Computers & Education 87 (2015) 1–9

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

The effect of regulation feedback in a computer-based formative assessment on information problem solving

Caroline F. Timmers [a,*], Amber Walraven [b], Bernard P. Veldkamp [c]

[a] *Saxion University of Applied Sciences, Enschede, The Netherlands*
[b] *Radboud University Nijmegen, Nijmegen, The Netherlands*
[c] *University of Twente, Enschede, The Netherlands*

ARTICLE INFO

Article history:
Received 28 May 2014
Received in revised form
5 March 2015
Accepted 9 March 2015
Available online 2 April 2015

Keywords:
Learning analytics
Assessment for learning
Information literacy
Self-assessment

ABSTRACT

This study examines the effect of regulation feedback in a computer-based formative assessment in the context of searching for information online. Fifty 13-year-old students completed two randomly selected assessment tasks, receiving automated regulation feedback between them. Student performance was (self-)graded by students and by experts. Expert, as well as student (self)grades showed a significant increase between Task 1 and Task 2. However, further analysis of the expert grades showed significant improvement in performance for girls only. Furthermore, the formative assessment system traced the number of searches and the number of websites consulted per student to complete the two assignments. On average, the results showed that students consulted significantly more websites for Task 2, compared to Task 1. The average number of searches did not differ significantly between Tasks 1 and 2. On the other hand, significant differences were found for those students who, during the evaluation of their performance on Task 1, explicitly stated that they would increase their searches.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Rapid technological developments during the past decades have led to changing requirements for today's and future workers, increasingly requiring them to apply what are now called 21st century skills (Murnane & Levy, 2004). These skills include critical thinking, problem solving and creativity, and they ‘are not new, just newly important’ (Silva, 2009, p.631). Information problem solving is such a 21st century skill (Voogt & Roblin, 2010). Information problem solving refers to the ability to identify information needs, locate corresponding information sources, extract and organise relevant information from each source, and synthesise information from a variety of sources (Walraven, Brand-Gruwel, & Boshuizen, 2008). Research shows that students are not nearly as information savvy as their substantial use of the Internet requires them to be (Connaway & Dickey, 2010; Kolowich, 2011; Thompson, 2011). In addition, students tend to overestimate their own information problem solving skills (Gross & Latham, 2012; Ivanitskaya, O'Boyle, & Casey, 2006). Previous research implies that students generally show deficiencies with regard to defining the information problem, searching and evaluating information (Head & Eisenberg, 2010; Monroe-Gulick & Petr, 2012; Pinto, 2012; Walraven et al., 2008; Walraven, Brand-Gruwel, & Boshuizen, 2009). Courses in and standards for information problem solving are focused on aspects such as search strategy and the evaluation of the relevance and accuracy of information. However, most of these courses are aimed at graduate students (e.g., Dirkx et al., 2011), although these skills are also relevant in primary and secondary education.

Walraven and Voogt (2014) developed a computer-based instrument that traces student behaviour while searching, selecting and using information to solve an information problem. This instrument is called the Digital Information Skills Measurement (DIM). The DIM combines the fast and large scale data collection of a survey with the accurate view of skills provided by an observation and can be used in educational settings by teachers. The instrument includes assignments which require students to search the Internet for relevant information. It is a computer-based test-like event which also generates feedback. As such, the DIM can be categorised as a computer-based formative assessment (CBFA). The main aim of formative assessment is to support and stimulate student learning. Previous research shows that

* Corresponding author.
E-mail address: c.f.timmers@saxion.nl (C.F. Timmers).

http://dx.doi.org/10.1016/j.compedu.2015.03.012
0360-1315/© 2015 Elsevier Ltd. All rights reserved.

feedback in CBFA can positively influence student learning, but does not necessarily do so (Van der Kleij, Timmers, & Eggen, 2011). The effect of feedback is influenced by its characteristics. Feedback included in the DIM can be characterised as feedback aimed at stimulating the self-regulation of learning or the so-called regulation level of feedback (Hattie & Timperley, 2007). This kind of feedback encourages learners to monitor or self-assess their performance in order to modify affective, cognitive and behavioural processes with the intention of improving performance (Sitzmann & Ely, 2011). Although regulation level feedback is suggested to be effective, there is hardly any research available on the effect of regulation level feedback in computer-based formative assessments (Van der Kleij et al., 2011). The aim of this study is to examine the effect of regulation feedback on student performance and information-seeking behaviour in a CBFA. We will focus on the effects of regulation feedback in the DIM and determine whether it can be recommended as a learning tool for educational practice.

2. Theoretical framework

2.1. Computer-based formative assessment

In this study, CBFA is defined as a purposefully designed instrument embedded within a learning process (Bennett, 2011). Although formative assessment is not univocally defined in previous research, researchers generally agree that the aim of formative assessment is to stimulate and direct student learning, for example, by providing students with feedback on performance. Feedback informs learners or teachers about the actual state of performance. Effective feedback includes information on current performance, the intended level of performance, and how to bridge a performance gap (cf. Hattie & Timperley, 2007).

In this article, CBFA is conceptualised using the five-stage model for the process of receiving feedback in the context of a test-like event (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). A schematic overview is presented in Fig. 1. The learners' initial state (stage 1) is characterised by cognitive aspects and motivational beliefs. Students' initial motivational beliefs (e.g., success expectancy and task-value beliefs) influence the effort that students invest in the subsequent stages of the model (Timmers, Braber-Van den Broek, & Van den Berg, 2013). When a test-like event is administered, students are encouraged to address relevant prior knowledge and skills (stage 2) and construct a response (stage 3). After generating one or more responses, students are provided with automated feedback. Next, students are expected to evaluate their performance by processing the feedback (stage 4). The evaluation of performance can result in adjustments in the students' cognitive state and motivational beliefs (stage 5), and, subsequently, lead to an adjusted initial state (stage 1). When, for example, the intended learning outcome of a CBFA is increased knowledge and understanding in a specific domain, the process is considered successful if learners add to, tune or restructure domain knowledge and understanding (stage 5). Additionally, students' experiences with the CBFA can result in adjustment or confirmation of the initial motivational beliefs (e.g., task-value beliefs). The effect of the CBFA (instrument, as well as process) can be defined as the differences between stages 1 and 5.

The effectiveness of CBFA is influenced by the characteristics of the CBFA instrument, including the feedback intervention. For example, CBFA instruments which include feedback providing students with correct answers and additional information about the task have been shown to be more effective than CBFA instruments that merely inform students whether their answer is correct or incorrect (Van der Kleij et al., 2011). Elaborated feedback refers to information in addition to information about the correctness of a response and the correct response and has proven to be the most effective (Narciss et al., 2014). Regulation feedback is considered a type of elaborated feedback. Furthermore, the effect of CBFA strongly depends on the effort students invest in the task, including processing corresponding feedback (Bangert-Drowns et al., 1991; Timmers et al., 2013). In addition, research by Narciss et al. (2014) showed gender differences in the effects of feedback. Boys tend to benefit less than girls from tutoring feedback, while practicing fractions.

2.2. Self-regulated learning and feedback

Hattie and Timperley (2007) distinguished between four levels of feedback: feedback about the self, the task, the process and self-regulation. Self-level feedback provides information about the person (e.g., well done). Task-level feedback provides learners with information about their performance on a certain task. Process-level feedback provides learners with information about (their performance in) processing the task. Feedback aimed at self-regulation, or regulation-level feedback, focuses on (greater skills in) self-evaluation or

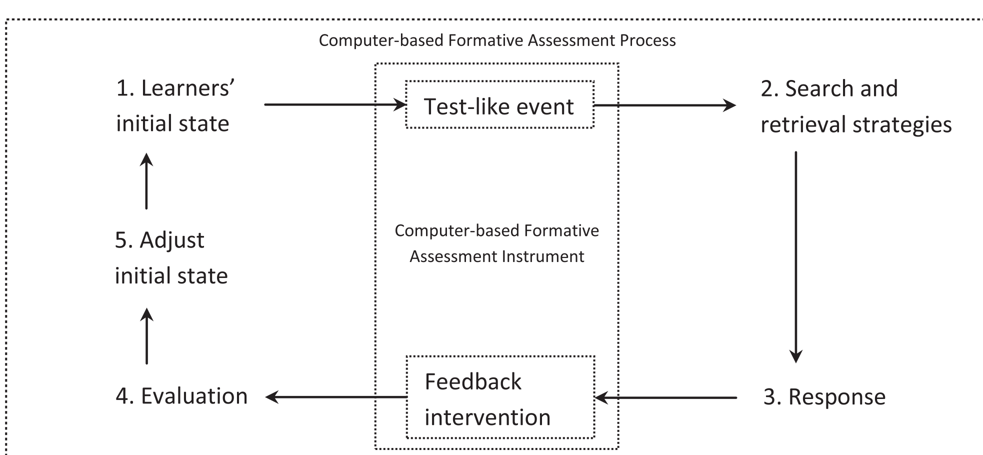

Fig. 1. Conceptualisation of CBFA based on the five-stage model proposed by Bangert-Drowns et al. (1991); adapted from Mory (2004, p. 752).

confidence to engage further in a task. In this study, regulation feedback is defined as feedback that explicitly stimulates students to actively engage in evaluating their performance, including the processes underlying their performance, which are aimed at the regulation of learning. From a meta-cognitivist perspective, the impact of regulation feedback on the modification or adjustment of goals, (meta)cognitive strategies and motivational beliefs is assumed to be higher, compared to feedback aimed at the task or the process (Hattie & Timperley, 2007; Thurlings, Vermeulen, Bastiaens, & Stijnen, 2013).

Self-regulating learners are learners that monitor and modify affective, cognitive and behavioural processes in order to manage and direct their own learning (Pintrich, 1999; Sitzmann & Ely, 2011). Sitzmann and Ely (2011) distinguished between various processes underlying successful self-regulated learning: 1) processes that initiate self-regulated learning, e.g., goal setting; 2) processes that learners apply to achieve their goals, e.g., planning and monitoring; and 3) processes that influence students' learning beliefs, e.g., self-efficacy. According to Butler and Winne (1995), feedback is an inherent catalyst for all self-regulatory processes underlying self-regulated learning. CBFAs can influence self-regulatory processes by generating feedback on performance and providing students with the opportunity to monitor or evaluate their performance by comparing their actual level of performance with the intended level of performance. The monitoring or self-assessment of performance may result in affective, cognitive and behavioural adjustments (consider stage 5 presented in Fig. 1) of the learner's initial state (consider stage 1 presented in Fig. 1). For example, when a student initially expects to be more successful on a task than he or she actually is, this is likely to result in an adjustment of the student's initial success expectancy (affective adjustment). In addition, learners might resolve a misunderstanding by carefully studying the feedback (cognitive adjustment). For example, when students learn about the existence of a new tool, they might start using this new tool (e.g., a subject-specific database with validated information) in addition to the old tool (e.g., Google) the next time they are confronted with a similar task or situation (behavioural adjustment).

Questions can be included in regulation feedback to encourage students to self-assess their performance on a task (De Bruijn-Smolders, Timmers, Gawke, Schoonman, & Born, 2014). An exemplary question is: What, in your opinion, is the quality of your performance on this task? With this exemplary question, students are asked to generate their own task or process level feedback through a self-assessment of their performance. However, students cannot be expected to be capable of creating valid feedback on a topic which they do not master (Hattie & Timperley, 2007). Therefore, information, such as worked-out examples or performance criteria, can be added as a supplement to the questions posed in regulation feedback to support the students' self-evaluation of performance. Kramarski and Gutman (2006) investigated the effect of what they call self-metacognitive questioning in a mathematical e-learning environment. The questions in self-metacognitive questioning can address 1) comprehension of the problem, 2) constructing a connection between previous and new knowledge, 3) using appropriate strategies to solve the problem, and 4) reflecting on the processes and the solution (Kramarski & Gutman, 2006). The treatment investigated in the study by Kramarski and Gutman (2006) included questions addressing all four categories of self-metacognitive questioning. The self-metacognitive questions were embedded in a mathematical e-learning environment and were provided during and after the solution process (see stages 2, 3 and 4 as presented in Fig. 1). The results showed a positive effect of the self-metacognitive questioning intervention on mathematical problem solving.

2.3. Information problem solving and information-seeking behaviour

There are several models of the process of information problem solving. Three prominent models are the search process model (Kuhltau, 1993), the research process model (Stripling & Pitts, 1988) and the Big6-model (Eisenberg & Berkowitz, 1990.). Although these models are prominent and successful, they are not primarily based on situations in which students use the WWW to search for information. Brand-Gruwel, Wopereis, and Walraven (2009) discussed a model based on empirical research that is aimed at information problem solving on the Internet: the IPS-I model (see Fig. 2). The model describes the constituent and sub-skills needed to solve an information problem when the Internet is used to search for information. Only the constituent skills are depicted in the model. The following description of the model in our article is based on the description of the model by Brand-Gruwel et al. (2009). For a complete description, please refer to their article.

The five constituent skills, 'Define information problem' (e.g., formulate main and sub-questions), 'Search information' (e.g., type a query and choose sites to open), 'Scan information' (e.g., form an impression of the kind of information and whether it is useful), 'Process

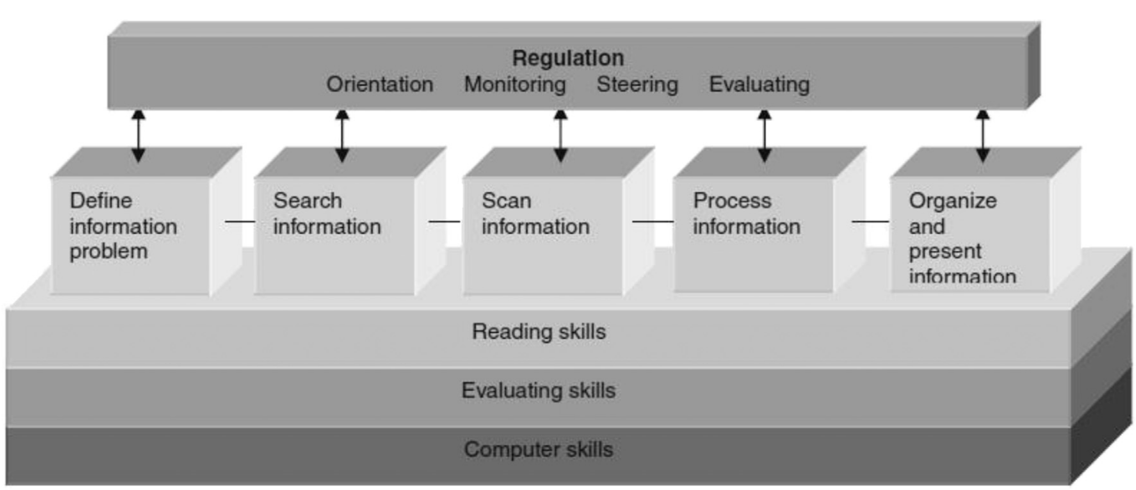

Fig. 2. The IPS-I model by Brand-Gruwel et al. (2009).

information' (e.g., process information to reach a deeper understanding and combine the new information found with relevant prior knowledge), and 'Organize and present information' (e.g., combine information to solve an information problem, resulting in a product) are linked blocks. Regulation of the process of solving an information problem happens during the execution of all skills and is visualised by the arrows. The five blocks rest upon three layers which represent the three most important conditional skills: (a) 'Reading skills', (b) 'Evaluating skills', and (c) 'Computer skills'.

3. Aim of the present study

The aim of this study is to examine the effect of regulation feedback in the DIM on student performance and information-seeking behaviour. The effect of regulation feedback is influenced by the presence of a performance gap, student acknowledgement of a performance gap and student intentions to bridge a performance gap (Sitzmann & Ely, 2011). Based on previous research (Head & Eisenberg, 2010; Monroe-Gulick & Petr, 2012; Pinto, 2012; Walraven et al., 2008), a performance gap in information-seeking behaviour is expected. Furthermore, it is likely that students will be unaware of an existing performance gap as they tend to overestimate their information-seeking behaviour (Gross & Latham, 2007; Ivanitskaya et al., 2006). Use of the DIM is expected to result in student awareness of existing performance gaps, due to the worked-out example and self-questions (Hypothesis 1). Subsequently, student awareness is expected to lead to student intentions to bridge the performance gap for example by setting goals (Hypothesis 2). Furthermore, based on previous research, regulation feedback is expected to have a positive effect on student performance (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Van der Kleij, Feskens, & Eggen, 2013). Therefore, it is expected that the regulation feedback in the DIM tool will lead to improved performance on Task 2 (Hypothesis 3). Additionally, as previous research suggests, girls tend to benefit more from feedback than boys (Narciss et al., 2014), and thus, the effect of regulation feedback is expected to be more positive for girls, compared to boys (Hypothesis 4). Finally, as regulation feedback in the DIM is intended and suggested to lead to behavioural adjustments (Hattie & Timperley, 2007; Sitzmann & Ely, 2011), it is expected that student information-seeking behaviour will change between Task 1 and Task 2 (Hypothesis 5).

4. Methodology

4.1. Respondents

Fifty students (24 males and 26 females) of 2 s grade secondary school classes at the same school were asked to solve two information problems using the DIM during history lessons. The respondents shared the same teacher. The mean age of the respondents was 13 (sd = 0.5).

4.2. Digital Information Skills Measurement instrument

The DIM is based on the model presented in Fig. 2 and was developed by Walraven and Voogt (2014) in order to gather data to examine student information skills and information-seeking behaviour. The DIM includes prose-task assignments, which require students to search the Internet for information in order to complete the assignments. The DIM included four comparable prose-task assignments to measure students' information skills. Mosenthal's (1998) constructs defining the difficulty level of prose-task processing were used in an endeavour to achieve comparability between the assignments, namely: 1) type of information requested, 2) type of match and 3) plausibility of distracters. All assignments required the identification of cause, effect, equivalence or difference (type of information requested). The assignments addressed an ambiguous topic and, as such, required the use of multiple sources. Furthermore, the assignment did not specify the available amount of perspectives on the matter, and it did not contain the most useful keywords in the assignment (type of match). The assignments addressed the following topics: influence of msn and sms on Dutch language skills, the existence of the greenhouse effect, the reputation of genetically modified food and radiation from cell phones. The assignment explicitly requested students to formulate their response in their own words, instead of copying one or more sections from the Internet.

The feedback included in the DIM is aimed at students' self-assessment of responses and information-seeking behaviour through "self-rating" and reflection (Brew, 1995) or self-questions (Kramarski & Gutman, 2006). Students were asked to compare their own response to a worked-out example and, subsequently, to grade their response on a scale of 1–10. Afterwards, the DIM presented students with a set of self-questions about their information-seeking behaviour. For example: '*Did you use more than one search term when searching for information? (yes/no/I don't remember)*' and '*What would you do similarly next time? Elaborate*'. Subsequently, students were asked to grade their own search process on a scale of 1–10. The worked-out example, the self-grading and the questions about information-seeking behaviour support student reflection on the process and the solution ((Kramarski & Gutman, 2006) and is intended to lead to behavioural adjustments (Hattie & Timperley, 2007; Sitzmann & Ely, 2011). To monitor actual behavioural adjustment, aspects of students' search behaviour were registered by the DIM, for example, the number of searches and the number of sites consulted. All data registered by the DIM system can be related to the various stages of the five-stage model presented in Fig. 1. Table 1 shows an overview of the data registered by the DIM for the various stages of the five-stage model.

4.3. Research design and procedure

The effect of regulation feedback on student performance and information-seeking behaviour was examined using one of the prose-task assignments included in the DIM (Task 1) as a pre-test and another as a post-test (Task 2). The assignments were randomly selected by the DIM system. Task 2 was selected from the remaining three assignments that had not yet been completed by a particular respondent. The intervention consisted of the regulation feedback included in Task 1. The data traced during Task 1 were used as an indication of the respondents' initial state (stage 1, as presented in Fig. 1). The data traced during Task 2 were used as an indication of the respondents' adjusted state (stage 5, as presented in Fig. 1). Task 1 was completed by the respondents on the 10th of January 2012. Two weeks later, on January 24th, the respondents were asked to complete Task 2.

Table 1
Overview of data registered by the DIM for the various stages of the five-stage model shown in Fig. 1.

Five-stage model of:	The process of receiving feedback in a test-like event	Data registered by the DIM
Stage 1	Students' initial state	
Stimulus	*Test-like event*	
Stage 2	Search and retrieval strategies	Number of searches Number of sites consulted
Stage 3	Construct a response	Student response
Stimulus	*Feedback Intervention*	
Stage 4	Evaluation	Self-grade response Self-grade process Student answer to: 'What would you do differently next time?'
Stage 5	Adjust initial state	

Before the students were asked to complete Task 1, the students were told that they were participating in a study on the way students search for information on the Internet, and that their information-seeking behaviour would be logged. The students were also informed about the random selection of the prose-task assignments. The students completed the assignments within a maximum timeframe of 30 min. The DIM provided a time indicator counting down from 30 to zero. At zero, the system automatically directed the students to the response page, where they could formulate their responses. The students were also informed that this activity would not influence their final grade for the history course.

4.4. Data analysis

Student responses to Task 1 and Task 2 were (self-)graded by the students and by the first two authors, both experts in the domain of information problem solving. Both the student and the expert grades were based on a comparison of student responses and the worked-out examples. The experts graded the student responses together, by discussing the response and using fixed criteria. To control for confirmation bias, nameless and dateless Task 1 and Task 2 responses were mixed. The responses were graded on a scale from 1 to 10, a commonly used scale in The Netherlands. In this study, the grades refer to "no sign of effort invested in the task" (1 or 2), "clearly insufficient" (3 or 4), "insufficient" (5), "sufficient" (6), "clearly sufficient" (7 or 8) and "outstanding" (9 or 10). The worked-out example was used as a reference level and was graded as "clearly sufficient" (8).

Student acknowledgement of a performance gap (Hypothesis 1) was tested by analysing whether the students graded their performances other than "clearly sufficient". Additionally, students' self-grades and expert grades for Task 1 were compared in order to examine the extent of acknowledgement of a performance gap. Herewith, the expert grades were used as an indication of the actual performance gap.

Student answers to the following self-question were used as an indication of the students' intentions to bridge a performance gap by setting goals (Hypothesis 2): 'What would you do differently the next time you are confronted with a similar task?' Student answers to this open-ended question were analysed with regard to student intentions to act differently the next time. Student answers were categorised as 1) 'I wouldn't change a thing or I don't know'; 2) 'I might act differently'; 3) 'I will act differently', and 4) 'I will act differently with regard to … '.

To examine the effect of the regulation feedback on student performance, the paired sample t-test was used to compare the Task 1 and Task 2 expert grades of all respondents (Hypothesis 3) and for boys and girls separately (Hypothesis 4).

Changes in student information-seeking behaviour between Task 1 and Task 2 (Hypothesis 5) were examined by comparing the number of searches and the unique websites consulted for Task 1 and Task 2, using a paired sample t-test. Changes in the number of searches and the number of websites consulted were examined for all respondents, and more specifically, for respondents that, during Task 1, explicitly set goals for changing their future information-seeking behaviour.

5. Results

5.1. Student acknowledgement of performance gaps

Table 2 provides an overview of all students' self-grades registered by the DIM for Task 1 and Task 2, as well as the expert-grades for the Task 1 and Task 2 responses. Students' self-grades for Task 1 responses, on a scale of 1–10, varied from 3 (clearly insufficient) to 9 (clearly sufficient). Three self-grades for Task 1 responses were missing. As can be seen in Table 3, 26 of the students graded their Task 1 response as "clearly sufficient" or "outstanding" (7–9). Ten students graded their Task 1 response as "sufficient" (6), eight as "insufficient" (5), and three as "clearly insufficient" (3–4). When using "clearly sufficient" as a reference level, 44.7% of the students acknowledged a performance gap or saw room for improvement in comparison to the worked-out example. When using the same reference levels for students' self-grades for the Task 1 information-seeking process, 28% of the students graded their Task 1 process lower than "clearly sufficient". Based on student self-grades, Hypothesis 1 was confirmed for 44.7% of the Task 1 responses and 28% of the Task 1 information-seeking process. As shown in Table 3, three out of 50 Task 1 responses were graded "clearly sufficient" by the experts. According to the experts, 94% of the responses showed room for improvement in comparison to the worked-out example.

5.2. Student goal-setting

Of all respondents, 84% ($n = 42$) answered the following question: 'What would you do differently next time? Thirty respondents (60%) indicated they would act differently next time. Twenty-two (44%) of the respondents explicitly indicated what they would do differently next time. Examples of student responses include: 'Next time, I would use more time to provide a more elaborate answer to the

Table 2
Students' self-grades and expert-grades for Task 1 and Task 2.

Grade	Expert grades		Students' self-grades			
	Task 1	Task 2	Task 1		Task 2	
	Response	Response	Response	Process	Response	Process
1						
2						
3	29	17	1	1		
4			2			
5	8	15	8	3	1	
6	10	8	10	10	9	9
7		4	18	20	19	22
8	3	6	6	10	18	16
9			2	3	3	3
10						
missing			3	3		

Table 3
Mean scores on Task 1 and Task 2 responses for males and females.

Gender	Task 1 response		Task 2 response	
	Expert-grade mean (sd)	Self-grade mean (sd)	Expert-grade mean (sd)	Self-grade mean (sd)
Males	4.29 (1.55)	6.83 (1.13)	4.50 (1.44)	7.33 (0.76)
Females	4.15 (1.64)	6.04 (1.33)[a]	5.46 (1.86)	7.19 (1.02)

[a] Three out of 26 self-grades were missing.

question'; 'next time, I would use more different search terms'; 'next time, I would visit more sites, to search for more different information'. The intention to visit more websites the next time was included in the answers of 10 students (20%). Five students answered that they would use (more) different search terms the next time. Based on the analysis of student answers to the question about what they would do differently, Hypothesis 2 was confirmed for 60% of the respondents, including 44% of the respondents that explicitly set goals for changing future information-seeking behaviour.

5.3. *Effect of regulation feedback on performance*

The paired sample t-test showed a significant difference ($t = -2.23$; $df = 49$; 2-tailed $p = 0.031$) between the quality of Task 1 responses ($M = 4.22$; $SD = 1.58$) and Task 2 responses ($M = 5.00$; $SD = 1.72$). Hypothesis 3 was confirmed.

Table 3 presents the mean Task 1 and Task 2 scores for male and female respondents. A t-test shows that the quality of Task 1 responses does not significantly differ between male and female respondents. In addition, paired sample t-tests show that the difference between the quality of Task 1 and Task 2 responses is not significant for the male respondents, while the difference between the quality of Task 1 and Task 2 responses is significant for the female respondents ($t = -2.75$; $df = 25$; 2-tailed $p = 0.011$). For girls, the quality of the Task 2 responses is significantly higher, compared to the Task 1 responses. Hypothesis 4 was confirmed.

5.4. *Changes in information-seeking behaviour*

Table 4 presents an overview of the number of searches and unique websites consulted for Task 1 and Task 2, distinguishing between all respondents and those who explicitly formulated a goal for changing future information-seeking behaviour. For all respondents, paired sample t-tests showed a significant increase only in the number of unique websites consulted ($t = -2.056$; $df = 49$; 2-tailed $p = 0.045$). Hypothesis 5 is confirmed for the number of unique websites only.

During Task 1, five respondents explicitly formulated that they intended to increase their number of searches the next time. If the significance level is raised to $p = 0.1$, instead of $p = 0.05$, a significant increase in the number of searches ($t = -2.269$; $df = 4$; 2-tailed $p = 0.086$) for the explicit goal-setters is found. Ten respondents explicitly formulated that they intended to consult more websites the next time ($n = 10$). A significant difference in behaviour was found for the respondents that indicated that they would consult more sites the next time ($t = -2.605$; $df = 9$; 2-tailed $p = 0.029$). For explicit goal-setters, Hypothesis 5 was confirmed for the number of searches, as well as the number of sites.

6. Discussion

The aim of this study was to examine the effect of regulation feedback on student performance and behaviour. Based on expert-grades, the quality of student Task 1 responses suggested a significant performance gap in relation to the worked-out example. The respondents perceived a much smaller performance gap. Students' self-grades were significantly higher, compared to expert-grades. For the quality of the Task 1 responses, for example, a significant difference ($t = -2.15$; $df = 46$; 2-tailed $p = 0.00$) was found between the expert-grades ($m = 4.22$; $sd = 1.58$) and the students' self-grades ($m = 6.45$; $sd = 1.28$). Based on students' self-grades for their Task 1 response, 44.7% of the respondents acknowledged a performance gap in comparison to the worked-out example (Hypothesis 1 was partly confirmed). When analysing student answers to the self-questions, it is striking that 60% of the respondents indicated that they would change their

Table 4
Number of searches and unique websites visited for Task 1 and Task 2.

	Task 1		Task 2	
	Nr. of searches mean (sd)	Nr. of sites mean (sd)	Nr. of searches mean (sd)	Nr. of sites mean (sd)
All respondents ($n = 50$)	4.12 (2.78)	4.10 (1.99)	4.66 (3.91)	5.12 (3.24)
Explicit goal-setters ($n = 5$; $n = 10$)	3.00 (2.35)	3.60 (2.22)	5.20 (1.10)	6.00 (3.02)

information-seeking behaviour the next time, including 44% of the respondents who explicitly formulated what they would do differently the next time (Hypothesis 2 was partly confirmed). On average, the results showed that the quality of Task 2 responses had significantly improved, compared to Task 1 responses (Hypothesis 3 was confirmed). However, when distinguishing between male and female respondents, a significant performance improvement was found for female respondents only (Hypothesis 4 was confirmed). Furthermore, for all respondents, a significant increase was found for the number of unique websites consulted for Task 2, compared to Task 1. The results of respondents who explicitly set goals for changing future information-seeking behaviour showed significant increases in the number of websites consulted, as well as the number of searches (Hypothesis 5 was confirmed). Overall, the results suggest that the regulation feedback in the DIM can positively affect student performance and behaviour and can be recommended as a learning tool for educational practice.

The student performance gap in information problem solving suggested by the expert-grades is in line with the observations of previous research (Connaway & Dickey, 2010; Pinto, 2012; Walraven et al., 2008). Furthermore, the discrepancy between the expert-grades and the student-grades indicates a lack of agreement on quality standards between the experts and the students. This lack of agreement on quality standards could be a symptom of the student tendency to overestimate their own information literacy (e.g., Gross & Latham, 2012). However, it could also be a symptom of student awareness of underachieving, because previous research shows that weaker students are more generous when grading themselves (Boud, 1989). In addition, at the beginning of Task 1, the students were unfamiliar with the assessment criteria used by the experts. The regulation feedback provided after the completion of Task 1 was intended to raise awareness about the quality standards and, consequently, the existing performance gaps. Based on the discrepancy between the Task 2 quality judgements of the experts ($m = 5.00$; $sd = 1.73$) and the students ($m = 7.26$; $sd = 0.90$), it appears that the worked-out example and the guiding questions did not sufficiently support an alignment of the quality standards used by the experts and the students. Previous research has shown that explicit assessment criteria used by teachers, e.g., via a grading rubric or a criteria sheet, aligns student and teacher grades (Baird & Northfield, 1992; Boud, 1989; Weaver & Cotrell, 1986). Although the results show that regulation feedback in the DIM can positively affect student performance, further improvement of student performance is desirable and might be realised through further alignment of the quality standards by communicating the assessment criteria more clearly.

In this study, the expert-grades showed far higher performance gaps than the students' self-grades. On the other hand, both the expert-grades ($t = -2.23$; $df = 49$; $p = 0.03$) and the students' self-grades ($t = -3.79$; $df = 46$; $p = 0.00$) indicated significant performance improvement in Task 2 responses, compared to Task 1 responses. In short, the students and the experts did not agree on the quality standards, but they did agree that the quality of Task 2 responses had improved. As the aim of the regulation feedback in the DIM was a positive effect on student performance, one can question whether the lack of alignment with regard to quality standards is problematic. The observed difference between the quality standards used by the experts and the students is problematic only when a negative relation exists between the students' self-grades and the improvement in performance. Furthermore, if the self-grades were negatively related to performance improvement, this might add to a possible explanation of the gender differences observed. On average, girls self-graded their Task 1 response as “sufficient”, while boys self-graded their performance as “clearly sufficient”. These self-grades suggests that female respondents perceived and acknowledged more room for improvement in performance than boys. Further research is needed to examine the relation between self-grades and performance improvement and the causes of gender differences in relation to the effects of regulation feedback. Variables that need to be taken into account in future research are student motivational beliefs and feedback behaviour (Timmers et al., 2013).

The gender differences encountered in this study are in line with previous research, which shows that girls benefit more from feedback than boys (Narciss et al., 2014). The results of the study by Narciss et al. (2014), however, referred to tutoring feedback. This type of feedback includes varying feedback prompts as the student is provided with multiple opportunities to respond correctly after an incorrect response to a question or completion of a task. In addition, tutoring feedback can be characterised as a combination of task, process and regulation level feedback. The main difference between tutoring feedback and the feedback included in the DIM is that tutoring feedback provides an expert-judgement about the quality of a response or a process, while the regulation feedback in this study did not include an expert-judgement.

While grading Task 1 and Task 2 responses, the researchers observed differences between performance on tasks with familiar and unfamiliar topics. The more familiar students were with the topic addressed in an assignment, the more they were inclined to base their response on their existing knowledge and opinions about that particular topic. The less familiar students were with the topic at hand, the more they based their responses on information searched for on the Internet. Student responses based mainly on existing knowledge and opinions about a particular topic resulted in lower expert grades than responses based on information searched for on the Internet. These observations require further examination of the comparability and validation of the prose-task assignments. With regard to the results of this study, the random selection of assignments led to a balanced distribution of the assignments across Task 1 and Task 2, and, as such, compensated for the observed differences in responses between the assignments.

A limitation of this study is the lack of a control group. As such, caution is warranted when attributing performance improvement to the regulation feedback included in the DIM. Furthermore, students received a total of 1 h of instruction on the evaluation of information, one of the conditional skills for information problem solving (see Table 2), during the two week period between the completion of Task 1 and Task 2. This instruction was aimed at another assignment that the respondents were asked to complete. For this assignment, the students were asked to rank several websites based on the reliability of the information provided by the website. The instruction might have resulted in an improved evaluation of information during Task 2, and consequently, might have improved the quality of the Task 2 responses. However, if this is at all the case, the instruction could have affected student performance only partly. The evaluation of information is only one of the skills required during information problem solving (see Fig. 2). Other skills, such as defining the information problem, search strategies and searching and scanning information were not addressed during the two week period between Task 1 and Task 2. In addition, the assessment criteria used for grading by the researchers did not address the quality of information used by the students. Finally, if the instruction had any effect on the quality of the Task 2 responses, it is still striking that this effect was not evident for the male respondents.

Another limitation is related to the sample. Fifty students from two classes who shared the same teacher participated in this experiment. Due to the small sample size, the power of the statistics was rather low. Only straightforward between-group comparisons could be made, and more complex models for examining student performance could not be used. On the other hand, although the power of the statistics was low, significant results were found. In addition, due to the small sample size and the convenience sampling strategy, the generalisability of the results, by themselves, is small. Fortunately, most of our findings correspond to the results reported in other papers. The triangulation of our results with existing literature further strengthens the conclusions.

Finally, modern technology can be used to quantitatively examine processes which were very hard to examine without it. For example, by tracing student information-seeking behaviour, it becomes possible to examine the relation between student performance and, for example, information-seeking behaviour. When student performance is predicted by traceable behaviour, computer-adaptive learning environments can be developed that provide students with relevant prompts based on traced behaviour. Further research is needed to examine the extent to which traceable information-seeking behaviour predicts the quality of solutions of information problems, and subsequently, the extent to which student performance can be improved by student-specific feedback based on logged information-seeking behaviour.

Acknowledgements

This research was funded by Kennisnet, the knowledge centre for information and communication technologies and education in the Netherlands.

References

Baird, J. R., & Northfield, J. R. (1992). *Learning from the PEEL experience*. Melbourne, Australia: Monash University.
Bangert-Drowns, R. L., Kulik, C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213–238.
Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25.
Boud, D. (1989). The role of self-assessment in student grading. *Assessment and Evaluation in Higher Education, 14*, 20–30.
Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using Internet. *Computers & Education, 53*, 1207–1217.
Brew, A. (1995). What is the scope of self-assessment? In D. Boud (Ed.), *Enhancing learning through self-assessment* (pp. 48–63) London: Kogan Page.
Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated Learning: a theoretical synthesis. *Review of Educational Research, 65*(3), 245–281.
Connaway, L. S., & Dickey, T. J. (2010). *The digital information seeker: Report of findings from selected OCLC, RIN and JISC user behaviour projects*. Retrieved 26 March 2013 from http://www.jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf.
De Bruijn-Smolders, M., Timmers, C., Gawke, J., Schoonman, W., & Born, M. (2014). Effective self-regulatory processes. *Studies in Higher Education, 39*(3), 0.1080/03075079.2014.915302.
Dirkx, A., Drent, M., Jonkers, J., Knippenberg, H., Vermijs, H., & Will, N. (2011). *Survey on policy and practice of instruction in informationskills at Dutch universities* [*Enquête over beleid en praktijk van instructies in Informatievaardigheden in Nederlandse universiteiten*]. Retrieved 26 March 2013 from http://hdl.handle.net/2066/91367.
Eisenberg, M. B., & Berkowitz, R. E. (1990). *Information problem-solving: The big six skills approach to library and information skills instruction*. Norwood, NJ: Ablex.
Gross, M., & Latham, D. (2007). Attaining information literacy: an investigation of the relationship between skill level, self-estimates of skill, and library anxiety. *Library & Information Science Research, 29*, 332–353.
Gross, M., & Latham, D. (2012). What's skill got to do with it?: Information literacy skills and self-views of ability among first-year college students. *Journal of the American Society of Information Science and Technology, 63*, 574–583. http://dx.doi.org/10.1002/asi.21681.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. http://dx.doi.org/10.3102/003465430298487.
Head, A. J., & Eisenberg, M. B. (2010). Truth be told: How college students evaluate and use information in the digital age. In *Project information literacy progress report*. Progress report. Retrieved 26 March 2013 from http://projectinfolit.org/pdfs/PIL_Fall2010_Survey_FullReport1.pdf.
Ivanitskaya, L., O'Boyle, I., & Casey, A. M. (2006). Health information literacy and competencies of information age students: results from the interactive online research readiness self-assessment (RRSA). *Journal of Medical Internet Research, 8*(2), 1–14.
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 245–284.
Kolowich, S. (2011). *What students don't know*. Retrieved 26 March 2013 from http://www.insidehighered.com/news/2011/08/22/erial_study_of_student_research_habits_at_illinois_university_libraries_reveals_alarmingly_poor_information_literacy_and_skills.
Kramarski, B., & Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments? *Journal of Computer Assisted Learning, 22*, 24–33. http://dx.doi.org/10.1111/j.1365-2729.2006.00157.x.
Kuhltau, C. C. (1993). *Seeking meaning: A process approach to library and information services*. Norwood, NJ: Ablex.
Monroe-Gulick, A., & Petr, J. (2012). Incoming graduate students in the social sciences: how much do they really know about library research? *Portal: Libraries & the Academy, 12*, 315–335.
Mory, E. H. (2004). Feedback research revisited. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745–783). Mahwah: Erlbaum Associates.
Mosenthal, P. B. (1998). Defining prose task characteristics for use in computer-adaptive testing and instruction. *American Educational Research Journal, 35*, 269–307.
Murnane, R., & Levy, F. (2004). *The new division of labor: How computers are creating the next job market*. Princeton, N.J: Princeton University Press.
Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, G., et al. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education, 71*, 56–76. http://dx.doi.org/10.1016/j.compedu.2013.09.011.
Pinto, M. (2012). Information literacy perceptions and behaviour among history students. *ASLIB Proceedings, 64*, 304–327.
Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*, 459–470.
Silva, E. (2009). Measuring skills for 21st-century learning. *Phi Delta Kappan, 90*(9), 630–634.
Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: what we know and where we need to go. *Psychological Bulletin, 137*, 421–442.
Stripling, B., & Pitts, J. (1988). *Brainstorms and blueprints: Teaching library research as a thinking process*. Littleton, CO: Libraries Unlimited.
Thompson, C. (2011, November). *Why kids can't search*. Retrieved 26 March 2013 from http://www.wired.com/magazine/2011/11/st_thompson_searchresults/.
Thurlings, M., Vermeulen, M., Bastiaens, T., & Stijnen, S. (2013). Understanding feedback: a learning theory perspective. *Educational Research Review, 9*, 1–15. http://dx.doi.org/10.1016/j.edurev.2012.11.004.
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. http://dx.doi.org/10.1016/j.compedu.2012.07.007.
Van der Kleij, F. M., Feskens, R., & Eggen, T. J. H. M. (2013). Effects of feedback in computer-based learning environment on students' learning outcomes: a meta-analysis. In *Paper presented at the NCME, San Franscisco*.

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*(1), 21–39. http://dx.doi.org/10.3280/CAD2011-001004.

Voogt, J., & Roblin, N. P. (2010). *21st century skills. Discussienota*. Enschede: Universiteit Twente iov Kennisnet.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2008). Information-problem solving: a review of problems students encounter and instructional solutions. *Computers in Human Behavior, 24*(3), 623–648. http://dx.doi.org/0.1080/03075079.2014.915302.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education, 52*, 234–246.

Walraven, A., & Voogt, J. (2014). *Integrating critical use of the Internet in secondary education: Teachers as designers and implementers* [*Integreren van kritisch Internetgebruik in de onderbouw van het VWO: Docenten als ontwerpers en uitvoerders*]. Enschede: Twente University for Kennisnet.

Weaver, R. L., & Cotrell, H. W. (1986). Peer evaluation: a case study. *Innovative Higher Education, 11*, 25–39.