http://mentep.eun.org/

D2.1 Literature Review
Online self-assessment with MENTEP:
A review of the literature

Janet Looney,
European Institute of Education and Social Policy

July 2015

This project has been funded with support from the European Commission.

Table of contents

This project has been funded with support from the European Commission.

1. Assessment and online learning

The Mentoring Technology Enhanced Pedagogy (MENTEP) project aims to develop a reliable, user-friendly and sustainable tool for teachers to self-assess progress in technology-enhanced teaching (TET) competence. TET competence refers to proficiency in using ICT in teaching, applying pedagogic and didactic judgment, and awareness of implications for learning. MENTEP will support teachers as they learn in complex, real world, or 'authentic' situations and where there are no predetermined or 'right' answers.

Teacher self-assessment is an integrated part of the learning process in this kind of authentic learning environment. Online self-assessment is iterative: learners set goals, test ideas, monitor progress and adjust strategies (a formative approach). Tools to facilitate online self-assessment include e-portfolios, rubrics and checklists setting out criteria to identify progress and areas for improvement.

This paper reviews relevant literature on online self-assessment. The next section describes the parameters used to identify relevant literature for this review. This is followed by a brief discussion of online self-regulated learning and Internet self-efficacy, which is fundamental for learners working in an autonomous online environment. The next sections, which are at the core of this review, discuss research on validity and reliability of online self-assessment in authentic learning environments. Self-assessment in MENTEP is particularly challenging because TET competence is such a new area and there are few exemplars against which learners will be able to gauge their progress. Peer assessment is one potential way strengthen the validity and reliability of learner self-assessments, so this aspect is an important focus for this review. The paper also describes online platforms that integrate different self- (and peer-) assessment methods. We include separate sections on rubrics to support self- and peer-assessment, learning analytics and game-based learning and assessment. The paper concludes with a brief summary of findings important for the MENTEP to consider in the design phase.

2. The search parameters

We consulted three online databases – EBSCO, ERIC and JSTOR – to identify relevant literature for this review. Keywords used for the searches included: teachers, online learning, self-assessment, self-evaluation, professional development, self-regulated learning, peer-assessment, rubrics, validity and reliability, learning analytics, online games, and quality. We then reviewed abstracts of articles identified in these searches to filter out non-relevant articles (e.g. articles focused on assessment of learning with pre-determined outcomes and on summative assessment). Additional studies were identified through relevant citations in the studies identified in the first round (snowball method).

This project has been funded with support from the European Commission.

Our main concern was to identify literature addressing the validity, reliability and usability of online tools for self-assessment of open, or authentic learning environments. Some interpretation regarding relevance of different studies was necessary, as we did not find any studies addressing the precisely the same population (in-service teachers) or the same study questions as in MENTEP (technology-enhanced teaching). However, we did identify studies on online learning programmes for pre-service teachers and other university-level learners, as well as a few studies of self-assessment among medical students, business leaders, and upper-secondary students engaged in authentic learning that provided useful insights. Although the focus of MENTEP is on self-assessment, as noted above we have extended our search to include peer-assessment. This approach can reinforce and strengthen teacher learning and self-assessment.

3. Learner self- regulation and self-efficacy in the online environment

The capacity to regulate one's own learning (i.e. self-regulated learning) is vital for success in online learning environments. While MENTEP is intended to build teacher's capacity to use technology to enhance learning, teachers using this platform will need to have a certain level of self-efficacy in order to work independently (Liang and Tsai, 2008, Steffens, 2006) as well as collaboratively in an online environment. 'Self-efficacy' refers to an individual's judgment about his/her ability to organize and actions necessary to attain goals. It is determined by previous positive experiences with similar challenges, and may be built up with time and experience (Bandura, 1994; Dweck, 2006].

Steffens (2006) notes that models of self-regulation typically include:

1) goal setting
2) monitoring of learning processes and strategies
3) feedback, and
4) self-assessment (in some models this includes the ability of the learner to interpret the outcomes of their efforts and adjust strategies).

Several empirical studies confirm the importance of these elements. For example, Elliot and Sheldon and Eliot (1998) have found that learners who define goals for learning that are personally meaningful are more effective at regulating their level of motivation and effort (see also see also Zimmerman, 1995, 1998a; Butler, 1998; Zeidner, Boekaerts & Pintrich, 2000). Lee, Choi and Kim (2013)[1] found that the most significant factors associated with completion of online courses were internal locus of control (i.e. ability to set personally meaningful goals) and metacognitive self-regulation skills (i.e. the second element of the above prototype -- learning processes and strategies).

[1] *N* = 169 adult online learners

This project has been funded with support from the European Commission.

Dettori et al. (2006), in a study of teacher trainees[2], found that these learners have strong skills for self-regulation of individual work, but often lack these skills in collaborative work (perhaps, the authors speculate, because of the lack of emphasis on collaboration in many school systems). They observed that the online social climate that developed in the context of their study evolved and matured; individuals developed skills for collaboration, were more likely to benefit from feedback (the third element of the model) and better able to manage their own learning. As they put it, "…not only individuals, but also the virtual community created in the course was able to self-regulate its own learning" (Dettori et al., 2006, p. 411).

The fourth element of Steffens' prototype highlights the importance of self-assessment (with support from peer- and expert-assessment). Mao and Peck (2013) suggest that self-regulated learning and assessment are mutually dependent. The approach is formative: learners track progress against criteria (e.g. in the form of a rubric outlining the characteristics of quality work) or exemplars, interpret feedback (i.e. from peers), identify areas for improvement and develop strategies to adjust their work (Butler, 2000; Black and Wiliam, 1998). In authentic learning contexts, learners approach problems iteratively, testing and revising assumptions regarding the most effective approach (Barak, 2013). This iterative approach may also scaffold the learning process, and support learners in building their self-efficacy and motivating them.

4. Validity, reliability of authentic assessments

'Validity' and 'reliability' are core principles of effective assessment. Validity refers to the degree to which assessments and evaluations measure what they are intended to measure. Reliability refers to the consistency and stability of results. To this, we might add 'usability', referring to the ease with which results may be interpreted and used to make improvements.

In the context of 'authentic' or 'open' learning (i.e. where there is no pre-determined answer), Gielen et al. (2003) argue that psychometric criteria to gauge validity should be expanded to include capacity to assess authentic and complex tasks, and the impact of tasks on students' learning or learning strategies. Linn, Baker and Dunbar (1991) suggest that the validity should be considered in terms of "consequence, impartiality, transference, content coverage, cognitive complexity, significance, judgment, cost, and efficiency" (p. 3). They argue that reliability (i.e. stability of results) is difficult to achieve with authentic assessment – although several more recent studies of online assessment of authentic learning have developed approaches that support strong inter-rater reliability (below).

There are some doubts, however, about the extent to which learners can be objective about their own performance. Kruger and Dunning (1999) found that low-achieving students were more likely to over-estimate their capacity, although students who received a short training

[2] *N*= 94 prospective teachers

This project has been funded with support from the European Commission.

to improve their metacognitive skills assessed their performances more accurately[3]. Sitzman et al. (2010) found that the quality of self-assessment is related to the learner's existing knowledge level, as well as opportunities to react to feedback[4]. Galbraith et al. (2008) found that medical students (who are presumably high-achieving students) frequently make inaccurate assessments of their competences.

It should also be noted that many of the studies on learners' capacity for accurate self-assessment are limited to single observations. Learners' self -assessment skills may improve with experience and over time (Brown et al., 1997 – cited in Jonsson et al., 1997; Galbraith et al. 2008). But, as Barber et al. (2015) suggest, skills for meaningful self-(and peer-) assessment are best developed through intentional learning processes[5]. For example, skills for effective feedback involve more than the 'good job' comments a novice assessor might give; rather, effective assessment is focused on the task at hand and includes specific ideas for improvement. Assessors also need a good understanding regarding levels of quality for the given task (Black and Wiliam, 1998). A shared understanding of criteria set out in rubrics and checklists is essential.

4.1 External Validation of Self-Assessment

Olfos and Zultaney (2007) recommend that online discussions, tests, and external assessments (by a tutor, peers or an expert) of the e-learners process and products[6] can help to improve the validity of assessments in web-based courses. Galbraith and colleagues (2008) recommend that self-assessments could be further strengthened through periodic external validation of the students' self-assessments. Barber et al. (2015) note that authentic learning environments include access to expert thinking and modeling of processes. Authentic context and tasks, opportunities to understand multiple perspectives and collaborative construction of knowledge opportunities for reflection, coaching and scaffolding are essential for effective online learning. Exemplars, tests with automated feedback and structured peer assessment (e.g. through rubrics and checklists) may also strengthen learners' self-assessment.

[3] Kruger and Dunning conducted four studies, each with tests to assess participants capacity to self-assess their performance on tests related to humour (Study 1 – *N* = 65 university students), logical reasoning (Study 2, *N* = 45 university students; and Study 4, *N* = 140 undergraduate students) and English Grammar (Study 3, *N* = 84 university students). In Study 4, half of the students received a short training to improve their logical reasoning skills. Students who received the training improved their metacognitive skills and their self-appraisals were more accurate.

[4] This meta-analysis included 166 studies, with 222 independent samples of 41,237 learners (75% of which were university students, 21% were employees, and 4% were military personnel.)

[5] *N* = 35 participants from including individuals from the fields of education, nursing and health care, gaming, and business.

[6] *N* = 28 teachers (13 primary school teachers and 15 secondary school math teachers – 5th to 10th grades)

This project has been funded with support from the European Commission.

4.1.1 Exemplars

Steffens (2006) describes references to external experts and exemplars as 'cognitive apprenticeship'. In an online learning environment, external guidance might include procedural facilitation, or suggestions about how to proceed if problems are encountered. We suggest that peers participating in online discussions might also perform this function. Exemplars also help to make criteria and standards delineated in rubrics more meaningful.

4.1.2 Online Tests (Or Quizzes) With Automated Feedback

Although interactive tests most frequently are used to bring learner self-assessment in line with summative assessments, we imagine that tools with automated feedback could also be used to improve learner understanding of quality criteria to assess authentic learning. For example, Jonsson et al. (2007) describe an interactive examination they have developed to help university learners to bring their self-assessments (based on a likert-like questions) more closely in line with their instructors' assessments or examination results[7]. Students who submitted an assessment of their competence received a document describing how an expert in the field would deal with the same task; students reported that they benefited from this exercise.

4.1.3 Peer Assessment

Peer assessment is perhaps the most robust approach to external assessment of competence levels in an authentic learning environment. Tsai (2012) notes that online peer assessment may improve cognition and metacognition, social and thinking skills (see also Topping, 1998; Tsai, Lin & Yuan, 2002[8])." Tsai (2012)[9] and Mao and Peck (2013) found that learners who used both self- and peer-assessment scored higher than those used only self-assessment[10]. Yu and Wu (2013) found that assessors also benefit as they use critical thinking skills to evaluate the quality of peers' work and also develop their social and argumentation skills[11]. Vickerman (2009) suggests that learners engaging in peer assessment may gain insight on the quality of their own work.

Interestingly, Li, Liu and Zhou (2012) report that there is no clear relationship between the quality of peer feedback that learners receive and their performance on final projects[12]. The learner being assessed may receive both good and misleading feedback. In their study, learners who accepted more constructive feedback and ignored the misleading comments

[7] The study included first year student teachers in science and mathematics (N = 171) and second year dental students (*N*= 34). (n=166). All participants were first-time users of the Interactive Examination.

[8] *N* = 24 pre-service teachers

[9] *N* = 45 graduate students in education.

[10] *N* = 264 participants in an online general education course.

[11] *N* = 233 students

[12] *N* = 42 undergraduate teacher education students

This project has been funded with support from the European Commission.

produced higher quality final projects. In other words, the learner's ability to distinguish between useful and less useful feedback is vital. Li, Liu and Zhou hypothesise that learners who receive peer feedback that does not match their own understanding will need to draw on their critical thinking skills, and as a result reinforce their own learning (see also Hattie, 2008). Strijbos et al. (2010) found that "concise general feedback from…low-competence assessors resulted in positive learning gains" for the assessee. The high-competence assessees were less reliant on specific feedback[13]. Yu and Wu (2013) suggest that when feedback contradicts previously held ideas ('cognitive conflict'), the learner is required to adjust or justify their choices.

Although feedback that provides specific suggestions for improvement is generally considered as more effective (Black and Wiliam, 1998), Mao and Peck (2013) found that self- and peer assessments providing either informational or detailed feedback did not immediately lead to improvements in learners' work. However, in subsequent collaborative exercises, learners who had received peer assessments outperformed those who relied on self-assessment only. Mao and Peck hypothesis that peer assessment may trigger more active reflection, or it may be that engagement with peers in collaborative tasks may reinforce the feedback dialogue.

At the same time, it's important to note that findings on the validity and reliability of peer assessment are mixed. Chang (2002) found that sources of score variation among expert raters (meaning that reliability is low) include subjective scoring methods, low number of portfolio entries and variation of portfolio contents, etc.[14] Chang and Wu (2012) suggest that the reliability of portfolio assessment may be improved with well-defined scoring criteria, improved familiarity of raters' with the scoring rubric[15].

As noted above, inter-rater reliability in authentic assessment contexts is difficult to achieve. A few researchers have suggested that it is possible to improve reliability in an authentic learning context by increasing the number of raters (Sulzen, Young and Hannifin 2008[16], Bouzidi and Jaillet (2009) found that having at least four peer assessments increased reliability, and comparing self-assessment with peer assessment improved the validity of feedback. Peer feedback in this study was *not* based on a rubric, so presumably, both validity and reliability could be increased with a well-designed tool[17]. Pollitt (2004) has developed a tool to facilitate progressive comparative judgements of pairs of portfolios (explored further below).

---

[13] *N* = 89 graduate student participants, four experimental and a control group. The experimental groups received a scenario with scenario with concise general (CGF), or one with more elaborated specific (ESF) feedback provided by peers with either high or low competence.

[14] Survey and interview of web-based portfolio users. *N* = 35 pre-service teachers participating in a course on "Computers and Instruction".

[15] *N* = 72 eleventh graders participating in a course on "Computer Application"

[16] *N* = 4 teacher educators who used a rubric to rate student teacher E-folios, with half the rater's area of subject expertise and half in a different area

[17] *N* = 242 students

This project has been funded with support from the European Commission.

Yu and Wu (2013) conducted a study on the individual and combined effects of quantitative and descriptive feedback in online peer-assessment systems. They found that both forms of feedback predicted the quality of work produced, and that the descriptive feedback explained more of the variance in the quality of work than quantitative ratings.

'Collective intelligence', which refers to the shared intelligence that emerges through collaboration, is also a source of information and feedback for learners. Tsai (2009) found that learners with more sophisticated epistemological beliefs were more likely to make significant improvements to their work and to also provide higher-quality comments to their peers[18]. In a separate study, Tsai (2012) investigated views on 'epistemic relativism' versus 'social relativism' (the former is defined as 'recognition of the diversity of knowledge perspectives in a research issue', and the latter as 'an understanding of the multiplicity of peer perspectives'). She found that while both perspectives contribute to learner performance, learners who valued 'collective' approaches were also better able to learn from peer-assessment. High Internet self-efficacy was associated with the development of 'social relativism' and learning from online peer assessment.

4.1.4 Learning analytics: tools to track progress and guide learning

There is a great deal of interest in the potential of 'learning analytic' tools to support online learning. The term 'learning analytics' refers to the '….measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" (Buckingham Shum & Ferguson, 2012, p. 4). These tools are still, however, in early stages of conceptualisation and development.

Spector (2013) describes 'annotated and dynamic concept map technologies', which prompt learners addressing challenging or ill-structured problems – for example in engineering design or technology integration – to conceptualise the 'problem space' and compare their approach with that of an expert addressing the same or similar problems. The dynamic concept map may point out similarities and differences in how learners are approaching problems, and encourage the learner to consider alternative approaches or to focus on overlooked aspects of the problem (see also Pirnay-Dummer, et al., 2010). Spector also refers to 'stealth assessment' for groups of learners working online to solve problems. With this method, an analysis of log data can be conducted to identify progress as well as gaps. The stealth assessment system might then prompt the learner(s) to consider alternatives. According to Spector and colleagues, both stealth assessment and dynamic concept maps have been shown to be effective for a variety of learning situations, although further development and clear data on impact are needed.

In another study, Wang et al. (2011) note that visual representations are useful for representing sequences, alternatives, choices and pathways that are necessary for knowledge

[18] *N* = 45 graduate education students.

This project has been funded with support from the European Commission.

construction. They describe a knowledge map they have developed to provide learners with automated feedback, correct answers and detailed explanations. The map is linked to relevant resources and assessment materials and discussions and related concepts. The system monitors learner progress and provides guidance on possible next steps.

The tools described by Spector (2013) and Wang et al. (2011) refer to learning where there are already expert exemplars or where there are 'right' answers, these approaches could conceivably be modified and adapted to MENTEP. Learners might find ways to link with peers working on similar pedagogical problems, and, as the platform develops, exemplary work of peers may also be shared online.

5. Rubrics

Rubrics are particularly well suited to assessment in 'authentic learning' contexts. They typically include:

- categories for which performance is to be rated
- definitions and examples illustrating the element being rated
- a rating scale for each component

(Perlman, 2002)

Evidence on the impact of different types of rubrics appears to be sparse. Chang and Wu (2012) suggest that although rubrics clearly increase reliability and validity of portfolios to some degree, more research is needed on the most effective approaches. They also suggest that multiple rubrics may be needed for different types of portfolios, and validity and reliability of the different approaches need to be examined.

A few rubrics designed to support different aspects of online learning and teacher professional development and with very different approaches are described below.

Holistic versus analytic, general versus specific approaches

Johnason et al. (2003) categorise rubrics as holistic or analytic. They set out four levels for the holistic approach: proficient, developing, emerging, and not yet. They also describe an analytic rubric they developed, which has a five-point scale and a detailed checklist of observable behaviours with examples of artifacts which, they note, improved inter-rater reliability. In another example of the analytic approach, Perlman (2000) describes a rubric developed for a science lab which includes sections for hypothesis, procedures, results, and conclusion. These categories may also be relevant for teachers piloting innovative methods.

Views on whether it is better to outline general or more specific skills in rubrics are mixed, and depend at least in part, on the purposes of the assessment. Saiz et al. (2015) describe

their efforts to strengthen an online programme to teach critical thinking skills[19] -- a subject where there are no 'right' answers. An early version of this programme was to found to have a positive but modest impact on learning. Saiz et al. subsequently modified the programme to include activities that were structured by rubrics, and which involved less direct instructional intervention. For example, a rubric on argumentation includes categories to rate the structure of the learner's argument and categories on the precision in drafting of ideas, identification of fundamental elements and relevant observations. The authors found that the greater specificity provided by this and other rubrics included in the programme, along with increased student autonomy, were the key variables associated with significantly improved outcomes.

Seery et al. (2012), on the other hand, in their discussion on development of an e-portfolio to support design students (also an area with no 'right' answers), argue that criteria that 'over-define' the design process undermine the experience of exploration and the development of strategies to address problems. They set out very general categories for skills (e.g. diversity of skills), criteria (e.g. within[20] and beyond the module) and problem solving (e.g. frequency of problems identified, sophistication of the problems and success in addressing them). These general categories are intended to capture the design process and to also value diversity of ways in which learners demonstrate capabilities and achieve excellence.

Seery et al. (2012) rely upon an adaptive comparative judgment (ACJ) system to rank portfolios based on peer assessment of overall portfolio quality (see also Pollitt, 2004). They see the peer assessment process of identifying quality work as a way of taking ownership of the assessment process. (Their approach to self- and peer-assessment with the ACJ is described in more detail in the section integrated approaches to online learning below.)

A rubric to structure online conversations

High-quality asynchronous online conversations may provide important opportunities for learners to explore ideas and concepts, as well as to give and receive feedback. For online learners, these conversations, which replace facilitated classroom discussions, may include email, listservs, discussion forums and bulletin boards. Gilbert and Dabbagh (2005) found that guidelines to assist facilitation and assessment may lead to higher quality student posting and more meaningful discourse[21]. Their study implements an assessment rubric to assign ratings (excellent, good, average and poor) and criteria (timeliness of contributions, responsiveness and demonstration of understanding of the assigned readings, and the ability to follow the discussion protocols) for postings in weekly discussions. Gilbert and Dabbagh found that

[19] *N* = 144 students first year undergraduate Psychology students.

[20] *N* = 137 pre-service teachers in Materials and Engineering Technology (Metal) and the Materials and Construction Technology (Wood) programmes

[21] *N* = 87 students participating in a graduate course on "Instructional Technology Foundations and Learning Theory". Twelve online transcripts across four semesters were analysed.

This project has been funded with support from the European Commission.

learners using the rubric were more deliberate about their participation in online conversations. They also found that timeliness of contributions was the most significant rubric criteria to impact quality of online discussions (and we might note that timeliness of feedback has also been identified as having a significant impact on learner performance in research on formative assessment methods (see Shavelson *et al.* 2008, Wiliam 2006)).

A rubric to support teacher self-assessment and reflective practice

Rubrics may also support teachers' reflective practice (a metacognitive approach). Matsumura et al. (2006) describe the Instructional Quality Assessment (IQA) rubric, which includes a number of features are relevant for MENTEP (although it was not initially developed for online use). The IQA rubric, which has been evaluated as having high validity and reliability, builds on evidence of the features of expert instruction and can be used to structure improvement and professional growth. The tool sets out standards for quality instruction based on general pedagogical principals and for different disciplines, and includes checklists with more detailed criteria for each score level and guidance on what to look for when rating performance (e.g., characteristics of quality instruction.) The toolkit also includes data collection protocols. These detailed descriptions support reliability of ratings (avoiding the need for extensive training of raters). The protocols also provide guidance on evidence for rubric ratings, and on qualitative difference between score levels.

6. Serious Games

While there has been a great deal of interest in the potential for serious games to support online learning and assessment, research and development are still only in the early stages. Potentially, serious games may provide opportunities to create more authentic contexts for learning (including visualisation and animation) to support self-regulated learning as well as collaboration, and challenge and competition and the development of problem-solving skills (Hess et al., 2013, Kickmeier-Rust and Albert 2010).

Games that provide immersive environments, simulations such as those used in professional training situations, or platforms for multi-player problem solving may be particularly relevant for MENTEP. For example, learners may use a game situation to simulate how they would address a demanding classroom situation, or challenge online peers to find ways to use technology to address different pedagogical challenges.

Kickmeier-Rust and Albert (2010) discuss the potential for games to adapt and tailor the learning experience for each individual learner (which they refer to as 'micro-adaptivity – an approach developed in the context of the Enhanced Learning Experience and Knowledge TRAnsfer (ELEKTRA) project). Adaptivity might include the presentation (layout, functionality, graphic design); curriculum sequencing (adapted to prior knowledge, goals, intervening when

there are misconceptions, etc.), and adaptive problem-solving support (feedback, hints to scaffold learning). Gee (2007) argues that high quality immersive games require players to think systematically and to consider relationships rather discrete events (a more holistic approach). Johnson et al. (2011) suggest that a well-designed game will include many possible decision points, ensuring that learners adapt to varying situations. Learners may assess compare their individual performance with other players, or may work collaboratively to assess and improve upon their approaches. De Freitas et al. (2010) suggest that Second Life (SL), a social virtual world may eventually used role play and mentoring. Better ICT capability within institutions will be needed, however.

Training simulations, which are less costly, may also be relevant for MENTEP learners. For example, ven der Spek and colleagues (2013) describe efforts to improve the effectiveness of an online simulation for triage professionals (Code Red Triage) through the introduction of a 'surprising event'. The researchers conducted pre- and post-tests of two groups of learners (one control group) (N = 41). They found although there were no differences in the experimental and control groups in regard to engagement and surface learning, learners using the 'surprising events' version of the game 'obtained significantly superior knowledge structures (van der Spek *et al.*, 2013, p. 156). As teachers are also frequently confronted with unanticipated situations, this approach may be worthwhile for MENTEP learners.

At present, there is not an extensive research base on the impact of different approaches on learning and professional development. More investment in research and development and on pedagogical value of different games is needed. The cost of different games and the sophistication of existing delivery systems also need to be considered.

7. Assessment as an integrated feature of online learning

This section explores in more detail how online assessment tools are integrated within online line learning platforms. These programmes may be considered as promising practices, as they are either relatively new and/or have only limited evidence of impact.

In a first example, Seery et al. (2012) describe an e-portfolio developed to support design students, to document and support their process of problem definition and solutions. With this e-portfolio, students are encouraged to use mobile technologies (e.g. their phones) to record and upload data. This 'data repository' provides opportunities for students to reflect on their experience of 'the design process' and to record learning and competence/capability development throughout the activity. Students tag data uploaded to the portfolio based on a colour coded system setting out the categories of 'having', 'growing' and 'proving' an idea. This tagging system is intended to provide assessors with information on how students developed their ideas.

A particular feature of this e-portfolio was a peer ranking system for student portfolios. The system is described as a 'holistic assessment interface' for adaptive comparative judgment, or ACJ (referenced above) (see also project e-scape, Kimbell et al., 2009). The ACJ model was first developed by Pollitt et al. (2004), and is grounded in Louis L. Thurstone's *Law of Comparative Judgement* (Thurstone, 1927). As Pollitt and colleagues describe, a judge who compares two performances uses his or her own personal standard to decide which is better. Each portfolio is judged by several different pairs (thus increasing reliability). The principle of adaptive comparative judgment, the authors argue, is very similar to competition in sport. Whenever two contestants or teams meet, the better of the two is likely to win, no matter what the absolute standard for the game, or the judge's own expectations for performance.

Following the Pollitt/Thurstone comparative judgement model, the e-portolio programme sets up assessors with two different portfolios, which they analyse for evidence of capability' and decides which is better. Assessors refer to criteria, and assign a single score to each portfolio. The rating rubric rates the complexity and impact of the work. A 'pairs engine' is used to rank portfolios, based on all assessors' scores. Outlier scores are identified and monitored.

The idea behind the ACJ model is that that there should be no predetermined assessment criteria by which to judge the effectiveness of designs that have been developed to address problems set out in the initial project phase. It is believed that the ACJ supports more risk taking, and allows mistakes to be valued and Seery et al. (2012) found that indeed students presented mistakes and weaknesses, along with a discussion on why these were considered as a weakness, as evidence of learning. Assessment by peers who have also participated in similar processes is also seen as important. Seery and colleagues conclude that the ACJ is important for providing learners with a proactive role in their own learning. Although assessment criteria may change in response to the evidence presented, the authors find that assessments in this ranking system are both valid and reliable, with a 'very high level of consensus. This heuristic approach to assessment is not necessarily incompatible with programmes that set out very broad guidelines and criteria (as recommended for MENTEP).

In a second example, Barber, King and Bucchanan (2015) discuss 'Digital Moments', a tool to facilitate assessment within a professional learning community', very much like the community envisioned for MENTEP. The tool supports problem-based learning strategies and authentic assessment. Barber et al. note that the tool supports trust among participants, and promotes motivation and creativity. The context is considered as 'authentic' in that e-learners choose tasks, collaborate, and consult with peers who have expertise in the particular technology they want to learn. Digital Moments logs individual and collective learning. Learners share stories of their learning each week, which Barber et al. note supports emergence of both tacit and explicit knowledge. Learners use this platform to design tasks that would be used to assess learning.

Learners using the 'Digital Moments' platform worked collaboratively to develop a variety of assessment tools, including rubrics, portfolios of course work, journals and peer comments.

The learners provided feedback and comments on the validity and reliability of the tools (not based on external validity/reliability tests), and adjusted them based on the comments. Barber et al. recount that this 'social constructivist' approach to developing tasks and assessment tools was meant to ensure that the learning was meaningful and relevant to learners and their professional contexts. Learners providing feedback on the problem-based learning strategy and digital moments platforms indicated that they appreciated constructivist approach to learning, choice and autonomy, collaboration, and flexibility and creativity. They also noted struggles in the lack of specific criteria for the end product.

8. Summary and Next Steps for MENTEP

This review has highlighted a number of elements that are important to consider in the next stages of the development of MENTEP. These include different approaches to online self- and peer-assessment that may be appropriate for authentic, collective learning. This is particularly pertinent in the domain of technology-enhanced teaching (TET) competence, an emerging area of interest. Teachers will be 'inventing' new approaches to integrating technology into classroom practice. Teachers will not only be assessing their own competence, but also the impact of innovative, technology-enhanced teaching methods.

Teacher self-efficacy the online environment and ability to self-regulate learning will be vital for the success of MENTEP. Some percentage of teachers participating in MENTEP will not have had prior experience in learning online. Therefore, the platform design, will need to be user-friendly.

Teachers joining MENTEP will become part of an online community of practice, so there is an important opportunity to enhance self-assessment with peer-assessment. As has been highlighted in this review, peer assessment is perhaps the most robust approach to external assessment of competence levels in an authentic learning environment. Feedback from a single peer assessor can help learners to improve the quality of their work, but of course the assessment process will be strengthened with more involved.

How peers interact and provide feedback will depend on the design of the platform. For example, it may support direct feedback on materials posted online, as with the 'digital moments tool, or, alternatively, peer ranking, as with the 'adaptive comparative judgment' tool. An advantage of the latter is that users will be identifying exemplars of good practice in the process of assessment.

Peer assessors may need some basic guidelines on how to use the tool, as well as on most effective approaches to providing feedback (i.e., timely and specific). The design of the rubric will also be important. In the authentic learning environment, where teachers are developing new approaches as they learn, very broad criteria are likely to be more appropriate.

We have also considered the importance of tools to help learners track their progress. Although we were not able to identify research on the impact of tracking tools on learning, these kinds of tools are widely used in many online apps and have proven to be very popular. They support formative approaches to assessment, allowing users to see their progress as well as areas where they will need to do more in order to reach goals.

MENTEP's support for open and authentic learning implies that a 'light touch' – broad guidelines and criteria for self- and peer –assessment – is likely to be more appropriate. This is also the case because the main users – teachers – already have a certain level of assessment literacy. In turn, feedback from MENTEP users regarding the usability and effectiveness for the platform will provide opportunities to fine tune the platform and better meet needs.

This project has been funded with support from the European Commission.

9. REFERENCES

Bandura, A. (1994), "Self-efficacy", in V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior*, Vol. 4, pp. 71-81, Academic Press, New York.

Barak, M. (2010), "Motivating self-regulated learning in technology education", International Journal of Technology and Design Education", Vol. 20, pp. 381–401, DOI 10.1007/s10798-009-9092-x

Barber W., S. King and S. Buchanan (2015), "Problem-Based Learning and Authentic Assessment in Digital Pedagogy: Embracing the role of collaborative communities" *The Electronic Journal of e-Learning*, Vol. 13, No. pp. 59-67, www.ejel.org.

Black, P. and D. Wiliam (1998), "Assessment and Classroom Learning", *Assessment in Education: Principles Policy and Practice*, Vol. 5, No. 1, pp. 7 – 74.

Boekaerts, M. and E. Cascallar (2006), "How Far Have We Moved Toward the Integration of Theory and Paractice in Self-Regulation?", *Education Psyhology Review*, Vol. 18, pp. 199 – 210, DOI 10.1007/s10648-006-9013-4.

Bouzidi, L., & Jaillet, A. (2009), "Can Online Peer Assessment be Trusted?," *Educational Technology & Society*, Vol. 12, No. 4, pp. 257–268.

Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing Student Learning in Higher Education*, Routledge, London.

Buckingham Shum, S., & Ferguson, R. (2012), "Social Learning Analytics", *Educational Technology & Society*, Vol. 15, No. 3, pp. 3–26.

Butler, D. (1998), "A strategic content learning approach to promoting self- regulated learning by students with learning disabilities", in: D. H. Schunk & B. J. Zimmerman (Eds) *Self-regulated Learning: From Teaching to Self- reflective Practice,* The Guildford Press, New York.

Butler, D. L. (2002), "Individualizing instruction in self-regulated learning" *Theory into Practice*, Vol. 41, No. 2, pp. 81-92.

Chang, C.-C. (2002), "Assessing and Analyzing the Effects of WBLP on Learning Processes and Achievements: Using the Electronic Portfolio for Authentic Assessment on University Students' Learning," World Conference on Educational Media and Technology, Denver, Colorado.

Chang, C.-C. and B.-H. Wu (2012), "Is Teacher Assessment Reliable or Valid for High School Students under a Web-based Portfolio Environment?," *Educational Technology & Society*, Vol. 15, No. 4, pp. 265–278.

Crosson, A., M. Boston, A. Levison, L.C. Matsumura, L. B. Resnick, M.K. Wolf and B. Junker (2006) "Beyond Summative Evaluation: The Instructional Quality Assessment as a Professional Development Tool", National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and Center for the Study of Evaluation (CSE), CSE Technical Report 691 UCLA, CA.

De Freitas, S. G. Rebolledo-Mendez, F. Liarokapis, G. Magoulas and A. Poulovassilis (2010), "Learning as immersive experiences: Using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world", *British Journal of Educational Technology*, Vol 41, No 1., pp. 69–85 doi:10.1111/j.1467-8535.2009.01024.x

This project has been funded with support from the European Commission.

Dettori, G., T. Giannetti and P. Donatella (2006), "SRL in Online Cooperative Learning: implications for pre-service teacher training", *European Journal of Education*, Vol. 41, Nos. 3, 4, pp. 397 – 414.

Dweck, C. (2006), *Mindset: The New Psychology of Success*, Random House, New York.

Galbraith, R.M., R.E. Hawkins, E.S. Holmboe (2008), "Making Self-Assessment More Effective*", Journal Of Continuing Education In The Health Professions*, Vol. 28, No. 1, pp. 20–24.

Gee, J. P. (2007), *Good video games and good learning: Collected essays on video games, learning, and literacy*, Peter Lang, New York.

Gielen, S., F. Dochy and S. Dierick (2003), "Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects", pp. 37-54 In Segers M., Dochy F., & Cascallar E. (Eds.), *Optimizing new modes of assessment: In search of qualities and standards*, Kluwer Academic Publishers, Dordrecht.

Gilbert, P. K. and N. Dabbagh (2005), "How to structure online discussions for meaningful discourse: a case study," *British Journal of Educational Technology*, Vol. 36, No. 1, pp. 5–18.

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

Hess, T. and G. Gunter (2013), "Serious game-based and nongame-based online courses: Learning experiences and outcomes" *British Journal of Educational Technology*, Vol. 44, No 3, pp. 372–385 doi:10.1111/bjet.12024

Johnason, R, L., S. Fisher, M.J. Willeke and F. Mcdaniel II (2003),"Portfolio assessment in a collaborative program evaluation: The reliability and validity of a family literacy portfolio, " *Evaluation and Program Planning*, Vol. 26, No. 4, pp. 367-377.

Johnson, L., R. Smith, R., H. Willis, A. Levine and K. Haywood (2011). The 2011 Horizon Report, The New Media Consortium, Austin, Texas.
http://net.educause.edu/ir/library/pdf/HR2011.pdf

Jönsson, A., Mattheos, N., Svingby, G., & Attström, R. (2007). Dynamic Assessment and the "Interactive Examination". Educational Technology & Society, 10 (4), pp. 17-27.

Kickmeier-Rust, M.D. and D. Albert (2010), ""Micro-adaptivity: Protecting immersion in didactically adaptive digital educational games," *Journal of Computer Assisted Learning*), Vol. 26, doi: 10.1111/j.1365-2729.2009.00332.x

Kimbell, R., T. Wheeler, K. Stables, K., Shepard, T., A. Pollitt, F. Martin, et al. (2009), "E-scape portfolio assessment phase 3 report", Teru, Goldsmiths University of London.

Kruger, J. and D. Dunning (1999), "Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments", *Journal of Personality and Social Psychology*, Vol. 77, No. 6, pp. 1121-1134.

Lee, Y., J. Choi and T. Kim (2013), "Discriminating factors between completers of and dropouts from online learning courses", *British Journal of Educational Technology*, vol. 44, No. 2, pp. 328 – 337, doi:10.1111/j.1467-8535.2012.01306.x

Liang, J.-C. and Tsai, C.-C. (2008), "Internet Self-efficacy and Preferences toward Constructivist Internet-based Learning Environments: A study of pre-school teachers in Taiwan. Educational Technology & Society: Vol. 11, No. 1, pp. 226-237.

This project has been funded with support from the European Commission.

Linn, R. L., E. Baker and S. Dunbar (1991), "Complex, performance-based assessment: Expectations and validation criteria", *Educational Researcher*, Vol. 20, No. 8, pp. 15-21.

Li, L., L. X. Liu and Y. Zhou (2012), "Give and take: A re-analysis of assessor and assessee's roles in technology-facilitated peer assessment," *British Journal of Educational Technology,* Vol. 43, No. 3, pp. 376–384.

Mao, J. and K. Peck (2013) "Assessment Strategies, Self-Regulated Learning Skills, And Perceptions Of Assessment In Online Learning", *The Quarterly Review of Distance Education*, Vol. 14, No. 2, pp. 75–95.

Matsumura, L.C. *et al.* (2006), "Using the Instructional Quality Assessment Toolkit to Investigate the Quality of Reading Comprehension Assignments and Student Work" CSE Report 669, January, Center for the Study of Evaluation (CSE), UCLA, Los Angeles, CA.

Olfos, R. and H. Zulantay (2007), "Reliability and Validity of Authentic Assessment in a Web Based Course," *Educational Technology & Society*, Vol. 10, No. 4, pp. 156-173.

Perlman, C. (2002), "An Introduction to Performance Assessment Scoring Rubrics". In C. Boston's (Eds.), *Understanding Scoring Rubrics* (pp. 5-13), ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, MD.

Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. M. (2010), "Highly Integrated Model Assessment Technology and Tools", *Educational Technology Research & Development*, Vol. 58 No. 1, pp. 3 – 18

Pollitt, A. , G. Elliott and A. Ahmed (2004), "Let's stop marking exams," Paper presented at the annual conference of the international association for educational assessment, Philadelphia. http://www.cambridgeassessment.org.uk/Images/109719-let-s-stop-marking-exams.pdf

Saiz, C., S.F. Rivas and S. Olivares (2015), "Collaborative learning supported by rubrics improves critical thinking," *Journal of the Scholarship of Teaching and Learning*, Vol. 15, No. 1, pp. 10 - 19.doi: 10.14434/josotl.v15i1.12905

Seery, N. C. Donal and P. Phelan (2012), "The validity and value of peer assessment using adaptive comparative judgement in design driven practical education", International Journal of Technology and Design Education 22:205–226 DOI 10.1007/s10798-011-9194-0

Shavelson, R., et al. (2008), "On the Impact of Curriculum-embedded Formative Assessment on Learning: A Collaboration between Curriculum and Assessment Developers", *Applied Measurement in Education*, Vol. 21, pp. 295-314.

Sheldon, K. M., & Elliot, A. J. (1998), "Not all personal goals are personal: Comparing autonomous and controlled reasons for goals and predictors of effort and attainment", Personality and Social Psychology Bulletin, Vol. 24, No. 5, pp. 546–557.

Sitzmann, T. K. Ely, K. G. Brown And K.N. Bauer (2010), "Self-Assessment of Knowledge: A Cognitive Learning or Affective Measure?, *Academy of Management Learning & Education*, Vol. 9, No. 2, pp. 169–191.

Spector, J. M. (2013), "Emerging Educational Technologies and Research Directions.", *Educational Technology & Society*, Vol. 16, No. 2, pp. 21–30.

Steffens, Karl, (2006) Self-Regulated Learning in Technology-Enhanced

This project has been funded with support from the European Commission.

Learning Environments: lessons of a European peer review, Vol. 41, No. 3, pp. 353 – 379

Strijbos, J. W., S. Narciss and K. Dünnebier (2010), "Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency?," *Learning and Instruction*, Vol. 20, No. 4, pp. 291-303.

Sulzen, J., M. Young and R. Hannifin (2008), "Reliability and validity of an ecologically-grounded student teacher electronic portfolio rubric", In K. McFerrin et al. (Eds.), Proceedings of Society for Information Technology & Teacher Education International Conference 2008, AACE, Chesapeake, VA.

Tsai, C.-C. (2012), "The Development of Epistemic Relativism versus Social Relativism via Online Peer Assessment, and their Relations with Epistemological Beliefs and Internet Self-efficacy," *Educational Technology & Society*, Vol. 15, No. 2, pp. 309–316.

Tsai, C.-C., S. S. J. Lin, and S-M Yuan (2002), "Developing science activities through a networked peer assessment system," *Computers & Education*, Vol. 38, pp. 241-252.

Topping, K. J. (1998), "Peer assessment between students in colleges and universities," *Review of Educational Research*, Vol. 68, No. 3, pp. 249-276.

van der Spek, E. D. H. van Oostendorp and J.J. Ch. Meyer (2013), "Introducing surprising events can stimulate deep learning in a serious game", British Journal of Educational Technology, Vol. 44, No. 1, pp. 156 – 169, doi:10.1111/j.1467-8535.2011.01282.x

Vickerman, P. (2009), "Student perspectives on formative peer assessment: An attempt to deepen learning?," *Assessment & Evaluation in Higher Education*, Vol. 34, No. 2, pp. 221-230.

Wang, M., J. Peng, J., B. Cheng, H. Zhou, H., and J. Liu, (2011), "Knowledge Visualization for Self-Regulated Learning", *Educational Technology & Society*, Vol. 14, No. 3, pp. 28–42.

Wiliam, D. (2006), "Formative Assessment: Getting the Focus Right", *Educational Assessment*, Vol. 11, pp. 283-289.

Yu, F.-Y., and C.-P Wu (2013), "Predictive Effects of Online Peer Feedback Types on Performance Quality," *Educational Technology & Society*, Vol. 16, No. 1, pp. 332–341.

Zeidner, M., M. Boekaerts and P. Pintrich (2000), "Self-regulation: Directions and challenges for future research" in: M. Boekarts, P. Pintrich & M. Zeidner (Eds) *Handbook of Self-regulation,* Academic Press, New York.

Zimmerman, B. J. (1995) "Self-regulation involves more than metacognition: A social cognitive perspective", *Educational Psychologist*, Vol. 30, No. 4, pp. 217–221.

Zimmerman, B. J. (1998) "Models of self-regulated learning and academic achievement in: B. J. Zimmerman & D. H. Schunk (Eds.) *Self- regulated Learning and Academic Achievement. Theory, Research and Practice*, Springer, New York.

This project has been funded with support from the European Commission.