

Winsor, P. J. T., Butt, R. L., & Reeves, H. (1998). Portraying professional development in preservice teacher education: Can portfolios do the job? *Teachers and Teaching: Theory and Practice, 5*, 9-31.

Wolf, K., & Dietz, M. (1998). Teaching portfolios: Purposes and possibilities. *Teacher Education Quarterly, 25*, 9-22.

Woodward, H., & Nanlohy, P. (2004a). Digital portfolios: Fact or fashion? *Assessment and Evaluation in Higher Education, 29*, 227-238. doi: 10.1080/0260293042000188492.

Woodward, H., & Nanlohy, P. (2004b). Digital portfolios in pre-service teacher education. *Assessment in Education, 11*, 166-178. doi**:** 10.1080/0969594042000259475.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage Publications.