Lifelong Learning Programme

Supporting Lifelong Learning with Inquiry-Based Education

Agreement no.: 2013 – 3806 /001 - 001

Project type: LLP- KA3MP

Start date of project: 1st January 2014   Duration: 24 months

Deliverable 4.4 - Review effective feedback and formative assessment in e-learning.

Due date of deliverable: 01 September 2014
Actual submission date: 01 September 2014

Deliverable Responsible: (Insert here Partner Acronym and code)

Project co-funded by the European Commission within the LifeLong Learning Programme	
Dissemination Level	
Public	**x**
Restricted to other programme participants (including the EC Services)	
Confidential, only for members of the consortium (including the EC Services)	

Document History

Version	Date	Comments
Review effective feedback in elearning3107141.docx	31/08/2014	
Review effective feedback in elearning310714	31/07/2014	
Review effective feedback in elearning240714	24/07/2014	
Review effective feedback in elearning170714	17/07/2014	Documents on the search process and data extraction form, etc. are former to the document history of the report

Authors

Section	Partner Organization	Author/s
Review report	UTWENTE	Maaike Heitink, Bernard Veldkamp
Revisions/feedback	LUC, UNIROMA TRE	Stephen Dobson, Gabriella Agrusti

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS

DT	Diagnostic Testing
DBDM	Data Based Decision Making
AFL	Assessment for Learning
KP	Knowledge of performance
KR	Knowledge of response
KCR	Knowledge of correct response
EF	Elaborated feedback
ICT	Information and Communication Technologies

EXECUTIVE SUMMARY

This systematic review is written in the frame of a European Union funded project called Supporting Lifelong learning with ICT Inquiry-Based Education (LIBE). Through this project, an e-learning environment will be developed for young low educational achievers (aged 16-24). A crucial part in every educational intervention is assessment and feedback. Hence, to contribute to this project, this study aims to review literature regarding computer based feedback and formative assessment in e-learning. Three different approaches are considered as formative assessment, namely: 'diagnostic testing' (DT), 'data based decision making' (DBDM) and 'assessment for learning' (AFL). The search and selection procedure for this review resulted in 19 studies about the effects of feedback and formative assessment in e-learning. Four interrelated themes emerged within this body of literature: (1) the effect on achievement, (2) the effect on motivation, (3) self-regulated learning and (4) the effect of learner characteristics. Results are elaborated around those themes and implications for the LIBE project are drawn.

1. INTRODUCTION

This systematic review is written in the frame of a European Union funded project called Supporting Lifelong learning with ICT Inquiry-Based Education (LIBE). Through this project, an e-learning environment will be developed for young low educational achievers (aged 16-24). This e-learning environment aims to enhance key information processing skills (literacy, numeracy and problem solving), enhance the motivation to learn and to promote lifelong learning. To contribute to this project, this study aims to review literature regarding computer based feedback and formative assessment in e-learning.

1.1. E-LEARNING

E-learning is defined as the use of electronic media and digital technology, such as laptops, tablets or smartphones, to deliver learning content and instruction (e.g. Clark & Mayer, 2011). As the general idea of e-learning is that technology is used to support learning, terms like computer assisted learning or computer based learning are common as well. Nowadays e-learning is typically integrated in online courses delivered through virtual learning environments. Often e-learning is focused on the individual learner but more and more courses support collaboration and sharing of knowledge with other users. The use of technology in learning makes time- and place-independent learning possible and can include multimedia to foster students' conceptual understanding. Because of the digital advantages, assessment and feedback in an e-learning environment can support total automation of administrating the teaching and learning processes, tailored to the individual user's needs.

1.2. FORMATIVE ASSESSMENT

A crucial part in every educational intervention is assessment. Assessment can have a formative and a summative function. Where summative assessment only focuses on assessing learning outcomes, formative assessment provides feedback to support learning and to gain insights in learning processes. Based on this support and insight, the learning process can be guided in the right direction (e.g. Van Der Kleij, Vermeulen, Schildkamp & Eggen, 2013). Formative assessment can allow a constant monitoring of learners' progress and verify the effectiveness of courses but most importantly, it is the main instrument to tailor pedagogy to the actual end-users' needs.

Several kinds of formative assessment are stated in literature which all have different strategies to assess and support learning. In this review three different approaches are considered as formative assessment, namely: 'diagnostic testing' (DT), 'data based decision making' (DBDM) and 'assessment for learning' (AFL). Diagnostic testing concerns mapping out learners' task response patterns to reveal their (inadequate) solution strategies and use this as an indication for the learners' development stage (Crisp, 2012). This type of formative assessment specifically focuses on identifying strengths and weaknesses of individual learners. Although DT was initially intended for identifying the needs of learners with limited abilities, it is more and more common to use this type of formative assessment for identifying educational needs of all learners (William, 2011).

Data based decision making regards the systematic collection and analysis of data to inform decisions that focus on educational improvement of teaching, curricula and (school) performance (Schildkamp & Kuiper, 2010). Data does not necessarily include achievement results but also other relevant information such as learners characteristics. DBDM can be used on class-, school level and individual level.

Assessment for learning is part of 'everyday' learning practices (Klenowski, 2009) and focuses on the quality of the learning process (Stobart, 2008). Feedback is incorporated to guide future learning, aiming at either class or individual level. AFL can be approached from a measurement perspective or an inquiry perspective (Hargreaves, 2005). AFL approached from a measurement perspective is characterized by the use of quantitative data to formulate feedback and to inform decisions on assessment activities that aim to determine to what extent a predetermined level has been achieved. Consequently, in this approach AFL concerns marking, monitoring and showing an ability level. Approaching AFL from an inquiry perspective results in the use of (usually) qualitative information (e.g. observations, demonstrations and conversations) to generate feedback and is viewed as a process of discovering, reflecting, understanding and reviewing.

Next to assessment for learning, literature distinguishes 'assessment as learning' and assessment of learning. In the context of e-learning, this review follows the suggestion of Webb, Gibson and Forkosh-Baruch (2013) to address assessment for learning and assessment of learning simultaneously by using strategies that include continuous (unobtrusive) measuring of performance while learners are engaged in computerised tasks that support learning. During this process users should have access to meaningful representations of evidence and arguments about their achievements. Assessment as learning is typical of self-assessment and peer-assessment, where the emphasis is upon the student's development of meta-cognitive skills (Dobson, Eggen, & Smith, 2009). Some authors include assessment as learning as a sub-set of assessment for learning since the gains in learning outcomes are often associated with forms of self-regulated learning and motivation. In this review we will view assessment as learning as a sub-set of assessment for learning as well.

1.3. FEEDBACK

A crucial element in this formative use of assessment is feedback. Hattie & Timperley (2007) define feedback as information provided by an agent regarding aspects of one's performance or understanding. Just as formative assessment, feedback can be presented in a formative or a summative way. In the context of e-learning, formative feedback can be defined as either (1) information specifically relating to the task or process of learning that fills a gap between what is understood and what is aimed to be understood (Sadler, 1989), (2) information providing insights into learners' learning which can be used to adapt instruction to the learners' needs or (3) help learners in making choices regarding the future learning path they are going to take. Next to these cognitive and metacognitive functions of feedback, feedback can also have a motivational function encouraging learners to maintain their effort and persistence (Narciss, 2008).

Feedback has been used in a large variety of types, defining their content (Hattie and Timperley, 2007; Narciss, 2008; Shute, 2008). These types can be classified as follows:

- Knowledge of performance (KP): summative feedback about learners' performance after completing a test/task or set of items (e.g. grade, percentage correct, etc.).
- Knowledge of response (KR): feedback on the correctness/quality of a response (e.g. correct/incorrect, error detection, etc.)
- Knowledge of correct response (KCR): providing the correct answer or solution of an item or task.
- Elaborated feedback (EF): providing more elaborated feedback additional to the former mentioned types. For example hints/clues, worked out examples, or information resources.

Other characteristics of feedback are found in its presentation. Feedback presentation refers to aspects of timing and scheduling. The timing of feedback is mostly distinguished in 'immediate feedback' delivered immediately after the learner answered a question, or 'delayed feedback' which is delivered at the end of a test or a certain set of items. Scheduling of feedback differs in the amount of times feedback will be presented (e.g. single try, multiple try, answer until correct) (Narciss, 2008).

Several studies have focussed on the effects of these different feedback types and its characteristics. Former reviews, for example Van der Kleij, Timmers & Eggen (2011), mostly focus on the effects of different feedback types on learner performance. However, not only the effects on performance directly play a part in learning. Feedback can have impact on learners' motivation and self-regulated learning as well (Narciss, 2013). According to Black and Wiliam (2009), activating learners' 'learning-ownership' involves motivation, metacognition and self-assessment. Since this review is written in frame of the LIBE project, which targets low achievers and aims to promote life long learning, these effects of feedback are important as well.

2. METHOD

2.1. PROCEDURE

This review was conducted according to the method for systematic review studies in the social sciences (Petticrew and Roberts, 2006). This stepwise process encompasses formulating research questions, defining search terms, selecting databases, conducting the search process, formulating inclusion criteria to select relevant publications and the extraction of data. Data was analysed using a data extraction form in order to collect similar data from each publication. Additionally, the scientific quality of each publication was determined and only studies with sufficient quality were selected.

2.2. DATABASES AND SEARCH TERMS

Four scientific databases were searched for relevant studies. These databases are: Education Resources Information Center (ERIC), Web of Science, Scopus and PsychINFO. The same search strategy was used for every database. The search process started with broad search terms like "formative assessment", "data-based decision making", "assessment for learning", "diagnostic testing" and related terms. Then this search was narrowed down by searching with additional terms related to "feedback" and again by terms related to e-learning. During this search process output was limited to studies that were published between 1998 and 2014, published within relevant research fields (e.g. social sciences, educational research, etc.) and limited to the languages English and Dutch. After searching these databases two extra steps were done to include studies relevant for this review's purpose: (1) additional search for common authors in the research field (e.g. Narciss, S., Shute, V.) and (2) snowballing of related review studies. More details on the search process can be found in Appendix 1.

2.3. INCLUSION CRITERIA

The following selection criteria were formulated for this review:

1) The study was published in a scientific, peer reviewed, journal or dissertation. Books, book chapters and conference proceedings are not selected to secure the scientific quality of the studies presented in the publications.

2) The publication concerns scientific research. Theoretical considerations, reviews and opinions will not be selected as this review study focuses on factors that have proven to be effective on using feedback or formative assessment in e-learning. Although there is a need to be careful to generalize the results of case studies, case studies were included in the selection for their practical examples in 'real life' contexts which is important for the LIBE project as well.

3) This review study focuses on the contexts of secondary education, vocational education and higher education - with specific attention paid to low achiever target

audiences - as these are contexts that overlap with the age of learners in the LIBE project. Primary education is not included, as is special education.

4) The study is focused on computer based feedback and formative assessment in e-learning.

2.4. SELECTION PROCESS

All publications were exported to Thomson Reuters Endnote X7 (2013) and duplicates were removed from the selection. Titles of the remaining publications were scanned to select the ones that fit the inclusion criteria. After the title scans, abstracts of the remaining publications were scanned for the same purpose. If it was unclear whether or not the study fully satisfied the inclusion criteria, studies remained in the selection. Documents were obtained of the studies that seemed to satisfy the inclusion criteria in order to use these in data extraction. When during data extraction the study showed not to satisfy the inclusion criteria after all, it was removed retroactively.

2.5. DATA EXTRACTION AND QUALITY DETERMINATION

The remaining publications were read and relevant results concerning effective feedback and formative assessment in e-learning were extracted through a data extraction form. The form was tested by using it on four articles and revised to make sure useful data would be collected. An instruction manual for using the data extraction form was formulated as well in order to foster comparable results among different users. The data extraction form, developed by the University of Twente, consists of the following parts:

- General information: author, publication year, title, context, etc.
- Research design: research question, formative assessment type, feedback types, methods, instruments, etc.
- Research population: number of respondents, sampling method.
- Results: answer on the research question, effects of…. Etc.
- Quality check: clear research goal, appropriate methods used, reliability, etc.

3. RESULTS

3.1. SEARCH AND SELECTION RESULTS

Going through the search process described above resulted in a total of 224 studies. The results of each search step in the different databases are shown in **Table 1**.

Table 1 Number of hits for each step during the search process in four different databases.

Search steps[a]	#hits Web of Science	#hits ERIC	#hits PsycINFO	#hits Scopus
1. Formative assessment[b]	62801	5383	5257	130217
2. Feedback[b]	1096	468	376	3156
3. e-learning[b]	34	26	12	232
4. Limitation to research fields	32	N/A	N/A	201
5. Limitation to language (English or Dutch)	32	26	12	200
6. Limitation to articles (inclusion criteria 1)	17	24	12	171

Note a: searched between publication years 1998 – 2014.

Note b: related search terms included, see appendix.

After removing duplicates, irrelevant journals, conference proceedings, books, etc. (inclusion criteria 1) 134 studies were left for further selection. Results of the title and abstract scan are shown in Table 2. Selection step 1 (referring to inclusion criteria 1) was not included in the table because this step was already done through scanning the sources in the Endnote database. Detailed results of the search process can be found in the Appendix (section 6.1).

Table 2 Number of studies selected and excluded during the title and abstract scan.

Selection step	Title scan		Abstract scan	
	selected	excluded	selected	excluded
2. The publication concerns scientific research	134	5	118	14
3. Context of secondary education, vocational education and higher education	129	3	104	6
4. The study is focused on feedback and formative assessment in e-learning.	126	8	98	43

This resulted in (98 - 43 =) 55 remaining studies. Related reviews and other non-empirical studies that were found during this process were collected in a separate Endnote database. Of the 24 studies in this Endnote database, 15 seemed interesting for snowballing. The actual documents of 12 studies could be attained. After this, an additional search was done for relevant studies through common author search. This resulted in an additional 5 studies and a total of 3 studies for data extraction. The actual documents could be attained from 43 studies.

3.2. DATA EXTRACTION AND ANALYSIS RESULTS

After data extraction 19 studies were left for analysis because 16 studies did not fit the inclusion criteria after all and 3 studies did not meet the scientific quality criteria. Studies were grouped into the following categories for analysis: formative assessment types, feedback types and emerging themes. Data extraction revealed four interrelated themes within the body of literature selected for this review: (1) the effect on achievement, (2) the effect on motivation, (3) self-regulated learning and (4) the effect of learner characteristics. Every theme is elaborated below. An overview of the results is shown in Table 3 (section 3.2.5). Only the first author and publication year will be mentioned in the in text references to keep it reading comprehensible.

3.2.1. Effects on achievement

Fourteen studies selected for this review focused on the effect of feedback or formative assessment on achievement. The effect of feedback on achievement has been reviewed before. The latest published systematic review found is a meta-analysis of Van der Kleij (2013) in which the latest study included is from 2010. Additionally, Gikandi et al. (2011) did a qualitative review on online assessment in higher education that included one study from 2011. To prevent double work, this review includes studies about the effect of feedback on achievement that were published after the meta-analysis or that were not included in both of these reviews. Additionally, studies of course had to fit the inclusion criteria formulated during the search process. As also described in the introduction, feedback and formative assessment are two closely related, even overlapping, themes. Especially when feedback is provided with a formative purpose. Therefore the results related to formative assessment in general are described first and will be followed feedback-specific results.

The effect of formative assessment on achievement

Ten studies in this review report significant effects on learners' achievement when using a computer based formative assessment (Einig, 2013; Wilson, 2011; de-Marcos 2010; González, 2010; Buchanan, 2000; Johnson-Glenberg, 2010; Bokhove, 2012; Ćukušić, 2014; McKenzie, 2013; Timmers, 2011). Except for one study, most studies regard to the formative assessment types assessment for learning, diagnostic testing or a combination of the two (see Table 3 as well). Data based decision making was only presented in one study (McKenzie, 2013), where the results of the pre-test were used to generate an individualized learning

plan. Significant differences on achievement were found between learners who used this environment and learners who did not.

With the AFL type, formative self-assessment in e-learning was specifically highlighted by three studies as an effective formative assessment strategy (de-Marcos 2010; Ćukušić, 2014; Ibabe, 2010). In a study of Ibabe (2010) formative self-assessment proved to be a highly significant predictor for learners' final grade, which was higher for learners that used the tool for self-assessment then learners that did not have that opportunity. Also the frequency of doing self-assessment exercises correlated positively with performance; the more self-assessment exercises learners did, the higher their final grade. Ćukušić (2014) found a similar significant difference between groups' exam pass rates in a study on a self-assessment tool.

The effect of feedback on achievement

Feedback types found in this study (either combined or single use) were knowledge of result (KR), knowledge of correct result (KCR), knowledge of performance (KP) and elaborated feedback (EF). Elaborated feedback, however, was ranging from information on the correct solution's strategy, to hints on how to proceed, to hints on location of mistakes, to information on sources where to look for further study to solve the problem. Overall, KR seems to be least effect type of feedback and EF seems to be the most effective type based on the studies selected for this review. Four studies in this selection specifically studied the differences between feedback types of which three found that receiving EF resulted in higher learning gains then learners that just received KR or KCR and only check their responses with the correct answers (Liminou, 2014; Narciss, 2004; Bokhove, 2012). Also Timmers (2011) found a significant positive relation between test score and the percentage of consulted EF for incorrectly answered questions. Although these studies did find significant learning effects, there is one study in which the effect of different types of feedback on performance was investigated but no significant effects on a summative post-test were found (Van der Kleij, 2012).

Regarding EF, a significant effect on performance was found when learners received feedback that included statements or clues aiming to assist in understanding why their answer was incorrect (McKenzie, 2013; Liminou, 2014; Bokhove, 2012; Timmers, 2011). Closely related to this kind of EF, and significantly effective as well, is providing hints on areas of weakness and how to proceed (Narciss, 2004).

Other characteristics of feedback studied in the selection for this review are timing and scheduling. Immediate feedback on individual responses and the opportunity to use that feedback in subsequence tries (multiple tries) seems to have a positive effect on learners' learning (Limniou, 2014; Buchanan, 2000; Narciss, 2004; Ibabe, 2010). Although van der Kleij (2012) did not find any significant effects on achievement for the timing of different feedback types, a significant difference between immediate and delayed feedback was found for the time spend on feedback. Learners who received immediate feedback spend more time on feedback then learner who did not. Furthermore Learners who received immediate KCR + EF were more likely to read the feedback whenever they guessed an item than learners who received delayed KCR + EF or KR only.

3.2.2. Effects on self-regulated learning

Feedback in the form of hints does not only help learners to overcome difficulties but also enables them to check whether they are on the correct solution path or not. This way they gained insight in their own knowledge and chosen strategy (Kleitman, 2014; Bokhove, 2012; Bälter, 2013; Wilson, 2011). If not on the right path, learners can decide to use a different strategy. This kind of feedback is not only focused on cognitive aspects but on meta-cognitive functions as well, and helps learners regulate their own learning process. This is also apparent in learners' perception on learning. Learners that received substantial feedback concerning hints about the reason why their answer would be incorrect felt encouraged to control their learning process in terms of what they already knew and what they needed to know. This helped learners to develop critical reflective skills on their learning outcomes, which stimulated independent study (Limniou, 2014; Einig, 2013). In a related study, Bälter (2013) found that formative quizzes could change learners' perception of their own knowledge and help them in guiding their study behaviour in the right direction. Another example is the study of Kleitman (2014). In this study, learners who used a formative assessment tool that asked for both an answer on the question and a prediction of the correctness of the answer (self-confidence rating) had a significantly lower prediction bias then learners who did not use the tool. Also, fewer attempts were needed when feedback included the opportunity to obtain hints (Bokhove, 2012), again suggesting self-reflective skills by gaining insight in the chosen strategy. All studies that revealed the effect of feedback on self-regulated learning strategies used AFL (3) or AFL in combination with DT (3).

Although these studies suggest the importance of using feedback to help learners regulate their own learning, Bokhove (2012) explored the effect of feedback 'regulated' by an e-learning environment. In their design the learning environment incorporated 'fading'; starting off with a lot of feedback to aid learning, but decreasing the amount towards the end to facilitate transfer. While fading occurred, learners' scores on the formative assessments decreased (although they were still significantly higher than pre-test results) but significant higher scores were found on the summative test. This is supported by a case study of Lafuente (2014) who identified patterns in the amount of provided support and different types of feedback. Lafuente (2014) found that in a virtual environment, higher progress correlated with less need for support. Additionally, feedback pointing out the correct parts in their work increased while feedback about incorrect answers and missing parts decreased.

3.2.3. Effects on motivation

Eight studies in the selection for this review found effects of feedback and/or formative assessment on learners' motivation. Motivation was studied in terms of effort, attention paid to feedback, engagement, persistence and satisfaction with performance.

Narciss (2004) found that the possibility to apply information gained from EF in subsequent tries kept learners more engaged on tasks and contributed to higher levels of satisfaction than providing correct/incorrect feedback only. Similarly, in Limniou (2014) learners that received substantial feedback in the form of hints seemed to actively participate in their learning activities constructing their own knowledge based on the received feedback. As a

result, while these learners were working on closing the gap between current and desired performance, the control group learners (who received only grades and correct answers) faced their e-assessments more as a way to practice for later summative tests (at the end of the course) rather than a motivator to study more. Additionally, instant EF and being able to try multiple times is less stressfull to learners then knowing that they will have to get it right the first time (Bälter, 2013).

Another indicator for motivation according to several studies was time spend on feedback. The more motivated learners are, the more time is spent on feedback resulting in more study time (Timmers and Veldkamp, 2011; Ibabe, 2010; Van der Kleij, 2012). Time spend on feedback was found to be significantly higher for learners who received immediate feedback compared to learners who received delayed feedback (Van der Kleij, 2012). However, Timmers (2011) found that the longer the test and the more difficult the test, the less attention was paid to feedback.

Next to attention paid to feedback, Timmers (2011) also shows that incorrectly answered items resulted in a more frequent use of available feedback then when a question was correctly answered. This suggests that the amount of attention paid to feedback is also influenced by learners' perceived usefulness of feedback and learners' actual need for feedback. Timmers (2013) found that learners' beliefs about the value of the task did predict learners' effort put into the formative assessment. Also, learners who received KCR + EF agreed that that the feedback was useful, while the opinions of learners who received KR were more diverse and more negative (Van der Kleij, 2012).

3.2.4. Feedback and learner characteristics

The way feedback and formative assessment effects learners' study behaviour is not only dependent on the feedback giver (in this case a computer based environment) but also on the feedback receiver (Narciss, 2004). Learners need to be able and willing to use the formative assessment in feedback in order for it to be affective. Six studies in this review focused on the way learner characteristics can influence feedback or formative assessment usage.

Immediate, multiple try formative assessment tasks that stimulate learners to focus on the difficulties and/or misconceptions (Limniou, 2014) seem to be especially beneficial to lower achieving learners as it allows them more time to go through basic concepts (Sorensen, 2013). This is also supported by Johnson-Glenberg (2010) who, on a related topic, found that the amount of prior knowledge significantly moderates the effect of formative quizzes on learning. Low experienced learners benefited more, learned significantly more, compared to 'experts'. In the same study, results even show that learners with high prior knowledge levels showed a lower learning gain when formative quizzes were used then to learners with high prior knowledge levels that did not use formative quizzes.

Next to prior knowledge and learners' achievement levels, self-confidence seems to act as a mediating variable as well. Low self-confidence was significantly related to lower persistence, less effort, poorer achievement, and less satisfaction with performance then high self-confidence (Narciss, 2004; Kleitman, 2014; Ibabe, 2010). This related to feedback types, using EF (hints) additional to KR was only an effective means of fostering task engagement and achievement for learners with high self-confidence, whereas it was less

effective for learners with medium or low self-confidence (Narciss, 2004). Controversially, Timmers (2013) found that learners' own expectation of success (which can be interpreted as a measure for self-confidence) did not predict learners' effort put into a formative assessment. Although, in the same study, 'feedback seeking' was positively related to learners' expectations of success. This indicates that learners who have higher self-confidence about their ability are more likely to be more involved in learning by seeking feedback.

3.2.5. Overview of the selected studies

Table 3 shows the studies selected for this review with their characteristics and focus. Most studies in this review were done in the context of higher education (17). Only two also included secondary education and two vocational education.

Table 3 overview of the studies selected for this review.

Author (year)	Formative assessment type[c]	Feedback type	Timing[d] and scheduling	achievement[a]	motivation	learner characteristics	Self regulated learning	Subject	Context
Bälter (2013)	AFL	KR	delayed, multi try (entire test)		X		X	Science, Humanities - Geography	Higher education
Bokhove, 2012	AFL, DT	KR vs. EF (hints)	Immediate, answer until correct	X*			X	Mathematics	Secondary
Buchanan 2000	AFL	KR + EF	Delayed, multi try (entire test)	**X***				Humanities	Higher education
Ćukušić, 2014	AFL (SA)	KR	Delayed, multi try (entire test)	**X*[b]**				Humanities - economics	Higher education
de-Marcos 2010	AFL (SA)	KR	Immediate, multi try	**X*** (for technology only)				Science, health	Secondary, higher education
Einig 2013	AFL, DT	KR + KCR + EF	Delayed, multi try	**X*** (in general)			X	Humanities - economics	Higher education
González (2010)	AFL	KR + KP	Immediate + delayed, multi try	X*				Mathematics	Higher education
Ibabe, 2010	AFL (SA)	KR + EF (hints) + KP	Immediate + delayed, multi try	**X**	**X**	X		Mathematics Mathematics	Higher education
Johnson-Glenberg	AFL	n/a		**X***			**X**	Language/literacy	Vocational education

(2010)									
Kleitman 2014	AFL, DT	KCR + EF + KP (+progress performance)	Immediate + delayed, multi try		**X**		**X**	Mathematics	Higher education
Lafuente (2014)	AFL	KR + EF	Delayed, single try (with revision possibilities)				**X**	Humanities	Higher education
Limniou, 2014	AFL	KCR + KP vs. KCR + EF (hints)	Immediate vs. delayed, single try vs. multi try	X$^?$	X		**X**	Science	Higher education
McKenzie, 2013	AFL, (automated) DBDM	KCR, EF (hints)	Delayed, multi try	**X***				Humanities	Higher education
Narciss, 2004	AFL	KR vs. KR + EF,	Immediate, multi try	X*	**X**	**X**		Not subject related	Higher education, vocational education
Sorenson 2013	AFL	KR	Immediate + multi try (with penalty)			X		Science	Higher education
Timmers (2013)	AFL	KR + KCR + EF	Delayed, single try		**X**	X		Language/literacy	Higher education
Timmers, 2011	AFL	KR + KCR + EF	Delayed, single try	X*	X	X		Language/literacy	Higher education
Van der Kleij, 2012	AFL	KCR + EF vs. KR	Immediate vs. delayed, single try	**X**	X			Humanities - economics	Higher education
Wilson 2011	DT	KR	Delayed, single try	**X***				Humanities - Geography	Higher education

Note a: * an asterisk indicates significant effects on performance were found.
Note b: significant difference on pass rates for learners that used the tool.
Note c: DT=diagnostic testing, AFL=Assessment for learning, DBDM=data based decision making, SA=self assessment.
Note d: "Delayed" means feedback is delivered directly after a whole test or item set unless mentioned otherwise.

4. DISCUSSION AND CONCLUSION

The search and selection procedure for this review found 19 studies about the effects of feedback and formative assessment in e-learning. Four interrelated themes emerged within this body of literature: (1) the effect on achievement, (2) the effect on motivation, (3) self-regulated learning and the (4) effect of learner characteristics.

Significant effects were found on learners' achievement when using a (computer based) formative assessment. Specifically highlighted was formative self-assessment which proved to be a highly significant predictor for learners' final grade. Although the majority of the studies in this review did find positive effects of formative assessment on learning, most studies only compared users versus non users. This makes it hard to identify what aspects of the formative assessments and different forms of feedback led to these positive results. However, four studies did specifically study these aspects and found that receiving 'elaborated feedback' resulted in higher learning gains then learners that just received feedback on 'knowledge of result' or 'knowledge on correct result'. Immediate feedback in the form of hints on individual responses and the opportunity to use that feedback in subsequence tries seems to result in positive achievement results. This is also found in related review studies that even emphasize that feedback regarding only 'knowledge of result' has virtually no effect on achievement at all (Bangert-Drowns, Kulik, Kulik, Morgan, 1991; Shute, 2008; Van der Kleij, 2011; Van der Kleij, 2013).

In every study, feedback is used to work on closing the gap between current and desired performance, therefore feedback should focus on criteria for performance that have been met and criteria for performance that not yet have been met (cf. Black & William, 2009; Welsh, 2012). Regarding these criteria it should be noted that not only cognitive but also meta-cognitive feedback is important as it helps learners to regulate their own learning process. Critical reflection is viewed as a key skill in self regulation. This is probably why 'elaborated feedback' is suggested by several studies as a good approach. Using 'elaborated feedback' to provide hints on why the learners answer is incorrect enables them to check whether they are on the correct solution path or not, by which they gain insight in their own knowledge and chosen strategy. Although 'elaborated feedback' is suggested as a good feedback approach, this feedback type has presented itself in a variety of ways within the studies of this review. Narciss (2013) makes a more subtle distinction within this category of feedback type: (1) knowledge about the task: information on task rules, task constraints and/or task requirements, (2) knowledge about concepts addressing conceptual knowledge, (3) knowledge about mistakes, for example flagging location of errors, (4) knowledge about the learning process addressing procedural knowledge (e.g. problem solving strategies), (5) knowledge about meta-cognition eliciting meta-cognitive knowledge and strategies necessary for self-regulating the learning process.

Since this review was written in the frame of the LIBE project, specific attention was supposed to be paid to the low achiever target audience. However, the results of this review show that very little seems to be published within this specific research field. Feedback related research tends to focus upon all learners, rather than upon the special needs of certain groups, such as low achievers or those with special needs. A case in point is the much quoted article by Hattie and Timperley (2007). Among the twelve meta-studies on feedback

discussed in their article, only one directly refers to pupils with special needs (namely: Skiba, Casey & Center, 1985-86). However, Hattie and Timperley do not describe this study in much detail while Skiba et al. actually make an important distinction between primary reinforces (praise, attention and physical closeness), action amplifiers (choose preferred activities) and feedback (information given to the learner which did not contain any reinforcement from the above list). The first two were more effective in group situations, and the latter more in individual-individual situations. These authors note that it is difficult to find any difference between the feedback and amplifiers since the first, to some extent, will always be a part of the second.

Hattie and Timperley's (2007) more generic model of feedback is interesting, but it is largely incentive to the (cultural) context in in which learning and assessment takes place. In their article, Hattie and Timperley are somewhat sceptical about the kind of feedback that puts an emphasis on 'the self' as a person ("you're a star") and emotional weighted ratings ("I am astonished over what you've written "), rather than 'subject directed feedback' only. It is however important to note that the feedback praising young under-achievers, such as those in the LIBE project, can be a significant springboard for motivation towards learning and task completion (Dobson & Søby, 2012).

Although research in this field lacks the focus on the special needs of specific learner groups, six studies in this review found results related to issues that have been known characteristics of low achieves such as self-confidence and motivation. Low self-confidence was significantly related to lower persistence, less effort, poorer achievement, and less satisfaction with performance then high self-confidence. As was concluded in the former paragraph, this result again points to the importance of boosting/maintaining the self-confidence of the learners participating in the LIBE project's virtual learning environment. Additionally it was found that the possibility to try formative assessments multiple times is especially beneficial to lower achieving learners as it allows them more time to go through basic concepts and reduces stress. On the other hand, time spend on feedback and formative assessment was influenced by motivation. The more motivated learners are, the more time is spend on feedback resulting in more study time. Time spend on feedback was found to be significantly higher for learners who received immediate feedback compared to learners who received delayed feedback (Van der Kleij, 2012), suggesting that this feedback characteristic influences learners' motivation. For the matter of maintaining motivation, it also seems important to keep feedback concise to prevent unnecessary feedback (e.g. overly directive feedback or too much detail) and feedback that does not match the learner's level or competencies.

Most studies regard to the formative assessment types 'assessment for learning', 'diagnostic testing' or a combination of the two. This is interesting as the implementation of formative assessment in the classroom often also involves DBDM (Schildkamp, Veldkamp, Dijkstra, van der Kleij, Heitink, Hoogland, 2014). In e-learning formative assessment and feedback seems mostly focused on the individual learner. Although DBDM can be employed on the individual level as well, AFL and DT might fit this aggregation level better. Additionally, the use of formative assessment in e-learning seems to be associated with AFL or DT, but not with DBDM. This might be because the definition of DBDM seems to be placed in a school context with teachers or principals collecting data for evaluation purposes. When looking from a broader perspective, automating DBDM might serve the overall goal of tailoring the

pedagogy used in e-learning to the learners' needs. A certain level of information is needed to provide this personalized experience (McKenzie, 2013) which for example can be achieved trough learning analytics.

A limitation of this study is that in systematic reviews it impossible to include all studies relevant to the subject of the review due to search and selection strategies. Furthermore a publication bias should be considered as often small studies with negative effects or small positive effects are not always published in scientific journals.

4.1. IMPLICATIONS FOR LIBE

Based on the finings and discussion points of this review, the following list summarizes the main points that are important for the design of LIBE's virtual learning environment.

- Use elaborated feedback on individual responses in the form of hints aiming to assist in understanding why the answer was incorrect or hints on areas of weakness, and hints on how to proceed.
- Next to cognitive and meta-cognitive feedback, Include feedback with an emotional function (praise and emotional weighted ratings) to keep young low achievers motivated towards learning and task completion and boost their self-confidence.
- Give learners the possibility to use feedback in subsequent tries and the possibility to try a formative assessment multiple times in order to allow them more time to go through basic concepts and to reduce stress.
- Time the feedback in an immediate way. This means that in the LIBE virtual learning environment feedback should be automated.
- Due to the automated nature of the feedback, the formative assessment should be focused on multiple choice or short answer questions.
- Give students the opportunity for self-assessment.
- More general feedback should focus on criteria for performance that have been met and criteria for performance that not yet have been met.
- Monitor the learner's progress trough the virtual learning environment in order adapt the learning path.
- Keep feedback concise in order to prevent demotivation by unnecessary feedback (e.g. overly directive feedback or too much detail). This can be done by 'cutting' the feedback in several portions of information of which more elaborated information can appear on the learner's demand. Both information used in feedback on individual responses and feedback about their (general) performance should be clear and meaningful.
- Keep the whole test/assessment concise and short to maintain the learner's attention. Rather use multiple short tests in stead of one long test.

- Adapt the test difficulty to the individual learner in order to maintain the learner's attention, motivation and self confidence. This includes both the difficulty of the questions as well as the difficulty of the texts etc. (e.g. use basic vocabulary) (cf. LIBE deliverable 3.1).

5. REFERENCES

5.1. INCLUDED IN THE REVIEW

Bälter, O., Enström, E. & Klingenberg, B. (2013). The effect of short formative diagnostic web quizzes with minimal feedback. *Computers and Education, 60*(1), 234-242. doi: 10.1016/j.compedu.2012.08.014

Bokhove, C. & Drijvers, P. (2012). Effects of feedback in an online algebra intervention. *Technology, Knowledge and Learning, 17*(1-2), 43-59. doi: 10.1016/j.compedu.2011.08.010

Buchanan, T. (2000). The efficacy of a World-Wide Web mediated formative assessment. *Journal of Computer Assisted Learning, 16*(3), 193-200. doi: 10.1046/j.1365-2729.2000.00132.x

Ćukušić, M., Garača, Ž., Jadrić, M. (2014). Online self-assessment and students' success in higher education institutions. *Computers and Education, 72*, 100-109. doi: 10.1016/j.compedu.2013.10.018

de-Marcos, L., Hilera, J. R., Barchino, R., Jiménez, L., Martínez, J.J., Gutiérrez, J.A., Gutiérrez, J.M., Otón, S. (2010). An experiment for improving students' performance in secondary and tertiary education by means of m-learning auto-assessment. *Computers & Education, 55*(3), 1069-1079.

Einig, S. (2013). Supporting Students' Learning: The Use of Formative Online Assessments. *Accounting Education 22*(5), 425-444. doi: 10.1080/09639284.2013.803868

González, J. A., Jover, L., Cobo, E., & Muñoz, P. (2010). A web-based learning tool improves student performance in statistics: a randomized masked trial. *Computers & Education, 55*(2), 704–713. doi: 10.1016/j.compedu.2010.03.003.

Ibabe, I. & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education, 59*(2), 243-258. doi: 10.1007/s10734-009-9245-6

Johnson-Glenberg & Mina, C. (2010). Embedded formative e-assessment: Who benefits, who falters. *Educational Media International, 47*(2), 153-171. doi: 10.1080/09523987.2010.492681

Kleitman, S. & Costa, D. S. J. (2014). The role of a novel formative assessment tool (Stats-mIQ) and individual differences in real-life academic performance. *Learning and Individual Differences, 29,* 150-161. DOI: 10.1016/j.lindif.2012.12.001

Lafuente, M., Remesal, A. & Álvarez Valdivia, I.M. (2014). Assisting learning in e-assessment: a closer look at educational supports. *Assessment and Evaluation in Higher Education, 39*(4), 443-460. doi: 10.1080/02602938.2013.848835

Limniou, M. & Smith, M. (2014). The role of feedback in e-assessments for engineering education. *Education and Information Technologies, 19*(1), 209-225. doi: 10.1007/s10639-012-9200-5

McKenzie, W.A., Perini, E., Rohlf, V., Toukhsati, S., Conduit, R. & Sanson, G. (2013). A blended learning lecture delivery model for large and diverse undergraduate cohorts. *Computers and Education, 64*, 116-126. doi: 10.1016/j.compedu.2013.01.009

Narciss, S. (2004). The Impact of Informative Tutoring Feedback and Self-Efficacy on Motivation and Achievement in Concept Learning. *Experimental Psychology, 51*(3), 214-228. doi: 10.1027/1618-3169.51.3.214

Sorensen, E. (2013). Implementation and student perceptions of e-assessment in a Chemical Engineering module. *European Journal of Engineering Education, 38*(2), 172-185. doi: 10.1080/03043797.2012.760533

Timmers, C. F., Braber-van den Broek, J. & van den Berg, S. M. (2013). Motivational Beliefs, Student Effort, and Feedback Behaviour in Computer-Based Formative Assessment. *Computers & Education, 60*(1), 25-31. 10.1016/j.compedu.2012.07.007

Timmers, C. & Veldkamp, B. P. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers and Education, 56*(3), 923-930. doi: 10.1016/j.compedu.2010.11.007

van der Kleij, F.M., Eggen, T. J. H. M., Timmers, C. F. & Veldkamp, B.P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education, 58*(1), 263-272. doi: 10.1016/j.compedu.2011.07.020

Wilson, K., Boyd, C., Chen, L. & Jamal, S. (2011). Improving student performance in a first-year geography course: Examining the importance of computer-assisted formative assessment. *Computers & Education, 57*(2), 1493-1500. doi: 10.1016/j.compedu.2011.02.011

5.2. OTHER REFERENCES

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. Review of Educational Research, 61(2), 213–238.

Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Clark, R.C. & Mayer, R.E. (2008). *E-learning and the science of instruction.* John Wiley & sons, Inc: USA.

Crisp, G. T. (2012). Integrative assessment: Reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education*, *37*, 33–43. doi:10.1080/02602938.2010.494234

Dobson, S., and Søby, K. (2012). Assessment as an expression of subject understanding and social recognition. A study of feedback to students with weak learning skills. (published in Norwegian: Vurdering som uttrykk for faglig forståelse og anerkjennelse. En nærstudie av feedback knyttet til elever med svake læreforutsetninger). In T. Nordahl (Ed.), *Improved*

learning for all pupils (Bedre læring for alle elever). Oslo: Gyldendal Publishers; pp134-149.

Dobson, S, Eggen, A. & Smith, K. (2009). *Vurdering, prinsipper og praksis (Assessment, principles and practice). Oslo: Gyldendal.*

J.W. Gikandi, J.W., Morrow, D. & Davis, N.E. (2011). Online Formative Assessment in Higher Education: A Review of the Literature. *Computers & Education, 57*(4), 2333-2351. doi: 10.1016/j.compedu.2011.06.004

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education, 35,* 213–224. doi:10.1080/03057640500146880

Hattie, J. & Timperley, H (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice, 16,* 263–268.

Narciss, S. (2008). *Feedback strategies for interactive learning tasks*. In J.M. Spector, M.D. Merrill, J.J.G. van Merrienboer, & M.P. Driscoll (Eds.), Handbook of Research on Educational Communications and Technology (3rd ed., pp. 125-144). Mahaw, NJ: Lawrence Erlbaum Associates.

Narciss, S. (2013). Designing and Evaluating Tutoring Feedback Strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *Digital Education Review, 23*.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education, 26,* 482–496.

Schildkamp, K., Veldkamp, B., Dijkstra, A., van der Kleij, F., Heitink, M. & Hoogland, I. (2014). Voorwaarden voor de implementatie van effectieve formatieve toetsing in de onderwijspraktijk [Conditions for the implementation of effective formative assessment in educational practice]. University of Twente: The Netherlands.

Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189.

Skiba, R.J., Casey, A., Center, B.A. (1985-86). Non-aversive procedures in the treatment of classroom behavioral problems. *Journal of Special Education, 19(*4), 459-481.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment.* Abingdon, England: Routledge.

Van der Kleij, F. M., Timmers, C. F., & Eggen, T. J. H. M. (2011). The effectiveness of methods for providing written feedback through a computer-based assessment for learning: A systematic review. *CADMO, 19*, 21–39.

Van der Kleij, F.M., Feskens, R.C.W. & Eggen, T.J.H.M. (2013). *Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis.* In: Annual meeting NCME, 26-04-2013 - 30-04-2013, San Francisco.

Van der Kleij, F., Vermeulen, J.A., Schildkamp, K., & Eggen, T. (2013). *Towards an Integrative Formative Approach of Data-Based Decision Making, Assessment for Learning, and Diagnostic Testing.* Paper presented at the International Congress for School Effectiveness and Improvement, January, Chili.

Webb, M., Gibson, D. & Forkosh-Baruch, A. (2013). Challenges for information technology supporting educational assessment. *Journal of Computer Assisted Learning, 29* (5), 451–462. doi: 10.1111/jcal.12033

Welsh, M. (2012). Student perceptions of using the PebblePad e-portfolio system to support self- and peer-based formative assessment*. Technology, Pedagogy and Education, 21*(1), 57-83.doi: 10.1080/1475939X.2012.659884

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, *37*, 3–14

6. APPENDICES

6.1. APPENDIX 1: SEARCH PROCESS

6.1.1. Search strategy and strings

SCOPUS

	Search terms	#hits
Formative assessment, diagnostic testing, data-based decision making en assessment for learning*	**"Formative assessment*"** OR **"diagnostic test*"** OR **"diagnostic assessment*"** OR **"data-based decision*"** OR **"data-driven decision*"** OR **"assessment for learning"**OR **"performance assessment*"** OR **"computer-based formative assessment*"**	130217
Feedback	AND ("feedback" OR "feed back" OR "formative feedback")	3156
e-learning	AND ("e-learn*" OR "electronic* learn*" OR "computer based learn*" OR "computer assist* learn*")	232
In the field of psychology, social sciences, computer sciences, decision sciences, multidisciplinary	LIMIT-TO(SUBJAREA, "DECI") OR LIMIT-TO(SUBJAREA, "SOCI") OR LIMIT-TO(SUBJAREA, "COMP") OR LIMIT-TO(SUBJAREA, "PSYC") OR LIMIT-TO(SUBJAREA, "DECI") OR LIMIT-TO(SUBJAREA, "MULT") OR LIMIT-TO(SUBJAREA, "MULT"))	201
English or Dutch	AND (LIMIT-TO(LANGUAGE, **"English"**) OR LIMIT-TO(LANGUAGE, **"Dutch"**))	200
Articles or en conference proceeding*	LIMIT-TO(DOCTYPE, "ar") OR LIMIT-TO(DOCTYPE, "cp"	171

*When selecting reviews, 14 were left.

Web Of Science

	Search terms	#hits
Formative assessment,	**"Formative assessment*"** OR **"diagnostic test*"** OR	62801

diagnostic testing, data-based decision making en assessment for learning*	**"diagnostic assessment*"** OR **"data-based decision*"** OR **"data-driven decision*"** OR **"assessment for learning"**OR **"performance assessment*"** OR "computer-based formative assessment*"	
Feedback	AND ("feedback" OR "feed back" OR "formative feedback")	1096
e-learning	AND ("e-learn*" OR "electronic* learn*" OR "computer based learn*" OR "computer assist* learn*")	34
In the field of psychology, social sciences, computer sciences, educational research	RESEARCH AREAS: (EDUCATION EDUCATIONAL RESEARCH OR COMPUTER SCIENCE OR PSYCHOLOGY OR SOCIAL SCIENCES OTHER TOPICS)	32
English or Dutch	AND (LIMIT-TO(LANGUAGE, **"English"**) OR LIMIT-TO(LANGUAGE, **"Dutch"**))	32
Articles or conference proceedings*	DOCUMENT TYPES: (ARTICLE)	17

* Only choice was article, meeting, other

ERIC

	Zoektermen	#hits
Formative assessment, diagnostic testing, data-based decision making en assessment for learning*	"Formative assessment*" OR **"diagnostic test*"** OR **"diagnostic assessment*"** OR **"data-based decision*"** OR **"data-driven decision*"** OR **"assessment for learning"**OR **"performance assessment*"** OR "computer-based formative assessment*"	5383
Feedback*	AND ("feedback" OR "feed back" OR "formative feedback")	468
e-learning	AND ("e-learn*" OR "electronic* learn*" OR "computer based learn*" OR "computer assist* learn*")	26
In the field of psychology, social sciences, computer sciences, educational research	nvt	-

English or Dutch	AND (LIMIT-TO(LANGUAGE, **"English"**) OR LIMIT-TO(LANGUAGE, **"Dutch"**))	26
Articles or conference proceedings*	Academic journals	24

* Only choice was academic journals, eric documents and books

PsycINFO

	Search terms	# hits
Formative assessment, diagnostic testing, data-based decision making en assessment for learning*	**"Formative assessment*"** OR **"diagnostic test*"** OR **"diagnostic assessment*"** OR **"data-based decision*"** OR **"data-driven decision*"** OR **"assessment for learning"** OR **"performance assessment*"** OR **"computer-based formative assessment*"**	5257
Feedback*	AND ("feedback" OR "feed back" OR "formative feedback")	376
e-learning	AND ("e-learn*" OR "electronic* learn*" OR "computer based learn*" OR "computer assist* learn*")	12
In the field of psychology, social sciences, computer sciences, educational research	nvt	-
English or Dutch	AND (LIMIT-TO(LANGUAGE, **"English"**) OR LIMIT-TO(LANGUAGE, **"Dutch"**))	12
Articles or conference proceedings*	Academic journals and dissertations	12

* Only choice was academic journals and dissertations

Total overview:

Selection steps	Number of publications
Start total (after search process)	224

After removing duplicates	192
After removing irrelevant journals	192
After removing conference proceedings, books, etc.	134

6.1.2. Inclusion criteria

5) The study was published in a scientific, peer reviewed, journal or dissertation. Books, book chapters and conference proceedings are not selected to secure the scientific quality of the studies presented in the publications. The scientific quality of the publications that remain will be explored in a later stadium of this review study (see section 2.5).

6) The publication concerns scientific research. Theoretical considerations, reviews and opinions will not be selected as this review study focuses on factors that have proven to be effective on using feedback or formative assessment in e-learning. Although there is a need to be careful to generalize the results of case studies, case studies were included in the selection for their practical examples in 'real life' contexts which is important for the LIBE project as well.

7) This review study focuses on the contexts of secondary education, vocational education and higher education - with specific attention paid to low achiever target audiences - as these are contexts that overlap with the age of students in the LIBE project. Primary education is not included, as is special education.

8) The study is focused on computer based feedback and formative assessment in e-learning.

6.1.3. Title scan

Titles of the remaining publications were scanned to select the ones that fit the inclusion criteria. When in doubt, the publication remained in the selection and was included in the abstract scan.

Selection steps	Selected studies
Start total	*134*
1. The study was published in a scientific, peer reviewed, journal or dissertation.	134 (-0)

2. The publication concerns scientific research.	129 (-5)
3. Context of secondary education, vocational education and higher education	126 (-3)
4. The study is focused on feedback and formative assessment in e-learning.	118 (-8)

6.1.4. Abstract scan

After the title scans, abstracts of the remaining publications were scanned to select the ones that fit the inclusion criteria.

Selection steps	Selected studies
Start total	*118*
1. The study was published in a scientific, peer reviewed, journal or dissertation.	118 (-0)
2. The publication concerns scientific research.	104 (-14)
3. Context of secondary education, vocational education and higher education	98 (-6)
4. The study is focused on feedback and formative assessment in e-learning.	55 (-43)

- After the abstract scan 55 articles were left for data extraction.
- During the search, 24 non-research (e.g. reviews) papers were collected. Since these papers could give insight in the field as well snowballing was applied. 15 of these studies seemed interesting for snowballing. From which 12 of these studies the actual paper was attained.

6.1.5. Additional search strategies

- Extra search for common authors in the field (e.g. Narciss, Shute): 52 articles left for data extraction.

In total, from 43 of these studies the actual paper was attained.