# Computer-Based Formative Assessment: Variables Influencing Feedback Behaviour

Caroline Timmers

# Computer-Based Formative Assessment:

# Variables Influencing Feedback Behaviour

Caroline Timmers

Graduation committee

| | |
|---|---|
| Chairman | Prof. Dr. K.I. van Oudenhoven-van der Zee |
| Promoter | Prof. Dr. C.A.W. Glas |
| Referee | Dr. W. Schoonman |
| Members | Prof. Dr. S. Brand-Gruwel<br>Prof. Dr. A.J.M. de Jong<br>Dr. D. Joosten-ten Brinke<br>Prof. Dr. S. Narciss<br>Dr. K. Schildkamp |

Timmers, Caroline Frieda
Computer-Based Formative Assessment: Variables Influencing Feedback Behaviour
PhD Thesis University of Twente, Enschede. – Met een samenvatting in het Nederlands.
ISBN: 978-90-365-0641-0
doi: 10.3990/109789036506410
Printed by Ipskamp Drukkers, Enschede
Cover designed by Ruben and Caroline Timmers


*This doctoral thesis was financially supported by Saxion University of Applied Sciences and facilitated by the Department of Research Methodology, Measurement and Data Analysis at Twente University*.

# COMPUTER-BASED FORMATIVE ASSESSMENT:

# VARIABLES INFLUENCING FEEDBACK BEHAVIOUR

DISSERTATION

To obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. H. Brinksma,
on account of the decision of the graduation committee,
to be publicly defended
on Friday, September 27, 2013 at 14.45 hrs.

by

Caroline Frieda Timmers
born on October 19th, 1976
in Noordwijk, The Netherlands

This dissertation has been approved by the promoter:

Prof. Dr. C.A.W. Glas

*“If we think of our children as plants... summative assessment of the plants is the process of simply measuring them. The measurements might be interesting to compare and analyse, but, in themselves, they do not affect the growth of the plants. On the other hand, formative assessment is the garden equivalent of feeding and watering the plants - directly affecting their growth.” Clarke (2001, p. 2).*

# Acknowledgements

The past seven years Saxion University of Applied Sciences has given me the opportunity to engage in a PhD project. I feel privileged and grateful to have been given this exceptional opportunity, which led me to encounter interesting researchers from all over the world. This thesis is a result of my PhD project. However, this thesis would not lie in front of you today without contributions of many others.

I would like to express my gratitude to my family, friends and colleagues at Saxion, OMD and RCEC for their interest and support, and to all respondents for providing the necessary data. Additionally, I thank Tim Logtenberg for encouraging me to take up this adventurous project despite the challenging conditions of recent motherhood. Wouter Schoonman, thank you for putting your trust in me, sharing life's joys and sorrows, and your location independent feedback delivery. Cees Glas, thank you for providing me with a supportive and instructive environment for my project. Lorette en Birgit, thank you for making me feel welcome and at home at OMD. Bernard, Jean-Paul, Stéphanie, Jannie, Fabienne, Theo, Paul, Wim, Laurens, Kommer, Nora, Rienk, Marjolein, Cindy and Lisanne, your cooperation, suggestions and feedback have been invaluable for my education and the completion of this thesis. I enjoyed working with you. Wil and Ruben, thank you for providing the images for my presentations.

Dear mum and dad, I especially want to thank you for looking after Daan, Quinten and Linus during my regular stays at Twente University and the occasional conferences. Knowing our precious boys would be safe and happy with your additional and loving care, made it possible for me to leave home and go to work with a peaceful mind. This also accounts for Marieke, Melissa, Linda, Andrea, Monique, Raquel and all the other dedicated employees of Babbeloes.

Pele, thank you for caring, sharing and all the love and effort invested in our fruitful business unit. Pele, Daan, Quinten and Linus, thank you for being wonderful and wising me up.

Although my PhD trajectory is almost at an end, it feels like my journey in the engaging world of research has only just begun. I'm looking forward to continue learning about computer-based formative assessment, feedback and learning analytics at Saxion University of Applied Sciences.

Enschede, August 19th

# CONTENTS


**CHAPTER 1 – INTRODUCTION ..... 1**

1.1 THE PROCESS OF RECEIVING FEEDBACK ..... 2
1.3 FEEDBACK BEHAVIOUR ..... 3
1.1 COMPUTER-BASED FORMATIVE ASSESSMENT ..... 4
1.4 OUTLINE ..... 5

**CHAPTER 2 - PATTERNS OF FEEDBACK SEEKING ..... 9**

2.1 INTRODUCTION ..... 9
2.2 FORMATIVE ASSESSMENT AND FEEDBACK ..... 10
2.3 METHODOLOGY ..... 13
*Research population and procedure ..... 13*
*Computer-based formative assessments on information literacy ..... 14*
*Data analysis ..... 15*
2.4 RESULTS ..... 16
*Patterns in attention paid to additional feedback and test length ..... 16*
*The relationship between task difficulty and attention paid to additional feedback ..... 19*
*Supervision and attention paid to additional feedback ..... 21*
2.5 CONCLUSION AND DISCUSSION ..... 21

**CHAPTER 3 - MULTILEVEL ANALYSES OF FEEDBACK BEHAVIOUR ..... 23**

3.1 INTRODUCTION ..... 23
3.2 FEEDBACK BEHAVIOUR IN COMPUTER-BASED FORMATIVE ASSESSMENTS ..... 24
3.3 AIMS OF THE PRESENT STUDY ..... 25
3.4 METHODOLOGY ..... 26
*Research population and procedure ..... 26*
*Computer-based formative assessments on information literacy ..... 27*
*Data analysis ..... 27*
3.5 RESULTS ..... 28
*Analysis of average feedback-use ..... 28*
*Multilevel modelling of feedback-use ..... 31*
*Multilevel modelling of item-feedback-time ..... 37*
3.6 CONCLUSION AND DISCUSSION ..... 40

**CHAPTER 4 - MOTIVATIONAL BELIEFS, STUDENT EFFORT, AND FEEDBACK BEHAVIOUR ..... 45**

4.1 INTRODUCTION ..... 45
4.2 THEORETICAL FRAMEWORK ..... 46
4.3 AIMS OF THE PRESENT STUDY ..... 48
4.4 METHODOLOGY ..... 49
*Research population and procedure ..... 49*
*Questionnaire on task-value beliefs, success expectancy and effort ..... 50*
*A computer-based formative assessment on information literacy ..... 50*
*Data analyses ..... 51*
4.5 RESULTS ..... 52
*Feedback behaviour ..... 52*
*Principal component analysis of task-specific motivation ..... 52*

*Descriptive statistics* ..... *54*
*Predictors of Effort* ..... *54*
*Predictors of feedback behaviour* ..... *54*
*Effort and Feedback Behaviour* ..... *55*
4.6 Conclusion and discussion ..... 55

**CHAPTER 5 - IMPLICATIONS OF PREVIOUS RESEARCH FOR THE DESIGN OF FEEDBACK INTERVENTIONS IN COMPUTER-BASED FORMATIVE ASSESSMENT ..... 59**

5.1 Introduction ..... 59
5.2 Effectiveness of formative assessment ..... 60
5.3 Methodology ..... 63
5.4 Results ..... 64
*An overview of variables influencing effectiveness of CBFA* ..... *64*
*Student characteristics affecting the effectiveness of CBFA* ..... *65*
*Instrumental characteristics affecting the effectiveness of CBFA* ..... *67*
5.5 Decision-making framework for the design of feedback interventions in CBFA ..... 70
5.6 Conclusion and discussion ..... 72

**CHAPTER 6 - DEVELOPING SCALES FOR INFORMATION-SEEKING BEHAVIOUR ..... 77**

6.1 Introduction ..... 77
6.2 Information literacy and information-seeking behaviour ..... 78
6.3 Previous research on information-seeking behaviour ..... 82
6.4 Development of a measurement instrument ..... 86
6.5 Evaluation of the measurement instrument ..... 89
6.6 Conclusion and discussion ..... 99

**REFERENCES ..... 103**

**APPENDIX A. AN EXEMPLARY ITEM WITH FEEDBACK ..... 111**

**APPENDIX B. KNOWLEDGE OF RESULTS WITH ADDITIONAL FEEDBACK PER ITEM ..... 113**

**APPENDIX C. INFORMATION-SEEKING BEHAVIOUR QUESTIONNAIRE ..... 115**

**SUMMARY ..... 119**

**SAMENVATTING ..... 123**

# Chapter 1 – Introduction

Today's society is complex and subject to rapid changes. Lifelong learning is emphasized as a means to keep up with change in so called learning societies. One of the items in the discourse on how to prepare students for lifelong learning is the role assessment should play in education (Boud, 2000). Assessment can be used to measure and acknowledge student performance, referred to as the summative function of assessment. Assessment can also be used to stimulate and direct student learning. This function is referred to as formative assessment. Formative assessment can refer to an instrument as well as a process. In this dissertation, formative assessment has been conceptualised as a purposefully designed instrument embedded within a learning process. Formative assessments contribute to learning by generating feedback. Here, feedback is conceptualised as information about learners actual state of performance intended to modify their thinking or behaviour for the purpose of improved performance (cf. Narciss, 2008; Shute, 2008). The formative assessment is forward looking and supplements the retrospective nature of summative assessment. Moreover, an emphasis on summative assessment creates dependence on external assessment, while lifelong learners benefit from assessments that foster self-regulatory skills, such as the capacity to judge one's own performance. Formative assessment fosters the capacity to judge one's own performance, especially, when feedback provides students with insight into their performance gaps and how to improve their performance. A shift in assessment thinking from testing students to fostering self-regulated learning is therefore advocated. The past two decades interest in the formative function of assessment has increased correspondingly (Bennett, 2011).

The relation between formative assessment and learning is complex as it is influenced by numerous variables, such as the nature of the feedback intervention and students motivational beliefs. Various researchers substantiate that effects of formative assessment foremost depend on whether students actively seek feedback and construct meaning from it (Ashford, Blatt, & VandeWalle, 2003; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Nicol & McFarlane-Dick, 2006). However, whether students seek and study feedback is mostly left out of scope in research examining the effects of feedback in formative assessment on learning outcomes (Van der Kleij, Timmers, & Eggen, 2011). To better understand the relation between formative assessment and learning, the extent to which students seek and study feedback and variables influencing this, so called, feedback behaviour needs further examination.

In computer-based environments feedback behaviour can be explored by capturing learner-produced data trails. These learner-produced data trails provide insight into students actions when asked to complete a computer-based formative assessment (CBFA). Such data can subsequently be used by educators to improve learning environments, for example to alter

the design of computer-based formative assessment. When traces of learners in educational settings are collected, analysed and reported for purposes of understanding and optimizing learning processes and the environments in which it occurs, this is referred to as learning analytics or educational data mining (Duval & Verbert, 2012; Siemens & Long, 2011). In the context of this dissertation, the method used to explore student feedback behaviour in a CBFA and variables influencing this behaviour is an example of learning analytics. Furthermore, previous research has been examined to determine the implications for the design of a CBFA. The dissertation also includes the development and evaluation of scales to measure information seeking behaviour. In the subsequent sections of this chapter, the theoretical framework for feedback behaviour in a CBFA is addressed. The chapter ends with an outline of the remaining chapters.

## 1.1 The process of receiving feedback

Valdez (2008) discusses the strengths and limitations of four leading conceptions of feedback proposed during the past decennium: Thorndike's law of effect, the three-cycle model of feedback (Kulhavy & Stock, 1989), the five-stage model for receiving feedback (Bangert-Drowns et al., 1991), and the feedback intervention theory (Kluger & DeNisi, 1996). Bangert-Drowns, Kulik, Kulik, and Morgan (1991) proposed a model for the process of receiving feedback. Their model conceptualizes a test-like event and feedback within a learning process. As such, the model can be used to visualise the conceptualisation of formative assessment as a purposefully designed instrument embedded within a learning process. The model of Bangert-Drowns et al. distinguishes itself from other conceptions of feedback by combining an active learner in a single act of a test-like event, while considering the influence of learners' motivational beliefs in use of feedback. Thorndike's law of effect and Kulhavy and Stock's three-cycle model of feedback do not recognize the learner as an active participant. Kluger and DeNisi's feedback intervention theory is multidimensional and does recognize the influence of learners' motivational beliefs.

A schematic overview of the five-stage model is presented in Figure 1.1. The model has been consummated to visualise the conceptualisation of formative assessment. The stages distinguished within this model represent the process of formative assessment. The test-like event together with the feedback intervention represent the formative assessment instrument. The model starts with appointing the learners' initial state (stage 1). The initial state is characterized by cognitive aspects (e.g. the degree of prior relevant knowledge) and motivational aspects (e.g. the degree of interest in a task). Characteristics of the initial state are assumed to influence the effort students invest in the subsequent stages of the process of receiving feedback. When a test-like event is administered, items activate the process of addressing relevant prior knowledge and skills (stage 2). Subsequently, a test taker constructs a response (stage 3). In this model, learners are provided with feedback after they have constructed a response (after stage 3). Next, learners evaluate their results or

responses supported by the feedback intervention (stage 4). This phase is crucial when assuming that the effect of formative assessment on learning foremost depends on whether, and to what extent, a student seeks and processes feedback. Feedback can be used, for example, to confirm, add to, correct, tune, or restructure knowledge and understanding about certain tasks and strategies. The adjustments made to characteristics of the initial state (stage 5) can be viewed as the learning outcomes of the formative assessment. When the purpose of a formative assessment is, for example, to support an increase in prior knowledge in a certain domain, the process is successfully completed when the degree of a students' prior knowledge present at the initial state increased. The five-stage model implies that variability in the relation between formative assessment and performance improvement is influenced by characteristics of the formative assessment instrument (the test-like event and feedback intervention) as well as characteristics of the formative assessment process.

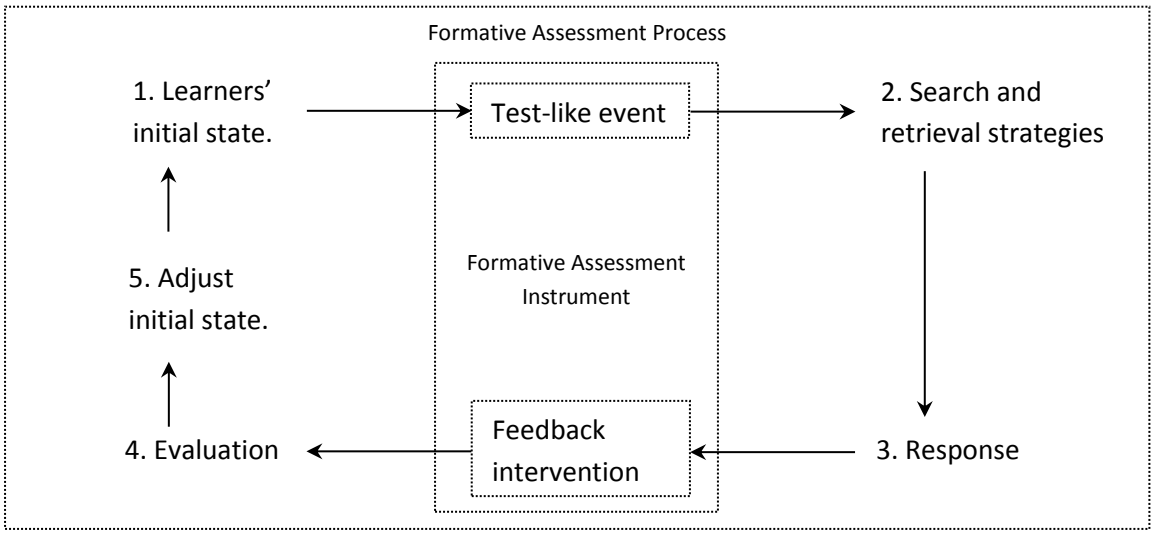


**Figure 1.1**. Conceptualisation of formative assessment based on the five-stage model proposed by Bangert-Drowns et al (1991); adapted from Mory (2004, p. 752)

## 1.3 Feedback behaviour

In the context of this dissertation, feedback behaviour refers to whether, and to what extent, students seek and study feedback. Feedback behaviour can be promoted or inhibited by various variables during the five-stages of receiving feedback described in the model of Bangert-Drowns et al (1991). For example, when the assessment task administered is perceived as too difficult, students might get demotivated and consequently quit investing effort in the task (Veldkamp, Matteucci, & Eggen, 2011). In that case, task difficulty impedes the process of receiving feedback at stage 2, inhibiting feedback behaviour that might lead to improved performance. In addition, when learners are provided with feedback that

merely informs them whether or not an answer is correct, the possibilities to construct meaning from feedback during the evaluation phase (stage 4) are limited (Hattie & Timperley, 2007) and, correspondingly, inhibit the effect of feedback on performance (Van der Kleij et al., 2011).

Previous research shows that adding feedback to computer-based environments does not guarantee that students will seek and process feedback (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003). Although formative assessments aim at providing feedback to support learners to improve performance, learners can react to feedback in ways that will not lead to enhanced performance or attainment of an intended level of performance. This is the case for example, when students respond to feedback by changing the intended level of performance, rejecting the feedback or abandoning commitment to the intended level of performance (Kluger & DeNisi, 1996). These behavioural options add to individual differences in the extent students seek and process feedback (Aleven et al., 2003).

Research on the relation between feedback and performance improvement shows heterogeneous results (Bangert-Drowns et al., 1991; Black & Wiliam, 1998; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Van der Kleij et al., 2011). Differences in feedback behaviour might add to an explanation of the variability in effectiveness of feedback. Remarkably enough, differences in feedback behaviour have mostly been left out of scope in research examining the relation between feedback and performance improvement. Researchers seem to assume that feedback is received and used to improve performance (e.g. Corbalan, Kester, & Van Merriënboer, 2009; Gordijn & Nijhof, 2002; Pridemore & Klein, 1995; Smits, Boon, Sluijsmans, & Van Gog, 2008; Van der Kleij et al., 2011; Wang, 2007). To better understand the relation between feedback and performance improvement, feedback behaviour needs to be explored, including variables influencing this behaviour.

## 1.1 Computer-based formative assessment

Using computers in formative assessments has several advantages (Buchanan, 1999; Wang, 2007). Computer-based formative assessment instruments can deliver individualized feedback in a timely manner, independent of time and place. Automated provision of feedback is a welcome solution to practical constraints such as time or workload pressure, especially when large numbers of students or lengthy pieces of work are involved. With a computer it is easy to generate immediate, objective and appropriate feedback using answering models constructed in advance. Computer-based formative assessment (CBFA) also cancels out the effect of credibility of the lecturer providing the feedback (Poulos & Mahony, 2008). Furthermore, students tend to perceive the computer as an attractive instrument for learning (Owston, 1997). Research shows students to be more likely to seek computer-mediated feedback than person-mediated feedback (Karabenick & Knapp, 1988; Kluger & Adler, 1993), probably because feedback seeking in computer environments often

remains unnoticed by others. As such, the costs of exposing one's uncertainty and need for help, so called *self-presentation cost*, does not come into play (Aleven et al., 2003).

Research has shown that automated feedback can improve performance, but not necessarily does (Van der Kleij et al., 2011). The effect of automated feedback on performance is influenced by type of feedback. Type of feedback refers to content-related classification of feedback components. A distinction is generally made between the following types of feedback: 1) knowledge of results (KR), which merely informs whether or not an answer is correct, 2) knowledge of correct response (KCR), where the correct response is provided, and 3) elaborated feedback (EF), where remedial information is provided (Shute, 2008). Different types of feedback have shown to be differentially effective on performance improvement. Overall, the effect of EF on performance improvement is higher than KCR. And the impact of KCR is higher than KR (Van der Kleij, Feskens, & Eggen, 2013).

## 1.4 Outline

In this introductory chapter, the conceptualizations of formative assessment and feedback behaviour have been addressed. The next three chapters address studies exploring whether, and to what extent, students paid attention to the additional feedback (KCR and EF) in a CBFA on information literacy and variables influencing this behaviour.

At Saxion University of Applied Sciences a CBFA on information literacy was developed and embedded in research skills training of various educational programs. Recent developments in Dutch higher education have led to an increased emphasis on research skills training. Research skills training at Saxion University of Applied Sciences generally include activities aimed at improving student information literacy. Information literacy or information problem solving refers to the ability to identify information needs, locate corresponding information sources, extract and organize relevant information from each source, and synthesize information from a variety of sources (Walraven, Brand-Gruwel, & Boshuizen, 2008). Previous research shows student deficiencies in information literacy and student tendency to overestimate their own information literacy (Ivanitskaya, O'Boyle, & Casey, 2006; Kuhlemeier & Hemker, 2005; Maughan, 2001). These findings corresponded with the practical experiences at various academies of Saxion University of Applied Sciences. The purpose of the CBFA was twofold, namely 1) to increase knowledge and understanding of what information literacy entails, and 2) to raise the degree of interest in education on this topic. Besides knowledge of results, the CBFA on information literacy included additional feedback (KCR and EF) to raise student knowledge on, for example, strategies to locate information sources. The next three chapters address the following research questions: To what extent do students pay attention to additional feedback in CBFA? What variables influence student attention paid to additional feedback?

Chapter 2 addresses the exploration of patterns in feedback seeking in a CBFA using descriptive statistics. In addition, feedback seeking has been related to student response (in-/correct), test length (10 or 20 items), achievement, and supervision (direct, indirect, and none).

The aim of the study presented in Chapter 3 was to explore individual and group differences in feedback seeking as well as feedback study time, through generalized and linear mixed models. Furthermore, the relations between feedback behaviour and the following person and item characteristics have been examined: student response (in-/correct), item difficulty, and achievement.

By examining students' motivational beliefs, researchers have learned much about the reasons why individuals choose to invest effort in learning activities or not (Eccles & Wigfield, 2002). Eccles and Wigfield (2002), for example, found that time and effort invested in a learning task is explained by success expectancy and task-value beliefs. The study presented in Chapter 4 addresses feedback seeking and feedback study time in relation to task-value beliefs, success expectancy, and student effort invested in completing a CBFA on information literacy.

Table 1.1 presents an overview of variables examined in this dissertation. The variables have been framed within the five-stage model of the process of receiving feedback (Bangert-Drowns et al., 1991). In brackets, reference has been made to the relevant Chapters of this dissertation (2, 3 and 4).

**Table 1.1**: An overview of variables examined in this dissertation.

| Five-stage model of: | The process of receiving feedback in a test-like event | Student characteristics** | Instrumental characteristics** |
|---|---|---|---|
| Stage 1 | Learners' initial state | *Success expectancy appraisal (C4)*<br>*Task-value beliefs (C4)* | |
| *Stimulus* | *Test-like event* | | *Length (C2)*<br>*Supervision (C2)*<br>*Item difficulty (C3)* |
| Stage 2 | Address relevant prior knowledge | | |
| Stage 3 | Construct a response | | |
| *Stimulus* | *Feedback Intervention* | | |
| Stage 4 | Evaluation of results by seeking and studying feedback:<br>*Feedback behaviour** | *Correctness of response (C2 and C3)*<br>*Achievement (C2 and C3)*<br>*Attributed success expectancy (C4)*<br>*Effort (C4)* | |
| Stage 5 | Adjust initial state | | |

* Dependent variable

** Independent variables

The relation between CBFA and performance improvement is influenced by numerous variables, such as task-value beliefs and type of feedback. These variables need to be taken into account by designers of CBFA for reasons of effectiveness. Chapter 5 draws an overview of variables substantially influencing the relation between CBFA and learning. Findings of previous research have been integrated into recommendations for the design of CBFA. In addition, the findings have been used to propose a theory based decision-making framework for the design of feedback interventions in CBFA.

The CBFA on information literacy embedded in education on research skills aimed at raising knowledge and understanding of what information literacy entails and, subsequently, supporting improvement of student information seeking behaviour. In absence of a reliable instrument to measure information seeking behaviour, scales on information seeking behaviour were developed and evaluated. The development and evaluation of the scales have been described in Chapter 6. These scales can be used to measure effects of interventions, such as research skills training, on information-seeking behaviour.

# Chapter 2 - Patterns of feedback seeking

**Abstract**

Three studies are presented on attention paid to feedback provided by a computer-based formative assessment on information literacy. Results show that the attention paid to feedback varies greatly. In general the attention focuses on feedback of incorrectly answered questions. In each study approximately fifty percent of the respondents paid attention to feedback of incorrect answers only. Approximately another twenty-five percent did not pay attention to feedback at all. Results suggest that differences in attention paid to feedback are influenced by task difficulty and test length. Supervision, however, does not seem to influence the average attention paid to feedback. On the other hand, results show that indirect and direct supervision lead to a greater impact of feedback in a computer-based formative assessment, as the number increases of students taking the test and, as a consequence, paying attention to feedback.

## 2.1 Introduction

The importance of information literacy within information societies is widely acknowledged. Information literacy refers to the ability to determine the information need, to access and critically evaluate information, and to use it effectively to solve problems. Results of previous research show insufficient information literacy and a tendency of students to overestimate their own information literacy (Ivanitskaya et al., 2006; Kuhlemeier & Hemker, 2005; Maughan, 2001). Computer-based formative assessment on information literacy can help students gain a more realistic view of their information literacy. It might also result in further development of information literacy. A key element of formative assessment is feedback. Feedback provides students with information that can be used to narrow the gap between actual performance and desired levels of performance. However, feedback can only be effective when the learner is willing and able to use it. This study researches the extent to which students pay attention to feedback provided by a computer-based formative assessment on information literacy and factors influencing attention paid to feedback.

---

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

## 2.2 Formative assessment and feedback

Formative assessment is a broad concept for which no univocal definition exists (Yorke, 2003). However, it is agreed that the central purpose of formative assessment is contributing to student learning. As opposed to summative assessments, which aim at determining the extent to which a student has achieved curricular objectives. In general, formative assessment is about gathering information about student performance and giving feedback in order to contribute to student learning. Other terms used to refer to formative assessment are assessments for learning and low-stakes tests.

Sadler (1989) states that formative assessment should facilitate 1) learners to gain a notion of the desired levels of performance and understanding, 2) teachers and learners to compare the actual level of performance with the levels desired, and 3) teachers and learners to tailor learning activities in order to narrow a performance gap.

Research shows that the effectiveness of formative assessment and feedback is influenced by various variables. Kluger and DeNisi (1996) suggest three classes of variables influencing effects of feedback interventions on performance: The nature of the task performed, situational variables, and properties of the feedback intervention. These classes of variables are discussed separately. Subsequently, a schematic overview is presented which includes the various variables that influence the process of formative assessment.

### *The nature of the task*

In this study, tasks are performed in the framework of computer-based formative assessment (CBFA). CBFA has several advantages in relation to other forms of formative assessment (Buchanan, 1999; Wang, 2007). CBFA is a general-purpose assessment system and can be used whenever and wherever. When feedback has been constructed in advance using answering models, it is very easy to generate immediate, objective and appropriate feedback with CBFA. A CBFA also gets round the effect of the credibility of the lecturer giving the feedback (Poulos & Mahony, 2008). Furthermore, students tend to find the computer an attractive instrument for learning (Owston, 1997).

Variables influencing the effectiveness of CBFA are task difficulty (or appropriate challenge) and test length or duration. To support learning, formative assessment activities should not be too easy or too difficult (Sadler, 1989; Wang, 2007). When a test is considered too easy, test takers might perceive only a small knowledge gap that is not worth any additional effort. On the other hand, when a test is considered too difficult, students can get frustrated and lose their motivation. Schoonman (1989) refers to the problem of diminishing motivation when students are asked to constantly perform at the top of their ability. Students might get frustrated due to insufficient positive reinforcement when easy items are not included in the assessment.

Length and duration of a formative assessment should be considered, since test takers have limited willingness to devote energy to test items of formative assessments or low-stakes tests (Wolf, Smith, & Birnbaum, 1995). Accordingly, Wise (2006) reports that rapid-guessing increases at the end of tests. Hence, Wise conjectures that more frequent shorter tests might be more effective than less frequent and longer ones.

### *Situational variables*

Situational variables are of a momentary nature (Crombach, 2002). They relate to a specific learning situation. Examples are students' perception of task utility and success expectancy. Another situational variable influencing the effectiveness of formative assessment is supervision of an assessment. Wellman and Marcinkiewicz (in Wellman, 2005) studied the impact of supervision of an assessment on learning. The assessment was part of a web-based and paper-based learning module on medical terminology. Results indicated that supervised assessment was more effective than no supervision in promoting learning. The supervision can be direct and indirect. The presence of a supervisor during an assessment is a form of direct supervision. An indirect form of supervision is asking students to hand in the results of an assessment after completing it on a self-selected point in time.

### *Properties of feedback interventions*

Three important properties of feedback interventions are type of feedback, level of feedback and timing of feedback. A wide variety of feedback types exist, which vary in specificity and purpose. Shute (2008) reviewed literature on formative feedback and listed feedback types encountered in previous research. The major variables of interest in previous studies are knowledge of results, knowledge of correct response, and elaborate feedback. Shute describes knowledge of response as a relatively unspecific type of feedback, since the test taker is merely told whether the answer is correct of incorrect. Knowledge of correct response implies that the test taker is told the correct answer, and is considered more specific. Most specific types of feedback are referred to as elaborate feedback. Elaborated feedback (EF) is a general term used for a wide variety of feedback, such as explanations of correct and incorrect responses, (links to) further reading materials, cues and suggestions, or a combination of the previous.

Shute (2008) studied the relation between feedback and learning. The relationship between feedback and the learning process is not necessarily positive. Previous research results show inconsistency and contradictions (Azevedo & Bernard, 1995; Kluger & DeNisi, 1996). According to Shute, correct delivery of feedback is the key to improved learning. She therefore developed guidelines for effective feedback from previous research, taking into account task complexity and student performance.

EF is crucial in deeper conceptual understanding (Bangert-Drowns et al., 1991). Hattie and Timperley (2007) also advocate the use of EF. They argue that feedback should answer the following questions: 1) where am I going?, 2) how am I going?, and 3) where to go next? The effectiveness of answers to these questions partly depend on the level at which the feedback operates. These levels relate to: feedback on the task, feedback on the process, feedback on self-regulation, and feedback on the self. Hattie and Timperley argue that feedback on the self, such as "well done, you are a great student", is least effective. This kind of feedback does not answer any of the three advocated feedback questions. Feedback aimed to move students from task to process to self-regulation is considered most effective.

With regard to timing a distinction is made between immediate and delayed feedback. The definitions of immediate and delayed feedback vary. Immediate is used for feedback after the student answers an item from a test. Both immediate and delayed can also be used to refer to feedback provided directly after finishing a test. The term delayed can also be used to refer to feedback which is provided a day or more after finishing a test. Shute (2008) reviewed literature on immediate and delayed feedback. She encountered conflicting results in the literature relating to the timing of feedback and the effects on learning outcome and efficiency. She notes that supporters of immediate feedback point out that immediacy of corrective feedback is more likely to result in efficient retention. Delayed feedback on the other hand is associated with facilitating transfer of what has been learned.

### *The process of formative assessment*

Figure 2.1 presents a schematic overview of the previously discussed variables influencing the process of formative assessment.

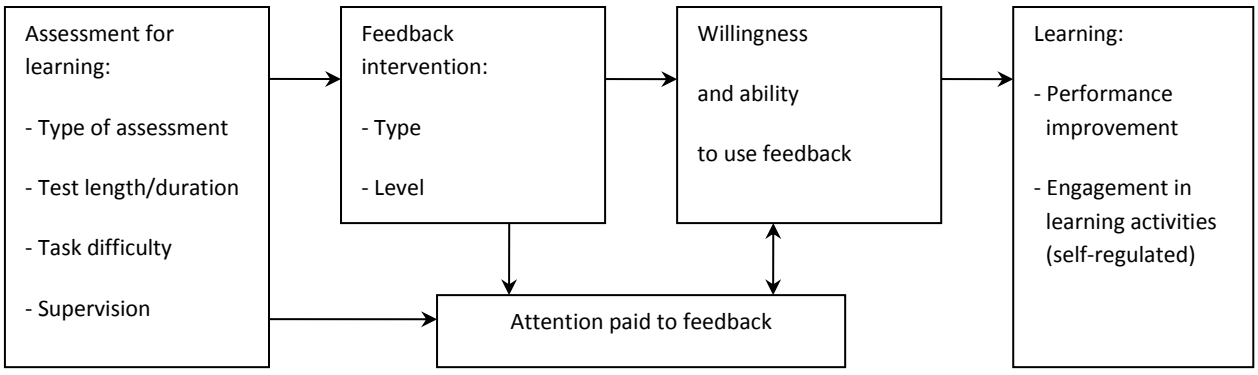


**Figure 2.1**: A schematic overview of the process of formative assessment and variables influencing attention paid to feedback.

Bangert-Drowns et al (1991) selected forty studies for their meta-analysis on instructional effects of feedback on test-like events, which measured post treatment performance on achievement tests. As such, willingness and ability to use feedback are treated as black-box variables. In this study it is assumed that feedback can only promote learning if a student is willing and able to use feedback. Or as Bangert-Drowns et al (1991) put it, feedback can only promote learning if it is received mindfully. No previous studies were found on actual attention paid to feedback and patterns of feedback behaviour. This study focuses on attention paid to elaborate and task related feedback provided by a CBFA on information literacy. The research questions are: To what extent do students pay attention to elaborated feedback provided by a CBFA on information literacy?; and To what extent is attention paid to elaborate feedback influenced by test length, task difficulty and supervision?

## 2.3 Methodology

### *Research population and procedure*

Three studies were conducted to examine attention paid to elaborate feedback provided by a CBFA on information literacy. In September and October 2009, first year bachelor students of Commerce (N = 200), Law (N = 200) and Health (N = 165) were asked to assess their information literacy using a CBFA. Supervision varied per group. Bachelors of Commerce were not supervised, bachelors of Law had indirect supervision, and bachelors of Health had direct supervision.

Bachelors of Commerce were sent an e-mail with a link to a ten-item test and the request to assess their own information literacy in preparation for training on this topic. The case of Commerce is characterized by no supervision of the CBFA

Bachelors of Law were notified about the CBFA on information literacy during the second session from a series of seven sessions on problem analyses. They were supervised by asking them to hand in the knowledge of results page of the assessment during the following session. The case of Law is characterized by indirect supervision of the CBFA. A link to a ten-item-test was send by e-mail. The test used for the Law students was similar to the test used for Commerce, except for several adaptations made for the context of Law.

Bachelors of Health were asked to complete a CBFA on information literacy during an in class session. The ninety-minute session was part of a learning trajectory on research skills. Lecturers of Health preferred the use of an extended CBFA. Nine of the items used for Law and Commerce were adapted to the context of Health. Eleven other items were developed in cooperation with the lecturers. Thus, six groups of Bachelors of Health were asked to assess their information literacy with a twenty-item CBFA. The sessions started with a brief

instruction about the test. A written instruction was also handed out. Students were told the outcomes of the assignment would be discussed in the forthcoming lesson. Furthermore, students were requested to at least stay the first forty-five minutes and work on a related assignment after completing the CBFA. Most students finished the test within the first forty-five minutes.

### *Computer-based formative assessments on information literacy*

A ten-item and a twenty-item CBFA on information literacy were used to study the attention paid to elaborate feedback by students in higher education. The multiple choice items were selected from eighty items developed over a period of three years in cooperation with five information specialists of Saxion University of Applied Sciences and the University of Twente in The Netherlands. An example of an item, including feedback, is presented in Appendix A.

The items are related to information literacy competence standards developed by the American College and Research Libraries (ACRL, 2000). These ACRL-standards describe the expected behaviour of students when seeking information within the context of higher education (see Table 2.1). These standards are widely used and were adopted and translated into Dutch in 2005 by a national committee (*Landelijk Overleg Omgaan met Wetenschappelijke Informatie,* LOOWI).

**Tabel 2.1**: ACRL-Standards.

| The information literate student… |
|---|
| 1. Determines the extent of information needed |
| 2. Accesses the information needed effectively and efficiently |
| 3. Evaluates information and its sources critically |
| 4. Uses information effectively to accomplish a specific purpose |
| 5. Understands the economic, legal, and social issues surrounding the use of information, and accesses and uses information ethically and legally |

The items for the CBFA were selected on the basis of the perceived difficulty and the attractiveness of the questions for the test takers. Lecturers of Commerce and Health were consulted in this process. For this study the stem and answers of the questions were adjusted to the field of Commerce, Law and Health.

After the administration of the CBFA is finished, an overview of correctly and incorrectly answered items is presented to the test taker. This overview links to additional feedback per

item. The additional feedback includes knowledge of correct response and an explanation of the various concepts used in the stem and the answering categories. In several cases, a reference to online study material was included and directly made available via a link.

There are various software programs to support CBFA. The available software package did not support monitoring of attention paid to additional feedback by test takers. Therefore, a new system was developed for the purpose of studying attention paid to additional feedback.

### *Data analysis*

The number of additional feedback pages opened by the test taker was used as an indication of attention paid to additional feedback. A distinction was made between additional feedback pages opened for correctly and incorrectly answered questions. A distinction was also made between opening all, several or none of the additional pages for correctly and incorrectly answered items. The category 'several' refers to opening a number of additional feedback pages between one and nine. For example, students might open all additional feedback pages for incorrectly answered items and no additional feedback pages for correctly answered items. Another possible pattern is students opening all additional feedback pages for both correctly and incorrectly answered items. This leads to nine possible patterns for attention paid to additional feedback.

The average test score was used as an indication of task difficulty. Mory (2004) stresses the corrective function of feedback. An incorrect answer can be viewed as a valuable opportunity to clarify misunderstanding. Therefore the analyses focused on additional feedback pages opened for incorrectly answered items.

The influence of supervision on opening additional feedback pages was also examined. The ten- and twenty-item CBFAs had nine items in common. These items were used for analyses of variance in the mean number of additional feedback pages opened. An overview with information on the items is presented in Table 2.2.

**Table 2.2**: The nine shared items: An overview of the subjects, formats and ACRL Standards covered per question.

| Question | Subject | Format | ACRL Standard |
|---|---|---|---|
| 1 | Domain specific database | Conventional | 2 |
| 2 | The scope of Google | Conventional | 1 |
| 3 | Google Scholar | Conventional | 2 |
| 4 | Using quotations marks | Conventional | 2 |
| 5 | Access to databases | Conventional | 1 |
| 6 | Search terms | Conventional | 2 |
| 7 | RSS | Conventional | 1 |
| 8 | Truncation | Conventional | 2 |
| 9 | Plagiarism | Conventional/ Multiple mark* | 5 |

* Conventional for Law and Commerce, and Multiple mark for Health.

## 2.4 Results

### *Patterns in attention paid to additional feedback and test length*

Patterns in attention paid to additional feedback were examined for those students of Commerce (N = 59), Law (N = 169) and Health (N = 154) that completed the CBFA. Tables 2.3a, 2.3b and 2.3c present an overview of the feedback patterns observed for Commerce, Law and Health, respectively. The patterns have been described by the extent (none, several, all) to which additional feedback pages were opened for correctly and incorrectly answered items. As can be seen in Table 2.3a to 2.3c, nine patterns of feedback behaviour have been distinguished. Per pattern the average test score and the average feedback pages opened have been presented. The patterns observed for Commerce, Law and Health are remarkably similar. For all three studies, the most frequently observed patterns were 1) opening all feedback pages for incorrectly and none for correctly answered items, 2) opening no feedback pages, and 3) opening several feedback pages for incorrectly and none for correctly answered items. The three patterns have in common that no feedback pages were opened for correct answers. Furthermore, these three patterns cover the behaviour of 78%, 71%, and 81% of the Commerce, Law and Health students, respectively.

**Table 2.3a**: Patterns of attention paid to additional feedback of incorrect and correct answers for Commerce.

| Feedback pages opened for incorrect answers | Feedback pages opened for correct answers | % test takers | Average test score | Sd | Average n feedback page opened | Sd |
|---|---|---|---|---|---|---|
| None | None | 22 | 3.00 | 1.16 | 0.00 | 0.00 |
| | Several | - | - | - | - | - |
| | All | 1 | - | - | - | - |
| Several | None | 25 | 3.33 | 1.45 | 2.47 | 2.23 |
| | Several | 3 | 2.50 | 0.71 | 4.50 | 0.71 |
| | All | 3 | 2.00 | 0.00 | 5.50 | 0.71 |
| All | None | 31 | 5.33 | 1.19 | 4.67 | 1.19 |
| | Several | 3 | 6.50 | 3.54 | 8.00 | 1.41 |
| | All | 10 | 4.33 | 2.25 | 10.00 | 0.00 |

**Table 2.3b**: Patterns of attention paid to additional feedback of incorrect and correct answers for Law.

| Feedback pages opened for incorrect answers | Feedback pages opened for correct answers | % test takers | Average test score | sd | Average n feedback page opened | sd |
|---|---|---|---|---|---|---|
| None | None | 25 | 4.67 | 1.98 | 0.00 | 0.00 |
| | Several | 2 | 5.75 | 1.50 | 1.00 | 0.00 |
| | All | 1 | - | - | - | - |
| Several | None | 19 | 4.14 | 1.69 | 1.47 | 0.91 |
| | Several | 6 | 4.40 | 1.08 | 4.00 | 2.00 |
| | All | 4 | 2.50 | 1.05 | 5.50 | 1.87 |
| All | None | 27 | 6.37 | 1.27 | 3.63 | 1.27 |
| | Several | 7 | 5.45 | 1.44 | 6.18 | 1.66 |
| | All | 8 | 4.85 | 1.95 | 10.00 | 0.00 |

**Table 2.3c**: Patterns of attention paid to elaborate feedback of incorrect and correct answers for Health.

| Feedback pages opened for incorrect answers | Feedback pages opened for correct answers | % test takers | Average test score | sd | Average n feedback page opened | sd |
|---|---|---|---|---|---|---|
| None | None | 26 | 8.38 | 2.46 | 0.00 | 0.00 |
| | Several | - | - | - | - | - |
| | All | - | - | - | - | - |
| Several | None | 34 | 9.21 | 2.55 | 5.62 | 3.56 |
| | Several | 12 | 9.11 | 2.21 | 7.58 | 3.89 |
| | All | 1 | 10.00 | - | 19.00 | - |
| All | None | 21 | 11.47 | 1.98 | 8.53 | 1.98 |
| | Several | 5 | 9.88 | 1.89 | 12.25 | 2.44 |
| | All | 1 | 10.00 | 2.83 | 20.00 | 0.00 |

Table 2.4a and 2.4b present overviews of the percentage of test takers opening feedback pages of questions answered incorrectly or correctly. For Commerce and Law the average percentage of opening feedback pages for incorrectly answered items are three to four times as high as the percentage for correctly answered items. For Health the difference is eight times as high. These results underline that test takers are mostly interested in additional feedback of incorrectly

**Table 2.4a**: Percentages for seeking additional feedback of items answered correctly or incorrectly for Commerce and Law.

| | Commerce | | | Law | |
|---|---|---|---|---|---|
| Item | Percentage feedback pages opened for correct answers | Percentage feedback pages opened for incorrect answers | item | Percentage feedback pages opened for correct answers | Percentage feedback pages opened for incorrect answers |
| 1 | 21.7 | 72.2 | 1 | 24.6 | 67.3 |
| 2 | 16.7 | 53.7 | 2 | 8.7 | 57.7 |
| 3 | 30.8 | 50.0 | 3 | 18.7 | 37.1 |
| 4 | 17.9 | 40.0 | 4 | 11.8 | 40.8 |
| 5 | 10.0 | 46.2 | 5 | 7.0 | 42.2 |
| 6 | 17.6 | 44.0 | 6 | 14.5 | 39.5 |
| 7 | 15.4 | 39.4 | 7 | 10.0 | 39.3 |
| 8 | 16.0 | 44.1 | 8 | 9.9 | 34.1 |
| 9 | 18.2 | 51.4 | 9 | 13.4 | 44.1 |
| 10 | 15.4 | 45.7 | 10 | 9.2 | 50.0 |
| Average | 18.0 | 48.7 | Average | 12.8 | 45.2 |

**Table 2.4b**: Percentages for seeking additional feedback of items answered correctly and incorrectly for Health.

| | Health | | | | |
|---|---|---|---|---|---|
| Item | Percentage feedback pages opened for correct answers | Percentage feedback pages opened for incorrect answers | item | Percentage feedback pages opened for correct answers | Percentage feedback pages opened for incorrect answers |
| 1 | 4.5 | 58.5 | 11 | 7.1 | 43.1 |
| 2 | 7.7 | 52.9 | 12 | 4.8 | 53.2 |
| 3 | 18.2 | 64.5 | 13 | 2.6 | 42.3 |
| 4 | 4.9 | 54.2 | 14 | 1.7 | 50.6 |
| 5 | 9.4 | 63.2 | 15 | 7.4 | 43.3 |
| 6 | 3.6 | 30.7 | 16 | 5.5 | 47.7 |
| 7 | 9.5 | 50.0 | 17 | 1.4 | 37.5 |
| 8 | 2.9 | 46.4 | 18 | 7.7 | 49.1 |
| 9 | 4.9 | 39.2 | 19 | 5.7 | 54.3 |
| 10 | 3.8 | 39.4 | 20 | 10.1 | 48.7 |
| | | | Average | 6.2 | 48.4 |

answered items. The results also suggest that the focus on additional feedback for correct answers decreases when test length increases. The average percentages of feedback pages opened for correct answers on the ten-item tests are 18.0% for Commerce and 12.8% for Law students. For the twenty-item completed by the Health students the average percentage of feedback pages opened for correct answers was only 6.2%.

Differences in patterns between the ten-item test (Commerce and Law) and the twenty-item test (Health) could well be caused by the difference in length. The influence of test length on observed differences in feedback patterns were further analysed as follows. Four groups were distinguished. First of all one finds that 22%, 25%, and 26% of Commerce (ten items), Law (ten items) and Health (twenty items) students, respectively, did not open any additional feedback pages. Secondly, differences are found when comparing percentages for opening several additional feedback pages for incorrectly answered items, 25%, 19%, and 34%, respectively. Thirdly, differences are also found for percentages of opening all additional feedback pages of incorrectly answered items, namely: 31%, 27%, and 21%, respectively. The other six patterns were combined as rest group and covered 22%, 29%, and 19%, of the feedback behaviour, respectively. A significant difference was found between length and difference in feedback pattern ($\chi^2 = 7.92$, $DF = 3$, $p = 0.048$). For a twenty-item test the number of students opening several additional feedback pages for incorrect answers and none for correct answers increases in comparison to ta ten-item test. On the other hand, the frequency of students opening all feedback pages for incorrect answers and none for correct answers decreases.

### *The relationship between task difficulty and attention paid to additional feedback*

Overviews of test score and average number of feedback pages opened for incorrectly answered questions are presented in Tables 2.5a and 2.5b. When the test score increases, the percentage of feedback pages opened for incorrectly answered items also increases. A possible explanation is that student's task-specific motivation might be a moderating variable. In the case of Commerce no supervision led to a smaller response. However, the students that did respond might well be the more task-motivated ones. As can be observed in table 2.5a the average number of feedback pages opened for both incorrect and correct answers are slightly higher for Commerce then for Law.

**Table 2.5a**: Relation between incorrect answers and the average number of feedback pages opened for incorrectly answered items for Commerce and Law.

| | Commerce | | Law | |
|---|---|---|---|---|
| Number of incorrect answers | N | Average number of feedback pages opened for incorrect answers | N | Average number of feedback pages opened for incorrect answers |
| 10 | 1 | 0.0 | 0 | - |
| 9 | 5 | 2.0 | 6 | 1.5 |
| 8 | 6 | 4.5 | 10 | 2.2 |
| 7 | 13 | 1.9 | 21 | 2.4 |
| 6 | 14 | 3.4 | 29 | 2.0 |
| 5 | 6 | 3.5 | 38 | 2.5 |
| 4 | 9 | 3.7 | 28 | 2.9 |
| 3 | 4 | 3.0 | 22 | 2.1 |
| 2 | 0 | - | 10 | 1.6 |
| 1 | 1 | 1.0 | 3 | 1.0 |
| 0 | 0 | - | 2 | 0.0 |

**Table 2.5b**: Relation between test score and percentage of feedback opened for incorrectly answered items for Health.

| Number of incorrect answers | N | Average number of feedback pages opened for incorrect answers | Number of incorrect answers | N | Average number of feedback pages opened for incorrect answers |
|---|---|---|---|---|---|
| 20 | - | - | 10 | 27 | 5.7 |
| 19 | - | - | 9 | 13 | 5.5 |
| 18 | 1 | 0.0 | 8 | 18 | 5.4 |
| 17 | 1 | 0.0 | 7 | 12 | 6.3 |
| 16 | 3 | 5.0 | 6 | 4 | 4.5 |
| 15 | 3 | 1.3 | 5 | 2 | 4.5 |
| 14 | 8 | 6.0 | 4 | 1 | 4.0 |
| 13 | 15 | 5.5 | 3 | - | - |
| 12 | 19 | 5.8 | 2 | - | - |
| 11 | 24 | 3.9 | 1 | - | - |

The percentage of feedback pages opened for incorrectly answered items were used to analyse the relation with the test score, since the maximum number of feedback pages that can be opened depends on the number of incorrect answers. A significant positive relation was found between test score and the percentage of feedback pages opened for incorrect answers for Commerce ($r = 0.86$, $df = 57$, $p < 0.00$), Law ($r = 0.81$, $df = 175$, $p < 0.00$), and Health ($r = 0.83$, $df = 153$, $p < 0.00$).

### *Supervision and attention paid to additional feedback*

Table 2.6 presents an overview of mean scores for the nine items shared between the ten- and twenty-item CBFA. Results suggest that the response rate is influenced by both direct and indirect supervision. Indirect and direct supervision lead to a greater impact of additional feedback provided by a CBFA. A one-way analysis of variance was used to find out if the mean feedback pages opened differ. No difference was found between mean feedback pages opened between the three studies. These results suggest that supervision does not influence attention paid to additional feedback.

**Table 2.6**: Overview of mean scores on shared items (n = 9) for Commerce, Law and Health.

| | N | Response rate | Mean test score | sd | Mean feedback pages opened | sd |
|---|---|---|---|---|---|---|
| Commerce (no supervision) | 59 | 29.5% | 3.32 | 1.70 | 3.39 | 2.93 |
| Law (indirect supervision) | 169 | 84.5% | 4.40 | 1.81 | 2.66 | 2.65 |
| Health (direct supervision) | 154 | 93.3% | 3.70 | 1.48 | 3.19 | 2.59 |

## 2.5 Conclusion and discussion

In this study, additional feedback provided by a computer-based formative assessment (CBFA) received considerable student attention. Two or more pages with additional feedback were opened by 61%, 56%, and 70% of, respectively, Commerce, Law and Health students. This supports the assumption of Kluger and DeNisi (1996) that feedback interventions: " … command, and often receive, considerable attention" (p.262). These findings suggest that it is worth the effort to add additional feedback to a CBFA.

The results of this study show that the attention paid to feedback mainly focuses on feedback of incorrectly answered items. This suggests that the respondents prefer the corrective function of feedback compared to the confirmatory function (Mory, 2004).

Within this study incorrect answers are viewed as an opportunity to clarify misunderstandings. The results show that when the amount of additional feedback grows for incorrectly answered items, as is the case when students´ test scores decrease, the relative attention paid to the additional feedback decreases. These findings imply that task difficulty should be taken into consideration during the development of formative assessments. Further research is needed to find out in what way task difficulty can be used to optimise attention paid to additional feedback.

Students' interest in additional feedback provided by a CBFA on information literacy varies considerably on an individual level. Overall, the three studies showed similar spacing in the various patterns of feedback behaviour. About a quarter of the students did not open any feedback pages, while another quarter opened all pages with additional feedback for incorrect answers.

The spacing over patterns in attention paid to additional feedback might be explained by differences in students' task-specific motivation. Also, the average test score of students that do not open any feedback pages for incorrect answers is lower than the average test score of students that opened all feedback pages for incorrect answers (see Tables 2.4a to 2.4c). This might indicate that students who did not open any pages with additional feedback consider the test too difficult. As a result this might have led to loss of motivation to seek additional feedback. Another explanation might be that the students with a low score did not try to answer the items correctly, possibly due to a lack of motivation for the task. One way or another, this suggests that motivation in relation to task difficultly and/or student ability is a moderating variable. Although the explanation of motivation is a challenging task (Eccles & Wigfield, 2002), motivation is widely recognized as a determinant for engagement with formative assessment and feedback (Kluger & DeNisi, 1996). Research shows that feedback both regulates and is regulated by motivational beliefs (Nicol & McFarlane-Dick, 2006). The influence of students' task-specific motivation on attention paid to feedback is a topic for further research.

The results show that supervision stimulates a greater impact of a CBFA and the feedback provided for. Responses are much higher in a supervised setting compared to a setting with no supervision. On the other hand, the results suggest that supervision does not influence the mean feedback pages opened.

A difference was found for patterns in opening feedback pages between the ten-item and the twenty-item test. This corresponds with the limited energy of test takers to devote to formative assessments mentioned by Wolf et al (1995). A lower percentage of the respondents on the twenty-item test opened all pages with additional feedback for incorrectly answered items compared to respondents on the ten-item test. On the other hand, the percentage of students opening several feedback pages for incorrectly answered items was higher for the twenty-item test. Also, the results suggest that the focus on additional feedback for correct answers decreases when test length increases. Further research is needed to find out in what way test length can be used to optimise attention paid to additional feedback.

In this study several effects, such as test length and task difficulty, have been studied independently. For further research it might be interesting to study the correlations between the various variables influencing attention paid to additional feedback.

# Chapter 3 - Multilevel analyses of feedback behaviour


## Abstract

Formative assessment can be used to generate feedback on performance to support student learning but individual differences in feedback behaviour complicate a straightforward efficient implementation. The aim of this study is to explore individual and group differences in feedback behaviour in a computer-based formative assessment (CBFA). Feedback behaviour per item of a CBFA is represented by whether a student seeks feedback and the time a student spends studying the feedback. Feedback behaviour is analysed through generalized and linear mixed models. Furthermore, the relations between feedback behaviour and the following person and item characteristics have been examined: student response (correct/incorrect), item difficulty, and achievement. Results show that feedback seeking and feedback study times were higher for incorrect responses, and among high and middle achieving students. In addition, when item difficulty increases the propensity to seek feedback increases for incorrect responses only.


## 3.1 Introduction

The purpose of formative assessments is to support and direct learning by generating feedback on student performance. In this study, feedback is defined as information about the actual state of learning or performance of learners, which is provided to learners after responding to a test-like event. The aim of feedback is to support subsequent actions in the learning process, such as clarification of misconceptions and misunderstanding. A considerable body of research exists on the effectiveness of feedback on student learning (Bangert-Drowns et al., 1991; Kluger & DeNisi, 1996; Narciss, 2008; Shute, 2008; Van der Kleij et al., 2011). This body of research shows heterogeneous results. Variability in effectiveness of feedback can be related to characteristics of the feedback intervention, such as type of feedback. For example, knowledge of correct response and elaborated feedback are more likely to lead to an average improvement of learning outcomes in comparison to no feedback or knowledge of results only (Bangert-Drowns et al., 1991; Van der Kleij et al., 2011).

On an individual level, however, effectiveness of feedback in formative assessment foremost depends on whether feedback is being sought and used to adjust prior knowledge and skills or motivational beliefs (Bangert-Drowns et al., 1991). Previous research shows that adding feedback to computer-based environments does not guarantee that students will seek and process feedback (Aleven et al., 2003). As it turns out, feedback seeking and student attention paid to feedback varies widely per individual (Aleven et al., 2003; Timmers & Veldkamp, 2011). To learn more about the effectiveness of feedback in formative

assessment, the relations between student characteristics, characteristics of the formative assessment and feedback behaviour need to be examined.

An advantage of using a computer in formative assessment is the possibility to record feedback behaviour by logging feedback study times and whether or not students seek feedback by linking to feedback pages. A CBFA can also log student response per item. As such, it becomes possible to examine feedback behaviour and variables influencing this behaviour.

## 3.2 Feedback behaviour in computer-based formative assessments

Feedback behaviour concerns whether students seek and process feedback. Models proposed for feedback seeking mostly focus on employees in organizations (Ashford et al., 2003; Park, Schmidt, Scheu, & DeShon, 2007; VandeWalle, Ganesan, Challagalla, & Brown, 2000) and (interactive) computer-based learning environments (Aleven et al., 2003). Previous research shows that students are more likely to seek computer-mediated feedback than person-mediated feedback (Karabenick & Knapp, 1988; Kluger & Adler, 1993). A plausible explanation is that feedback seeking in computer environments often remains unnoticed by others. As a consequence, the cost of exposing one's uncertainty and need for help, so called self-presentation cost, does not come into play (Aleven et al., 2003). Self-presentation cost corresponds with the *image-based motive* underlying feedback behaviour. Other motives underlying feedback behaviour are the *instrumental motive* to achieve a goal or perform well, and the *ego-based motive* to defend or enhance one's ego (Ashford et al., 2003). The aim of a formative assessment is to support learning. As such, a formative assessment aims at addressing the instrumental motive for feedback behaviour.

Feedback study time can be viewed as an indication of student reception or processing of feedback. From an instrumental perspective on feedback behaviour, students are expected to spend more time studying feedback of incorrectly answered items compared to correct answers. From an ego-based perspective, students are expected to avoid studying feedback of incorrect answers. Instead, feedback study time, if at all observed, would focus on feedback of correct answers to enhance the ego. Assuming an instrumental motive for feedback behaviour, Kulhavy and Stock (1989) argue discrepancy between student response (correct or incorrect) and response certitude (low or high) to influence feedback study times. Overall, their research suggests feedback study times for incorrect responses to be longer in comparison to correct responses.

Within a CBFA the topic, complexity of, and variation in feedback are determined by the developer of the CBFA. Shute (2008) describes various types of feedback in terms of specificity, complexity, length and timing. She stresses that many learners will not pay attention to feedback when it lacks specificity or when it is too long and complex, as learners

will view the feedback as useless, frustrating, or both. The least specific form of feedback merely tells the student whether their answer is correct or incorrect and is referred to as knowledge of results (KR). The correct answer is not provided. Previous research shows that this type of feedback is not very effective in supporting learning (Bangert-Drowns et al., 1991; Van der Kleij et al., 2011). Types of feedback that are more likely to have an effect on student learning are knowledge of correct response (KCR) and elaborated feedback (Bangert-Drowns et al., 1991; Pridemore & Klein, 1991; Van der Kleij et al., 2011). KCR refers to feedback that represents the correct answer. Elaborated feedback refers to feedback that provides additional information besides outcome-related information (Narciss, 2008). As such, elaborated feedback comes in many shapes and sizes, which vary in length and complexity.

In computer-based environments, a distinction can be made between the availability of on-demand help or feedback before as well as after a student formulates a response to a task or an item (e.g. Kluger & Adler, 1993; Narciss, Körndle, Reimann, & Müller, 2004) and post response help or feedback only (e.g. Karabenick & Knapp, 1988; Mory, 1994). This study limits itself to a context in which additional feedback is made available after students have provided a response to all items of a CBFA. In a CBFA which provides the opportunity to seek additional feedback after KR is presented, users can decide whether to seek feedback for correctly or incorrectly answered items or both. Research on feedback seeking patterns in CBFA shows that, in general, students tend to focus on feedback of incorrect responses (Timmers & Veldkamp, 2011; Van der Kleij, Eggen, Timmers, & Veldkamp, 2012). However, this does not imply that an increase in difficulty of a CBFA automatically leads to an increase in feedback seeking and study time for incorrect responses. Research shows that a task, such as a CBFA, should not be too difficult (Sadler, 1989; Schunk, Pintrich, & Meece, 2008). In correspondence, Schoonman (1989) refers to the problem of diminishing motivation when students are asked to constantly perform at the top of their ability. When easy activities or items are excluded, this might lead to insufficient positive reinforcement and a need to defend or enhance one's ego by ignoring feedback for incorrect answers or quit engagement in the CBFA altogether. As such, difficult activities might lead to frustration and loss of motivation. Easy activities, on the other hand, might lead students to perceive a small performance gap that is not worth any additional effort, such as seeking and studying feedback.

## 3.3 Aims of the present study

The aim of the present study is to explore individual and group differences in feedback behaviour in a CBFA and to examine the relations between student response, achievement, item difficulty, and feedback behaviour. All variables included in this study are measured by the CBFA system. Feedback behaviour is represented by 1) the observation per item whether a student opens a pop-up page with additional feedback, referred to as feedback-use, and 2)

the time between opening and closing a pop-up page with additional feedback, referred to as item-feedback-time.

When the CBFA addresses the instrumental motive for feedback behaviour as intended, it is expected that feedback for incorrect answers is consulted more frequently and longer compared to correct answers (Ashford et al., 2003; Mory, 2004). As such, a negative relation is expected between answering an item correct and the probability of feedback-use (Hypothesis I). Furthermore, in correspondence with findings of Kulhavy & Stock (1989), item-feedback-times for incorrect responses are expected to be higher compared to item-feedback-times for correct responses (Hypothesis II).

If, in general, feedback is sought more frequently for incorrect answers, feedback-use can be expected to differ between difficult and easy items. As such, the probability of feedback-use is expected to positively relate to item difficulty (Hypothesis III). No relation is expected between item difficulty and item-feedback-time, as the feedback is designed similarly for all items and does not increase in complexity for, for example, the more difficult items (Hypothesis IV). In addition, assessments should be difficult enough to perceive a need for feedback, but, on the other hand, not too difficult to prevent frustration or loss of motivation. As such, a nonlinear (higher-order) relation is expected between student achievement and the probability of feedback-use (Hypothesis V). Finally, the relation between student achievement and item-feedback-time is explored, as it has not been studied extensively. One could reason that an increase in correctly answered items leads to a decrease in perceived need to study feedback. On the other hand, students with higher levels of achievement could be the ones that put more time and effort in studying feedback to further improve their levels of achievement.

## 3.4 Methodology

### *Research population and procedure*

In October 2009 and September 2010, first year bachelor students of a Dutch university of applied sciences studying Health (N = 151), Law (N = 218) and Business Administration (N = 241) completed a CBFA on information literacy. Information literacy refers to the ability to identify information needs, locate corresponding information sources, extract and organize relevant information from each source, and synthesize information from a variety of sources (Walraven et al., 2008). The CBFA was embedded in learning activities aimed at developing student research skills. The bachelors of Health and Law completed the assessment under supervision during class. The bachelors of Business Administration (BA) were sent an e-mail asking them to prepare for information literacy training by completing the assessment at home without supervision. Previous research by Timmers and Veldkamp (2011) showed that supervision did not influence student attention paid to feedback. The response rates,

however, are influenced by supervision. The CBFA was completed by all bachelors in the supervised conditions, Health and Law respectively. Only 54.5% of the BA students completed the CBFA. Another 9.3% started the assessment without completing it.

### *Computer-based formative assessments on information literacy*

A twenty-item and two fifteen-item CBFA's on information literacy were used to gather the observational data on feedback behaviour. The multiple choice items were selected from a series of items developed over a period of three years in cooperation with five information specialists of two Dutch universities. The items were selected in correspondence with lecturers involved. This resulted in CBFA's varying in items and length. Furthermore, the items were adjusted to the context of the different subject domains. An example of an item, including the additional feedback, is presented in Appendix A. When all items in the CBFA have been answered, a KR page is automatically generated and presented (see Appendix B). The KR page links to pop-up pages with additional feedback per item. The purpose of the additional feedback is to provide students with information for knowledge acquisition and clarifying misunderstanding. The students decided themselves if, how much and which additional feedback to consult. As such the provision of additional feedback was student-initiated instead of system-initiated. The additional feedback included KCR and an explanation of the various concepts used in the answering categories and the original item (elaborated feedback). In order to continue after opening a pop-up page with additional feedback, the pop-up page first needs to be closed. The instrument recorded whether or not a student linked to additional feedback. And when a student linked to a pop-up page with additional feedback, the time was recorded until the student closed the pop-up.

### *Data analysis*

For each item of the CBFA feedback behaviour is represented by two observations: 1) a binary observation for feedback-use, which is one when the feedback is consulted and zero when it is not, and 2) a continuous observation for item-feedback-time. Each feedback-use observation corresponds to a person and an item. First, feedback-use is analysed and related to student response, student achievement, and item difficulty. Second, item-feedback-time is analysed and related to student response, student achievement, and item difficulty. Consequently, feedback-use and item-feedback-time are the two dependent variables in this study. To compensate for extreme values of item-feedback-time and to induce a positivity restriction, the logarithm of the item-feedback-time is used.

Item difficulty is defined as the proportion of students answering an item incorrect. This measure is easier to interpret than the more obvious measure of proportion correct, which

translates to the easiness of an item. The proportion of correct answers on the CBFA per student is used as an estimate of the student's achievement.

The explanatory information, student response, item difficulty, and student achievement are defined at different hierarchical levels. The student response is a level-1 variable, which is the observational level. Item difficulty and student achievement are level-2 variables, where items and persons define a level. That is; at level 2, student responses are nested within items on the one hand, since responses to one item are more alike than responses to different items. On the other hand, student responses are nested within persons, since responses of one person are more alike than responses of different persons. To take into account this correlation structure and to handle observations at different hierarchical levels, use is made of multilevel analysis techniques (e.g. McCullogh, Searle, & Neuhaus, 2008; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). At level 1, it is hypothesized that feedback behaviour is a function of the student response, which is represented by either a correct or incorrect answer. At level 2, feedback behaviour is a function of item difficulty as well as student achievement. Furthermore, at level 3, a check is made for variation in feedback behaviour by clustering students per subject domain, since first-year Bachelor students across different subject domains completed the CBFA.

The lme4 package (Bates & Maechler, 2011) in R was used to test hypotheses related to feedback-use (I, III, and V). Item-feedback-times are also clustered in students and items. Therefore, a linear mixed model can be used to relate the non-zero item-feedback-times with the student response, item difficulty, and student's achievement, while accounting for the clustered structure of the data. The linear mixed models routine of PASW statistic 18 was used to test hypotheses related item-feedback-times (II and IV).

## 3.5 Results

### *Analysis of average feedback-use*

Two different item characteristics were defined: One represents the usual difficulty of an item and the other represents average feedback-use per item. The item difficulty is quantified as the observed proportion-incorrect. Average feedback-use per item is quantified as the observed proportion of additional feedback pages opened per item. In Figure 3.1, the observed proportions representing item difficulty and average feedback-use per item are plotted for all test items. It can be seen that the maximum use of feedback is around 50%, for an item with a difficulty level over 90%. This means that approximately half of the respondents answered the item incorrect, but did not use the feedback. Furthermore, the average feedback-use per item increases when the item difficulty increases. The minimum average feedback-use per item approximates 5% of the respondents for the item that was answered incorrectly by 10% of the respondents.

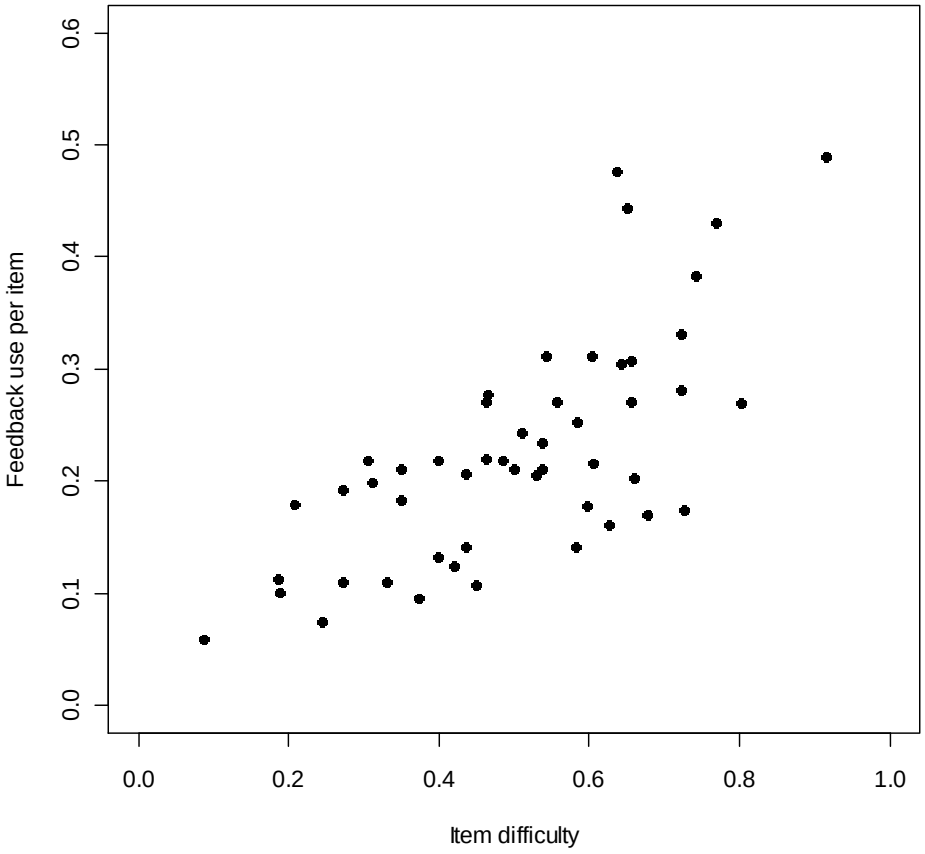


**Figure 3.1:** For each item, the observed proportion incorrect, as a measure of item difficulty, versus the observed proportion average feedback-use per item is plotted.

Besides investigating average feedback-use across items, feedback-use can also be explored across students. Student achievement is defined as the proportion correct over items and average feedback-use per student is computed as the proportion feedback opened over items.

In Figure 3.2, the estimated student achievements are plotted against average feedback-use per student. It follows that the tendency to seek feedback varies greatly between students.

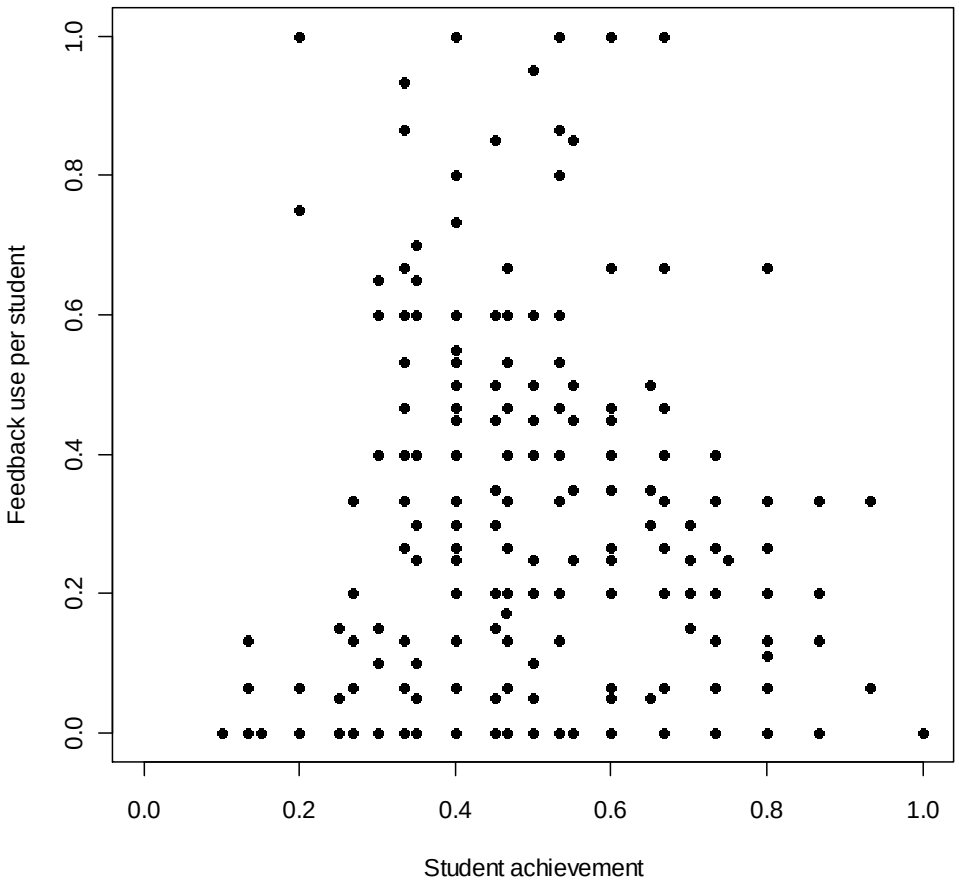


**Figure 3.2:** For each student, achievement versus proportion feedback-use per student is plotted.

Figures 3.1 and 3.2 show different relations that seem to contradict one another. In Figure 3.1, it can be seen that average feedback-use per item increases when items are more difficult. As a result, it is to be expected that students with low achievement levels have a higher tendency to use feedback. However, Figure 3.2 shows a wide variation in average feedback-use between students. On average low and high achieving students tend to use feedback pages less often than students with an average achievement. This will turn out to be an example of Simpson's Paradox (Simpson, 1951) or the ecological fallacy. The relations cannot be separately analysed since student achievement interferes with feedback-use and item difficulty. In Figure 3.1 and 3.2, the effect of student achievement and item difficulty is ignored, respectively. In a joint (multilevel) analysis both relationships can be correctly identified.

The differences in feedback-use over items and persons lead to specific relations that need to be analysed simultaneously in a cross-classified multilevel model. Then, differences in feedback-use can be explored over items and persons, while accounting for the fact that the observations are nested within items and persons. Furthermore, the within-person relationship between feedback-use and the response can be different from the between-person relation between feedback-use and an aggregate response per student. This phenomenon is known as the ecological fallacy. It covers the fact that erroneous findings can be obtained when only aggregate data are used for specifying lower-level characteristics (e.g. Snijders & Bosker, 1999).

### *Multilevel modelling of feedback-use*

The observed feedback-use is considered to be the outcome variable in a generalized multilevel model or generalized linear mixed model (GLMM). A generalized linear modelling approach is needed since the outcome variable is discrete. Conditional on the item and student, the probability of feedback-use is modelled. In generalized linear (mixed) modelling this success probability is transformed to a log odds, by computing the logarithm of the odds, which is the success probability divided by the failure probability. The log-odds is modelled as a linear combination of explanatory variables with fixed and random effects. The lme4 routine in R was used for parameter estimation. Table 3.1 shows descriptive statistics for the different variables.

**Table 3.1:** Descriptive statistics for feedback-use

| Variable name | Level | N | Mean | Sd | Minimum | Maximum |
|---|---|---|---|---|---|---|
| Feedback-use | 1 | 9905 | 0.21 | 0.41 | 0 | 1 |
| Item response | 1 | 9905 | 0.50 | 0.50 | 0 | 1 |
| Student achievement | 2 | 610 | 0.50 | 0.17 | 0.10 | 1 |
| Item difficulty | 2 | 50 | 0.50 | 0.16 | 0.09 | 0.91 |

Various models were used to study feedback-use (see Table 3.2). In Model 1, the log-odds of feedback-use is assumed to be linearly related to an intercept and a random student and random item effect. The intercept represents the baseline on the log-odds scale where the random effects are zero. The random item and student effect are assumed to be independent and represent the variation in feedback-use in log-odds across items and students, respectively. In Table 3.2, the parameter estimates are given under the label

Model 1. The estimated intercept corresponds to a student, with a random effect of zero, and an item, with a random effect of zero, and equals -2.42 on a log-odds scale, which corresponds to a success probability of .08; that is, 1/(1 + exp(2.42)). In the generalized linear model the intercept refers to the average use of feedback which is about 21%. However, this model ignores the clustering of the data and contains only fixed effects parameters.

It can be seen that students vary a lot in their feedback-use since the student random effect is relatively large and significant. The unconditional variation in feedback-use across items is relatively small and not significant.

**Table 3.2:** Effects of student and item information on feedback-use; Model information criteria and GLMM mean and standard error estimates of fixed and random effects.

| | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|
| Parameter | Est. | SE | Est. | SE | Est. | SE | Est. | SE |
| *Fixed effects* | | | | | | | | |
| Intercept | -2.42 | 0.15 | -1.43 | 0.17 | -1.52 | 0.27 | -0.92 | 0.20 |
| Student response | | | -4.93 | 0.14 | -4.93 | 0.14 | -4.24 | 0.23 |
| Item difficulty | | | 0.62 | 0.28 | 0.61 | 0.28 | 1.05 | 0.29 |
| Student achievement | | | 6.14 | 0.67 | 6.03 | 0.69 | 5.60 | 0.75 |
| Bsc Law | | | | | 0[a] | | | |
| Bsc Health | | | | | 1.31 | 0.41 | | |
| Bsc BA | | | | | -0.88 | 0.40 | | |
| Stud response * achievement | | | | | | | -7.43 | 1.13 |
| *Random Effects* | | | | | | | | |
| *Between students*: Intercept | 5.32 | 2.31 | 12.71 | 3.56 | 12.37 | 3.52 | 18.45 | 4.30 |
| Student response | | | | | | | 9.90 | 3.15 |
| *Between items*: Intercept | 0.54 | 0.73 | | | | | | |
| *Information Criterion* | | | | | | | | |
| Df | 3 | | 5 | | 7 | | 8 | |
| AIC | 7933 | | 5279 | | 5257 | | 4871 | |
| BIC | 7955 | | 5315 | | 5308 | | 4929 | |
| Log likelihood | -3964 | | -2635 | | -2622 | | -2427 | |
| Chisquare (df) | | | | | 26 (2) | | 414 (3) | |
| Chisquare (prob) | | | | | 2.124e-06 | | <2.2e-16 | |

[a] this parameter is set to zero because it is redundant

In Model 2, the random variation in feedback-use across items is restricted to zero, but item difficulty is included as an explanatory variable. Furthermore, the student achievement and the student response are added as explanatory variables. The explanatory variables item difficulty and student achievement are grand-mean centred such that the explanatory variable value of zero corresponds to the mean population level.

The results in Table 3.2 show that the effects of the explanatory variables are all significant. In Model 2, the log-odds of the probability of feedback-use of student *i* and item *k* is represented by

$$\log\left(\frac{p_{ik}}{1-p_{ik}}\right) = -1.43 - 4.93(\text{response})_{ik} + .62(\text{difficulty})_k + 6.14(\text{achievement})_i + \text{error}_{ik,}$$

where the error term represents the residual variation in the random effect of achievement. For the moment, assume an error term of zero to interpret the effects of the explanatory variables. Then, the estimated intercept of -1.43 represents the log-odd of feedback-use for an incorrect response, average item difficulty, and average student achievement, which corresponds with a probability of feedback-use of 19% $(\text{i.e.}, .19 = (1+\exp(1.43))^{-1})$. The log-odd of feedback-use increases when the item becomes more difficult. Consider an item of above average difficulty which is answered incorrectly by an additional 10% of the students. Then, the probability of feedback-use increases around .062 in log-odds, -1.43 + .062 = -1.368. A log-odd of -1.368 represents a probability of 20.3%, and shows an increase of around 1.3% in probability compared to an item of average difficulty given student achievement and response.

Student achievement has a much higher impact on feedback-use than item difficulty. Consider a student with a 10% above average level of achievement, the log odd increases to -.816=-1.43+.614, which represents a probability of 30.6% $(\text{i.e.}, .306 = (1+\exp(.816))^{-1})$ given an incorrect response to an item of average difficulty . The 10% increase in achievement leads to around 12% increase in the probability of feedback-use given item difficulty and response. The probability of feedback-use is around 0.3% when this student answers an item of average difficulty correct.

The feedback-use of the incorrectly answered items is much higher than of the correctly answered items. For an item of average difficulty and student with an average level of achievement, the feedback-use in log-odds is -1.43 when answered incorrectly and -6.4 when answered correctly. It follows that the probability of feedback-use is almost zero when the item is answered correctly.

Finally, the increase in random variation in feedback-use across students is much larger compared to the unconditional Model 1. In Model 2, it represents the variation in log-odds across students (with average achievement) for the incorrectly answered items (of average difficulty), where in Model 1 it is the variation in log-odds across all item responses. The variation in log-odds of feedback-use across students is around zero for the correctly answered items, which downsizes the random effects variance.

In Model 3, the additional grouping of students per subject domain was tested. It can be seen that Health students are significantly more likely to use feedback than Law and BA students. Note that the intercept represents Law students and that the intercept is smaller than that of Model 2. For an average student scoring an average item incorrect, the

probability of feedback-use is around 8.3%, 17.9%, and 44.9% for bachelors of Law, BA, and Health, respectively.

Finally, in Model 4, the effect of the student response was allowed to vary across persons without controlling for subject domain differences. This follows from the fact that the clustering of students in subject domains hardly influenced the effects of the other explanatory variables. From Table 3.2 it follows that the effect of student response does vary significantly across students. The average effect of student response is around -4.24 but this effect varies across students. The random intercept is negatively related to the random effect of student response and is around -.76. This means that the effect of student response on using feedback is significantly smaller for a student that uses feedback above average.

The average student response effect is slightly smaller and the average item difficulty effect is slightly higher than the corresponding estimated effects of Model 2. The students did not respond to exactly the same items in the test since an incomplete design was used. The level-1 variable student response can therefore explain variation in feedback-use between persons and between items. The intercept represents the log-odds of feedback-use for an average student (with random effects of zero) that answers an item of average difficulty incorrect. The item difficulty and the student achievement have a positive effect on feedback-use when the item is incorrectly answered. When the item is answered correctly the positive effect of achievement is cancelled due to the interaction effect of student response and achievement. It follows that the effect of student achievement is negatively related with feedback for the correctly answered items.

In Figure 3.3, the probability of feedback-use is plotted as a function of student achievement for various students and items. That is, for twenty randomly selected students the fitted probability of feedback-use for various items is given as a function of student achievement. A distinction is made between the correct and incorrect responses. It can be seen that an increase in achievement has a significant positive impact on the probability of feedback-use, when the item is answered incorrectly. An increase in achievement has a small decreasing effect on the feedback-use when the item is answered correctly. The opposite response-specific effects of achievement on feedback-use are captured by the main effect of student achievement and its interaction effect with student response.

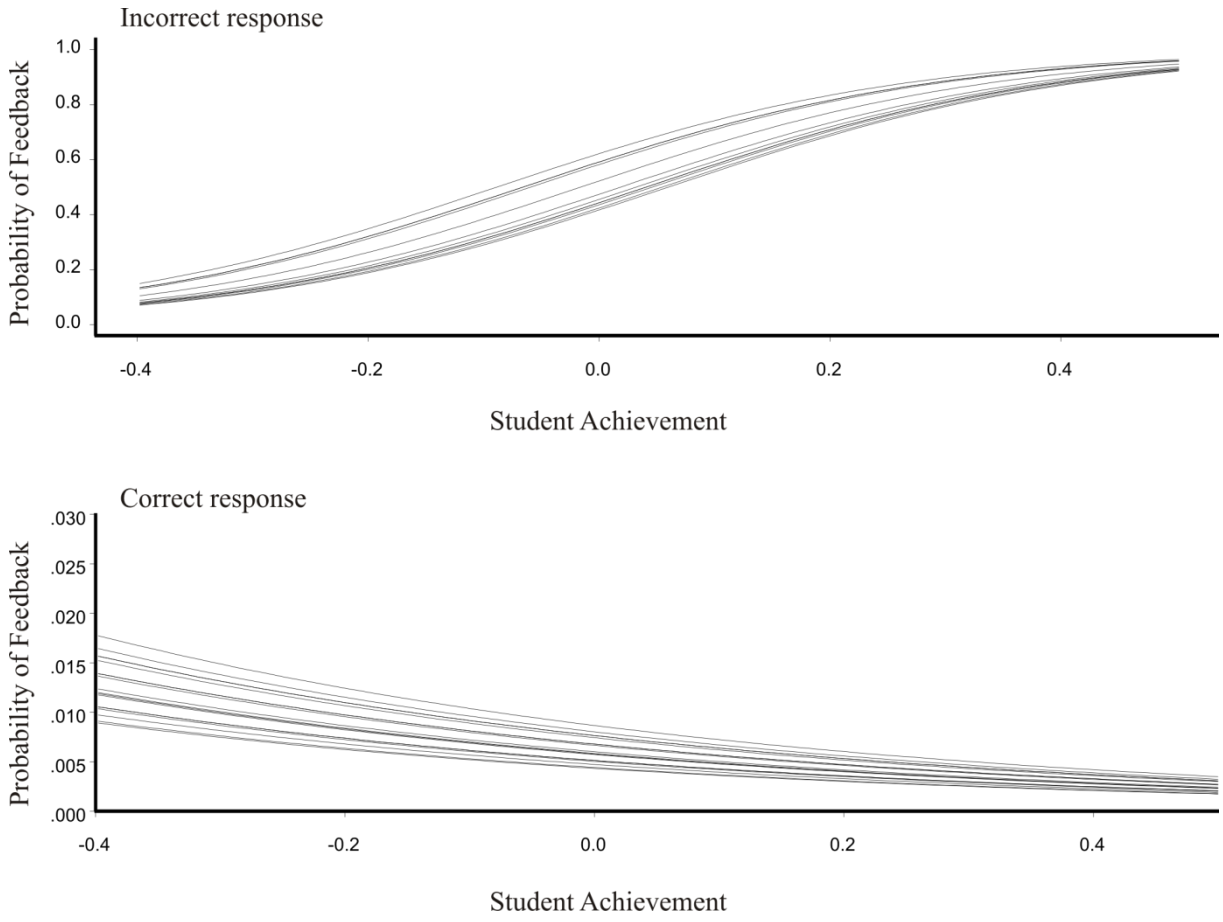


**Figure 3.3:** Probability of feedback-use for various students and items given the estimated fixed and random effects parameters.

A likelihood-ratio test was performed to compare nested models. The test statistic is computed as twice the difference in log-likelihood, which, in general, is asymptotically chi-square distributed with the difference in the number of parameters as the degrees of freedom. This test statistic does not have a chi-square distribution when testing a variance component to be zero, which concerns a null hypothesis on the border of the parameter space. Following Snijders and Bosker (1999), in that case the corresponding significance probabilities will be divided by two as a remedy for their conservative behaviour. The non-nested models are compared using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), The model with smaller information criteria values is to be preferred with respect to model accuracy and model complexity.

In Table 3.2, the results of the model comparison are given. Each chisquare value represents the likelihood-ratio test-statistic value related to the comparison of the current model and Model 2, where the difference in the number of parameters is given in brackets. It can be seen from the significance probabilities, labeled chisquare(prob), that the extensions in Model 3 and 4 are supported by the data. When comparing the non-nested models Model 1 and Model 2, the random effect of the factor item is not significant but the inclusion of the

explanatory variables (student response, achievement, and item difficulty) does lead to model improvement, given the AIC and BIC values. The grouping of students in Bachelor programs was tested, and the average feedback-use differs across programs. When comparing Model 4 with the other models, using the AIC and BIC, it follows that the random effect of the student response variable and the interaction effect of student response with student achievement are both important, and that Model 4 is the preferred model. Note that both significance probabilities are less than .001, even without dividing by two.

Table 3.3 shows average expected feedback-use, under Model 4, for correct and incorrect answers on items with a low, medium, and high level of difficulty of low, medium, and high achieving students. The probabilities show that student achievement and student response strongly influence feedback-use. Achievement has a positive effect on feedback-use for incorrect answers. In general, students hardly use feedback of correctly answered items. They prefer feedback of incorrect answers. The probability of using feedback of incorrect answers is highest for high achieving students. However, since they have answered less items incorrectly than students with low or medium achievement the actual number of feedback pages opened does not differ all that much between medium and high achieving students. Item difficulty influences feedback-use for incorrect responses only. Expected feedback-use increases when item difficulty increases.

In general, the results confirm a significant negative relation between answering an item correct and the probability of feedback-use. Therefore, hypothesis I is not rejected. Hypothesis III is not rejected, since a positive relation between item difficulty and the probability of feedback-use was found. Hypothesis V is not rejected, but can be refined. Results showed a positive non-linear relation between student achievement and the probability of feedback-use for incorrectly answered items. Furthermore, the results showed a negative non-linear relation between student achievement and the probability of feedback-use for correct answers.

**Table 3.3:** Probability of feedback-use for correct and incorrect response on items with a low, medium, and high level of difficulty for low, medium, and high student achievement.

| Student achievement | Item difficulty | Student response | Probability of feedback-use | Estimated nr. of feedback pages used of 20-item test |
|---|---|---|---|---|
| Low (25% correct) | Low (25% incorrect) | Correct | 1% | 0.15 |
| | | Incorrect | 7% | 1.05 |
| | Medium (50% incorrect) | Correct | 1% | 0.15 |
| | | Incorrect | 9% | 1.35 |
| | High (75% incorrect) | Correct | 1% | 0.15 |
| | | Incorrect | 11% | 1.65 |
| Medium (50% correct) | Low (25% incorrect) | Correct | 1% | 0.10 |
| | | Incorrect | 24% | 2.40 |
| | Medium (50% incorrect) | Correct | 1% | 0.10 |
| | | Incorrect | 28% | 2.80 |
| | High (75% incorrect) | Correct | 0% | - |
| | | Incorrect | 34% | 3.40 |
| High (75% correct) | Low (25% incorrect) | Correct | 0% | - |
| | | Incorrect | 55% | 2.75 |
| | Medium (50% correct) | Correct | 0% | - |
| | | Incorrect | 62% | 3.10 |
| | High (75% incorrect) | Correct | 0% | - |
| | | Incorrect | 68% | 3.40 |

### *Multilevel modelling of item-feedback-time*

The feedback time was recorded in seconds for each pop-up page with additional feedback opened per student per item. The logarithm of item-feedback-time conditional on positive feedback-use was computed such that the item-feedback-times of zero were ignored. In total 2123 observations were examined of which 87.5% related to incorrect responses and 12.5% related to correct responses. These log item-feedback-times were considered an outcome variable in a multilevel study due to the nested structure of the data.

The Mixed models routine in PASW 18 was used to analyse the log item-feedback-times using the linear multilevel modelling framework for continuous outcomes. For the various models that were considered, the convergence criteria were evaluated and showed no problems, and restricted maximum likelihood estimates were reported. Table 3.4 shows descriptive statistics for the different variables.

**Table 3.4:** Descriptive statistics for item-feedback-time

| Variable name | Level | N | Mean | Sd | Minimum | Maximum |
|---|---|---|---|---|---|---|
| Log item-feedback-time | 1 | 2093 | 3.01 | 3.62 | 0.69 | 6.80 |
| Item response | 1 | 2093 | 0.12 | 0.33 | 0 | 1 |
| Student achievement | 2 | 610 | 0.50 | 0.14 | 0.13 | 0.93 |
| Item difficulty | 2 | 50 | 0.45 | 0.17 | 0.09 | 0.91 |

First, the cross-classification structure of the log-feedback times was investigated, where the non-zero feedback times are nested in items and students. That is, it was investigated using a random item intercept and a random student intercept whether variation in feedback times is explained across items and students, respectively. In Table 3.5, the parameter estimates of the empty multilevel model with a random student and item effect are reported under the label Model 1. In this multilevel model, a random intercept that varied across items and a random intercept that varied across students were defined, besides the residual variation in log item-feedback-times. This means that the logarithmic observations were assumed to be independently and normally distributed given the random item and student effect.

The average item-feedback-time of the opened feedback pages in the population is around 9.412 seconds, which is the exponent of 2.242. There are three sources of variation modelled. Conditional on the student and item effect, the unexplained residual variation is around .975. The between-student variation in log item-feedback-time is around .277 and significantly different from zero. The between-student variation shows that students differ from each other with respect to the time that the feedback page was opened. Given the item effect, differences between students explain around 22% of the total variation. The estimated variance of the random item intercept is around .052 and also significantly different from zero. Thus, a cross-classified structure was detected with an item and student clustering. The variation across items is small and only explains 5% of the total variance conditional on the student effect.

Various explanatory variables were used to explain the detected variation in the logarithmic item-feedback-times. In Model 2, the student response and the student achievement were used as level-1 predictors for feedback time with fixed effects, where student achievement was grand-mean centred. The effects of the explanatory variables were not allowed to vary across items or students. It follows that the feedback-time of the incorrectly answered items are significantly higher than those of the correctly answered items. The feedback-times of the correctly answered items are on average 0.53 seconds lower than those of the incorrectly answered items. The estimated student achievement effect is around 0.45, which means that an increase of one point in student achievement leads to an increase of 0.45 in

the logarithmic feedback-time, which translates to 1.57 seconds. This means that the more able students spend more time on a feedback page than the less able students. In Model 2, the fixed intercept represents the average feedback-time of incorrectly answered items of students with an average achievement, which is around 10.04 seconds.

**Table 3.5:** Multilevel analysis of log item-feedback-times using student achievement and student response as explanatory information.

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| Parameter | Est. | SE | Est. | SE |
| *Fixed Effects* | | | | |
| Intercept | 2.242 | .051 | 2.307 | .139 |
| Student Response | | | -.643 | .074 |
| Student Achievement | | | .450 | .249 |
| *Random Effects* | | | | |
| Residual Variance | .975 | .033 | .928 | .031 |
| Intercept Variance (student) | .277 | .036 | .305 | .038 |
| Intercept Variance (item) | .052 | .015 | .053 | .015 |
| *Information Criteria* | | | | |
| Log likelihood | 6338.429 | | 6266.075 | |
| AIC | 6344.429 | | 6272.075 | |

It can be seen that the predictor variables added in Model 2 did explain around 5% of the variation at level 1. The level-2 variable student achievement was expected to explain between-student variation and the variance of random student effect decreased from 0.277 to 0.273 when only including student achievement as a predictor. The inclusion of the level-1 explanatory variable student response led to a slight increase of the variance of the random student effect. The information criteria show that Model 2 fits the data better than Model 1, which supports the inclusion of both predictors.

The effects of the student response and achievement did not vary across students or items. Furthermore, the item difficulty did not influence the item-feedback-time. The interaction effect representing a different effect of student achievement for correctly and incorrectly answered items was neither significantly different from zero. The feedback times did not differ significantly across subject domain. It was not possible to explain a lot of variation in

the item-feedback-times and more research is needed to identify factors that can explain differences in feedback-time between students and items.

Table 3.6 shows probabilities of item-feedback-time for correct and incorrect answers for low, medium, and high student achievement calculated with Model 2. The expected item-feedback-time increases slightly, for both incorrect and correct responses, when achievement increases. The expected item-feedback-time approximately doubles for incorrect responses compared to correct responses.

**Table 3.6:** Probabilities of item-feedback-time for correct and incorrect answers for low, medium, and high student achievement conditional on feedback-use.

| Student achievement | Student response | Expected item-feedback-time |
|---|---|---|
| Low (25% correct) | Correct | 4.72 seconds |
| | Incorrect | 8.98 seconds |
| Medium (50% correct) | Correct | 5.28 seconds |
| | Incorrect | 10.04 seconds |
| High (75% correct) | Correct | 5.91 seconds |
| | Incorrect | 11.24 seconds |

Based on the results, the hypotheses II and IV were not rejected. The results show a positive relation between answering an item incorrect and item-feedback-time. No significant relation between item difficulty and item-feedback-time was found. Furthermore, the results show a positive relation between student achievement and item-feedback-time. An increase in achievement leads to an increase in item-feedback-time.

## 3.6 Conclusion and discussion

In this study, individual and group differences in feedback behaviour in CBFA were explored and relationships between student response, student achievement, item difficulty, and feedback behaviour were examined. Feedback behaviour is represented by 1) the observation per item whether a student opens a pop-up page with additional feedback, referred to as feedback-use, and 2) the time between opening and closing a pop-up page with additional feedback, referred to as item-feedback-time. The (generalized) multilevel modelling approach was used to explore effects of student response (correct or incorrect), item difficulty, and student achievement on feedback-use and item-feedback-time.

Results show that variation in feedback-use is significantly explained by student response. As expected, students were mostly inclined to consult additional feedback for incorrect responses (Hypothesis I is not rejected). Furthermore, item-feedback-time increases for incorrect responses compared to correct responses (Hypothesis II is not rejected). In fact, the item-feedback-time approximately doubles for incorrect responses compared to correct ones. These findings for feedback-use, item-feedback-times, as well as the overall ratio between feedback-use of correct (12.5%) and incorrect (87.5%) responses suggest the CBFA to predominantly address the instrumental motive for feedback behaviour.

Various significant results were found for feedback-use of incorrect responses. Results showed that feedback-use for incorrect responses is significantly related to item difficulty. The propensity of feedback-use, given student achievement, increases when item difficulty increases (Hypothesis III not rejected). On the other hand, as expected, no significant relation was found between item difficulty and item-feedback-time (Hypothesis IV not rejected).

For incorrect responses, student achievement also explained variation in feedback-use. The propensity of feedback-use for incorrect responses is highest for high achieving students. However, the actual number of pop-up pages with additional feedback consulted does not differ much between middle and high achieving students (see Table 3.3). As a consequence, Hypothesis V is not rejected. The decrease in probability of feedback-use for low achieving students compared to middle and high achieving students suggests the CBFA to increasingly address the ego-based motive as achievement decreases. From an ego-based motive, feedback for incorrect responses and low levels of achievement can be unpleasant to receive (VandeWalle & Cummings, 1997) and lead to other strategies than putting extra effort in attaining the standard (Kluger & DeNisi, 1996; Narciss, 2008). Instead, students may respond to feedback reporting lacunas or standard discrepancies by ignoring or rejecting the feedback, by changing or reinterpreting the standard to match present performance, or by escaping the situation, for example by abandoning the standard. In addition, Wolf, Smith, & Birnbaum (1995) argue that the amount of energy students are willing to invest in low-stakes tasks such as formative assessments should be taken into account as it is suggested to be limited. In accordance, previous research showed test length, in this case the number of items in a CBFA, to moderate feedback seeking patterns (Timmers & Veldkamp, 2011). The number of additional feedback pages consulted by middle achieving students being comparable to the number consulted by high achieving students could be related to the theory of limited amount of energy. However, this would be remarkable as the CBFA's used included fifteen or twenty items only.

Furthermore, the expected item-feedback-time increases slightly, for both incorrect and correct answers, when achievement increases. These findings seem to support the suggestion that students have a limited amount of energy they are willing to invest in low-stakes tasks such as formative assessments (Wolf et al., 1995). However, these findings

could also imply that students with higher levels of achievement could be the ones that put more time and effort in studying feedback to further improve their levels of achievement.

Results showed subject domain to influence the probability of feedback-use. Bachelors of Health were most likely to consult additional feedback compared to Bachelors of BA and Law. A possible explanation for the differences in feedback behaviour between subject domains might be different proportions in gender. Previous research shows that girls are more likely than boys to seek help when they need it (Aleven et al., 2003). An underlying explanation for this feedback behaviour might be that girls tend to be more uncertain and doubtful about their abilities than boys. Health students can be expected to be nearly all females while both BA and Law students can be expected to represent both males and females. An alternative explanation for the differences in the search for feedback might be the discrepancy in uncertainty, irrespective of gender. This would mean that Law and Business students, in general, are more certain and less doubtful about their abilities than Health students. Differences in learning conditions between Bachelor programs, such as supervision and the timing of the CBFA, might add to the explanation of the differences in feedback-use encountered.

The feedback included in the CBFA combined elaborated feedback, such as explanations why the alternatives are incorrect, with KCR. Narciss and Huth (2004) advise against the combination of elaborated feedback and KCR by suggesting that the inclusion of KCR feedback increases the risk of superficial learning by preventing students from studying the elaborate feedback. The differences encountered in item-feedback-times for incorrect versus correct responses suggest that, in general, elaborated feedback is used as a complement to KCR feedback. Students tend to study elaborated feedback (longer) for incorrect responses. As such KCR does not seem to prevent students with instrumental feedback behaviour from studying elaborated feedback. Note that this study focuses on delayed informative feedback, while Narciss and Huth (2004) proposed guidelines for immediate informative tutoring feedback. The findings suggest added value of the combination of elaborated feedback and KCR, for example, when providing delayed informative feedback. However, further research is needed to examine the effects of the combination of these types of feedback on student learning.

Differences in item-feedback-times between correct and incorrect responses might be explained by a variance in motives for consulting additional feedback as well as response certitude. In this study, student response has been discussed as either correct or incorrect. However, this distinction is ambiguous as a correct response might just be a lucky guess and an incorrect response might result from a misconception (Kulhavy & Stock, 1989). Kulhavy and Stock, therefore, integrated response certitude into their model of text-based feedback. Their model explains feedback behaviour by relating student responses with student's initial response confidence level. In accordance with Kulhavy and Stock's certitude model, research shows high-confidence correct responses to yield the shortest feedback study times, and

high-confidence errors to yield the longest feedback study times (Mory, 2004). Feedback study times for low-confidence responses, either correct or incorrect, fall somewhere in between. Besides consulting additional feedback for low-confidence correct answers (instrumental motive), additional feedback for correct answers can be consulted to enhance the ego (ego-based motive) and to confirm correct answers (Mory, 2004). For the latter cases, a focus on KCR feedback only can be considered sufficient. These cases could then have contributed to lower item-feedback-times for correct responses. In addition, from an instrumental perspective, studying elaborated feedback can be considered relevant for low-confidence correct responses as well as high- and low-confidence incorrect responses. The high-confidence incorrect responses in particular might have contributed to higher item-feedback-times for incorrect responses as previous research showed high-confidence errors to yield the longest feedback study times.

There are limitations to the measure used to represent the time that a student spent studying additional feedback. The time measure may not accurately represent the time students study the feedback as students, for example, might have been interrupted between opening and closing a feedback page. This would lead to an extended item-feedback-time. Supervision differed for the three groups of bachelor students. One group completed the CBFA without supervision while the other two groups were supervised. The difference in supervision did not seem to have influenced the item-feedback-times, as item-feedback-times did not differ significantly between the three groups of Bachelor students. Even though, the logarithm of the item-feedback-time was used to compensate for extreme values of item-feedback-time and to induce a positivity restriction.

The variables examined only explained part of the residual variance in feedback-use and item-feedback-time. For example, in the model for the item-feed-back times, where the residual variance is directly parameterized, only 4.8% of the residual variance was explained by incorporating student response and student achievement. Further research is needed to identify factors that explain the residual variance in feedback-use between students that remained unexplained. Variables that would, for example, be worthwhile to include in future research are gender, response certitude, goal orientation, task value, and success expectancy (Bangert-Drowns et al., 1991; Mory, 1994).

Researchers studying the relation between feedback and learning outcomes tend to leave feedback behaviour out of scope (e.g. Corbalan et al., 2009; Gordijn & Nijhof, 2002; Pridemore & Klein, 1995; Smits et al., 2008; Van der Kleij et al., 2011; Wang, 2007). Since feedback behaviour varies greatly between students, it is recommended to distinguish between students that seek and attend to feedback and those who don't. As such, possible effects of feedback interventions on learning outcomes can be examined more precisely.

# Chapter 4 - Motivational beliefs, student effort, and feedback behaviour


## Abstract

Feedback can only be effective when students seek feedback and process it. This study examines the relations between students' motivational beliefs, effort invested in a computer-based formative assessment, and feedback behaviour. Feedback behaviour is represented by whether a student seeks feedback and the time a student spends studying the feedback. The motivational beliefs examined in this study are success expectancy and task-value beliefs. Results show that the effort invested in the formative assessment was predicted by task-value beliefs, but not by success expectancy beliefs. Furthermore, feedback seeking was predicted by success expectancy as well as task-value beliefs, while feedback study time was not. In addition, feedback seeking was predicted by student effort invested in the formative assessment.


## 4.1 Introduction

Assessments are often used to formulate judgments about performance in order to select, classify or certify. These purposes are referred to as the summative function of assessment. Assessments can also be aimed at contributing to student learning, which is referred to as the formative function of assessment. The formative function of assessment is realized by communicating information on student performance, in the form of feedback, to teachers and/or students to stimulate and enhance teaching and learning processes. Feedback comes in many shapes and sizes. A distinction is often made between the following types of feedback: 1) knowledge of results (KR), where the student is merely informed whether or not an answer is correct, 2) knowledge of correct response (KCR), and 3) elaborated feedback (EF), where more information is provided (Shute, 2008).

Previous research has shown that KCR and EF are more likely to lead to improved learning outcomes than KR (Bangert-Drowns et al., 1991; Van der Kleij et al., 2011). However, simply adding KCR or EF feedback to computer-based formative assessments (CBFA) does not guarantee that students seek and process the feedback (Aleven et al., 2003; Timmers & Veldkamp, 2011). Foremost, students need to be willing to invest time and effort in seeking


Adapted from: 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. doi: 10.1016/j.compedu.2012.07.007

and processing feedback. Many factors influence this willingness. These factors can be categorized as characteristics of the feedback intervention itself (e.g. type of feedback), characteristics of the task and setting (e.g. test length), and student characteristics (e.g. motivational beliefs) (Narciss & Huth, 2004; Timmers & Veldkamp, 2011).

The present study examines the relations between students' motivational beliefs, effort invested in a CBFA and feedback behaviour. An advantage of using a computer in formative assessment is the possibility to record feedback behaviour by logging feedback study times and whether or not students seek feedback by linking to feedback pages. The motivational beliefs examined in this study are success expectancy and task-value beliefs.

## 4.2 Theoretical framework

Formative assessment is a term that is conceptualized in various way (Bennett, 2011; Yorke, 2003). However, it is generally agreed that its main purpose is to contribute to student learning by providing feedback. In this study, feedback as part of formative assessment is defined as information regarding student performance on a test-like event that is provided by an agent (e.g. teacher, computer) to teachers and/or learners with the purpose of stimulating and enhancing teaching and learning processes (cf. Hattie & Timperley, 2007; Sadler, 1989). This information can pertain to actual levels of

performance as well as desired levels of performance. Additionally, it can include hints and means for reducing the gap between the actual and desired levels of performance or understanding (Hattie & Timperley, 2007; Sadler, 1989).

Mindful processing of feedback is considered crucial for feedback to result in improved learning outcomes (Bangert-Drowns et al., 1991). Students have to play an active and self-regulating role in seeking and constructing meaning from feedback information (Nicol & McFarlane-Dick, 2006). But of course, students then need to be willing to invest time and effort in a formative assessment as well as seeking and processing feedback. Such willingness can be promoted, but not guaranteed.

Models proposed for feedback seeking mostly focus on employees in organizations (Park et al., 2007; VandeWalle et al., 2000) and (interactive) computer-based learning environments (Aleven et al., 2003). Previous research shows that students are more likely to seek computer-mediated feedback than person-mediated feedback (Karabenick & Knapp, 1988; Kluger & Adler, 1993), probably because feedback seeking in computer environments often remains unnoticed by others. As such, the cost of exposing one's uncertainty and need for help, so called *self-presentation cost*, does not come into play (Aleven et al., 2003).

Bangert-Drowns et al (1991) describe the process of receiving feedback in a test-like event in a five-stage model. Within this model, motivational beliefs are positioned as both affecting

and being affected by feedback seeking and processing. The model starts with describing cognitive (degree of prior relevant knowledge) and motivational aspects (degree of self-efficacy, degree of interest, and kind of goal orientation) of the leaners initial state (stage 1). The initial state is assumed to influence the effort students invest in the subsequent stages of the model or cycle. When a test-like event, in this case a CBFA, is administered, items activate the process of addressing relevant prior knowledge (stage 2). Subsequently, the test takers construct a response (stage 3). In this model, the option of on-demand help or feedback during stage 2 and 3 has not been taken into account. Instead, learners are provided with feedback information after they have constructed a response (stage 3). The next stage is the evaluation of results (stage 4). When the purpose of a test-like event is knowledge acquisition the cycle is successfully completed when adjustments are made to the degree of prior knowledge (stage 5). Test-like events are assumed to influence motivational aspects of the initial stage, e.g. degree of interest, for future (comparable) test-like events. As such, test-like events can also be used for the purpose of influencing motivational aspects of the initial stage. This could, for example, be relevant when students under- or overestimate their own abilities. A test-like event could then be used to monitor (limitations of) certain abilities and, as a consequence, could lead to adjustments of students' degree of self-efficacy or interest in the task or topic. The adjustments made in reaction to the test-like event (stage 5), or differences between stage 1 and 5, can be viewed as learning outcomes of the test-like event.

Previous research shows that student time and effort invested in a test-like event, or learning task, as well as seeking and processing feedback can be promoted or inhibited by various variables for the various stages mentioned by Bangert-Drowns et al (1991). For example, if feedback is available when learners are requested to address prior knowledge and formulate responses (stage 2 and 3), students might copy feedback answers and, as such, short-circuit the searching-response-evaluation- adjustment process (Bangert-Drowns et al., 1991). Furthermore, possibilities for evaluation (stage 4) are limited if learners are provided with KR only (Hattie & Timperley, 2007). In addition, a positive attitude towards the test-like event can promote feedback study time (stage 4) (Van der Kleij et al., 2012). Previous research shows that feedback study time is also influenced by the correctness of the response and response certitude (stage 4) (Mory, 2004). Correct answers on items for which students were confident that they answered them correctly (low discrepancy) yield the shortest feedback time. Incorrect answers on items for which students were confident that they answered correctly (high discrepancy) yield the longest feedback study time.

According to the expectancy-value-theory of Eccles and Wigfield (2002), student willingness to invest time and effort in a task is explained by success expectancy and task-value beliefs, that is, so called motivational beliefs. Success expectancy beliefs refer to students' judgements about their capabilities to complete certain academic tasks. Beliefs about the importance of, interest in, and value of the task are referred to as task-value beliefs. Furthermore, motivational beliefs are viewed to be task-specific as they vary for different

activities (Crombach, Boekaerts, & Voeten, 2003). Students' task-specific motivational beliefs result from initial judgments about a task based on previous experiences with similar tasks and content. Crombach et al (2003) distinguish between two kinds of task-specific judgments: Judgments about the task before and judgments about the task after completing a learning task, appraisals and attributions, respectively.

## 4.3 Aims of the present study

The present study examines the relations between students' motivational beliefs, effort invested in a CBFA and feedback behaviour. A schematic overview of the relations examined is presented in Figure 4.1. In terms of the five-stage model described by Bangert-Drowns et al (1991), this study focuses on the relation between the leaners initial state (task-value beliefs and success expectancy), addressing relevant prior knowledge and constructing a response (effort students invest in a test-like event), and evaluation (feedback behaviour). The test-like event in this study is a CBFA on information literacy. Previous research shows that students have shortcomings in information literacy and at the same time tend to overestimate their information literacy (Ivanitskaya et al., 2006; Kuhlemeier & Hemker, 2005; Maughan, 2001). As such, the purpose of the CBFA is twofold, namely 1) knowledge acquisition, and 2) increase the degree of interest in the topic.

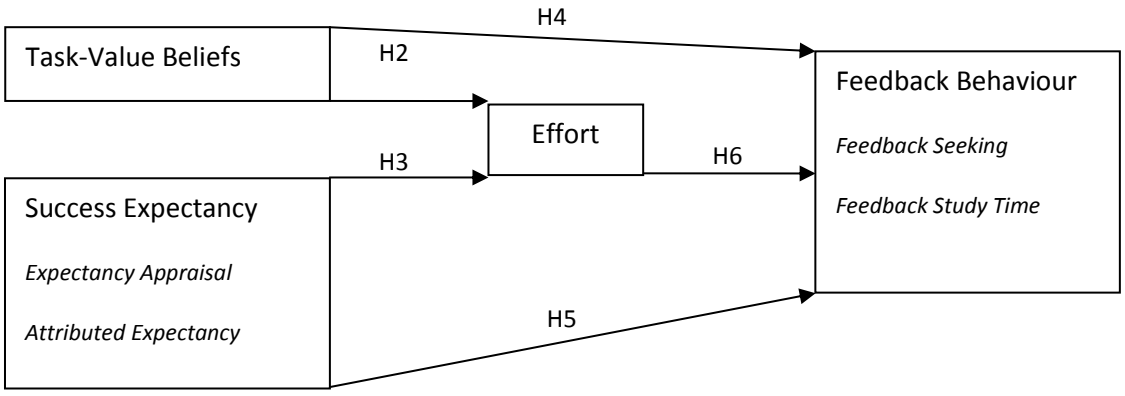


**Figure 4.1**: A schematic representation of the six hypotheses tested in this paper.

Given the tendency of students to overestimate their information literacy, students are expected to experience the CBFA as more difficult than expected in advance (H1). This assumption is related to 'success expectancy' and referred to as 'expectancy discrepancy'.

Based on the expectancy-value theory of Eccles and Wigfield (2002), Effort invested in a CBFA is assumed to be predicted by Task-Value Beliefs (H2) and Success Expectancy (H3). Pintrich (1999) found that both success expectancy and task-value beliefs predict the

employment of follow-up activities, such as seeking and processing feedback, in learning processes. Therefore, it is expected that Feedback Behaviour is predicted by Task-Value Beliefs (H4) and Success Expectancy (H5). Furthermore, the relation between Effort invested in a test-like event and Feedback Behaviour is explored, as it has not been studied extensively. Effort can be viewed as an indirect measure of motivation (Crombach et al., 2003). Therefore, a positive relation is expected between Effort and Feedback Behaviour (H6).

## 4.4 Methodology

### *Research population and procedure*

First-year bachelors of Health (N=151) were requested to assess their own knowledge and understanding related to information literacy with a 20-item CBFA. The vast majority of bachelors of Health is female and their age ranges between 17 and 19 as the CBFA was embedded in a regular four-year training. Information literacy refers to the ability to identify information needs, locate corresponding information sources, extract and organize relevant information from each source, and synthesize information from a variety of sources (Walraven et al., 2008). Questions on task-value beliefs, success expectancy and effort invested in the CBFA were integrated in the CBFA system. The questions related to task-value beliefs and expectancy appraisal were presented before the items of the CBFA. Students were shown two exemplary items of the CBFA beforehand to get an impression of the task. Questions related to attributed expectancy and effort were presented after all items of the CBFA had been completed and before the KR page was presented. In addition, students were asked to estimate the number of correct answers just before the first item and directly after answering the last item of the CBFA to measure expectancy discrepancy.

The CBFA was programmed during one of a series of lessons of a research skills training program for first-year bachelors of Health. No consequences were linked to skipping the lesson, as attending lessons is recommended to students but not compulsory. In general, all first-year bachelors of Health (N=190) were recommended to attend lessons of the research skills program. The students (N=151) that attended the lesson in which the CBFA was programmed were divided over six separate and proctored sessions of ninety minutes. The surveillant started the sessions with a brief verbal instruction about the CBFA and handed out a written instruction. Students were told the outcomes of the CBFA would be discussed in the forthcoming lecture. Furthermore, after completing the CBFA students were requested to stay at least for the first forty-five minutes. They were encouraged to use the remaining time to work on an assignment for the next lesson of the research skills training program. Most students finished the test within forty-five minutes. All students (N=151) that attended one of the sessions completed the formative assessment. Only 141 completed the entire questionnaire on task-specific motivational beliefs.

### *Questionnaire on task-value beliefs, success expectancy and effort*

Crombach, Boekaerts, and Voeten (2003) developed a measurement instrument to study student's task-specific motivational beliefs when faced with specific curricular tasks, the online-motivation questionnaire (OMQ). The OMQ identifies cognitive judgments students make to determine the amount of effort they intend to spend on a task. The original version of the OMQ consists of seven dimensions. Only the dimensions concerning task-value beliefs, success expectancy (attributions and appraisals) and effort were selected for this study. All items of the OMQ are 4-point Likert-scales.

The questionnaire used in this study, included dimensions referred to as task-value beliefs, expectancy appraisal, attributed expectancy, and effort. Effort is an indirect measure of task-specific motivational beliefs as it is considered a result of motivational beliefs. In addition, students were asked to indicate how many items they expected to (have) answer(ed) correct. This question was asked twice, both 1) preceding the CBFA and the first part of the questionnaire, and 2) directly after completing the CBFA and before the second part of the questionnaire. The answers to this question give an impression of the discrepancy between expectancy appraisal and attributed expectancy. As such, success expectancy encompassed the following three dimensions: expectancy appraisal, attributed expectancy and expectancy discrepancy.

### *A computer-based formative assessment on information literacy*

In computer-based environments, a distinction can be made between on-demand help or feedback before KR is provided on a task (e.g. Kluger & Adler, 1993; Narciss et al., 2004) and additional feedback after KR is provided (e.g. Karabenick & Knapp, 1988; Mory, 1994). In a test-like event which provides multiple tries or on-demand help or feedback before KR is provide on a task, hints or worked-out examples without KCR are preferred. The exclusion of KCR prevents the undesirable effects of pre-search availability (Narciss & Huth, 2004). This study limits itself to a context in which additional feedback is available for students after they have been provided with KR on a CBFA.

The CBFA presented students with 20 multiple-choice items on information literacy. The items were selected from an existing pool of items developed by information professionals working at university libraries. The lecturers involved in the research skills training program selected the items based on estimated relevance. Furthermore, the final selection of items covered the various aspects of information literacy, such as source selection, the application of search strategies, and referencing. By including KCR and EF, students were given an opportunity to engage in knowledge acquisition activities. The CBFA used in this study provided students with a KR page after the student completed all items of the CBFA. The KR page presented information whether items were answered correct or incorrect. The KR page

is automatically generated and includes hyperlinks to pop-up pages with additional feedback per item. The additional feedback included KCR and an explanation of the various concepts used in the stem and answering categories (EF). In order to continue after opening an additional feedback page, this pop-up page first needs to be closed. The CBFA registered feedback behaviour by logging 1) whether students linked to additional feedback pages, and 2) the time between a pop-up pages with additional feedback was opened and closed.

### *Data analyses*

The dependent variable feedback behaviour per student is represented by feedback seeking, and feedback study time. Feedback seeking was measured by the proportion of additional feedback pages opened per student. Feedback study time was measured per item by logging the time between opening and closing an additional feedback page. Measures per item were added to calculate total feedback study time per student. Although multitasking during the CBFA was discouraged by surveillance, it could not be excluded completely. To compensate for extreme values of feedback study time, the logarithm of total feedback study time was used.

An adapted version of the OMQ was used to measure task-value beliefs, expectancy appraisal, attributed expectancy, and effort. The questions about the number of items students expected to have answered correctly, were added to the questionnaire. The expected number of correctly answered items, out of twenty, was divided by five to create a metric in accordance with the 4-point Likert-scales of the OMQ. To confirm the various dimensions in the questionnaire, a varimax rotated principal component analysis was used.

For the various constructed subscales found in the principal component analysis unweight means were calculated (within a range of 1 to 4). Furthermore, a variable was added for expectancy discrepancy. This variable provides an indication of the extent to which students experienced the CBFA to be more or less difficult than expected in advance. This variable was calculated by subtracting the student estimation about the expected number of correctly answered items made directly before completing the CBFA and the estimation made directly after completing the CBFA. This variable pertains to Hypothesis 1.

A linear regression analysis, using PASW statistic 18, tested whether effort was predicted by task-value beliefs (H2) and the three dimensions of success expectancy (H3). Subsequently feedback seeking and total feedback study time were regressed on task-value beliefs (H4) and the three dimensions of success expectancy (H5) using PASW statistic 18 probit-analysis (Liao, 1994) and linear regression analysis, respectively. Finally, probit and linear regression analysis were used to explore the relation between effort and feedback behaviour (H6), feedback seeking and total feedback study time, respectively.

## 4.5 Results

### *Feedback behaviour*

Out of 151 students, 113 opened a total number of 871 additional feedback pages. As such, 75% of the students consulted additional feedback. On average this group of students opened 7.7 (SD=4.10) additional feedback pages.

For those students that opened additional feedback pages, the average feedback study time per item was 17.2 seconds (SD=10.8). The average total feedback study time was 137.5 seconds (SD=129.8) per student.

### *Principal component analysis of task-specific motivation*

The principal component analysis with varimax rotation presented in Table 4.1 shows four dimensions of students' task-specific motivation. Based on the Eigenvalues and the changes in explained variance, the analysis with four components was preferred over other solutions. The four components explained 55.9% of total variance. A fifth component explained an additional 6% of variance but was difficult to interpret. One item (I did well on the test, because I was lucky) was removed based on the results of the analyses as this question had a low factor loading on all four components.

Table 4.1 shows a 6-item scale for Effort (Cronbach's $\alpha$=.77). For Expectancy Appraisal a 4-item scale ($\alpha$=.75) was encountered. The third aspect is a 5-item scale referred to as Attributed Expectancy ($\alpha$=.67). The fourth aspect is referred to as Task-Value Beliefs and contains only 2 items. These items showed a Spearman correlation coefficient of 0.47 ($p < 0.01$).

**Table 4.1:** Principal component loadings for Effort, Expectancy Appraisal, Attributed Expectancy, Effort, and Task-Value Beliefs after varimax rotation.

| Dimensions | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| *Effort* | | | | |
| How much effort did you spend completing the task? | **.44** | .03 | -.40 | .11 |
| How much time did you spend answering the items? | **.62** | -.13 | -.08 | .27 |
| How much effort did you spend to reach a high test score? | **.69** | .06 | -.17 | .00 |
| I did (not do) well, because I (do not) like this kind of tasks | **.74** | .11 | .23 | -.02 |
| I did (not do) well, because I (did not) tried my best | **.80** | .00 | -.20 | .01 |
| I did (not do) well, because the task was (not) pleasant | **.69** | -.00 | .12 | .18 |
| *Expectancy Appraisal* | | | | |
| How well do you expect to perform on this task? | .18 | **.73** | .10 | .08 |
| How good are you in tasks which call on information literacy? | .03 | **.70** | .09 | -.09 |
| What is your opinion on this kind of tasks generally? | -.08 | **.77** | -.10 | -.23 |
| How many of the 20 items do you expect to answer correct? * | -.08 | **.79** | .16 | .13 |
| *Attributed Expectancy* | | | | |
| How many of the 20 items do you think you answered correct? * | -.04 | .39 | **.58** | .21 |
| How difficult do you perceive this task to be? | -.21 | -.13 | **.74** | -.02 |
| I did (not do) well, because I'm not good at this kind of tasks | .30 | .28 | **.51** | -.33 |
| I did (not do) well, because I almost knew nothing about it | -.10 | .00 | **.70** | -.21 |
| I did (not do) well, because I found the task to be difficult | .11 | .20 | **.57** | -.02 |
| *Task-Value Beliefs* | | | | |
| I consider the task to be valuable | .07 | -.05 | -.21 | **.79** |
| How important is it for you to perform well on this kind of tasks? | .22 | .02 | .02 | **.80** |

* *used to measure expectancy discrepancy*

### *Descriptive statistics*

Descriptive statistics for Task-Value Beliefs, the three dimensions of success expectancy and Effort are presented in Table 4.2. The means for Task-Value Beliefs, Expectancy Appraisals, Attributed Expectancy and Effort represents population average of unweight means per student. The mean for expectancy discrepancy is the average discrepancy between a students' estimate of expected number of items of the CBFA answered correct (within a range of 0 to 20) before and after completing the CBFA. The descriptive statistics for expectancy discrepancy show that on average students' estimates before and after completing the CBFA differed 3.54 points. This confirms that students, on average, experienced the CBFA to be more difficult than expected in advance (H1 is confirmed).

**Table 4.2:** Descriptive statistics

| Variable name | N | # items | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|---|---|
| Task-Value Beliefs | 151 | 2 | 2.79 | 0.44 | 1 | 3.5 |
| Expectancy Appraisals | 151 | 4 | 2.32 | 0.22 | 2 | 2.9 |
| Attributed Expectancy | 151 | 5 | 2.08 | 0.45 | 0.6 | 3.2 |
| Expectancy Discrepancy | 151 | 1 | 3.54 | 3.08 | -10 | 10 |
| Effort | 141 | 6 | 2.61 | 0.46 | 1 | 3.5 |

### *Predictors of Effort*

Effort was regressed on Task-Value Beliefs, and the three dimensions of success expectancy using linear analyses (N=141). Results showed a positive relation between Task-Value Beliefs and Effort (H2 is confirmed). That is, the regression coefficient β was equal to 0.27 with a significance probability p=0.02. No significant relations were found between the three dimensions of success expectancy and Effort (H3 is confirmed).

### *Predictors of feedback behaviour*

Feedback seeking was regressed on Task-Value Beliefs and the three dimensions of success expectancy using probit analyses (N=113). Results showed a positive relation between feedback seeking and Task-Value Beliefs (H4 is confirmed for feedback seeking). That is, the regression coefficient β was equal to 0.14 with a significance probability p=0.02. Significant relations were also found for feedback seeking and Expectancy Appraisals (β=0.30, p=0.01), Expectancy Attributions (β=0.16, p=0.01), and expectancy discrepancy (β=-0.02, p=0.02) (H5

is partly confirmed for feedback seeking). Total feedback study time per student was regressed on Task-Value Beliefs and the three dimensions of success expectancy using a linear analysis (N=113). Results showed no significant relations (H4 and H5 are disconfirmed for total feedback study time).

### *Effort and Feedback Behaviour*

Feedback seeking was regressed on Effort using probit analysis (N=113). Results showed a positive relation ($\beta=0.16$, $p < 0.01$). No significant relation was found for total feedback study time and Effort. As such, H6 is confirmed for feedback seeking and disconfirmed for total feedback study time.

## 4.6 Conclusion and discussion

The purpose of this study was to examine the relations between students' motivational beliefs, effort invested in a CBFA, and feedback behaviour. The motivational beliefs studied are success expectancy and task-value beliefs. Success expectancy encompassed the following three dimensions: expectancy appraisals, attributed expectancy and expectancy discrepancy. A considerable amount of Health students (75%) showed willingness to invest time and effort in seeking additional feedback pages (on average eight out of twenty pages) included in a CBFA on information literacy. Measures for expectancy discrepancy showed that on average students experienced the CBFA to be more difficult than expected in advance (H1 was confirmed).

Based on Eccles and Wigfield (2002) expectancy-value-theory it was expected that student effort invested in a CBFA would be explained by their success expectancy and beliefs about the value of a task. Students who perceived the CBFA to be more valuable reported more effort invested in completing the CBFA (H2 was confirmed). No relation was found between success expectancy and student-reported effort invested in the CBFA (H3 was disconfirmed). Students' goal orientation might add to an explanation for the results encountered. Students with a more intrinsic orientation tend to focus on learning and understanding. This *learning goal orientation* is assumed to lead to the use of more appropriate learning strategies compared to performance goal orientation (Aleven et al., 2003). The respondents attended a non-compulsory lesson, a learning strategy which suggests an inclination towards learning goal orientation. From a learning goal orientation perspective, the influence of success expectancy on effort is expected to be limited as the focus is on mastery instead of performance. In addition, the introduction of the CBFA included an explanation of the formative function of the CBFA, which aimed at reassuring students that the assessment aimed at provided an opportunity to learn instead at judging performance. After all,

students were confronted with the CBFA for the first time and without the opportunity to prepare themselves. This could have limited the effect of success expectancy on effort. Task-value beliefs are still expected to influence effort within the context of a learning goal orientation, as students have to perceive the task as valuable enough to invest effort in.

Feedback behaviour was assumed to positively relate to task-value beliefs and success expectancy. Feedback behaviour was represented by feedback seeking as well as feedback study time. Results showed that feedback seeking was indeed positively related to task-value beliefs (H4 was confirmed for feedback seeking). Furthermore, higher levels of success expectancy before completing the CBFA (expectancy appraisals) positively related to feedback seeking. Higher levels of success expectancy after completing the CBFA (attributed expectancy) also related positively to feedback seeking. As such, the results are largely in line with the results of Pintrich (1999), who found that students who feel more efficacious about their ability to do well are more likely to be cognitively involved to learn, for example by seeking additional feedback. On the other hand, results showed a small negative relation between expectancy discrepancy and feedback seeking. This means that an increase in underestimation of the difficulty of the CBFA led to a decrease in feedback seeking. As such, an increase in expected and perceived difficulty seems to lead to an *ego-based motive* to defend or enhance one's ego by rejecting or ignoring the feedback (Ashford et al., 2003; Kluger & DeNisi, 1996). The aim of a formative assessment is to support learning, and, as such, to address an instrumental motive, to achieve a goal or perform well, for feedback behaviour instead of an ego-based motive (Ashford et al., 2003). In conclusion, a positive relation between success expectancy and feedback seeking is largely confirmed, but needs to be refined by noting that a high expectancy discrepancy seems to lead to an opposite effect and should therefore be limited (H5 is largely confirmed for feedback seeking, but needs to be refined).

Results showed that differences in the logarithm of total feedback study time per student could not be attributed to task-value beliefs nor the three dimensions of success expectancy (H4 and H5 were disconfirmed for feedback study time). Time measures can be expected to highly vary due to random events of various sorts, for instance, individual differences in reading speed. This results in a power problem which hampers finding significant results for the relation between feedback study time and variables such as success expectancy and task-value beliefs. Furthermore, previous research has shown response certitude to influence feedback study time (Kulhavy & Stock, 1989; Mory, 1994). Response certitude contrasts success expectancy per item against the KR of the respective item. Unfortunately, the CBFA system used did not include the option for students to indicate their certitude about their response. For further research, it is recommended to study differences in feedback study time while controlling for response certitude.

The negative relation between expectancy discrepancy and feedback seeking suggests that difficult test-like events can lead to loss of motivation and, as a consequence, rejection of

feedback. Easy activities, on the other hand, might lead students to perceive a small performance gap that is not worth any additional effort, such as seeking and studying feedback. As a task should not be too easy or too difficult, the relation between student achievement and feedback behaviour cannot be expected to be linear (Sadler, 1989; Schunk et al., 2008). Further research is needed to examine motivational beliefs, such as success expectancy and goal orientation, and feedback behaviour in relation to student achievement.

The effects observed for motivational beliefs and feedback behaviour were rather small. However, the measures used did not have very high reliability coefficients so the effect sizes may well have been underestimated. In addition, it is reasonable to expect the non-compulsory setting to have influenced the findings. Skipping lessons of the research training program was not linked to any consequences. As such, the setting led to self-selection of respondents. A total of 190 first-year bachelors of Health were expected to attend the lesson in which they were requested to complete the CBFA. Instead a total of 151 students showed up. The motivational beliefs of students (N=39) skipping the lesson can be expected to differ from those students attending the lesson. Inclusion of all first-year bachelors of Health would have led to a wider variation in motivational beliefs and, possibly, stronger effects. On the other hand, if completing the CBFA had been compulsory, students' task-value beliefs might have increased extrinsically, as students would be influenced by the impression that lecturers consider the CBFA to be an important aspect of their research training.

There are limitations to the measures used to represent feedback study time and student effort invested in completing the CBFA. In the present study it is assumed that the time between opening and closing an additional feedback page represents the time that a student spends studying the additional feedback. However, student might have been interrupted between opening and closing a feedback page. As a consequence, the time measures may not accurately represent the time students spent studying the feedback. Furthermore, effort was measured through self-report. It remains unclear to what extent the self-reported measure approaches actual effort invested in the CBFA.

Based on the results developers of and lecturers using CBFAs are recommended to take into account students task-value beliefs and success expectancy. Motivational interventions should be considered to create attractive CBFAs as well as a context which positively fosters students' task-value beliefs and success expectancy, and, as a consequence, effort invested in a CBFA, including feedback seeking. An exemplary intervention is asking students to set specific goals, related to the respective topic, before using a CBFA (Boekaerts & Corno, 2005). In addition, a CBFA should not be experienced as begin too difficult to prevent frustration or loss of motivation. On the other hand, a CBFA should be difficult enough to perceive a need for feedback.

# Chapter 5 - Implications of previous research for the design of feedback interventions in computer-based formative assessment

## 5.1 Introduction

Assessments can be used to measure and appraise student performance. This is referred to as the summative function of assessment. Assessments can also be used to generate feedback on performance to support and direct student learning. This is referred to as the formative function of assessment or formative assessment. Summative assessments has a retrospective nature, while formative assessment is forward looking. The past two decades interest in formative assessment has increased in societies that are complex and subject to rapid changes, so called learning societies. Lifelong learning is emphasized as a mean to keep up with change in these increasingly complex societies. Formative assessment fits the process of preparing students for lifelong learning, especially, when formative assessment aims at providing students with insight in their performance gaps and how to improve their performance (Boud, 2000).

Feedback is a crucial element of formative assessment. Here, feedback is conceptualised as information about learners' actual state of performance that is provided to learners in the context of a test-like event, intended to modify their thinking or behaviour for the purpose of improved performance (cf. Narciss, 2008; Shute, 2008). Generating individualized feedback to support and direct learning is a time consuming task for teachers. The use of computers in formative assessments has several advantages as it provides the opportunity to generate immediate, objective and appropriate feedback based on response models constructed in advance.

Previous research shows that feedback in computer-based formative assessment (CBFA) can improve student learning (Van der Kleij et al., 2011). Previous research also shows that the relation between CBFA and learning is complex. The effectiveness of CBFA is influenced by numerous variables, such as the intended learning outcomes, types of feedback and student motivational beliefs (Timmers, Braber-Van den Broek, & Van den Berg, 2013; Van der Kleij et al., 2011). These variables need to be taken into account by designers of CBFAs for reasons of effectiveness. In the context of this study, CBFA is considered effective when fostering student effort invested in a CBFA for the purpose of improving performance. The aim of this study is to support the design of effective CBFAs by drawing an overview of variables substantially influencing the effectiveness of CBFA. Findings of previous research have been translated to recommendations for the design of CBFAs. In addition, the findings have been used to propose a theory based decision-making framework for the design of feedback interventions in CBFA.

Although CBFA can have a positive effect on student performance, this effect cannot be guaranteed. The effectiveness of CBFA is influenced by numerous instrumental variables (e.g. types of feedback) and student characteristics (e.g. motivational beliefs). The aim of this study is to draw an overview of instrumental and student characteristics that substantially influence the relation between CBFA and learning. Additionally, the aim is to determine the implications of previous research for the design of feedback interventions in CBFA.

## 5.2 Effectiveness of formative assessment

In previous research, formative assessment has been construed as either an instrument or a process (Bennett, 2011). The instrumental approach emphasizes the design and evaluation of formative assessments (Sadler, 1998). In this context, an assessment is defined as formative when its specific aim is to generate feedback to support student learning. The process approach emphasizes the relation between formative assessment and learning (Black & Wiliam, 2003). From this perspective, it is claimed that assessment is formative only when feedback information is actually used to adapt teaching and learning to improve student learning. As Bennett argues (2011) both conceptualisations of formative assessment exclude one another and are therefore limited. In this study, formative assessment has been conceptualised as a purposefully designed instrument embedded within a learning process. In other words, formative assessment has been conceptualized as combination of an instrument and a process.

Variability in the effectiveness of formative assessment can be explained by 1) characteristics of the formative assessment instrument and 2) student characteristics, which moderate the process of formative assessment. Bangert-Drowns, Kulik, Kulik, and Morgan (1991) described the process of receiving feedback in a five-stage model. The stages distinguished within this model represent the process of formative assessment. The test-like event together with the feedback intervention represent the formative assessment instrument. A schematic overview of the conceptualisation of formative assessment based on of the five-stage model of Bangert-Drowns et al (1991) is presented in Figure 5.1.

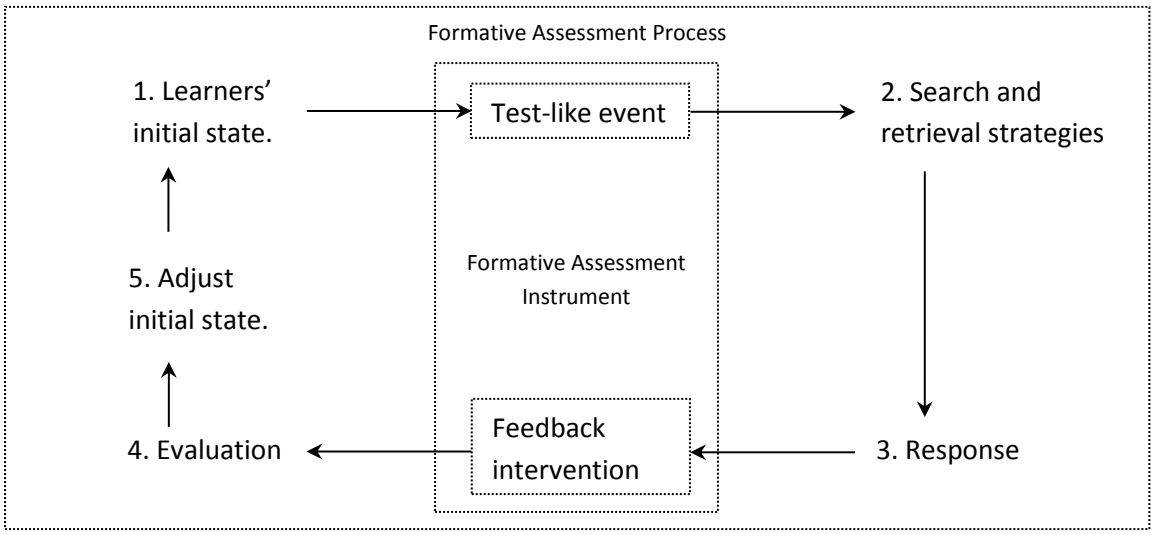


**Figure 5.1:** Conceptualisation of formative assessment based on the five-stage model of Bangert-Drowns et al (1991); adapted from Mory (2004, p. 752)

The model starts with describing the learners' initial state (stage 1). The initial state is characterized by cognitive (e.g. degree of prior knowledge) and motivational aspects (e.g. task-value beliefs). The initial state is assumed to influence the effort learners invest in the subsequent stages of the process of receiving feedback. When a test-like event is administered, items activate the process of addressing relevant prior knowledge and skills (stage 2). Subsequently, the test takers construct a response (stage 3). In this model, the option of on-demand help or feedback during stage 2 and 3 has been left out of scope. Instead, learners are provided with feedback after they have constructed one or more responses (after stage 3). The next stage is the evaluation of the response(s) by studying feedback (stage 4). When, for example, the intended learning outcome of a test-like event is further knowledge and understanding in a specific domain, the process is successfully completed when adjustments are made to the degree of prior knowledge and understanding in the corresponding domain (stage 5). The adjustments made in reaction to the test-like event (stage 5), or differences between stage 1 and 5, are considered learning outcomes of the test-like event.

The process of receiving feedback can be promoted and inhibited by various variables. For example, the evaluation phase (stage 4) is limited when students are provided with corrective feedback only (Hattie & Timperley, 2007). Furthermore, a positive attitude towards the test-like event is associated with an increase in feedback study time (stage 4) (Van der Kleij et al., 2012).

The feedback intervention is essential in formative assessment. The information included in feedback interventions can vary from information on whether or not an answer is correct to various types of remedial information, such as a worked example or references to further instructional materials. Feedback in formative assessment aims at stimulating subsequent

actions in the learning process, such as clarification of misconceptions and misunderstandings or fostering further study and practice, with the aim of improved performance.

To learn more about the relation between CBFA and learning, the relation between feedback and learning needs to be explored. Feedback interventions can have different functions which affect the learning process at various levels (Butler & Winne, 1995; Narciss, 2008). A distinction is made between cognitive, metacognitive and motivational functions. When feedback provides information about the nature of a performance gap and strategies to resolve it, it serves a cognitive or a meta cognitive function. Depending on the errors reflected in students' incorrect answers, this information can be used to confirm, add to, overwrite, tune, or restructure knowledge about tasks and strategies (Butler & Winne, 1995). Feedback can also lead to higher motivation and more efficient task strategies when, for example, it is used to reduce the cognitive load of a learner while faced with complex problem-solving tasks. In this case feedback serves a motivational function (Narciss, 2008). Feedback functions typically come into play simultaneously.

A considerable body of research on the effectiveness of feedback interventions shows that automated feedback can result in increased performance (Van der Kleij et al., 2011). However, the effects of feedback interventions vary greatly (Bangert-Drowns et al., 1991; Black & Wiliam, 1998; Hattie & Timperley, 2007; Kluger & DeNisi, 1996). The relation between feedback and learning is mediated and moderated by numerous variables (Bangert-Drowns et al., 1991; Van der Kleij et al., 2011). Type of feedback is an example of an instrumental variable, which has shown to influence effectiveness of feedback (Shute, 2008; Van der Kleij et al., 2011). A distinction is often made between the following types of feedback: 1) knowledge of results (KR), which merely informs whether or not an answer is correct, 2) knowledge of correct response (KCR), where the correct response is provided, and 3) elaborated feedback (EF), where information other than KR and KCR is provided (Shute, 2008). Research shows that different types of feedback are differentially effective. Overall, EF has shown to be more effective in realizing performance improvement than KCR and KR.

Instrumental variables such as type of feedback only partly influence the complex relation between feedback interventions and learning. The relation between feedback and learning is strongly influenced by student characteristics. Students must need and feel the need to receive feedback to invest effort in processing feedback (Bangert-Drowns et al., 1991). When feedback signals a performance gap, it cannot be assumed that learners will increase their effort to attempt to improve performance, for example to reach a certain standard (Kluger & DeNisi, 1996; Narciss, 2008). Students can also choose to decrease their effort in reaching a certain standard. This is particularly common when a students' success expectancy is low. Another strategy is lowering the standard, rather than abandoning it altogether. Students can also choose to deny that the performance gap exists. For feedback interventions to be

effective, it is conditional that students invest effort in processing feedback to improve performance.

## 5.3 Methodology

In order to determine the implications of previous research for the design of feedback interventions in CBFA, an inventory was made of instrumental and student variables substantially affecting the effectiveness of feedback interventions. A considerable body of research exists on the effectiveness of formative assessment and feedback interventions on learning. This body of research has been reviewed thoroughly by several researchers (e.g. Bangert-Drowns et al., 1991; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008; Van der Kleij et al., 2012). Seminal reviews have been selected and studied for the purpose of this study. Reviews examining effectiveness of non-automated feedback interventions on learning were also selected, as these were considered informative for this study.

Scopus and the Web of Science were searched for seminal reviews discussing variables influencing the effectiveness of formative assessment and feedback on learning. The completeness of the results was checked by consulting an expert on feedback research. This consult led to the inclusion of an unpublished meta-analysis on effects of feedback in computer-based learning environments on students' learning outcomes (Van der Kleij et al., 2013). In addition, citations to the three most recent reviews, were checked through Scopus and Web of Science for relevant supplements. The search was confined to reviews written in English and published in academic peer-reviewed journals with performance improvement as dependent variable.

Initially, fourteen reviews were selected and further appraised. Further appraisal included a check on whether the reviews specified findings on instrumental and student variables substantially affecting the effectiveness of formative assessment or feedback interventions in instructional or test-like events. The initial selection included two reviews focussing on formative assessment and learning (Black & Wiliam, 1998; Nicol & McFarlane-Dick, 2006). However, after further appraisal these reviews appeared to be too generic for the purpose of this study. The reviews discussed the importance of formative assessment and characteristics of effective feedback practice in general, without specifying variables mediating and moderating the relation between formative assessment, feedback and learning. The findings presented in these reviews were too generic for the purpose of this study or could not be translated to implications for the design of feedback interventions in CBFA. As a result, the reviews were discarded.

The five-stage model of Bangert-Drowns, Kulik, Kulik, and Morgan (1991) was used to position the variables encountered in the reviews in an active process of receiving mediated intentional feedback in a test-like event, such as a CBFA (see Figure 5.1).

## 5.4 Results

### *An overview of variables influencing effectiveness of CBFA*

Table 5.1 presents an overview of the seminal reviews selected for the purpose of this study. The final selection of reviews all focused on the relation between feedback and learning. Three reviews focused on computer-based environments (Narciss, 2008; Van der Kleij et al. 2011; Van der Kleij et al. 2013). Five reviews proposed a model for feedback provision (Bangert-Drowns et al, 1991; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Kulhavy & Stock, 1989; Narciss, 2008).

**Table 5.1:** an overview of reviews selected for the purpose of this study

| First author (publication date) | Focus on … | Focus on computer-based environment | Dependent variable(s) |
|---|---|---|---|
| Kulhavy & Stock (1989) | Feedback in written instruction | No | Feedback study time and correctness of post-test responses |
| Bangert-Drowns et al (1991) | Feedback in test-like events | No | Correction of erroneous response |
| Kluger & DeNisi (1996) | Feedback interventions | No | Performance |
| Mory (2004) | Feedback | No | Learning outcomes |
| Hattie & Timperley (2007) | Levels of feedback | No | Reducing the discrepancy between current and desired understanding |
| Narciss (2008) | Informative tutoring feedback | Yes | Efficiency of external feedback |
| Shute (2008) | Different types of formative feedback | No | Skill revision or improved understanding |
| Van der Kleij et al (2011) | Methods for providing written feedback | Yes | Learning outcomes |
| Van der Kleij et al (2013) | Methods for providing written feedback | Yes | Learning outcomes |

Table 5.2 presents an overview of student and instrumental characteristics substantially influencing the effectiveness of CBFA, positioned within the five stages of Bangert-Drowns et al (1991) model of receiving feedback. Next, the implications of the influence of these variables for the design of feedback interventions in CBFA have been discussed. Subsequently, a decision-making framework is proposed for the design of feedback interventions in CBFA that aim at maintaining or increasing student effort and, consequently, foster performance improvement.

**Table 5.2**: An overview of student and instrumental variables influencing the effectiveness of CBFA

| | Process of receiving feedback in a test-like event | Student characteristics | Instrumental characteristics |
|---|---|---|---|
| Stage 1 | Learners initial state | *Motivational beliefs*<br>*Prior Knowledge and skills* | |
| *Stimulus* | *Test-like event* | *Perceived task difficulty* | *Intended learning outcomes*<br>*Task complexity* |
| Stage 2 | Address relevant prior knowledge and skills | *Effort* | |
| Stage 3 | Construct a response | *Response certitude*<br>*Effort* | |
| *Stimulus* | *Feedback Intervention* | | *Type of feedback*<br>*Timing of KCR feedback* |
| Stage 4 | Evaluation of results by processing feedback | *Correctness of response(s)*<br>*Effort* | |
| Stage 5 | Adjust initial state | *Effort* | |

## *Student characteristics affecting the effectiveness of CBFA*

### *Prior knowledge and skills and perceived task difficulty*

Research shows that instructional or test-like events that are experienced or perceived as (too) difficult, can cause student motivation to diminish and, consequently, lead to a decrease in effort invested in the task (Sadler, 1989; Veldkamp et al., 2011). Easy activities, on the other hand, might lead students to perceive a small gap between current and desired levels of performance that is not worth any additional effort, such as studying feedback. Correspondingly, Schunk, Pintrich, and Meece (2008) argue that students tend to be most motivated when they perceive a moderate level of achievement.

Based on the previous, designers of CBFAs are recommended to determine students' levels of prior knowledge and skills or initial ability in order to provide students with a test-like event which results in a moderate achievement. When, for example, a moderate achievement is constituted as a percentage of correct responses of approximately fifty or sixty, it is recommended to present a student with items for which he or she has an average probability of 0.5 to 0.6 of answering them correct. As student levels of prior knowledge and skills tend to vary designers should consider using mechanisms for adaptive selection of items in CBFA. Mechanisms used for computer adaptive testing (CAT) support automated matching of item or task difficulty to student levels of prior knowledge and skills or initial ability (Wauters, Desmet, & Van den Noortgate, 2010).

In addition, Shute (2008) substantiates a focus on directive feedback (providing corrective information) for learners with low levels of prior knowledge and skills and a shift towards facilitative feedback (providing guidance and cues) for learners with high prior levels of knowledge and skills. As such, it is recommended that level of prior knowledge and skills is accounted for in the design of feedback interventions. The decision-making framework proposed for the design of feedback in CBFA accounts for, amongst others, level of prior knowledge and skills.

### *Motivational beliefs and effort*

Previous research suggests that effort invested in, and consequently effectiveness of, CBFA is dependent on students motivational beliefs, such as success expectancy and task-value beliefs (Eccles & Wigfield, 2002). Success expectancy refers to students' judgements about their capabilities to complete certain academic tasks. Students who feel more efficacious about their ability to do well are more likely to be cognitively involved to learn, for example by investing effort in a CBFA (Pintrich, 1999; Timmers et al., 2013). Beliefs about the importance of, interest in, and value of the task are referred to as task-value beliefs. Previous research showed that attention paid to automated written feedback in a CBFA positively related to task-value beliefs (Timmers et al., 2013). The above implies that the effectiveness of CBFA can be fostered by actions that lead to an increase in students motivational beliefs. Previous research has shown that learners who try to attain a certain standard or goal are more likely to increase their effort when feedback signals that their performance falls short of the standard (Kluger & DeNisi, 1996; Narciss, 2008). Correspondingly, research has shown goal setting interventions to increase effort and persistence invested in tasks such as a CBFA. A goal setting intervention presents the learner with an explicit, proximal, challenging goal and includes summary (end of task) feedback that tells the learner how closely they have approximated that goal (Kluger & DeNisi, 1996).

Careful design of CBFA cannot guarantee an increase in student effort invested in CBFA. Motivational beliefs that increase student effort invested in a CBFA can, however, be supported by careful design of CBFA. Motivational beliefs can be fostered by goal-setting interventions and by including an introduction of the task explaining the value of the task in general and the alignment with the summative assessment (Bennett, 2011). Furthermore, as discussed in the previous paragraph, student effort invested in a CBFA can be maintained or increased by providing students with a test-like event that leads to a moderate achievement.

### *Correctness of response and response certitude*

Kulhavy and Stock (1989) discuss the limitations of categorizing a response as either correct or incorrect. A learner may provide a correct response, without being sure about the

correctness of the response or vice versa. Hence, they introduced the concept of response certitude. The relation between response certitude and feedback study time was studied by various researchers (Mory, 2004). Results showed that feedback study time is influenced by the discrepancy between correctness of the response and response certitude. Correct responses on items for which students were confident that they responded correctly (low discrepancy) yield the shortest feedback study time. Incorrect responses on items for which students were confident that they responded correctly (high discrepancy) yield the longest feedback study time. In addition, Bennett (2011) discusses the limitations of categorizing a response as incorrect. An incorrect response might, among other things, be '*a slip – that is, a careless procedural mistake; or a misconception, some persistent conceptual or procedural confusion (or naive view); or a lack of understanding in the form of a missing bit of conceptual or procedural knowledge, without any persistent misconception* (p. 17)'. Bennett argues that each of these causes implies a different instructional action or, in the context of CBFA, error-specific feedback. An advantage of error specific feedback is that it helps regulate cognitive load (Narciss, 2008). On the other hand, determining the error underlying an incorrect response is rather complicated. Shute (2008) refers to research comparing error-specific feedback to reteaching content, which showed reteaching content was found to be as effective as feedback based on more expensive error analyses. Error-specific feedback requires the development of a detailed model underlying the automated feedback. A less detailed model is required when selecting the strategy of reteaching content. Further research is needed to compare effects of error-specific feedback and reteaching content, preferably in contexts of lower-level learning (LLL) as well as higher-level learning (HLL).

As indications of correctness and incorrectness of responses are limited, it is recommended to provide students with KCR and EF after incorrect as well as after correct responses. Furthermore, to regulate cognitive load, designers of CBFAs are recommended to design feedback interventions that present feedback in manageable pieces, for example, by using sequential presentation of feedback (e.g. informative tutoring feedback) or providing feedback with hyperlinks to extra information (e.g. definitions). In the latter case, a student can control the amount of feedback information and adapt the feedback information to his or her own needs.

## *Instrumental characteristics affecting the effectiveness of CBFA*

### *Types of feedback*

Reviews used for this study all discuss types of feedback in one way or another. Here, type of feedback refers to content-related classification of feedback components. A first distinction can be made between verification (KR) and informational components (KCR and EF) of

feedback interventions. Informational components presented as EF have been further categorized by Shute (2008) as well as Narciss (2008). In the context of this study, the categorization of Shute has been adopted as it is arrayed by complexity and explicitly refers to informative tutoring feedback, which Narciss focuses on. Shute (2008) distinguishes between the following six types of EF. Table 5.3 presents an overview of the six types of EF distinguished by Shute (2008, pp. 160).

In feedback research, feedback is often classified as KR, KCR, EF, or a combination of these three categories. Claims about the effectiveness of different types of feedback are generalized accordingly. When learners are merely told whether an answer is correct or incorrect (KR), feedback has hardly any effect on performance (Bangert-Drowns et al., 1991; Van der Kleij et al., 2011). KCR has shown a moderate effect on performance. Overall, EF has shown to be most effective in realizing performance improvement. (Jaehnig & Miller, 2007; Narciss, 2008; Van der Kleij et al., 2011).

However, as Shute substantiated: "*there is no "best" type of formative feedback for all learners and learning outcomes* (p. 183)." Further claims about the effectiveness of different types of feedback need to be nuanced by taking into account variables influencing the relation between feedback and learning, such as student level of prior knowledge and skills and level of intended learning outcomes.

**Table 5.3**: Types of elaborated feedback distinguished by Shute.

| Type of elaborated feedback | Description |
|---|---|
| Attribute isolation | Feedback that presents information addressing central attributes of the target concept or skill being studied. |
| Topic contingent | Feedback providing the learner with information relating to the target topic currently being studied. May entail simply reteaching material. |
| Response contingent | Feedback that focuses on the learner's specific response. It may describe why the incorrect answer is wrong and why the correct answer is correct. This does not use formal error analysis. |
| Hints, cues, prompts | Feedback guiding the learner in the right direction, e.g., strategic hint on what to do next or a worked example or demonstration. Avoids explicitly presenting the correct answer. |
| Bugs, misconceptions | Feedback requiring error analysis and diagnosis. It provides information about the learner's specific errors or misconceptions (e.g., what is wrong and why) |
| Informative tutoring | The most elaborated feedback (from Narciss & Huth, 2004), this presents verification feedback, error flagging, and strategic hints on how to proceed. The correct answer is usually not provided. |

### *Intended learning outcomes and task complexity*

A CBFA can address different types of cognitive learning outcomes (Krathwohl, 2002). A distinction is made between lower-level learning outcomes (LLL), e.g. knowledge and comprehension, and higher-level learning outcomes (HLL), e.g. application, analysis, synthesis and evaluation. It is assumed that mastery of the less complex LLL is prerequisite to mastery of the more complex HLL (Krathwohl, 2002). Smith and Ragan (2005) claim different types of feedback to be differentially effective for LLL and HLL. Their claim is supported by results of a systematic review on methods for providing written feedback through CBFA (Van der Kleij et al., 2011). Here, feedback interventions have been categorized as KR, KCR, and EF or a combination of these types of feedback. The results showed that KCR can be beneficial for LLL. EF interventions in CBFA proved to be potentially beneficial for LLL as well as HLL (Van der Kleij et al., 2011). In the context of HLL (e.g. problem solving) Smith and Ragan (2005) advise hints or guiding questions, without KCR, as initial feedback after an incorrect response. The initial feedback may include data on which information has been used or misused, the appropriateness of selected solutions, or whether individual phases of the solution have been correctly performed. Next, students can apply the initial feedback as they are provided with another opportunity to respond to a particular item of a CBFA. KCR or a worked solution is provided after a second or a third try. This type of feedback is referred to as informative tutoring feedback (Narciss, 2008).

In short, the level of intended learning outcomes needs to be accounted for in the design of effective feedback interventions. The decision-making framework proposed, further on in Figure 5.2, accounts for level of intended learning outcomes.

### *Timing of feedback*

Various researchers have studied effects of immediate and delayed feedback. In general, research on effects of feedback timing shows that students can benefit from immediate as well as delayed feedback (Bangert-Drowns et al., 1991; Van der Kleij et al., 2011). However, results lack clarity as the terms immediate and delayed have not been operationalized consistently (Mory, 2004). For example, feedback provided directly after a response and directly after completion of a task have both been categorized as immediate. Here, in the context of CBFA, immediate feedback refers to feedback directly following a student response (Van der Kleij et al., 2011). Feedback provided after (several or) all items of a CBFA is referred to as delayed feedback. A meta-analysis by Van der Kleij et al (2013) showed effects of immediate feedback on LLL to be higher than effects of delayed feedback. The effect sizes of immediate and delayed feedback on HLL did differ significantly. In addition, research by Van der Kleij et al (2012) showed feedback study time in a CBFA to be higher for immediate feedback than delayed feedback. Furthermore, they also found students to prefer immediate feedback over delayed feedback.

Within the context of immediate feedback, timing of KCR is an issue that needs consideration. CBFAs can include one or more attempts to provide a correct response, single try feedback (STF) and multi try feedback (MTF) respectively. Informative tutoring feedback is an example of a MTF strategy providing KR and EF that guide the learner toward successful task completion without immediately offering KCR (Narciss, 2008). Facilitative feedback, such as MTF, is suggested to be specifically beneficial for HLL and for students with high levels of prior knowledge and skills (Shute, 2008). Informative tutoring feedback requires learners to analyse their erroneous responses and to identify error correction steps. MTF hardly seems beneficial for LLL when students lack certain declarative knowledge (Clariana & Koul, 2006).

Although further research is needed to find out how many attempts in MTF are effective under which conditions (Narciss, 2008), it can already be recommended to adapt the number of attempts in MTF (or timing of KCR) to the level of prior knowledge and skills. Students with low levels of prior knowledge and skills are expected to get demotivated sooner, while presented with extra attempts to provide a correct response than students with high levels of prior knowledge and skills.

Based on previous research on feedback-study time and student preferences, it is recommended to provide students with feedback immediately after they have constructed a response. Additionally, it is recommended to adapt try-again feedback strategies to intended level of learning outcomes as well as level of prior knowledge and skills. The proposed decision-making framework, further on in Figure 5.2, explicitly accounts for timing of KCR feedback and, consequently, number of attempts in MTF.

## 5.5 Decision-making framework for the design of feedback interventions in CBFA

Research shows that the relation between CBFA and learning is influenced by students prior knowledge and skills, motivational beliefs, (perceived) task difficulty, response certitude, intended learning outcomes, type of feedback, timing of (KCR) feedback, and correctness of response. The decision-making framework for feedback in CBFA presented in Figure 5.2, accounts for the enumerated instrumental and student characteristics. The framework focuses on fostering student effort invested in a CBFA for the purpose of increasing performance.

CBFAs can be used at various stages of learning trajectories aimed at mastering certain LLL or HLL with the aim of creating awareness about the nature of performance gaps and ways to address these. The level of prior knowledge and skills typically increases throughout such learning trajectories. The framework presented in Figure 5.2 proposes different feedback interventions for the following four situations: 1) Use of CBFAs focusing on LLL by students with low levels of prior knowledge, 2) Use of CBFAs focusing on LLL by students with high levels of prior knowledge, 3) Use of CBFAs focusing on HLL by students with low levels of

prior knowledge and skills, and 4) Use of CBFAs focusing on HLL by students with high levels of prior knowledge and skills. Although the framework proposes immediate feedback for all four situation, the feedback interventions vary with regard to type of feedback, timing of KCR feedback, and correctness of response.

The EF components in the framework differ for LLL and HLL. The EF components are based on Shute's categorization of EF which is arrayed by complexity. This corresponds with the difference in complexity of tasks that focus on LLL and HLL. The framework proposes MTF or delayed KCR feedback for incorrect responses, with the exception of CBFAs focusing on LLL used by students with low levels of prior knowledge. The framework proposes EF after incorrect as well as after correct responses (for MTF feedback the timing of KCR is delayed) to account for response certitude and the limits of observing correctness of response. In addition, the framework proposes a combination of KR, EF and (delayed) KCR. This is in line with the findings of Bangert-Drowns et al that effective types of feedback both signal the correctness or incorrectness of an answer (KR) and provide information to confirm, add to, overwrite, tune, or restructure knowledge about tasks and strategies (KCR and EF).

It should be noted that, contemporary assessment software often does not support the design of MTF, such as informative tutoring feedback. In such cases, designers of feedback in CBFAs are recommended to immediately provide KCR together with KR and EF components that correspond with the level of learning outcomes.

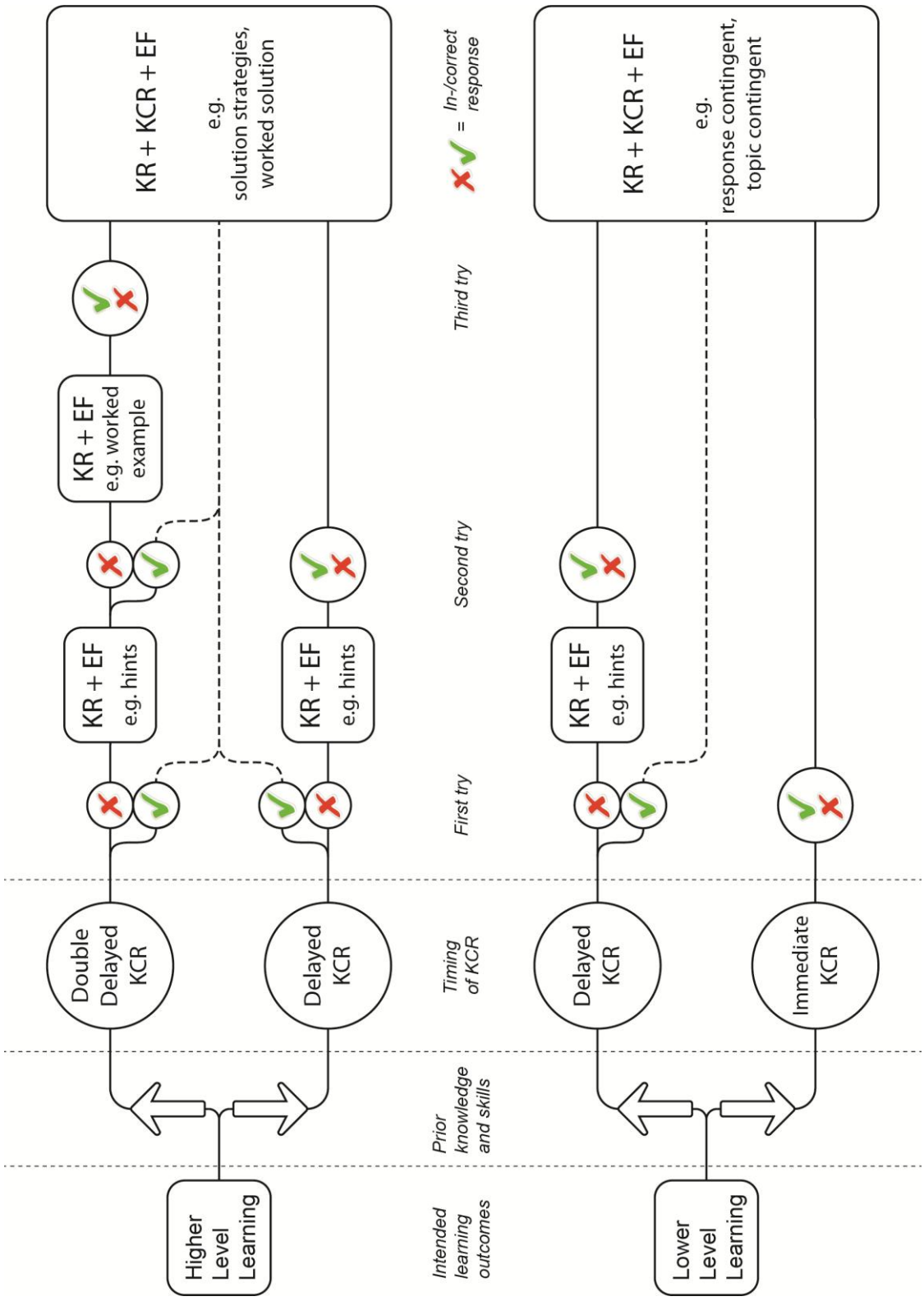


**Figure 5.2**: Decision-making framework for the design of feedback interventions in a CBFA with a cognitive function

## 5.6 Conclusion and discussion

Although careful design of CBFA does not guarantee student effort invested in a CBFA, it can, however, foster student effort invested in a CBFA. The purpose of this study was to 1) draw an overview of student and instrumental characteristics substantially influencing the relation between CBFA and learning based on previous research, and 2) to determine the implications for the design of feedback interventions in CBFA. A distinction was made between student and instrumental characteristics influencing effectiveness of CBFA. The following student characteristics have been discussed: Student prior knowledge and skills, motivational beliefs, perceived task difficulty, correctness of response, and response certitude. Results of previous research on these student characteristics have been translated to recommendations for the design of CBFA. The following instrumental characteristics have been discussed: Intended level of learning outcomes, task difficulty, and timing and type of feedback. Results of previous research on these instrumental characteristics have also been translated to recommendations for the design of CBFA. Subsequently a decision-making framework has been proposed for feedback interventions in CBFA that foster student effort invested in a CBFA aimed at performance improvement. The framework explicitly accounts for intended level of learning outcomes, student prior knowledge and skills, correctness of response, and timing and type of feedback. The framework implicitly accounts for motivational beliefs, perceived task difficulty, and response certitude.

Theoretically based guidelines for the design of effective feedback interventions have been developed before (Narciss & Huth, 2004; Shute, 2008). Shute (2008) proposed guidelines for generating formative feedback. Formative feedback refers to information communicated to the learner that is intended to modify his or her thinking or behaviour for the purpose of improving learning. Narciss and Huth (2004) proposed guidelines which focus on a particular type of feedback namely, informative tutoring feedback. Informative tutoring feedback includes multiple-try feedback strategies providing EF that guide the learner toward successful task completion without immediately offering KCR. This study built on the work of Narciss and Huth and Shute and recent research on automated feedback(e.g. Van der Kleij et al., 2013).

Shute (2008) refers to a decision-making framework for the provision of feedback in computer-based instruction (CBI) proposed by Mason and Bruning (2001). Their framework is based on literature regarding the use of feedback in CBI, excluding, remarkably enough, literature focusing on use of computer-based feedback in assessment. This study proposes a theory-based decision-making framework for feedback in CBFAs. In this study, CBFA is viewed as a manifestation of CBI (Aleven et al., 2003). As such, previous research on effectiveness of feedback on learning in instructional as well as test-like events have been considered relevant for the purpose of this study. The decision-framework for CBFA and the framework Mason and Buning (2001) proposed for CBI differ. The framework proposed by Mason and Bruning starts with a distinction between low and high student achievement,

while the framework for CBFA starts with distinguishing between levels of learning outcomes. Mason and Bruning (2001) concluded immediate feedback to be superior to delayed feedback for students with low achievement levels in the context of either simple (lower level) or complex (higher level) tasks. Delayed feedback is suggested for students with high achievement levels confronted with complex tasks. The framework for feedback in CBFA proposes immediate feedback with a variation in timing of KCR feedback for MTF. The framework for feedback in CBFA accounts for student achievement in terms of correctness of response. Furthermore, the framework for feedback in CBFA distinguishes between a wider variety of EF components. When viewing CBFA as a specific form of CBI, the decision-making framework for feedback in CBFA might also be considered useful for the design of feedback interventions in CBI.

This study did not elaborate on the medium for the presentation of feedback. Previous research on effectiveness of feedback interventions have focused mostly on feedback presented in written form, while feedback can also be presented via audio, images, video or a combination of various media. Further research is needed to explore differences in effectiveness of feedback presented via various media to various populations in various conditions.

The strategy used to select studies relevant for the purpose of this study is limited. Every review deals with the impossibility of including all relevant studies. This study included reviews of research on the relation between formative assessment, feedback and learning only. A systematic review of reviews as well as primary studies is likely to result in a more complete overview of variables affecting the relation between CBFA and learning. For example, task length and duration of the instructional or test-like event was not discussed in the reviews studied. The reviews focused on the relation between feedback and learning, leaving task characteristics out of scope. Research shows that student willingness to invest effort in a formative assessment or low-stakes tests is limited (Wolf et al., 1995). Accordingly, research shows test length to increase rapid-guessing (Wise, 2006) and negatively affecting feedback study time (Timmers & Veldkamp, 2011). Hence, frequent use of shorter CBFA is likely to be more effective than less frequent use of longer CBFAs.

A variable which has been discussed in previous research by various researchers, but has not been addressed in this study so far is presearch availability (Bangert-Drowns et al., 1991; Shute, 2008). Presearch availability refers to the availability of KCR before a student constructs a response to an item. When presearch availability is uncontrolled, students can simply copy KCR without paying any effort to construct a response. Presearch availability has shown to mediate effects of feedback. Shute (2008), among others, recommends to avoid showing answers before students have tried to solve a problem on their own or, in other words, to control for presearch availability. However, in this time and age learners are increasingly online and, therefore, able to search the Internet for correct responses. As such, controlling for presearch availability becomes difficult to realize. Further research is needed

to examine the limitations, but also the opportunities, of the availability of Internet to search for correct responses in relation to CBFA.

Heterogeneous findings confirm the need for further research on the relation between feedback interventions in instructional and test-like events and learning. However, the existing body of research can already, and therefore should, be used to enhance contemporary educational practice. A contemporary trend in learning societies is the availability and increased use of serious games and educational apps at school as well as at home. These computer-based learning environments often include information that is provided to learners to modify their thinking or behaviour for the purpose of improving learning, or what we refer to as feedback interventions. The design of feedback interventions is preferably based on well-founded design principles. The framework proposed in this study can be used as starting-point for the design of feedback interventions or the evaluation of existing interventions.

# Chapter 6 - Developing scales for information-seeking behaviour


## Abstract

The main purpose of this chapter is to describe the development of an instrument designed to measure information-seeking behaviour of undergraduate students during study assignments. Previous research was examined to determine characteristic information-seeking behaviour of undergraduates and draw up questions for the questionnaire. The next phase consisted of an actual field test, followed by analyses of the structure and reliability of the instrument. Four scales were found within a 46-item survey on information-seeking behaviour: a 10-item scale for applying search strategies ($\alpha$ = 0.68), a 14-item scale for evaluating information ($\alpha$ = 0.74), a 6-item scale for referring to information ($\alpha$ = 0.81) and a 12-item scale for regulation activities when seeking information ($\alpha$ = 0.75). The four scales for information-seeking behaviour can be used to monitor and evaluate this behaviour of students in higher education.


## 6.1 Introduction

Students in higher education are expected to be information literate. However, many students seem to rely on one-keyword searches to find the answers to all their questions at the top of the search result pages, using Google as an oracle of modern times. Various studies and reports stress the importance of information literacy as a key for success in information societies and indicate the consequences of a lack of information literacy (ALA, 1989; AWT, 2005; Drucker, 1993; Van Deursen & Van Dijk, 2008b).

This study focuses on an aspect of information literacy, that is, information-seeking behaviour. Little is known about information-seeking behaviour of Dutch undergraduates. To determine the need for support of Dutch undergraduates in acquiring desired information-seeking behaviour it is necessary to be informed about both current undergraduate behaviour and desired undergraduate behaviour. This research aims to develop a reliable and valid measurement instrument to study information-seeking behaviour of undergraduates. The instrument should be cost-efficient and usable for undergraduates with a wide variety of backgrounds. Data collected with the instrument can be used to decide whether learning support is necessary and if so, which support is needed. The instrument might also be used to monitor student progress over time and to investigate the effects of interventions.


Adapted from: Timmers, C. F., & Glas, C. A. W. (2010). Developing scales for information seeking behaviour. *Journal of Documentation, 66*(1), 46-69. doi: 10.1108/00220411011016362

The next paragraph describes the conceptual framework of this study. Subsequently, findings of previous research on information-seeking behaviour of undergraduates are described, followed by the research methods used to develop and evaluate the measurement instrument. The chapter concludes with a summary of the findings and a discussion.

## 6.2 Information literacy and information-seeking behaviour

Information literacy is a broad and multidimensional concept. This literacy is based on a wide variety of knowledge, skills and actual behaviour relating, amongst others, localisation, evaluation and effective use of information. Furthermore, information literacy encompasses personal, social and ethical dimensions of interacting with information (Boon, Johnston, & Webber, 2007).

Information literacy is a concept for which there is no satisfactory translation in Dutch. Instead a term that translates to information skills is mostly used. The term information problem solving skills is used in more recent Dutch research (Brand-Gruwel, Wopereis, & Vermetten, 2005). The term information skills is used to refer to different sets of skills. These can be summarized by the following three conceptualizations (Boekhorst, 2003). *The ICT conceptualization* refers to skills needed to use ICT to retrieve and disseminate information. These skills are called operational and formal skills by van Deursen and van Dijk (2008a). *The information (re)sources conceptualization* refers to skills needed to find and use information independently or with the aid of intermediaries. *The information process conceptualization* refers to the process of recognizing the need for information, retrieving, evaluating and using information to acquire or extend knowledge. Initially the term information skills was mostly used to refer to the ICT skills. Accordingly, the focus of policy makers was on providing citizens access to computers and the Internet. Van Deursen and van Dijk (2008b) have tested operational, formal and information skills of 109 subjects by giving them several assignments. They found that the 'digital generation' (age 18-29) performed significantly better on the operational and formal skills assignments than older age groups. However, this was not the case for the information skills assignments. They also found that information skills did not grow with years of Internet experience and amount of time spent online weekly. This leads to the assumption that being on and connected to the Internet does not guarantee the development of information skills. These findings support the shift in focus of policy makers towards the information process concept. This concept is the most comprehensive as it includes both ICT and the information (re)sources concept. The three concepts mentioned above correspond with five out of the seven faces of information literacy described by Bruce (1997). The explanation of the information process conception given by Boekhorst (2003) also corresponds with the knowledge construction and the knowledge extension conceptions, due to the addition of using information "to acquire or

extend knowledge". The information process conception and the wisdom conception were not addressed.

Several frameworks have been developed that describe information literacy within the context of higher education. Boon, Johnston and Webber (2007) compared three frameworks developed by the Society of College, National and University Libraries (SCONUL, 1999), the Association of College and Research Libraries (ACRL, 2000) and Australian and New Zealand Institute for Information Literacy (ANZIIL) (Bundy, 2004). They found that these frameworks share the following elements:

- the identification of an information need;
- access, acquisition, evaluation and manipulation of information; and
- ethical issues of information use.

The framework used within this study is the Association of College and Research Libraries' "Information Literacy Competency Standards for Higher Education" (ACRL, 2000). This framework has been adopted and translated in Dutch in 2005 by a national committee (*Landelijk Overleg Omgaan met Wetenschappelijke Informatie,* LOOWI). The main difference between the ACRL-standards and the other frameworks is that the ACRL-standards do not cover the creation of new knowledge as an element of information literacy and do not relate information literacy to lifelong learning.

Information-seeking behaviour is an aspect of information literacy. Boon et al (2007, p. 207) describe the relation between these two concepts as: "one step towards becoming information literate is to acquire an appropriate information-seeking behaviour". Frameworks describing information literacy in higher education, such as the ACRL-standards, can be used as a source to determine information-seeking behaviour desired of undergraduates. Within behavioural science a distinction is made between physiological, cognitive and affective behaviour. Wilson (1999) has studied models that are used in information behaviour research. He suggests a nested model for the research area of information behaviour. This model is presented in Figure 6.1. Research on information behaviour relates to the totality of human behaviour in relation to sources and channels of information, including both active and passive information-seeking, and information use. A sub-set of this field of investigation is research on information-seeking behaviour. This area is concerned with the purposive seeking of information to solve a certain information problem. Wilson (1999) discusses the following information-seeking models: Dervin's (1983) sense-making theory; the behavioural model of information seeking strategies (Ellis, 1989; Ellis, Cox, & Hall, 1993); Kulthau's (1991) model of the stages of information-seeking behaviour; and his own model (Wilson, 1999). Wilsons model describes four modes of information-seeking behaviour: passive attention, passive search, active search and on-going search. The models of Ellis and Kulthau focus on the mode of active search, as does this

study. The features and stages distinguished by Ellis and Kulthau, respectively, can all be related to one or more of the ACRL-standards and underlying performance indicators.

The information-seeking models vary in their emphasis on physiological, cognitive and affective aspects of information-seeking behaviour. This study focuses on the physiological and cognitive behaviour that can be observed by the undergraduates themselves. Instruments for affective information (seeking-) behaviour already exist (Kuhlthau, 2004; Kurbanoglu, Akkoyunlu, & Umay, 2006; Monoi, O'Hanlon, & Diaz, 2005; Norman & Skinner, 2006; Van Kampen, 2004).

Research on information searching behaviour is seen as a sub-set of information-seeking behaviour. It focuses on the interactions between information user and computer-based information systems and thereby corresponds with to the previously mentioned ICT concept of information skills (Boekhorst, 2003) and operational and formal skills (Van Deursen & Van Dijk, 2008b).

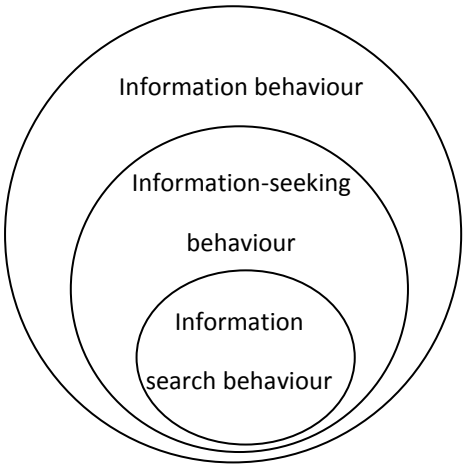


**Figure 6.1:** Wilson's nested model of the information-seeking and information searching research areas (1999).

The focus on actual behaviour was inspired by Miller's pyramid of competence, a model frequently used in higher education to classify tests (Van der Vleuten, Driessen, & Van Tartwijk, 2006). Figure 6.2 represents Miller's pyramid of competence.

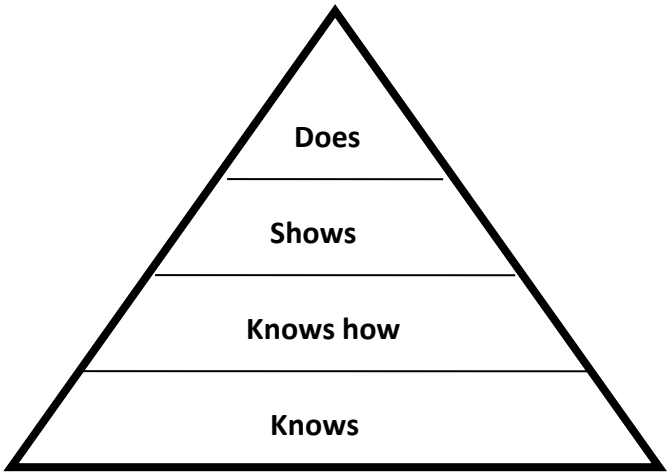


**Figure 6.2.** Miller's pyramid of competence (1990)

The bottom layer refers to declarative knowledge (*knows*) required for competence. The layer above this refers to procedural knowledge (*knows how*). This layer signifies the step towards the next layer, which refers to actual skills (*shows*). The top of the pyramid refers to the actual behaviour of a competent person (*does*). This model implies knowledge and skills underlying actual competent behaviour. Previous research indicates the tendency of undergraduates to overestimate their own knowledge and skills supporting information literacy (Gross & Latham, 2007; Ivanitskaya et al., 2006; Maughan, 2001). Therefore this study focuses on the top layer of Miller's pyramid of competence.

Information-seeking behaviour relates to activities, which can be applied in any context, whether seeking information to solve economic, health or technical problems. Activities related to information-seeking behaviour distinguished within this study have been derived from the ACRL-standards (see Table 6.1). All but one of the ACRL-standards have been used to define the research components. The aim of this research is to develop an instrument to measure general information-seeking behaviour of undergraduates while seeking information for study assignments. Behaviour related to standard four of the ACRL-standards is indissolubly related to specific assignments. Therefore standard four has been discarded.

Furthermore, one component was added to the research components derived from the ACRL-standards. The ACRL-standards refer to regulation activities in the performance criteria underlying the various standards. Since regulation activities play a crucial role in competent information-seeking behaviour, regulation activities have been added as a separate component (Edwards, 2006). Brand-Gruwel et al (2005) have analysed problem solving skills by identifying and describing main and sub skills. They describe regulation activities as a series of activities that take place continuously and consist of orientation towards a task or

problem, steering the process, monitoring the process, and testing the product. Table 6.1 provides an overview of the research components in relation to the ACRL-standards.

However, it should be stressed that the scales developed within this study cover elements shared by all three frameworks outlined in the previous section on information literacy and can therefore also be used within the context of the other frameworks.

**Tabel 6.1:** ACRL-Standards and components studied within the research

| ACRL-standards: The information literate student… | components |
|---|---|
| 1. Determines the extent of information needed | a) Defining information problems |
| 2. Accesses the information needed effectively and efficiently | b) Using sources |
| | c) Applying search strategies |
| 3. Evaluates information and its sources critically | d) Evaluating information |
| 4. Uses information effectively to accomplish a specific purpose | |
| 5. Understands the economic, legal, and social issues surrounding the use of information, and accesses and uses information ethically and legally | e) Referring to information |
| - *Underlying standards 1 to 5* - | f) Regulation activities |

## 6.3 Previous research on information-seeking behaviour

Results of previous research on information-seeking and searching behaviour were studied to determine characteristic information-seeking behaviour of undergraduates. The characteristic information-seeking behaviour encountered was used to develop an instrument to measure this behaviour. Table 6.2 provides an overview of the studies reviewed. The characteristic information-seeking and searching behaviour of undergraduates encountered within the studies are presented in de following sections per research component, as summarized in Table 6.1.

The main terms used to locate the information were information literacy, information problem solving, and information-seeking behaviour. The search engines and databases consulted were Academic Search, Picarta, Scopus, Web of Science and Google Scholar. Various articles were found by checking references of articles found via databases. Articles dated before 1997 were excluded from the review, since current research focuses on generation Y, also known as millennials or echo boomers. Generation Y consists of people that have grown up during the rise of instant communication technologies made possible

through use of the Internet. In the United States, Canada, Japan and certain parts of Europe, these people are usually the people born between 1980 and 1994. From 1997 onwards people from generation Y started their undergraduate studies.

The research methods used in these studies vary considerably. Of a total of eleven studies, five were based on quantitative research techniques and six on qualitative techniques.

**Table 6.2:** Studies on characteristic information behaviour of undergraduate students

| | |
|---|---|
| Aula & Nordhausen (2006) | 20 students from various disciplines (Finland) |
| Brand-Gruwel et al (2005) | 5 PhD-students and 5 psychology freshmen (The Netherlands) |
| Callinan (2005) | 73 first year biology & 23 final year biochemistry (Ireland) |
| Ivanitskaya et al (2006) | 308 students of Midwestern college of health sciences (U.S.) |
| MaKinster et al (2002) | 7 students from a large Midwestern university (U.S.) |
| Metzger et al (2003) | 365 + 436 undergraduates of a West Coast public university (U.S.) |
| Mittermeyer (2005) | 3003 students of 15 Quebec universities (Canada) |
| O'Brien et al (2007) | 249 survey respondents and 9 students for a focus group (U.S.) |
| Pan et al (2007) | 16 undergraduates of various disciplines (U.S.) |
| Twait (2005) | 13 students of various disciplines (U.S.) |
| Whitmire (2002) | 5175 students of 38 4-year institutions (U.S.) |

Besides the studies listed in Table 6.2, two literature reviews were found (Kuiper, Volman, & Terwel, 2005; Walraven et al., 2008). The literature review of Kuiper et al (2005) focuses on information literacy of primary and secondary school children. Their results showed that most children find it difficult to locate relevant information on the Web and to assess the relevance of the information. The literature review of Walraven et al (2008) describes problems which children, teenagers and adults experience when solving information problems using the World Wide Web for searching information. Their results show that students in all age groups experience problems when specifying search terms and evaluating search results and sources. The regulation of the search process is also found to be problematic within all age groups.

Five articles describing measurement instruments were found, but none fitting the theoretical framework of this research. One of the instruments was designed to gather data on information-seeking behaviour. More specific, ten items questioned students on their library use behaviour. This is only a small component of information-seeking behaviour as defined within this study. The other four instruments all covered emotional aspects

influencing cognitive behaviour. Van Kampen (2004) developed and evaluated a multidimensional library anxiety scale. Monoi et al (2005) developed an inventory to assess self-efficacy for online searching skills. Kurbanoglu et al (2006) developed and evaluated an information literacy self-efficacy scale. The instrument of Norman & Skinner (2006) measures comfort at seeking information and also perceived knowledge and skills at finding, evaluating, and applying electronic health information to health problems.

***Defining the information problem***

None of the studies addressed characteristic behaviour for defining the information problem. Defining the information problem is addressed in the literature review by Walraven et al (2008). Based on the lack of problems encountered for this skill in their review, they stated that adults (above the age of 18) do not seem to have any problems relating this skill. Another and probably more realistic explanation could be the lack of focus on skills underlying defining the information problem.

***Sources used***

The overall impression is that undergraduates mainly search the World Wide Web when seeking information. The most popular search engines today are Google and Yahoo. All undergraduates studied by Pan et al (2007) and Aula & Nordhausen (2006) reported the use of Google as their primary search engine. However, the use of sources by undergraduates changes over time. Callinan (2005) studied sources used by first- and final-year biology undergraduates. Only 1% of the first-year respondents used journal articles as opposed to 48% of the final-year respondents. She also found that students do not use the library's electronic databases, the primary reason being lack of awareness of the existence of these databases. Mittermeyer (2005) also states that it cannot be assumed that incoming undergraduates are familiar with, amongst other things, the library catalogue and the scholarly journal.

Other sources studied were teachers, peers, librarians, books and journals. Metzger et al (2003) found that students were likely to consult teachers rather than librarians, and that the people they turn to most frequently are peer students. Remarkably, the undergraduates studied by Metzger et al reported using books significantly more often than the Internet.

The sources that undergraduates turn to, differ for various disciplines (O'Brien & Symons, 2007; Whitmire, 2002). O'Brien et al (2007) studied students of humanities and social science and science students. The students of humanities and social science stated that they often go to the library to locate books and journals and many science students reported that they never engage in this behaviour. Whitmire (2002) found students of 'soft' disciplines

(humanities, business, social sciences and education) to consult the on-line catalogue, librarians and books significantly more often than students from 'hard' disciplines (physical sciences and engineering).

### *Applying search strategies*

For most of the undergraduates MaKinster et al (2002) observed, the first step when searching for information is to enter the keyword for their search in the search engine. Most of the undergraduates studied by Pan et al (2007) "…input two or three terms in search engines" . According to Mittermeyer (2005) and Aula & Nordhausen (2006) undergraduates find it difficult identifying significant terms or keywords. The use of significant terms seems to be related to success, since "successful students tended to use well-composed keyword phrases and often put their keywords in quotes" (MaKinster et al., 2002, p. 161). The twenty students and two researchers studied by Aula & Nordhausen (2006) rarely used Boolean operators and term modifiers. Additionally, in most of the cases the use of these advanced search features was erroneous. Phrase search (quotation marks around the terms) was used by only four of the participants.

Ivanitskaya et al (2006) found that few students use advanced search features or limit their searches in some other way. She found that many students rely on basic searches even though most of them understand that a one-word search is likely to return too many documents. She also found that "few of the students are able to narrow a search by using multiple search categories simultaneously or by employing the Boolean operators" (Ivanitskaya et al., 2006, p. 8). She asked students to perform an advanced search for a book, knowing the book's author (with a very common last name), general topic, and publication date. Few students were able to perform the advanced search. This search is advanced because imprecise book specifications make it hard to find the book without performing a search that takes into account all or nearly all of the available information.

### *Evaluating information*

An increasingly important activity is the evaluation and verification of information, due to the widespread use of the World Wide Web. Metzger et al (2003) studied the evaluation behaviour of undergraduates and found the following:

> "The most frequent verification strategies used by students are checking to see if the website information is current, complete, and comprehensive, considering whether the views represented by the author are facts or opinions, seeking out other sources to validate online information and considering the author's goals or objectives in posting information to the web. Less-often used strategies include checking to see

> who the author is, looking for a stamp of approval or recommendation for the site, making sure contact information is provided by the author, and verifying the Website author's qualification or credentials." (p. 285)

The overall impression derived from the articles studied is that most undergraduates do not evaluate information satisfactorily while selecting information. Most undergraduates only view and click the first few links returned by Google or another general search engine and rarely visit the second page of results (Ivanitskaya et al., 2006; Pan et al., 2007). Furthermore, results of Pan et al (2007) indicated that undergraduates are heavily influenced by the order in which the results were presented and, to a lesser extent, the actual relevance of the abstracts. Twait (2005) suggested that students seek evidence to support their own arguments.

***Referring to information***

Research results of Brand-Gruwel et al (2005) indicated that PhD-students refer to sources more often than undergraduates. Ivanitskaya et al (2006) found students committing plagiarism and being unaware of it. These students did not cite a specific source as they believed it to be appropriate to present other people's ideas as their own when slightly modifying the original wording.

***Regulation activities***

Little information was found on characteristic behaviour related to the regulation of the information-seeking process. Aula & Nordhausen (2006) studied twenty undergraduates searching information for clearly defined tasks. "The less successful searchers … do not seem to systematically modify their next query based on the results they receive. …successful searchers often systematically reformulated queries" (p.1689). MaKinster et al (2002) studied seven undergraduates searching for a web page on Newton's Third Law. Several of the most successful students demonstrated higher levels of self-regulation, such as articulating goals for the task, planning how to best use the time and continually monitoring their progress.

## 6.4 Development of a measurement instrument

The aim of this research was to develop a reliable and valid measurement instrument for information-seeking behaviour. The instrument is a questionnaire existing of various scales for information-seeking behaviour. Oosterveld (1996) distinguishes six different methods for the development of questionnaires. Based on his research he concluded that deductive

approaches are preferable to external and prototyping methods. The usability of deductive methods depends on the state of knowledge concerning the constructs of interest. In this study, a deductive approach could be used to develop the questionnaire due to the presence of substantial knowledge about the content and structure of information-seeking behaviour.

The development process consisted of three different phases. First, previous research on information-seeking behaviour of undergraduates has been studied. The second phase consisted of drawing up questions for a questionnaire using the results of the literature review and evaluating the questionnaire by consulting experts, undergraduates and lecturers. The third phase consisted of an actual field test, followed by analyses of the structure and reliability of the instrument.

Previous research indicates that information behaviour depends on the context in which it is displayed (Agosto, 2002). Therefore, students are asked to report on their information-seeking behaviour when searching information for a study assignment. Characteristics used for items in the questionnaire focused on acts of students when seeking information for an assignment. Furthermore, students have to be able to reflect on these characteristics, without them having to indicate whether they possess certain knowledge or abilities. This is important since undergraduates tend to have an inflated image of their own information literacy (Ivanitskaya et al., 2006; Maughan, 2001). Students are asked to indicate which actions they 'always, often, sometimes and seldom or never' undertake when seeking information for a study assignment, instead of estimating their knowledge and skill related to information literacy. Also the opportunity is provided within the measurement instrument to indicate if a certain concept in the items is not clear.

Initially six components underlying information-seeking behaviour were defined. However, after a qualitative evaluation of an initial version of the measurement instrument by experts, lecturers and undergraduates too few items for *defining information problems* were left to study this component. The component *defining information problems* was therefore discarded from the second version of the measurement instrument. Table 6.3 presents an overview of the number of items per component in the two versions of the measurement instrument.

**Table 6.3:** Overview of the number of items per component

| Research components | Number of items Initial version | Number of items Second version |
|---|---|---|
| a) Defining information problems | 6 | - |
| b) Sources | 20 | 14 |
| c) Applying search strategies | 13 | 13 |
| d) Evaluating information | 20 | 15 |
| e) Referring to information | 6 | 6 |
| f) Regulation activities | (13*) | (13*) |

* The items covering regulation activities also relate to one of the other research components.

The items are four-point Likert-scales, with the following response categories: 3 = always, 2 = often, 1 = sometimes, 0 = rarely or never. A fifth response category was added, based on the presumption that certain terms used would be unfamiliar to the respondent (9 = what's that?).

The initial version of the questionnaire was evaluated by experts, students and lecturers. These evaluation activities focussed on the validity of the items. Three experts, that is, one information specialist and two PhD students, were consulted to evaluate completeness, focus on actions and the items being univocal. Three students and two lecturers have been consulted to evaluate the connection to linguistic usage of the respondent and the items being univocal.

The experts and lecturers reported their feedback on the items both orally and in writing. The students were consulted individually via the talk-through-method. This involved having the respondents verbalizing their thoughts while undertaking a task, in this case while filling in the questionnaire. After filling it in the students were interviewed about the questionnaire. The feedback of the experts, lecturers and students led to shortening the questionnaire, adjusting the order of the items and to reformulating several items. Comments on the items were mainly about use of jargon and about items being too broad and general, which made the answers ambiguous or obvious. All reviewers agreed that the questionnaire was too long. Therefore the number of items was reduced from 65 to 48. Items removed overlapped other items or were unspecific (e.g. when I search for information for an assignment for school: … I surf from one webpage to another; … I consult web pages). The second version of the research instrument covered the following five components *using sources*, *applying search strategies*, *evaluating information*, *referring to information* and *regulation activities*. Table 6.4 presents examples of items for the remaining components. The 48-item questionnaire is added as an Appendix.

**Table 6.4:** Examples of items per component

| | |
|---|---|
| | *When I seek information for a study assignment* |
| *b)* | *…, I use Wikipedia.* |
| *c)* | *…, I examine the number of results my search produces.* |
| *d)* | *…, I examine the results on the subsequent result pages.* |
| *e)* | *…, and write down information in my paper literally, I refer to the source(s) used within the text.* |
| *f)* | *…, and a search produces many results, I refine my search to decrease the number of results.* |

## 6.5 Evaluation of the measurement instrument

### *Validity and reliability*

Allen and Yen (1979) describe various methods to determine validity. In the present study, content validity was addressed by deriving items from previous research results and consulting experts and respondents such as described in the previous paragraph. Construct validity was addressed by conducting a field test and performing several psychometric analyses which are described in this paragraph. To get a global impression of the extent to which the components of information-seeking behaviour could be retrieved from the data of the field test, first an explorative factor analysis with a varimax rotation was performed. After this explorative phase an item response theory (IRT) model was used to further investigate in detail the scalability of the factors found. Also the reliability of the different proposed scales was computed using Cronbach's Alpha.

### *Field test*

The questionnaire was completed by 345 undergraduate students of Saxion University of Applied Sciences (n = 193) and the University of Twente (n = 152). The respondents of Saxion were students of the School of Marketing & International Management (MIM). The respondents of the University of Twente studied Communication Science (CS). The

questionnaires were handed out during lectures. An overview of the respondents is shown in Table 6.5.

**Table 6.5:** Overview of the respondents

| Study | N | Response rate | 1st year | 2nd year | other |
|---|---|---|---|---|---|
| **MIM** | 193 | 49% | 95 | 83 | 15 |
| **CS** | 152 | 73% | 66 | 71 | 15 |

The analyses carried out to develop the instrument focussed on two aspects: the scalability of the items and the reliability of the scales found. The scalability of the items pertains to the factor structure of the items and the interpretation of the scores. Scalability can be assessed using factor analysis or IRT analysis. When scales are found, the next question pertains to the reliability of the scale scores. This topic is related to, but not entirely concurrent with scalability. For instance, Wood (1978) gives an example of perfectly scalable observations with a reliability of zero (i.e., coin tosses). On the other hand, it makes little sense to search for scales that are a priori unreliable. Therefore, an initial reliability analysis was performed first.

***Initial Reliability analysis***

The maximum of Cronbach's Alpha reliability coefficient (0.87) was reached for 46 out of 48 items. The two items lowering the reliability of the information-seeking behaviour scale were discarded from further analyses.

***Factor analysis***

Information-seeking behaviour is a multidimensional construct. To explore the main factors of information-seeking behaviour, a factor analysis was carried out using the method of maximum likelihood. After reconsidering the content of the items related to *using sources*, it was decided that the amount in which the various sources are used does not sufficiently describe the level of information-seeking behaviour. Therefore the items related to *using sources* were discarded from the factor analysis. An overview of these items is presented in Table 6.6.

The factor analysis was done with the remaining 32 items. The response category *what is that* was considered missing data. Factor analysis followed by Varimax rotation indicated the presence of two factors. An overview of the factor loadings of this analysis is presented in Table 6.7. The first column gives item labels related to the classification as given in Table 6.1. Thus item C1 refers to component *applying search strategies*. When an item is also related to the component *regulation activities*, this is indicated with (F). Out of 32 items, 26 items loaded with factor loadings higher than or very close to 0.30 on two factors. In Table 6.7, these factor loadings are shaded (the factor loadings of item C22 and D38, which both are 0.299, were also shaded, because these factor loadings were considered very close to 0.30). Factor 1 contained 20 items with factor loadings higher then or very close to 0.30. These items relate to the components *applying search techniques*, *evaluating information*, and *regulation activities*. These three research components could not be distinguished from one another via factor analysis. Factor 2 corresponded very well with the research component *refer to information*.

**Table 6.6:** Overview of items related to b) using sources

| Item | When I seek information for a study assignment … |
|---|---|
| B 1 | I use studybooks handed by my teacher |
| B 2 | I go to the library |
| B 3 | I consult librarians |
| B 4 | I consult the library catalogue |
| B 5 | I consult databases |
| B 6 | I consult Google |
| B 7 | I consult other search engines |
| B 8 | I consult my teacher |
| B 9 | I consult my fellow students |
| B 10 | I consult newspapers |
| B 11 | I consult books |
| B 12 | I consult journals |
| B 13 | I consult Wikipedia |
| B 14 | I consult metasearch engines |

* A Dutch version of the questionnaire was used to gather the data.

### *Item Response Theory analyses*

The explorative factor analysis presented in Table 6.7 gave a first impression of the factor structure of the items. The factor structure was further tested in a confirmatory mode using IRT analyses. We distinguished between the expected components as listed in Table 6.1 and Factors 1 and 2, such as found with the explorative factor analysis. The Generalized Partial Credit Model (GPCM) (Muraki, 1992) was used as an IRT model. The fit was evaluated by distinguishing between first- and second-year students of the University of Twente and Saxion University of Applied Sciences.

**Tabel 6.7:** Factor analysis. Factor loadings over or very close to .300 are shaded to emphasize the factors found.

| Item | When I seek information for a study assignment … | 1 | 2 |
|---|---|---|---|
| C 15 (F) | I search for general background information on the topic | .477 | .137 |
| C 16 (F) | I formulate sub questions | .353 | .179 |
| C 17 | I use words from my questions as search terms | .198 | .101 |
| C 18 (F) | I make a list with search terms before I start my search | .267 | .226 |
| C 19 (F) | I determine the best places to search for this information | .406 | -.003 |
| C 20 | I determine new search terms during the search process | .315 | .030 |
| C 21 | I use the option 'advanced search' | .328 | .229 |
| C 22 | and find little or no information on the topic, I adjust the question | .299 | .121 |
| C 23 (F) | I examine the number of results found | .410 | .075 |
| C 24 (F) | and a search produces many results, I narrow my search | .346 | .247 |
| C 25 (F) | I manage the information found so as to easily find it again later on | .370 | .047 |
| D 26 | I scan through the information found | .548 | -.047 |
| D 27 | I use the top list results | .229 | -.048 |
| D 28 | I examine the results on subsequent result pages | .387 | .080 |
| D 29 (F) | I examine the URL to evaluate the information | .500 | .142 |
| D 30 | I select information that corresponds with my own opinion | .106 | -.124 |
| D 31 | I select information which brings new thoughts to mind | .369 | .210 |
| D 32 | I select information which takes an effort (e.g. copying, visit library) | .266 | .063 |
| D 33 | I select information that is accessible right away | .245 | .024 |
| D 34 (F) | on the www, I examine the date of the last update | .414 | .110 |
| D 35 (F) | on the www, I examine the rest of the webpage to judge the reliability of the info | .497 | .133 |
| D 36 (F) | on the www, I determine whether the info consists of facts or opinions | .585 | .163 |
| D 37 (F) | on the www, I examine who makes or edits the website | .451 | .203 |
| D 38 | I carefully read the information found | .299 | .153 |
| D 39 | I use more than one source to answer my question | .453 | .151 |
| D 40 | I formulate the answer to the question in my own words | .416 | .096 |
| E 41 | and write a report, I add an overview of sources used | .215 | .484 |
| E 42 | and write a report, I refer to information within the text | .007 | .728 |
| E 43 | and write a report, I use reference standards to refer to sources used | .084 | .751 |
| E 44 | and use information literally in my report, I use quotation marks | .289 | .535 |
| E 45 | and use information in my report, I refer to the source in the text | .169 | .655 |
| E 46 | and use information put into my own words, I refer to the source in the text | .109 | .634 |

* A Dutch version of the questionnaire was used to gather the data.

Table 6.8 presents an overview of the GPCM analysis for *applying search strategies*. The first column gives item labels related to the classification as given in Table 6.1. Fit to the GPCM was calculated using LM statistics (Glas, 1999). These statistics are based on the difference between observed item scores and expected item scores under the GPCM. The columns labelled *p* give the significance probabilities of the LM tests. A value under 0.01 indicates poor item fit. The columns labelled *d* give the effect size of the fit, that is, the absolute difference between the observed and expected mean item scores across the latent continuum. High values (say over 0.20) indicate a serious misfit. No items needed to be discarded due to lack of model fit. However, one item was discarded based on the subsequent reliability analysis for this scale. Two other items related to this component had already been discarded as a consequence of the reliability analysis described in section 6.2. The results presented in Table 6.8 show an acceptable model fit. The conclusion is that the items within this scale and the students' responses regarding applying search strategies are scalable on a one-dimensional latent scale. The reliability of the 10-item scale was calculated with Cronbach's Alpha ($\alpha$ = 0.68).

**Table 6.8:** GPCM analyses for *applying search strategies*. Significance probabilities of LM tests *p* and effect sizes *d*. Significance probabilities below .01 and effect sizes over .20 are an indication of lack of item fit and are therefore shaded.

| | MIM1 | | MIM2 | | CS1 | | CS2 | |
|---|---|---|---|---|---|---|---|---|
| Item | *p* | *d* | *P* | *d* | *p* | *d* | *p* | *d* |
| C15 | .69 | .06 | .30 | .10 | .01 | .23 | .05 | .14 |
| C16 | .44 | .06 | .51 | .08 | .93 | .07 | .26 | .08 |
| C18 | .30 | .05 | - | .05 | .14 | .18 | - | .11 |
| C19 | .32 | .09 | .04 | .12 | .42 | .19 | .00 | .23 |
| C20 | .58 | .10 | .50 | .09 | .41 | .14 | .92 | .07 |
| C21 | .76 | .06 | .76 | .07 | .01 | .29 | .57 | .15 |
| C22 | .83 | .03 | .29 | .12 | .66 | .19 | .05 | .24 |
| C23 | .14 | .11 | .21 | .09 | .41 | .07 | .30 | .09 |
| C24 | .89 | .07 | .26 | .08 | .24 | .08 | .05 | .13 |
| C25 | .72 | .14 | .83 | .09 | .41 | .14 | .10 | .22 |

**Table 6.9:** GPCM analyses for *evaluating information*. Significance probabilities of LM tests *p* and effect sizes *d*. Significance probabilities below .01 and effect sizes over .20 are an indication of lack of item fit and are therefore shaded.

| | MIM1 | | MIM2 | | CS1 | | CS2 | |
|---|---|---|---|---|---|---|---|---|
| Item | *p* | *d* | *p* | *d* | *p* | *d* | *p* | *d* |
| D26 | .88 | .04 | .71 | .05 | .93 | .05 | .24 | .07 |
| D27 | .09 | .16 | .38 | .08 | .17 | .12 | .13 | .12 |
| D28 | .02 | .11 | .00 | .14 | .18 | .12 | .06 | .26 |
| D29 | .66 | .07 | .85 | .09 | .89 | .05 | .57 | .10 |
| D31 | .35 | .10 | .22 | .09 | .03 | .17 | .53 | .07 |
| D32 | .09 | .15 | .01 | .15 | .01 | .27 | .08 | .18 |
| D33 | .95 | .04 | .35 | .08 | .02 | .10 | .64 | .07 |
| D34 | .76 | .06 | .70 | .05 | .82 | .06 | .36 | .14 |
| D35 | .64 | .08 | .13 | .16 | .96 | .10 | .69 | .01 |
| D36 | .62 | .08 | .86 | .02 | .45 | .08 | .52 | .10 |
| D37 | .02 | .08 | .42 | .10 | .53 | .05 | .36 | .16 |
| D38 | .16 | .11 | .00 | .12 | .01 | .17 | .25 | .09 |
| D39 | .44 | .06 | .27 | .08 | .12 | .13 | .63 | .07 |
| D40 | .77 | .05 | .60 | .05 | .20 | .10 | .14 | .08 |

**Table 6.10:** GPCM analyses for *refer to information* and Factor 2. Significance probabilities of LM tests *p* and effect sizes *d*. Significance probabilities below .01 and effect sizes over .20 are an indication of lack of item fit and are therefore shaded.

| | MIM1 | | MIM2 | | CS1 | | CS2 | |
|---|---|---|---|---|---|---|---|---|
| Item | *p* | *d* | *P* | *d* | *p* | *d* | *p* | *d* |
| E41 | .69 | .09 | .60 | .11 | .21 | .12 | .00 | .17 |
| E42 | .29 | .09 | .01 | .13 | .06 | .15 | .73 | .04 |
| E43 | .97 | .01 | .37 | .07 | .05 | .21 | .00 | .13 |
| E44 | .42 | .14 | .02 | .15 | .00 | .41 | .66 | .05 |
| E45 | .54 | .07 | .04 | .13 | .41 | .06 | .83 | .04 |
| E46 | .26 | .09 | .84 | .05 | .50 | .10 | .07 | .13 |

**Table 6.11:** GPCM analyses for *regulation activities*. Significance probabilities of LM tests *p* and effect sizes *d*. Significance probabilities below .01 and effect sizes above .20 are an indication of lack of item fit and are therefore shaded.

| | MIM1 | | MIM2 | | CS1 | | CS2 | |
|---|---|---|---|---|---|---|---|---|
| Item | *p* | *d* | *p* | *d* | *p* | *d* | *p* | *d* |
| C15 | .19 | .09 | .43 | .08 | .22 | .17 | .54 | .10 |
| C16 | .48 | .07 | .94 | .02 | .55 | .10 | .57 | .06 |
| C18 | .21 | .07 | - | .08 | - | .07 | - | .07 |
| C19 | .84 | .05 | .33 | .09 | .69 | .13 | .14 | .16 |
| C23 | .45 | .08 | .07 | .11 | .75 | .12 | .02 | .20 |
| C24 | .30 | .06 | .10 | .14 | .08 | .15 | .01 | .22 |
| C25 | .21 | .13 | .77 | .06 | .88 | .07 | .31 | .15 |
| D29 | .24 | .11 | .14 | .11 | .61 | .04 | .48 | .13 |
| D34 | .01 | .21 | .86 | .02 | .62 | .11 | .03 | .14 |
| D35 | 1.00 | .05 | .08 | .14 | .35 | .09 | .51 | .05 |
| D36 | .09 | .11 | .56 | .05 | .95 | .02 | .18 | .11 |
| D37 | .05 | .17 | .40 | .09 | .65 | .05 | .23 | .14 |

**Table 6.12:** GPCM analyses for Factor 1. Significance probabilities of LM tests *p* and effect sizes *d*. Significance probabilities below .01 and effect sizes over .20 are an indication of lack of item fit and are therefore shaded.

| | MIM1 | | MIM2 | | CS1 | | CS2 | |
|---|---|---|---|---|---|---|---|---|
| Item | *P* | *d* | *p* | *d* | *p* | *d* | *p* | *d* |
| C15 | .44 | .06 | .43 | .07 | .00 | .14 | .56 | .13 |
| C16 | .01 | .12 | .55 | .06 | .50 | .10 | .35 | .07 |
| C18 | .07 | .14 | - | .05 | - | .20 | - | .11 |
| C19 | .53 | .12 | .10 | .11 | .15 | .13 | .14 | .16 |
| C20 | .27 | .14 | .05 | .11 | .31 | .20 | .04 | .11 |
| C21 | .18 | .13 | .78 | .09 | .00 | .32 | .07 | .22 |
| C22 | .04 | .19 | .00 | .17 | .47 | .19 | .00 | .30 |
| C23 | .78 | .05 | .23 | .09 | .02 | .23 | .01 | .21 |
| C24 | .25 | .10 | .01 | .18 | .56 | .12 | .00 | .28 |
| C25 | .20 | .12 | .30 | .09 | .25 | .16 | .05 | .16 |
| D26 | .19 | .05 | .73 | .03 | .67 | .07 | .15 | .08 |
| D27 | .03 | .22 | .07 | .17 | .21 | .12 | .02 | .17 |
| D28 | .05 | .12 | .51 | .06 | .23 | .14 | .00 | .23 |
| D29 | .64 | .07 | .92 | .04 | .19 | .17 | .89 | .03 |
| D31 | .02 | .11 | .15 | .10 | .00 | .12 | .34 | .07 |
| D32 | .01 | .15 | .02 | .09 | .08 | .15 | .08 | .20 |
| D33 | .16 | .05 | .47 | .09 | .04 | .12 | .26 | .10 |
| D34 | .70 | .07 | .96 | .03 | .74 | .06 | .52 | .09 |
| D35 | .46 | .08 | .07 | .16 | .83 | .05 | .58 | .07 |
| D36 | .03 | .11 | .93 | .03 | .51 | .08 | .65 | .10 |
| D37 | .43 | .07 | .65 | .05 | .61 | .12 | .15 | .15 |
| D38 | .05 | .12 | .00 | .10 | .04 | .16 | .10 | .11 |
| D39 | .28 | .08 | .54 | .05 | .04 | .13 | .36 | .13 |
| D40 | .21 | .08 | .27 | .10 | .30 | .09 | .41 | .08 |

Table 6.9 gives an overview of the MIRT results for the items related to *evaluating information*. The organization of Table 6.9 is analogous to the organization of Table 6.8. Again, the GPCM fits well and no items needed to be discarded due to lack of model fit. However, one item was discarded based on the reliability analysis that followed. The reliability of the 14-item scale was calculated with Cronbach's Alpha ($\alpha$ = 0.74). The conclusion is that the items within this scale and the students' responses regarding evaluating information are scalable on a one-dimensional latent scale.

Table 6.10 gives an overview of the MIRT results for the items related to *referring to information* and also for Factor 2, since they are the same. The organization of Table 6.10 is analogous to the organization of Table 6.8. The GPCM fits well and no items have been discarded. The reliability of the scale was calculated with Cronbach's Alpha ($\alpha$ = 0.81). The items within this scale and the students' responses regarding *referring to information* are scalable on a one-dimensional latent scale.

Table 6.11 presents an overview of MIRT results for *regulation activities* performed during the information-seeking process. The results of the factor analyses did not indicate a scale for regulation activities. An explanation may be that the items of this scale are also related to the *applying search strategies* and *evaluating information*. However, from the results presented in Table 6.11 it can be inferred that items related to regulation activities fitted the GPCM well and are scalable on a one-dimensional latent scale. The reliability of the scale was calculated with Cronbach's Alpha ($\alpha$ = 0.75).

Table 6.12 presents an overview of MIRT results for Factor 1 covering the items related to *applying search strategies*, *evaluating information* and *regulation activities*. The organization of Table 6.12 is analogous to the organization of Table 6.8. The GPCM analysis for Factor 1 results in a less convincing fit than the GPCM analyses of the separate research components. Therefore it is preferred to use the scales for *applying search strategies*, *evaluating information* and *regulation activities* separately instead of combined as one factor.

***Correlation matrix***

For future application of the instrument it is interesting to describe the relationship between the four scales underlying information-seeking behaviour. A Pearson correlation matrix is presented in Table 6.13. The mean scores on the four scales all positively correlate. The highest Pearson correlation values are found for the mean scores for regulation activities and both applying search strategies and evaluating information (0.832 and 0.818 respectively). This might be expected, since the items related to regulation activities are also items belonging to the other two scales. The lowest Pearson correlation is found when relating the mean scores of referring to applying search strategies, evaluating information and regulation activities (0.360, 0.360 and 0.381 respectively). For applying search strategies and evaluating information a Pearson correlation value was found of 0.589.

**Table 6.13.** Pearson correlation matrix for the scales underlying information-seeking behaviour.

| | Search strategies | Evaluation of information | Referring | Regulation activities |
|---|---|---|---|---|
| *Search strategies* | | | | |
| Pearson Correlation | 1.000 | | | |
| *Evaluation of information* | | | | |
| Pearson Correlation | .589 | 1.000 | | |
| *Referring* | | | | |
| Pearson Correlation | .360 | .360 | 1.000 | |
| *Regulation Activities* | | | | |
| Pearson Correlation | .832 | .818 | .381 | 1.000 |

### *Items relating sources used*

The items related to using sources were considered insufficiently scalable. They can, however, be used to describe the extent to which various sources are being used by different groups of undergraduates. Items relating to this research component are presented in Table 6.6. Figure 6.3 gives an overview of the mean scores for the extent to which marketing students and communication students use various sources when seeking information for study assignments. The values on the Y-axis relate to the answering categories ‘rarely or never’ ‘sometimes’, ‘often’ and ‘always’ respectively. The results from this study confirm the widespread use of Google by Dutch undergraduates and correspond to findings of Pan et al (2007) and Aula & Nordhausen (2006) reporting Google being the primary search engine of undergraduates. Search engines other than Google are consulted ‘sometimes’. Study materials handed out by the teachers are ‘often’ consulted for study assignments, just like fellow students, books and Wikipedia. Librarians are ‘rarely or never’ consulted. This corresponds to the findings of Metzger et al (2003) that undergraduates are likely to consult teachers rather than librarians and most frequently turn to fellow or peer students. Marketing students ‘rarely or never’ consult the library catalogue and databases. Communication students consult the databases ‘sometimes’. Metzger et al (2003) found that undergraduates did not use libraries’ electronic databases, primarily due to the lack of awareness of the existence of these databases. Within this study the undergraduates were asked whether they consulted certain discipline specific databases. A large number of students - for some databases over fifty percent of the respondents - answered these questions using the ‘what is that’ response category.

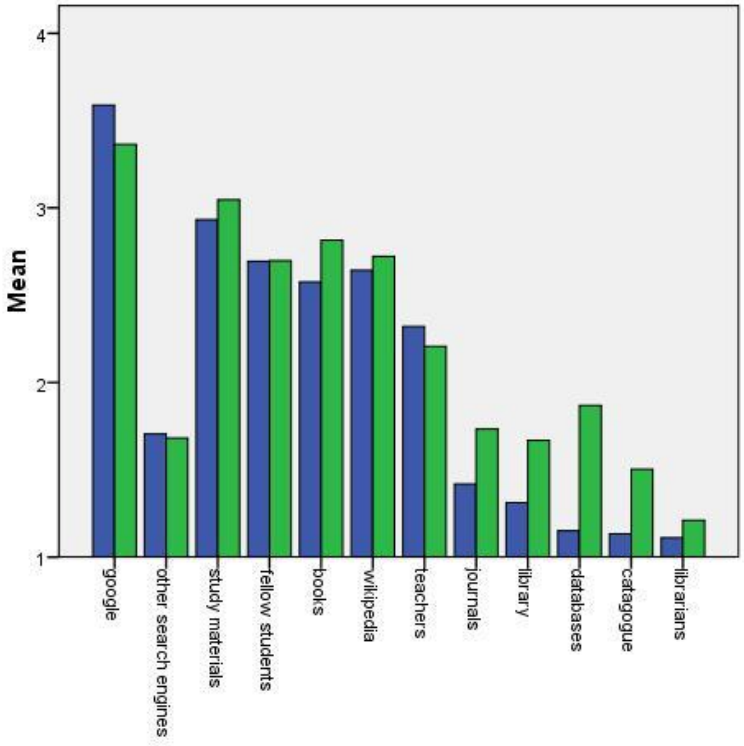


**Figure 6.3.** Mean scores for the extent to which marketing and communication science students, respectively, use various sources of seeking information for a study assignment.

## 6.6 Conclusion and discussion

The aim of this study was to develop a valid and reliable measurement instrument for information-seeking behaviour. Information-seeking behaviour is a multidimensional construct for which several underlying components have been studied within this study. A thorough development and evaluation process resulted in support for the assumption that the following four scales underlie information-seeking behaviour: a 10-item scale for applying search strategies (alpha 0.68), a 14-item scale for evaluating information (alpha 0.74), a 6-item scale for referring to information (alpha 0.81) and a 12-item scale for regulation activities (alpha 0.75). Items related to using sources were considered insufficiently scalable. So the extents to which various sources are used may insufficiently describe the level of information-seeking behaviour. As is demonstrated in the previous paragraph they might be used to describe the extent to which various sources are being used by various groups of undergraduates.

The scales were evaluated using explorative factor analysis and an item response theory (IRT) model to investigate the scalability of the found factors in detail. Two underlying constructs were found using the factor analysis, while the four constructs were confirmed

with the IRT model. Differences between the findings of the factor and IRT analyses can be explained by the strong correlation between three of the identified scales.

The scales can be used within the various frameworks that describe information literate behaviour in higher education, since the scales cover elements common to the various frameworks. The scales can be used to compare groups, monitor populations and to determine effects of interventions, for example, aimed at developing desirable information-seeking behaviour. The scales can be used separately, depending on information needed and time available.

The set of items relating the use of sources can also be used separately to gather information on sources used by a certain population. Or a selection of these items can be used. These items can also be adjusted with sources that match a specific context or recent developments (e.g. World Cat).

The reliability of the scale for applying search strategies is fairly low. This scale could use further development and improvement. Furthermore, information-seeking behaviour in relation to defining the information need requires further research. That is, the question should be addressed whether undergraduates experience difficulties while defining their information needs and whether there is any characteristic behaviour that can be translated to proper items. These items should require students to reflect on their own behaviour without them having to indicate whether they possess certain knowledge and abilities.

Constant and rapid changes characterize the field of information science. However, the scales are expected to preserve their value in the near future. The items are rather generic and it seems unlikely that they will become outdated soon.

The scales have been developed and evaluated for higher education. Such scales could also be useful in primary and secondary education to monitor the interventions and development of information-seeking behaviour. A certain overlap between scales for undergraduates and other population is expected. The evaluation of information is relevant and problematic for all populations mentioned (Walraven et al., 2008). Differences are also expected. Fewer requirements apply to, at least, the primary school population. For example, references to information used are generally not yet relevant. It would be recommended to further analyse the overlap and differences in current and desired information-seeking behaviour of the various populations in order to evaluate the extent in which the scales can be used for other populations.

For this study the results show a widespread use of Google by undergraduates seeking information for a study assignment. Google is an easily accessible source which searches a vast number of web pages on all kinds of subjects. The widespread use of Google for study assignments does not necessarily have to be problematic, especially if Google Scholar is used. However, to find information of acceptable quality it is important that search

strategies are applied and information is evaluated. If this is done sufficiently, information found and used can be of acceptable quality. The question is whether this is being done sufficiently. Results from the literature review indicate that undergraduates seem to rely on basic searches to find the answers to their questions and that they find it difficult to identify terms or keywords (Mittermeyer, 2005; Aula & Nordhausen, 2006). In this study the responses to the items C21 (I use the option 'advanced search') and C24 (I narrow my search, when a search produces many results) indicate a lack of applying search strategies. Large percentages of the MIM and CS students 'rarely or never' use advanced search options (39% and 35% respectively) or limit their searches (51% and 32% respectively) when seeking information for their study assignments.

Future research will focus on gathering and comparing data on information-seeking behaviour of undergraduate students of various disciplines and on first and final year students using the measurement instrument developed in this study. Furthermore, the predictive validity of the measurement will be evaluated by correlating information-seeking behaviour to results on other assignments.

# References


ACRL. (2000). Information literacy competency standards for higher education. Chicago: American Library Association.

Agosto, D. E. (2002). A model of young people's decision-making in using the web. *Library & Information Science Research, 24*, 311-341.

ALA. (1989). Presidential committee on information literacy. Washington, D.C.: American Library Association.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research, 73*, 277-320. doi: 10.3102/00346543073003277

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Belmont: Wadsworth.

Ashford, S. J., Blatt, R., & VandeWalle, D. (2003). Reflections on the looking glass: A review of research on feedback-seeking behavior in organizations. *Journal of Management, 29*, 773-799. doi: 10.1016/s0149-2063(03)00079-5

Aula, A., & Nordhausen, K. (2006). Modelling successful performance in web searching. *Journal of the American Society for Information Science and Technology, 57*, 1678-1693.

AWT. (2005). Tijd voor een opkiQer: Méér investeren in onderwijs en onderzoek. Rijswijk: Quantes: Adviesraad voor het Wetenschap- en Technologiebeleid.

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*, 111-127.

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.

Bates, D. M., & Maechler, M. (Producer). (2011). lme4: Linear mixed-effects models using S4 classes. *R package*. Retrieved from http://lme4.r-forge.r-project.org/

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*, 5-25.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7-74.

Black, P., & Wiliam, D. (2003). 'In praise of educational research': formative assessment. *British Educational Research Journal, 29*(5), 623-637. doi: 10.1080/0141192032000133721

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*, 199-231. doi: 10.1111/j.1464-0597.2005.00205.x

Boekhorst, A. K. (2003). Becoming information literate in the Netherlands. *Library Review, 52*, 298-309.

Boon, S., Johnston, B., & Webber, S. (2007). A phenomenographic study of English faculty's conceptions of information literacy. *Journal of Documentation, 63*, 204-228.

Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151-167. doi: 10.1080/713695728

Brand-Gruwel, S. S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices. *Computers in Human Behavior, 21*, 487-508.

Bruce, C. (1997). *The seven faces of information literacy*. Blackwood, Australia: Auslib.

Buchanan, T. (1999). Using the world wide web for formative assessment. *Journal of Educational Technology Systems, 27*, 71-80.

Bundy, A. (Ed.). (2004). *Australian and New Zealand Information Literacy Framework* (2nd ed.): Australian and New Zealand Institute for Information Literacy, Adelaide.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245-281.

Callinan, J. E. (2005). Information-seeking behaviour of undergraduate biology students: A comparative analysis of first year and final year students in University College Dublin. *Library Review, 54*, 86-99.

Clariana, R. B., & Koul, R. (2006). The effects of different forms of feedback on fuzzy and verbatim memory of science principles. *British Journal of Educational Psychology, 76*, 259-270.

Corbalan, G., Kester, L., & van Merriënboer, J. G. (2009). Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction, 19*, 455-465. doi: 10.1016/j.learninstruc.2008.07.002

Crombach, M. J., Boekaerts, M., & Voeten, M. J. M. (2003). Online measurement of appraisals of students faced with curricular tasks. *Educational and Psychological Measurement, 63*, 96-111.

Dervin, B. (1983). *An overview of sense-making research: concepts, methods and results to date*. Paper presented at the International Communications Association Annual Meeting, Dallas, TX.

Drucker, P. F. (1993). *Post-capitalist society*. New York: HarperBusiness.

Duval, E., & Verbert, K. (2012). Learning analytics. *eleed, 8*. doi: urn:nbn:de:0009-5-33367

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values and goals. *Annual Review of Psychology, 53*, 109-132.

Edwards, S. L. (2006). *Teaching users how to find that needle in a haystack: Enabling end user information literacy.* Paper presented at the 4th International Lifelong Learning Conference: Partnerships, Pathways and Pedagogies, Rydges Capricorn Resort, Yeppoon, Qld.

Ellis, D. A. (1989). A behavioural approach to information retrieval design. *Journal of Documentation, 45*, 171-212.

Ellis, D. A., Cox, D., & Hall, K. (1993). A comparison of the information-seeking patterns of researchers in the physical and social sciences. *Journal of Documentation, 49*, 356-369.

Glas, C. A. W. (1999). Modification indices for the 2-pl and the nominal response model. *Psychometrika, 64*, 273-294.

Gordijn, J., & Nijhof, W. J. (2002). Effects of complex feedback on computer-assisted modular instruction. *Computers & Education, 39*, 183-200. doi: 10.1016/S0360-1315(02)00025-8

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.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81-112. doi: 10.3102/003465430298487

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*, 1-14.

Jaehnig, W., & Miller, M. L. (2007). Feedback types in programmed instruction: A systematic review. *Psychological Record, 57*, 219-232.

Karabenick, S. A., & Knapp, J. R. (1988). Effects of computer privacy on help-seeking. *Journal of Applied Social Psychology, 18*, 461-472. doi: 10.1111/j.1559-1816.1988.tb00029.x

Kluger, A. N., & Adler, S. (1993). Person- versus computer-mediated feedback. *Computers in Human Behavior, 9*, 1-16. doi: 10.1016/0747-5632(93)90017-m

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*, 245-284.

Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An overview. *Theory into Practice, 41*, 212-218.

Kuhlemeier, H., & Hemker, B. (2005). Computergebruik thuis en Internetvaardigheid in het voortgezet onderwijs. *Pedagogische Studiën, 82*, 115-136.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*, 361-371.

Kuhlthau, C. C. (2004). *Seeking meaning : A process approach to library and information services* (2nd ed.). Westport, Connecticut: Libraries unlimited.

Kuiper, E., Volman, M., & Terwel, J. (2005). The web as an information resource in K-12 education: Strategies for supporting students in searching and processing information. *Review of Educational Research, 75*, 285-328.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*, 279-308. doi: 10.1007/bf01320096

Kurbanoglu, S. S., Akkoyunlu, B., & Umay, A. (2006). Developing the information literacy self-efficacy scale. *Journal of Documentation, 62*, 730-743.

Liao, T. F. (1994) Interpreting probability models: Logit, probit, and other generalized linear models. *Vol. 07-101. Sage university paper series on quantitative applications in the social sciences*. Thousand Oaks, CA: Sage.

MaKinster, J. G., Beghetto, R. A., & Plucker, J. A. (2002). Why can't I find Newton's third law? Case studies of students' use of the web as a science resource. *Journal of Science Education and Technology, 11*, 155-172.

Mason, B. J., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us*. Center for Instructional Innovation. University of Nebraska–Lincoln: 14. Retrieved from http://dwb.unl.edu/Edit/MB/MasonBruning.html

Maughan, P. P. D. (2001). Assessing information literacy among undergraduates: A discussion of the literature and the university of California-Berkeley assessment experience. *College and Research Libraries, 62*, 71-86.

McCullogh, C. E., Searle, S. R., & Neuhaus, J. M. ( 2008). *Generalized, Linear, and Mixed Models* ( 2nd ed.). Hoboken, New Jersey: Wiley.

Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). Student Internet use, perceptions of information credibility, and verification behavior. *Computers & Education, 41*, 271-290.

Mittermeyer, D. (2005). Incoming first year undergraduate students. *Education for Information, 23*, 203-232.

Monoi, S., O'Hanlon, N., & Diaz, K. R. (2005). Online searching skills: Development of an inventory to assess self-efficacy. *Journal of Academic Librarianship, 31*, 98-105.

Mory, E. H. (1994). Adaptive feedback in computer-based instruction: Effects of response certitude on performance, feedback-study time, and efficiency. *Educational Computing Research, 11*, 263-290.

Mory, E. H. (2004). Feedback research revisited. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745-783). United Kingdom: Routledge Ldt.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

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

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. M. Niegemann, D. Leutner & R. Brunken (Eds.), *Instructional design for multimedia learning* (pp. 181-195). Munster, New York: Waxmann.

Narciss, S., Körndle, H., Reimann, G., & Müller, C. (2004). *Feedback-seeking and feedback efficiency in web-based learning - How do they relate to task and learner characteristics?* Paper presented at the EARLI SIG Instructional Design and Learning and Instruction with Computers, Tuebingen.

Nicol, D., & McFarlane-Dick, D. (2006). Formative assessment and self-regulated learning : A model and seven principles of good feedback practice *Studies in Higher Education, 31*, 199-218.

Norman, C. D., & Skinner, H. A. (2006). eHEALS: The ehealth literacy scale. *Journal of Medical Internet Research, 8(4)*: e27.

O'Brien, H. L., & Symons, S. (2007). The information behaviors and preferences of undergraduate students. *Research Strategies, 20*, 409-423.

Oosterveld, P. (1996). *Questionnaire design methods.* Universiteit van Amsterdam, Amsterdam.

Owston, R. D. (1997). The world wide web: A technology to enhancing teaching and learning? *Educational Researcher, 26(2)*, 27-33.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication, 12*, 801-823.

Park, G., Schmidt, A. M., Scheu, C., & DeShon, R. P. (2007). A process model of goal orientation and feedback seeking. *Human Performance, 20*, 119-145. doi: 10.1080/08959280701332042

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*, 459-470.

Poulos, A., & Mahony, M. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education, 33*, 143-154. doi: 10.1080/02602930601127869

Pridemore, D. R., & Klein, J. D. (1991). Control of feedback in computer-assisted instruction. *Educational Technology Research and Development, 39*(4), 27-32. doi: 10.1007/bf02296569

Pridemore, D. R., & Klein, J. D. (1995). Control of practice and level of feedback in computer-based instruction. *Contemporary Educational Psychology, 20*, 444-450.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.): Newbury Park, California: Sage.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144.

Sadler, D. R. (1998). Formative assessment: revisiting the territory. *Assessment in Education, 5*, 77-84.

Schoonman, W. (1989). *An applied study on computerized adaptive testing*. Lisse: Swets & Zeitlinger.

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education: Theory, research and applications* (3rd ed.). Upper Saddle River, New Jersey: Pearson.

SCONUL. (1999). *Information skills in Higher Education*. Task Force on Information Skills. National and University Libraries, London.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189. doi: 10.3102/0034654307313795

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 30-32.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Ser. B*(13), 238–241.

Smith, P. L., & Ragan, T. J. (2005). *Instructional design*. Upper Saddle River, New Jersey: Merrill.

Smits, M., Boon, J., Sluijsmans, D. M. A., & van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: Effects on learning as a function of prior knowledge. *Interactive Learning Environments, 16*, 183-193. doi: 10.1080/10494820701365952

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

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*, 25-31. doi: 10.1016/j.compedu.2012.07.007

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

Twait, M. (2005). Undergraduate students' source selection criteria: A qualitative study. *The Journal of Academic Librarianship, 31*, 567-573.

Valdez, A. J. (2008). *Encouraging mindful feedback processing: Computer-based instruction in descriptive statistics.* University of New Mexico Albuquerque, NM.

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*, 263-272. doi: 10.1016/j.compedu.2011.07.020

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*. 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*, 21-39. doi: 10.3280/CAD2011-001004

Van der Vleuten, C., Driessen, E., & Van Tartwijk, J. (2006). Toetsprogramma's. In H. van Hout (Ed.), *Vernieuwing in het hoger onderwijs* (pp. 119-130). Assen, The Netherlands: Van Gorcum.

Van Deursen, A. J. A. M., & Van Dijk, J. A. G. M. (2008a). Digitale vaardigheden van Nederlandse burgers. Een prestatiemeting van operationele, formele, informatie en strategische vaardigheden bij het gebruik van overheidswebsites. Enschede: Universiteit Twente.

Van Deursen, A. J. A. M., & Van Dijk, J. A. G. M. (2008b). *Measuring digital skills*. Paper presented at the Conference of the International Communication Association, Montreal, Canada.

Van Kampen, D. J. (2004). Development and validation of the multidimensional library anxiety scale. *College and Research Libraries, 65*, 28-35.

VandeWalle, D., & Cummings, L. L. (1997). A test of the influence of goal orientation on the feedback-seeking process. *Journal of Applied Psychology, 82*, 390-400. doi: 10.1037/0021-9010.82.3.390

VandeWalle, D., Ganesan, S., Challagalla, G. N., & Brown, S. P. (2000). An integrated model of feedback-seeking behavior: Disposition, context, and cognition. *Journal Of Applied Psychology, 85*, 996-1003. doi: 10.1037/0021-9010.85.6.996

Veldkamp, B., Matteucci, M., & Eggen, T. H. M. (2011). Computerized Adaptive Testing in Computer Assisted Learning? In S. Wannemacker, G. Clarebout & P. Causmaecker (Eds.), *Interdisciplinary approaches to adaptive learning. A Look at the neighbours* (Vol. 126, pp. 28-39): Springer Berlin Heidelberg.

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*, 623-648. doi: Urn:nbn:nl:ui:28-60705

Wang, T. H. (2007). What strategies are effective for formative assessment in an e-learning environment? *Journal of Computer Assisted Learning, 23*, 171-186. doi: 10.1111/j.1365-2729.2006.00211.x

Wauters, K., Desmet, P., & Van den Noortgate, W. (2010). Adaptive item-based learning environments based on the item response theory: Possibilities and challenges. *Journal of Computer Assisted Learning, 26*, 549-562. doi: 10.1111/j.1365-2729.2010.00368.x

Wellman, G. S. (2005). Comparing learning style to performance in on-line teaching: Impact of proctored versus un-proctored testing. *Journal of Interactive Online Learning, 4*, 20-39.

Whitmire, E. (2002). Disciplinary differences and undergraduates' information-seeking behavior. *Journal of the American Society for Information Science and Technology, 53*, 631-638.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation, 55*, 249-270.

Wise, S. L. (2006). An investigation of the differential effort received by items on a low-stakes computer-based test. *Applied Measurement in Education, 19*, 95-114. doi: 10.1207/s15324818ame1902_2

Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test motivation, and mentally taxing items. *Applied Measurement in Education, 8*, 341-351. doi: 10.1207/s15324818ame0804_4

Wood, R. (1978). Fitting the Rasch model - a heady tale. *British Journal of Mathematical and Statistical Psychology, 31*, 27-32.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education 45*, 477-501.

# Appendix A. An exemplary item with feedback

***Item***

You are using Google to find information about sustainable entrepreneurship.

Which search command will lead to the smallest number of hits?

A sustainable AND entrepreneurship

B sustainable OR entrepreneurship

C "sustainable entrepreneurship"

***Additional Feedback***

The correct answer is "sustainable entrepreneurship".

When searching information using " ", you will only find those web pages on which sustainable and entrepreneurship are positioned right next to one another and sustainable put first.

When searching information using AND, you will find web pages which contain sustainable and entrepreneurship, but not necessarily positioned right next to one another.

When searching information using OR, you will find web pages which contain sustainable and/or entrepreneurship. Thus, you will find the web pages which you would have found using AND plus the web pages which contain sustainable only or entrepreneurship only.

# Appendix B. Knowledge of results with additional feedback per item

You can consult extra feedback by clicking the link.
*Note! You can't consult the extra feedback after leaving this page.*
The number of correctly answered items: 7

| Item nr. | Topic | Result | |
|---|---|---|---|
| 1 | Databases | ✗ | Additional feedback |
| 2 | Search engines | ✓ | Additional feedback |
| 3 | Search engines | ✗ | Additional feedback |
| 4 | Boolean operators | ✓ | Additional feedback |
| 5 | Access information sources | ✓ | Additional feedback |
| 6 | Search engines | ✓ | Additional feedback |
| 7 | Define search terms | ✗ | Additional feedback |
| 8 | RSS | ✗ | Additional feedback |
| 9 | Plagiarism | ✓ | Additional feedback |
| 10 | Truncate | ✓ | Additional feedback |
| 11 | Type of publication | ✗ | Additional feedback |
| 12 | Boolean operators | ✗ | Additional feedback |
| 13 | Define search terms | ✗ | Additional feedback |
| 14 | Search engines | ✓ | Additional feedback |
| 15 | Databases | ✗ | Additional feedback |

# Appendix C. Information-seeking behaviour questionnaire

**Antwoorden zo aangeven: ●**
**Niet zo: ⊘ ⊕**

Beste student,

Wat doe jij als je op zoek bent naar informatie voor een opdracht van school? Met de vragenlijst die hier volgt proberen we een zo goed mogelijk beeld te krijgen van het zoekgedrag van studenten van Saxion Hogescholen. Je antwoorden worden anoniem gebruikt voor een onderzoek. Er zijn **geen** goede en foute antwoorden en we willen je dan ook vragen om de vragenlijst zo waarheidsgetrouw mogelijk in te vullen. Mochten er vragen (of begrippen) niet helder zijn dan kun je dat aangeven met behulp van de antwoordoptie "wat is dat?". Het invullen van de vragenlijst duurt 15 minuten. Alvast hartelijk dank voor je medewerking.

In welk jaar van de opleiding zit je? O 1e O 2e O 3e O 4e

Geslacht O Man O Vrouw

Nationaliteit O Nederlands O Duits O Anders

Leeftijd O jonger dan 18 jaar O 18 tot 22 jaar O 22 jaar of ouder

| **Als ik op zoek ben naar informatie voor school ...** | altijd | vaak | soms | zelden tot nooit | *wat is dat?* |
|---|---|---|---|---|---|
| ... maak ik gebruik van de door mijn docent aangereikte studieboeken en -materialen | O | O | O | O | O |
| ... ga ik naar Saxion bibliotheken | O | O | O | O | O |
| ... raadpleeg ik medewerkers van Saxion bibliotheek | O | O | O | O | O |
| ... maak ik gebruik van de catalogus van Saxion bibliotheek | O | O | O | O | O |
| ... maak ik gebruik van de databanken van Saxion bibliotheek | O | O | O | O | O |

| **Als ik op zoek ben naar informatie voor school raadpleeg ik...** | altijd | vaak | soms | zelden tot nooit | *wat is dat?* |
|---|---|---|---|---|---|
| ... Google | O | O | O | O | O |
| ... Google Scholar | O | O | O | O | O |
| ... andere zoekmachines dan Google | O | O | O | O | O |
| ... PiCarta | O | O | O | O | O |
| ... Europmaat | O | O | O | O | O |
| ... docenten | O | O | O | O | O |
| ... medestudenten | O | O | O | O | O |
| ... kranten | O | O | O | O | O |
| ... boeken | O | O | O | O | O |
| ... tijdschriften | O | O | O | O | O |
| ... Wikipedia | O | O | O | O | O |
| ... metazoekmachines | O | O | O | O | O |

0064508474

| **Als ik op zoek ben naar informatie voor school ...** | altijd | vaak | soms | zelden tot nooit | *wat is dat?* |
|---|---|---|---|---|---|
| ... zoek ik algemene achtergrond informatie over het onderwerp | O | O | O | O | O |
| ... formuleer ik deelvragen | O | O | O | O | O |
| ... maak ik, voordat ik begin met zoeken, een lijst met woorden om mee te zoeken | O | O | O | O | O |
| ... bepaal ik waar ik die informatie het beste kan zoeken | O | O | O | O | O |
| ... gebruik ik meer dan drie verschillende woorden per zoekactie | O | O | O | O | O |
| ... bedenk ik al zoekende nieuwe woorden om verder mee te zoeken | O | O | O | O | O |
| ... gebruik ik de optie 'geavanceerd zoeken' | O | O | O | O | O |
| ... en ik weinig tot geen informatie over het onderwerp vind, dan pas ik mijn vraag aan | O | O | O | O | O |

| **Als ik op zoek ben naar informatie voor school ...** | altijd | vaak | soms | zelden tot nooit | *wat is dat?* |
|---|---|---|---|---|---|
| ... bekijk ik hoeveel hits mijn zoekactie heeft opgeleverd | O | O | O | O | O |
| ... en een zoekactie levert veel hits op, dan pas ik de zoekactie aan zodat het aantal hits afneemt | O | O | O | O | O |
| ... orden ik de gevonden informatie, zodat ik deze later gemakkelijk terug kan vinden | O | O | O | O | O |
| ... lees ik de informatie globaal door | O | O | O | O | O |
| ... gebruik ik de hits die bovenaan de pagina staan | O | O | O | O | O |
| ... bekijk ik de hits die op de volgende pagina('s) staan | O | O | O | O | O |
| ... bekijk ik het webadres (URL) om de informatie te beoordelen | O | O | O | O | O |

2232508473

| **Als ik op zoek ben naar informatie voor school ...** | altijd | vaak | soms | zelden tot nooit | *wat is dat?* |
|---|---|---|---|---|---|
| ... selecteer ik informatie die me op nieuwe gedachten brengt over het onderwerp | O | O | O | O | O |
| ... selecteer ik informatie waarvoor ik moeite moet doen (zoals reserveren, kopieren, naar de bibliotheek gaan) | O | O | O | O | O |
| ... selecteer ik informatie die direct toegankelijk is | O | O | O | O | O |
| ... op een website, bekijk ik de datum van de laatste update | O | O | O | O | O |
| ... op een website, bekijk ik ook de rest van de website om de informatie te beoordelen | O | O | O | O | O |
| ... op een website, bepaal ik of de informatie uit feiten of meningen bestaat | O | O | O | O | O |
| ... op een website, dan ga ik na wie de maker en/of uitgever van de website is | O | O | O | O | O |
| ... neem ik de tijd om de gevonden informatie nauwkeurig door te lezen | O | O | O | O | O |
| ... gebruik ik meerdere bronnen om mijn vraag te beantwoorden | O | O | O | O | O |
| ... formuleer ik het antwoord op mijn vraag in mijn eigen woorden | O | O | O | O | O |

| **Als ik op zoek ben naar informatie voor school ...** | altijd | vaak | soms | zelden tot nooit | *wat is dat?* |
|---|---|---|---|---|---|
| ... en een verslag schrijf, voeg ik een overzicht van de gebruikte bronnen toe | O | O | O | O | O |
| ... en een verslag schrijf, verwijs ik in de tekst naar de gebruikte informatiebronnen | O | O | O | O | O |
| ... en een verslag schrijf, gebruik ik richtlijnen voor het verwijzen naar gebruikte informatiebronnen | O | O | O | O | O |
| ... en tekst letterlijk overneem in een verslag, zet ik die tekst tussen aanhalingstekens | O | O | O | O | O |
| ... en tekst letterlijk overneem in een verslag, verwijs ik in de tekst naar de gebruikte informatiebron(nen) | O | O | O | O | O |
| ... en informatie in mijn eigen woorden opneem in een verslag, verwijs ik in de tekst naar de gebruikte informatiebron(nen) | O | O | O | O | O |

8505508472

# Summary

Assessment can be used to stimulate and direct student learning. This refers to the formative function of assessment. To prepare students for lifelong learning in increasingly complex societies that are subject to rapid changes, a shift in assessment thinking from testing students to fostering self-regulated learning is advocated. In this context, formative assessment is viewed as a means to help students take control of their own learning process. Formative assessments contribute to learning by generating feedback. Here, feedback is conceptualised as information about learners actual state of performance that is provided to leaners to modify their thinking or behaviour for the purpose of improved performance (cf. Narciss, 2008; Shute, 2008). The use of computers in formative assessments has several advantages as it provides the opportunity to generate immediate, objective and appropriate feedback based on response models constructed in advance. Research shows that adding knowledge of correct response and remedial feedback to computer-based formative assessment (CBFA) is associated with performance improvement (Van der Kleij et al., 2011). However, for feedback to result in improved performance on an individual level, it is conditional that students actively seek feedback and construct meaning from it (Ashford et al., 2003; Bangert-Drowns et al., 1991; Nicol & McFarlane-Dick, 2006). To better understand the relation between CBFA and performance improvement, the extent to which students seek and study feedback has been explored as well as variables influencing this, so called feedback behaviour.

ICT provides the opportunity to capture learner-produced data trails to gain insight into what happens when learners are asked to engage in a CBFA. Collecting, analysing and reporting traces of learners in educational settings for purposes of understanding and optimizing learning processes and the environments in which it occurs, is referred to as learning analytics or educational data mining (Siemens & Long, 2011). In the context of this dissertation, the method used to explore student feedback behaviour in a CBFA and variables influencing this behaviour is an example of learning analytics. In absence of a systems that could capture feedback behaviour data in CBFA, a system was developed especially for the purpose of this research.

The learner-produced data trails were explored to determine patterns of feedback behaviour in a CBFA using descriptive statistics (Chapter 2). In addition, feedback behaviour was related to student response (in-/correct), test length (10 or 20 items), achievement, and supervision (direct, indirect, and none). The results show that feedback behaviour varies greatly between students. In general feedback behaviour focuses on feedback of incorrectly answered questions. Approximately fifty percent of the respondents paid attention to feedback of incorrect answers only. Another twenty-five percent did not pay attention to feedback at all. Results suggest that differences in attention paid to feedback are influenced by achievement and test length. Supervision, however, did not seem to influence the average attention paid to extra feedback.

Individual and group differences in feedback behaviour as well as feedback study time, have been explored through generalized and linear mixed models (Chapter 3). Furthermore, the relations between feedback behaviour and student response (in-/correct), item difficulty, and achievement have been examined. Results show that feedback seeking and feedback study times were higher for incorrect responses, and among high and middle achieving students. In addition, when item difficulty increases the propensity to seek feedback increases for incorrect responses only.

Feedback behaviour has also been related to task-value beliefs, success expectancy, and student-reported effort invested in a CBFA (Chapter 4). Results show that feedback seeking was predicted by success expectancy as well as task-value beliefs, while feedback study time was not. In addition, feedback seeking was predicted by student effort invested in the formative assessment.

The relation between CBFA and performance improvement is influenced by numerous variables, such as type of feedback and student motivational beliefs. These variables need to be taken into account by designers of CBFA for reasons of effectiveness. Findings of previous research have been examined and used to draw an overview of variables substantially influencing the relation between CBFA and performance improvement (Chapter 5). Subsequently, findings of previous research have been translated to recommendations for design of CBFA and used to propose a theory based decision-making framework for design of feedback interventions in CBFA.

The CBFA used to explore student feedback behaviour aimed at increasing knowledge and understanding of what information literacy entails and, subsequently, support improvement of student information-seeking behaviour. In absence of a reliable instrument to measure information-seeking behaviour, scales were developed and evaluated. The development and evaluation of scales for information-seeking behaviour is described in Chapter 6. The scales can be used to measure effects of interventions, such as research skills training, on information-seeking behaviour. The scales can also be used to increase knowledge and understanding of what information-seeking behaviour or information problem solving entails.

In conclusion, learning analytics or educational data mining techniques provide valuable methods to further examine the complex relation between CBFA and learning. In addition, designers of feedback interventions in computer-based environments, such as a CBFA, should account for variables influencing the effectiveness of feedback on learning. Results of previous research already provide numerous recommendations for the design of automated feedback in varying situations, such as presented in the Framework for the design of feed in CBFA.

# Samenvatting

Wanneer een toets gericht is op het sturen en stimuleren van een leerproces, wordt deze formatief genoemd. In maatschappijen waar belang wordt gehecht aan een leven lang leren en zelfregulerend leren, wordt in toenemende mate gepleit voor meer aandacht voor de formatieve functie van toetsing. Formatieve of leertoetsen worden beschouwd als een geschikt middel om studenten inzicht te geven in hun eigen leerproces en zodoende eigenaar te maken van hun leerproces. Formatieve toetsing draagt bij aan leerprocessen door middel van het genereren van feedback. Feedback wordt dan ook beschouwd als een essentieel onderdeel van leertoetsen. Feedback is een breed begrip dat, in de context van formatieve toetsing, verwijst naar alle informatie die gegeven wordt aan een student over diens staat van leren met als doel het leerproces te voeden en leeropbrengsten zodoende te vergroten.

Het gebruik van computers bij formatieve toetsing kent als voordeel dat feedback direct gegenereerd kan worden, objectief is en afgestemd kan worden op de resultaten van de student. Uit onderzoek blijkt dat geautomatiseerde feedback tot een significante verbetering van leeropbrengsten kan leiden (Van der Kleij et al., 2011). Echter, de invloed van feedback op individuele leeropbrengsten hangt met name af van de mate waarin de lerende aandacht besteedt aan feedback en feedback gebruikt om van te leren (Ashford et al., 2003; Bangert-Drowns et al., 1991; Nicol & McFarlane-Dick, 2006). Dit gedrag wordt hier aangeduid met de term feedbackgedrag. Om beter inzicht te krijgen op de relatie tussen computergestuurde formatieve toetsen en leeropbrengsten, is onderzocht in welke mate studenten gebruik maken van feedback in een computergestuurde leertoets en welke variabelen hierop van invloed zijn.

ICT biedt de mogelijkheid om informatie over gedrag van studenten in computergestuurde leeromgevingen te verzamelen. Dergelijke data kan bijvoorbeeld gebruikt worden om zicht te krijgen op het feedbackgedrag van studenten als ze de opdracht krijgen om een computergestuurde leertoets te maken. Het meten, verzamelen, analyseren en rapporteren van data over studenten en hun leeromgeving met als doel inzicht in en optimalisatie van leerprocessen en –omgevingen, wordt aangeduid met de term *learning analytics* (Siemens &

Long, 2011). De in dit proefschrift gehanteerde methode voor onderzoek naar feedbackgedrag en de variabelen die hierop van invloed zijn, is een voorbeeld van *learning analytics*. Om informatie te kunnen verzamelen over het feedbackgedrag van studenten in een computergestuurde leertoets, is een systeem ontwikkeld dat het feedbackgedrag registreert.

Om patronen in feedbackgedrag te achterhalen is de verzamelde data geanalyseerd met behulp van beschrijvende statistiek (Hoofdstuk 2). Daarnaast is onderzocht in hoeverre het feedbackgedrag relateerde aan de juistheid van een antwoord (on-/juist), de lengte van de leertoets (10 of 20 items), de totaalscore, en de aanwezigheid van toezicht bij het maken van de toets (direct, indirect, of geen). Uit het onderzoek blijkt dat de aandacht die studenten besteden aan de feedback onderling sterk verschilt. In het algemeen laten de resultaten zien dat de aandacht voor feedback zich concentreert op feedback voor onjuist beantwoorde items. Circa vijftig procent van de respondenten besteedt enkel aandacht aan feedback van onjuist beantwoorde items. Circa vijfentwintig procent van de respondenten besteedden geen aandacht aan feedback. Resultaten suggereren dat het feedbackgedrag wordt beïnvloed door de totaalscore en de lengte van de toets. Toezicht leek niet van invloed op het feedbackgedrag.

Verschillen in feedbackgedrag tussen individuen en groepen zijn geanalyseerd met behulp van multilevel modellen (Hoofdstuk 3). Daarnaast, zijn de relaties tussen feedbackgedrag, de juistheid van het antwoord (on-/juist), de moeilijkheid van een item en de score op de toets onderzocht. De resultaten laten zien dat de aandacht voor feedback toeneemt wanneer een item onjuist is beantwoord én de desbetreffende student een gemiddeld tot hoge score haalt op de toets. Verder laten de resultaten zien dat de kans toeneemt dat een student extra feedback raadpleegt wanneer de moeilijkheid van een item toeneemt. Dit laatste geldt alleen voor onjuist beantwoorde items.

Hoofdstuk 4 behandelt onderzoek naar de relatie tussen feedbackgedrag en motivationele drijfveren van studenten. Voor en na afloop van de toetsvragen beantwoorden studenten vragen over hun succesverwachtingen, de waardering voor de leertoets en de moeite die ze namen om de toets te maken. Uit de resultaten blijkt dat hogere succesverwachting, en hogere waardering voor de toets positief correleren met de mate waarin studenten

aandacht besteden aan extra feedback. Ook een hogere inzet correleerde positief met feedbackgedrag.

De relatie tussen computergestuurde leertoetsen en leren wordt beïnvloed door verschillende variabelen, zoals de inhoud van de feedback interventie en de motivationele drijfveren van de student. Op basis van eerder onderzoek is een overzicht gemaakt van variabelen die de relatie tussen formatieve toetsing en leren substantieel beïnvloeden (zie Hoofdstuk 5). De bevindingen uit eerder onderzoek zijn vertaald naar aanbevelingen voor het ontwerp van computergestuurde leertoetsen. Ook is een raamwerk ontwikkeld voor het ontwerp van feedback in computergestuurde leertoetsen.

De computergestuurde leertoets gebruikt voor dit onderzoek had tot doel om de kennis en het inzicht van studenten betreffende informatievaardigheden te vergroten en, zodoende, het gewenst informatiegedrag van studenten te bevorderen. Om effecten van onderwijs in informatievaardigheden op het informatiegedrag van studenten in het hoger onderwijs te kunnen meten is een vragenlijst ontwikkeld en geëvalueerd. De ontwikkeling en evaluatie van de vragenlijst wordt beschreven in Hoofdstuk 6. De vragenlijst kan gebruikt worden om effecten te meten van interventies, zoals een training onderzoekvaardigheden, op informatiegedrag. De vragenlijst kan ook gebruikt worden om inzicht te vergroten in wat het beteken om informatievaardig te zijn.

Tot slot, *learning analytics* biedt waardevolle mogelijkheden om de complexe relatie tussen computergestuurde leertoetsen en leeropbrengsten nader te onderzoeken. Verder, ontwerpers van feedback in computergestuurde leeromgevingen dienen rekening te houden met variabelen die de effectiviteit van feedback op leeropbrengsten beïnvloeden. Resultaten uit eerder onderzoek leiden tot aanbevelingen zoals beschreven in het raamwerk voor het ontwerp van feedback in computergestuurde leertoetsen.