Computers & Education 71 (2014) 56–76

Contents lists available at ScienceDirect

Computers & Education

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

Exploring feedback and student characteristics relevant for personalizing feedback strategies

Susanne Narciss [a,*], Sergey Sosnovsky [b], Lenka Schnaubert [c], Eric Andrès [b], Anja Eichelmann [a], George Goguadze [d], Erica Melis [b,†]

[a] Technische Universität Dresden, Fachrichtung Psychologie, Psychologie des Lehrens und Lernens, 01062 Dresden, Germany
[b] Deutsche Forschungszentrum für Künstliche Intelligenz GmbH, Campus D3.2, 66123 Saarbrücken, Germany
[c] Universität Duisburg-Essen, Fachgebiet Psychologische Forschungsmethoden, Abteilung für Informatik und Angewandte Kognitionswissenschaft, Lotharstraße 65, 47057 Duisburg, Germany
[d] Leuphana Universität Lüneburg, Kompetenzzentrum Hochschuldidaktik Mathematik, Fakultät Bildung, Scharnhorststraße 1, 21335 Lüneburg, Germany

A R T I C L E I N F O

Article history:
Received 21 June 2013
Received in revised form
11 September 2013
Accepted 13 September 2013

Keywords:
Tutoring feedback strategies
Intelligent tutoring system
Evaluation of CAL systems

A B S T R A C T

Personalized tutoring feedback is a powerful method that expert human tutors apply when helping students to optimize their learning. Thus, research on tutoring feedback strategies tailoring feedback according to important factors of the learning process has been recognized as a promising issue in the field of computer-based adaptive educational technologies. Our paper seeks to contribute to this area of research by addressing the following aspects: First, to investigate how students' gender, prior knowledge, and motivational characteristics relate to learning outcomes (knowledge gain and changes in motivation). Second, to investigate the impact of these student characteristics on how tutoring feedback strategies varying in content (procedural vs. conceptual) and specificity (concise hints vs. elaborated explanations) of tutoring feedback messages affect students' learning and motivation. Third, to explore the influence of the feedback parameters and student characteristics on students' immediate post-feedback behaviour (skipping vs. trying to accomplish a task, and failing vs. succeeding in providing a correct answer). To address these issues, detailed log-file analyses of an experimental study have been conducted. In this study, 124 sixth and seventh graders have been exposed to various tutoring feedback strategies while working on multi-trial error correction tasks in the domain of fraction arithmetic. The web-based intelligent learning environment ActiveMath was used to present the fraction tasks and trace students' progress and activities. The results reveal that gender is an important factor for feedback efficiency: Male students achieve significantly lower knowledge gains than female students under all tutoring feedback conditions (particularly, under feedback strategies starting with a conceptual hint). Moreover, perceived competence declines from pre- to post-test significantly more for boys than for girls. Yet, the decline in perceived competence is not accompanied by a decline in intrinsic motivation, which, instead, increases significantly from pre- to post-test. With regard to the post-feedback behaviour, the results indicate that students skip further attempts more frequently after conceptual than after procedural feedback messages.

1. Introduction

Development of personalized learning environments is among the most important research areas of computer-based education for the next decade (Spada et al., 2012; Wolf, 2010). Such environments should be capable of tracing accurately learners' activity, monitor their individual characteristics, and generate timely adaptive interventions according to effective pedagogical strategies. Over the years, many different technologies have been developed for optimizing instructional interventions to the needs, goals, and knowledge of individual

* Corresponding author. Psychology of Learning and Instruction, Technische Universität Dresden, 01062 Dresden, Germany. Tel.: +49 351 4633 6059; fax: +49 351 4633 7294.
E-mail addresses: susnarciss@gmail.com, susanne.narciss@tu-dresden.de (S. Narciss).
† Dr. Erica Melis was one of the initiators and the principal investigators of the AtuF project. She passed away during the implementation of AtuF.

0360-1315/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compedu.2013.09.011

learners. For example, cognitive tutors rely on rigorous rule-based models of cognitive tasks and step-by-step tracing of students' progress through these tasks in order to provide just-in-time assistance (Anderson, Corbett, Koedinger, & Pelletier, 1995). Constraint-based tutors try to achieve a similar goal by defining the boundaries of correct knowledge in the domain, restricting the possible solution space and reacting when a student's solution violates it (Mitrovic, 2011). Personalized course generation technologies are used to plan sequences of learning objects optimized for individual student's knowledge and learning goals (Brusilovsky & Vassileva, 2003). Tutorial dialogue systems engage in educational conversations with students: They track students' behaviour with the system and use these behavioural traces for monitoring what students know and do not know, and generate dialogue moves that help students to gradually construct their knowledge (Olney, Graesser, & Person, 2010). Adaptive educational hypermedia combines a wide range of techniques in order to tailor the presentation of hypermedia learning content based on student characteristics and navigate students through large corpora of learning documents by generating hyper-links between them and/or decorating these links with personalized cues (Brusilovsky & Henze, 2007).

One more way to organize personalized learning is to provide adaptive formative feedback (Shute & Zapata-Rivera, 2008) or tutoring feedback strategies (Narciss, 2008) to the learners working on exercises. Formative feedback is a key element of formative assessment systems; it provides learners with information about their current state of knowledge in order to improve their learning. Tutoring feedback strategies combine elaborated formative feedback with tutoring and mastery learning strategies. In doing so, they provide such formative feedback that makes learners aware of important gaps existing between their current state of knowledge and their learning goal. Additionally, they provide assistive elaborated feedback (e.g. hints, explanations, and attribute-isolation examples) that is aimed at helping students to detect errors, overcome obstacles, and try more efficient solution paths. In doing so, tutorial feedback strategies offer strategically useful information for task completion without immediately providing the correct solution; they also prompt the learner to apply this information to solve the learning task in the next trial. Furthermore, after successful task completion, they provide confirmatory positive feedback components (cf. Narciss, 2006, 2008, 2012a, 2012b, 2013; Narciss & Huth, 2006). Personalized tutoring feedback has been identified as a very powerful method that expert human tutors apply when helping students to resolve learning difficulties, to monitor their progress, and to optimize the overall learning process (Merrill, Reiser, Ranney, & Trafton, 1992; VanLehn, 2011). Such tailoring of tutoring feedback based on learners' characteristics, and/or parameters of the environments is a promising way to implement adaptive computer-based learning.

Unfortunately, while there is a large body of empirical research on the effectiveness of different types of non-adaptive feedback (Narciss, 2008, 2012, 2013; Hattie & Gan, 2011; Hattie & Timperley, 2007; van der Kleij, Eggen, Timmers, & Veldkamp, 2012; Shute, 2008; Thurlings, Vermeulen, Bastiaens, & Stijnen, 2012), automatic feedback adaptation has received much less attention in empirical and/or theoretical feedback research. Several researchers have suggested frameworks providing guidelines or methods for developing adaptive feedback (Choe, Bae, Kim, & Lee, 2004; Chuang & O'Neil, 2006; Economides, 2006; Gimeno Sanz & De-Siqueira, 2009), yet, these frameworks are only partly rooted in thorough educational theories, and their empirical evaluation with students has been rather scarce. However, a few attempts to implement and test adaptive feedback functionality in a real adaptive educational system (AES) have been made and the results look promising. For example, Vasilyeva and colleagues (Vasilyeva, De Bra, Pechenizkiy, & Puuronen, 2008; Vasilyeva, Pechenizkiy, & De Bra, 2008) as well as Parvez and Blank (2008) have demonstrated the possibility of automatically tailoring instructional feedback to students' learning styles and found a positive effect of adaptive feedback on students' performance. Another study has shown that students themselves believe that adapting feedback may be not needed for successful students, but can be very important if a student is under-achieving (Dennis, Masthoff, & Mellish, 2012). Conati and Manske (2009) investigated adaptive feedback in a serious learning game; their findings indicate that the effectiveness of adaptive feedback can decrease, if it is presented too frequently and provides more information than the learner needs at the moment. Goldin and his colleagues (Goldin, Koedinger, & Aleven, 2012) have data-mined several models measuring individual feedback effectiveness and have found, that for different students the effectiveness of feedback messages varies. This allowed them to conclude that there may be a feedback processing skill or proficiency that depends on the level of details provided by the feedback messages.

These rather mixed results indicate that design and evaluation of adaptive feedback strategies is a challenging task, because so many individual and situational variables can facilitate or hinder the effect of feedback on learning process. Unfortunately, most studies have focused only on one or two of such parameters and investigated their effects mainly on aggregated outcomes (e.g., performance, learning gain). Studies analysing in detail how individual learner and feedback characteristics influence not only the global outcomes, but also more local variables characterizing the immediate effect of feedback on learning process (e.g., post-feedback behaviour) are sparse (Timmers, Broek den, & van den Berg, 2012). Thus, despite the growing attention to feedback-based learning technologies (Author, 2008, 2012, 2013; Hattie & Gan, 2011; van der Kleij et al., 2012; Shute, 2007; Thurlings et al., 2012), many problems in this field remain unresolved. And the big questions, such as "How to design effective tutoring feedback strategies?", "Which factors of learning process should such strategies address in the first place?", and "How these factors influence each other when combined in adaptive feedback components?" have not been answered yet by the research community.

In this paper, we focus on the two latter questions by examining sets of (a) content-related feedback characteristics and (b) learner characteristics influencing learners' problem solving behaviour in computer-based learning tasks supported with tutoring feedback. In many respects, these factors define how students would react to feedback messages and how much they would learn when practicing with an AES providing tutoring feedback, and thus, represent critical information for feedback adaptation. This direction of research is highly important; because even the most thoroughly designed adaptive feedback strategy can be inefficient if students do not use the feedback content in a mindful way in order to improve their learning (see also Timmers et al., 2012). To address these issues, in the remainder of this section, we further synthesize current and past feedback research on the basis of an integrative, multi-dimensional framework for designing and evaluating tutoring feedback strategies (Narciss, 2008, 2012, 2013) and specify our research questions.

1.1. Design of adaptive feedback strategies

In light of recent reviews of feedback research (Narciss, 2008, 2013; Hattie & Gan, 2011; Shute, 2007; Thurlings et al., 2012) and research on computer-based adaptive learning environments (Vandewaetere, Desmet, & Clarebout, 2011), the design and examination of adaptive feedback strategies and, in particular, tutoring feedback strategies require adopting a multidimensional view of feedback. According to Narciss (2012), the nature and quality of a feedback strategy is determined by at least three facets of feedback:

1. Functions of a feedback message or strategy which have to be derived from the instructional goals and objectives relevant in the instructional context (cognitive functions, such as promoting information processing; metacognitive functions, such as fostering self-evaluation and reflection; and motivational functions, such as reinforcing correct responses or encouraging effort and persistence).
2. Content of a feedback message which may include information for verifying if the learner's current state of understanding meets the desired state of understanding, and or information helping to overcome gaps between the current and the desired state of understanding.
3. Formal and technical aspects related to the presentation of the feedback content (e.g., feedback timing and scheduling, sequencing tactics for complex feedback content, and modes of feedback representation and delivery).

Taking into account these facets, instructional designers have to define situational and individual conditions under which the feedback will be provided. Situational conditions help to identify the instructional context in which the feedback strategy will be used (e.g., learning goals and tasks, method of instruction, organization and communication of learning material, as well as the sources of potential learning problems and typical errors). Individual conditions define the characteristics of a learner that might be critical for adaptation, including individual learning objectives, prior knowledge, learning strategies, procedural and meta-cognitive skills, as well as individual motivational prerequisites, such as value of academic achievement (i.e., attainment value), intrinsic task values, academic expectancies (e.g., academic self-concept, self-efficacy), and meta-motivational or volitional skills.

Based on this multidimensional view on feedback, Author has proposed the Interactive-Tutoring-Feedback (ITF) model that encapsulates the state of the art in developing feedback strategies for interactive learning tasks (Narciss, 2006, 2008, 2013). According to the ITF model, a feedback strategy can be defined as a coordinated plan integrating clear and decisive statements specifying at least the following aspects of a learning process with feedback (Narciss, 2012):

- *scope and function* – what (instructional) goals or purposes the feedback serves,
- *content* – what information is included in the feedback,
- *presentation* – in which form and modes the feedback content is presented to a learner,
- *conditions* – under which situational and individual conditions the feedback is provided,
- *timing and schedule* – which events within the learning process trigger feedback messages.

Such definition of a feedback strategy supports a broad range of possibilities in terms of designing personalized and adaptive feedback strategies. For example, one or several feedback functions, simple or complex feedback content, uni-modal or multi-modal feedback types can be tailored to one or even a set of learner characteristics. From the point of interactivity, feedback adaptation can be implemented in a static or dynamic way. In the case of static adaptation, feedback settings are adjusted once according to the global task and/or learner characteristics. Dynamic adaptation implies that the decision about feedback settings for the current learning interaction is made on the fly based on varying parameters of the instructional context (knowledge state of the learner, history of interaction, motivational factors, etc.).

Design of effective strategies for adapting tutoring feedback based on such a multitude of options requires thorough empirical research that should help to identify which of the strategy factors warrant adaptation (Shute & Zapata-Rivera, 2008; Vandewaetere et al., 2011). To be more specific, it should be investigated which learner characteristics and actions are crucial for effective learning with feedback.

1.2. Content-related factors affecting feedback efficiency

A large body of research comparing the effects of feedback with different contents reveals a variety of non-adaptive feedback types and strategies originating from different instructional approaches (Narciss, 2008, 2012, 2013; Hattie & Gan, 2011; Hattie & Timperley, 2007; van der Kleij et al., 2012; Mory, 2004; Shute, 2008; Thurlings et al., 2012). The following types of feedback (and accompanying interaction patterns) have been identified in the literature (in the order of elaboration):

1. Knowledge of performance provides summative feedback after a set of tasks (e.g., percentage of correctly solved tasks, number of errors).
2. Knowledge of result offers information on the correctness or quality of a response (e.g., correct/incorrect; excellent/poor).
3. Knowledge of the correct response provides the correct response to the given task.
4. Answer until correct provides knowledge of result and the opportunity to make further attempts until the task is solved correctly.
5. Multiple try feedback (knowledge of result and the opportunity to make a limited number of further attempts to solve the task).
6. Elaborated feedback (additional information besides knowledge of result or knowledge of the correct response). There is a variety of elaborate information that might be added to knowledge of result. Narciss (2008) suggests differentiating between at least five categories of elaborated feedback:
   - Kowledge about task constraints, offering information on task rules, task constraints and/or task requirements;
   - knowledge about concepts, addressing conceptual knowledge (e.g., response-contingent hints on concept attributes, or attribute-isolation examples);
   - knowledge about mistakes, offering information on errors or mistakes (e.g., flagging location of errors, providing hints on types or sources of errors),
   - knowledge about how to process the task, addressing procedural knowledge (e.g., task-contingent hints about procedural skills or problem solving strategies),
   - knowledge helping to apply and develop meta-cognitive skills and strategies necessary for self-regulated learning processes (e.g., topic-contingent hints about useful sources of information).

Given such a large variety of feedback content types, it is not surprising that existing findings on the effectiveness of providing learners with different feedback content are rather inconclusive. Some studies have demonstrated positive effects of certain feedback types, while others show no effect or even negative effects of feedback (for further details, see Narciss, 2008; Kluger & DeNisi, 1996; Mory, 2004; Shute, 2008; Thurlings et al., 2012).

In this study, we experiment with the two content-related factors of tutoring feedback: feedback specificity (concise hints versus elaborate explanations), and the type of knowledge communicated by feedback (procedural versus conceptual). Feedback specificity refers to the level of elaborated information provided by a feedback message (e.g., Goodman, Wood, & Chen, 2011). Recently, Serge and his colleagues have found that a higher degree of feedback specificity contributes to faster learning (Serge, Priest, Durlach, & Johnson, 2012). Furthermore, a study by Lin, Atkinson, Christopherson, Joseph, and Harrison (2013) reveals that the degree of feedback specificity matters particularly if the feedback is provided by an animated agent.

During the last decade the importance of procedural skills and conceptual understanding has been actively debated, especially in the mathematics education literature (Bokhove & Drijvers, 2012). However, until now, there have not been enough experimental studies investigating if and how feedback hints or explanations providing procedural or conceptual information influence learners' post-feedback behaviour and performance.

1.3. Learner characteristics affecting feedback efficiency

Learners differ from each other in many ways including prior knowledge, meta-cognitive skills, motivational and affective state, or learning strategies and styles. Thus, there are many individual factors that may influence how feedback is processed by each learner. Consequently, a variety of individual factors can be used to support the design of personalized feedback strategies and implementation of AESs which can deliver feedback messages that are tailored to the characteristics of an individual learner or a category of learners.

Traditionally, prior knowledge is considered to be the core factor for adapting instruction to an individual learner (see, for example, the reviews by Wittwer & Renkl, 2008; Wulfeck, 2009). Accordingly, the learner characteristics, which have been addressed the most by feedback research, especially in the field of AES, are the learners' current and/or prior level of knowledge (e.g., Gertner, Conati, & VanLehn, 1998; Hancock, Thurman, & Hubbard, 1995). Research on adapting instruction to the level of knowledge has shown that, in general, learners with higher levels of knowledge need less guidance and support than learners with lower levels of knowledge (e.g., Tobias, 1989, 1994). Interestingly, Smits, Boon, Sluijsmans, and Van Gog (2008) found that learners characterized by high levels of prior knowledge were able to learn more with less elaborate feedback, even though more specific feedback was perceived more positively. Moreover, that study also showed that for learners with lower levels of prior knowledge, feedback specificity and timing did not affect performance and feedback perceptions at all (Smits et al., 2008).

Learner motivation comprises another set of characteristics that have a great impact on learning in general. Expectancy-value theories of learners' motivation (e.g., Eccles & Wigfield, 2002; Pintrich, 2003) suggest that expectancies or beliefs regarding ones competencies (hereafter referred to as perceptions of competence), achievement-related value (i.e., attainment value), intrinsic values (i.e., interest, activity-related enjoyment), and fear of failure are critical motivational learner variables. Some of these variables have been investigated with regard to their impact on feedback processing and efficiency. Variables that have been investigated so far include expectancies (e.g., Narciss & Huth, 2004; Timmers et al., 2012; Tuckman & Sexton, 1992), goal orientation (e.g., Senko & Harackiewicz, 2005), and task-value beliefs (e.g., Timmers et al., 2012). Altogether, these studies indicate the importance of including several motivational variables into the research on feedback processing and efficiency.

Research in mathematics education has also identified gender as an important factor influencing students' performance, strategies, and activities (Gallagher & Kaufman, 2004). Empirical evidence suggesting that gender does not only matter in classroom contexts, but also when students use tutoring systems, has been provided by Arroyo and her colleagues (Arroyo, Beck, Beal, Wing, & Woolf, 2001; Arroyo, Woolf, & Beal, 2006; Arroyo et al., 2012). The findings of their studies using the tutoring systems AnimalWatch, and WayangOutpost indicate that in general, male K-12 students show more maladaptive behaviours than girls (e.g., gaming the system, skipping, not engaging seriously into problem solving). Accordingly, the cognitive and affective/motivational benefits of learning with these tutoring systems have been larger for girls than for boys. Yet, in a recent study taking learners' level of prior knowledge into account, low achieving students benefitted more – irrespective of their gender. This indicates that gender differences do not occur consistently, but rather under certain conditions. Thus, Arroyo concludes that, based on these findings, it is still too early to establish clear principles regarding the issue of how to tailor feedback content specifically to male and female students (Arroyo et al., 2012).

Some metacognitive skills, such as self-assessment, are also considered to be crucial factors for feedback processing (e.g., Narciss, 2008). Therefore, a substantial amount of psychological research has focused on the relation between feedback effectiveness and the learner's metacognitive state, generally measured by his/her response certitude (e.g., Hancock, Stock, & Kulhavy, 1992; Hancock et al., 1992; Mory, 1991, 2004). Recently, another set of metacognitive skills, such as help-seeking (Roll, Aleven, McLaren, & Koedinger, 2011) and general feedback processing (Goldin et al., 2012), have received attention by feedback researchers from the Intelligent Tutoring System community (instead of eliciting response certitude, they relied on educational data mining techniques). Finally, individual learning style has been considered for feedback adaptation. Even though there has been a lot of controversy and debate about the value of this learner characteristic (e.g., Akbulut & Cardak, 2012; Cook, 2012; Pashler, McDaniel, Rohrer, & Bjork, 2008; Popescu, 2009; Wulfeck, 2009), several recent AES papers demonstrate promising results (e.g., Mavroudi, 2012; Parvez & Blank, 2008).

To recapitulate, a variety of learner characteristics might be critical for feedback processing and efficiency. Vandewaetere and colleagues (2011) classify the individual characteristics which have been addressed by empirical research on the effectiveness of adaptive systems into three groups of factors: (a) cognitive factors (e.g., knowledge, working memory capacity, intelligence, cognitive style, and goal orientation), (b) affective factors (e.g., motivation, mood, certainty), and (c) behavioural factors related closely to the cognitive and affective factors (e.g., help seeking, self-regulation, or behavioural parameters such as number of tries, exercises worked on etc.). Furthermore, they point to the necessity of distinguishing between (relatively) stable learner characteristics (e.g., gender, intelligence, learning or cognitive styles), and learner characteristics that develop over time, and thus may change during learning (e.g., knowledge, skills, motivation, and behaviour).

The latter distinction draws attention to the issues of how relatively stable learner characteristics are linked to dynamic learner characteristics, and may have combined, complementary, and/or compensatory effects on students' learning, motivation, and behaviour when they practice with an AES providing tutoring feedback strategies. There are only a few studies addressing these issues by investigating the role of gender for learning with tutoring systems (e.g., Arroyo et al., 2006, 2012). The results of these studies reveal that the interplay between gender and dynamic learner factors is rather complex and disserves further investigation.

1.4. Purposes and research questions

The overview of prior work indicates that there are still major gaps in feedback literature, particularly in relation to interactions between (content-related) feedback aspects, and learner characteristics that potentially influence feedback processing and effects (Shute, 2008). Furthermore, the overview suggests that the focus of prior studies has been on investigating the effects of the few selected variables on learning outcomes. Modern multidimensional views on feedback strategies indicate, however, that a broader range of individual learner characteristics and various feedback parameters may influence feedback efficiency.

The purposes of this work are threefold: First, to identify relatively stable learner characteristics that affect learning outcomes, namely students' knowledge gain and motivation when practicing fraction tasks with feedback. To address this issue we investigated the relations between and the impact of specific learner characteristics for which prior research provided partial evidence that they can be relevant for developing personalized or adaptive feedback strategies, namely gender, knowledge, and several motivational variables (intrinsic value, attainment value, perceived competence, and fear of failure).

Second, to investigate the impact of identified learner characteristics on how tutoring feedback strategies varying in content (procedural vs. conceptual) and specificity of feedback messages (concise hints vs. elaborated explanations) affect a student's learning and motivation.

Third, to explore if and how feedback content and identified learner characteristics influence a student's post-feedback behaviour. To address this issue we conducted detailed log-file analyses of an experimental study in which learners were exposed to various tutoring feedback strategies while working on multi-trial error correction tasks presented by the web-based AES ActiveMath (cf. Section 2). The focus of these log-file analyses is on two distinct features of learners' problem solving behaviour before and after they were provided with feedback: first, how often do they skip versus try to solve the task (quantity of input); second, how often do they enter a correct versus incorrect response (quality of input). We are particularly interested in the amount of post-feedback skipping and solving for the following reasons: If learners with certain profiles of individual characteristics often skip further correction attempts after receiving a feedback message, such feedback messages have to be considered ineffective for this category of learners, because they neither help the learners to find the correct response nor do they succeed in encouraging learners to try again. If these learners make another attempt (correct or not) after a feedback message, such a message can be considered effective in terms of keeping learners on-task. On the other hand, if a feedback message results in regular correct responses from a certain category of learners, this means it effectively supports such learners in transitioning from not being able to solve a task to being able to solve it. Thus, we assume that the analyses of post-feedback solving and skipping behaviour can provide valuable information for developing personalized feedback strategies.

To summarize, this study aims to contribute empirical findings to the following research questions:

1. Which relatively stable learner characteristics are related to learning outcomes when students are provided with various feedback strategies?
2. Are there differential effects of feedback strategies on students' knowledge gain from pre- to post-test due to relatively stable learner variables?
3. Are there differential effects of feedback strategies on changes in students' motivation due to relatively stable learner variables?
4. Which learner characteristics are related to students' post-feedback behaviour when they work on solving error correction tasks?
5. Are there differential effects of feedback content and relatively stable learner variables on post-feedback skipping rates?
6. Are there differential effects of feedback content and relatively stable learner variables on post-feedback success rates?

Since there is empirical evidence that gender matters in mathematics education (Gallagher & Kaufman, 2004) and when students learn with tutoring systems (e.g., Arroyo et al., 2012) we are particularly interested in the role of gender and its interplay with students' prior knowledge and motivational characteristics (i.e., perceived competence, intrinsic motivation, attainment value, and fear of failure).

The rest of the paper is structured as follows. Section 2 presents the AES ActiveMath used in our experiment; it especially focuses on the exercise subsystem of ActiveMath and its features that allow for dynamic composition of feedback strategies. Sections 3–5 describe the empirical study including its design, the collected data, and the applied statistical methods; special attention is given to the tasks with typical errors used during the treatment phase of the study. Finally, Section 6 summarizes the results of the experiment and concludes the paper with a discussion and implications for future work.

2. Problem-solving/exercise environment

In this section we introduce ActiveMath, the Web-based AES which the reported study is based on. It maintains adaptive access to mathematical learning content, is highly customizable, and offers a broad range of tools for both learners and teachers. ActiveMath allows learners to work on learning material at their own pace, and provides various degrees of support: from tutoring feedback during problem solving to personalized course generation as well as adaptive navigation through learning material (for a more detailed description of ActiveMath features see Melis et al., 2006; Sosnovsky et al., 2013). The architecture of ActiveMath contains all of the traditional components of intelligent tutoring systems, including the semantic domain model, the overlay learner model, and a set of adaptation modules.

A distinctive feature of ActiveMath is its sophisticated exercise framework. It supports a variety of interaction types and elements. ActiveMath exercises can consist of one or multiple steps; they can combine various kinds of questions, from single and multiple choice to mapping and fill-in-blank; they can ask a student to calculate a numeric answer, or derive a formula; exercise feedback can vary from mere signalling of correct and incorrect attempts to detailed hints and explanations. In addition, ActiveMath features a unique mechanism that allows to apply different feedback strategies, modifying the runtime behaviour of an exercise. In the described experiment (see Section 3), this mechanism was used to implement several feedback strategies, altering the sequence and content of feedback messages within every training exercise depending on the assigned condition.

2.1. Representation of exercises

Exercises in ActiveMath are represented as finite state machines (FSM) consisting of nodes and transitions between them (Goguadze, 2009). The nodes represent stages of exercise solutions, and can be one of the three following types:

1. Task (or problem statement for the next step);
2. Interaction (soliciting learner's input);
3. Feedback (providing the system's reaction to the learner's input).

The transitions between the nodes define the interaction pattern of the exercise by connecting system-generated stages with correct and incorrect learners' inputs. To automatically assess the correctness of user inputs, ActiveMath either relies on its own exercise subsystem, or connects to external tools such as Computer Algebra Systems (CAS). These tools can be used to evaluate complex mathematical expressions. Exercises and their individual steps can be associated with the elements of ActiveMath domain models: concepts that the exercise trains or tests, or misconceptions that the exercise helps to identify. By solving exercises and their steps, learners provide ActiveMath with evidence of mastering corresponding concepts and/or having corresponding misconceptions.

2.2. Exercise strategies

Each problem statement can be combined with different tutorial strategies in order to produce versions of an exercise with different interaction patterns. These strategies can vary in terms of feedback type (e.g., hint vs. explanation), feedback timing (delayed vs. immediate), feedback initiative (solicited by the learner vs. unsolicited help generated by the system), number of attempts allowed per step, etc. ActiveMath provides means for the automatic enrichment of exercise representations with different tutorial strategies that change the exercise behaviour depending on the current state of the learner model, or on other required parameters.

Consider the following example: Fig. 1 visualizes two simple tutorial strategies defining different interaction paths for the same exercise. The initial task nodes of both exercise graphs ask a learner to compute the sum of two fractions: $1/2 + 1/3$. In terms of assessment, both versions of the exercise process the learner's responses in a similar way:

1. A learner can provide the final correct solution right away: $1/2 + 1/3 = 5/6$.
2. A learner can find the common denominator of the fractions and then compute the solution: $1/2 + 1/3 = 3/6 + 2/6 = 5/6$.
3. Any other response (on either of the two steps) is recognized as a mistake.

However, if the learner submits an incorrect response, the reaction of the exercise differs depending on the applied strategy. In Fig. 1, the version on the left simply provides knowledge of result (KR) by signalling that the answer is incorrect, while the version on the right illustrates an answer-until-correct strategy, which provides KR and asks the learner to attempt the step again and again until the correct answer is given.

From a technical perspective, a tutorial strategy is a script-based transformation of one exercise automaton into another exercise automaton, which expresses the tutorial decisions required to achieve the instructional goals of the strategy.

2.3. Structured Templates for Exercise Progressive Solutions

In the domain of fraction arithmetic (as in many other domains), only the most trivial exercises can be solved in one step. Typically, learners are required to apply a series of operations in order to solve a fraction exercise. Moreover, there can be several paths leading to the

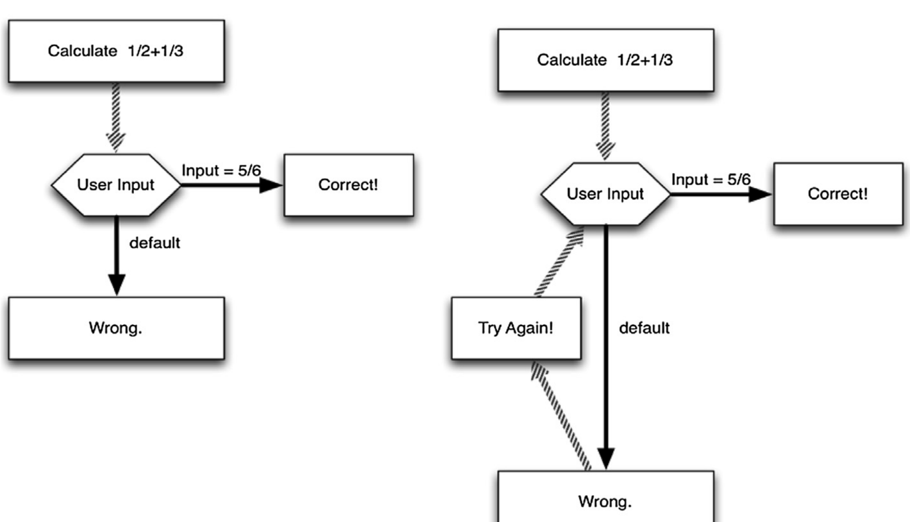

Fig. 1. Two simple exercise strategies.

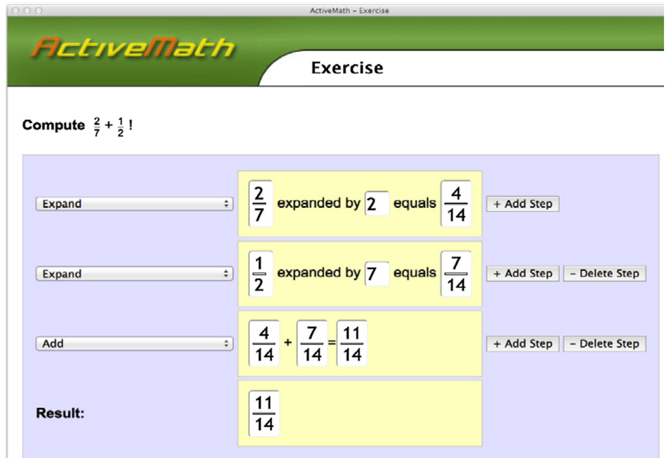

Fig. 2. Screenshot of fraction task with the STEPS-Interface (translated to English).

final answer of an exercise. Predicting the best granularity of the solution path, as well as accounting for all the possible paths, is challenging and often impractical – the exercise graph would explode. Yet, providing the learners with an opportunity to work with exercises in a progressive manner, approaching the solution step-by-step, is instructionally effective.

We have developed, evaluated, and iteratively improved a special interface allowing students to design their own solution paths based on a predefined set of operations (Andrès, 2012; Eichelmann, Narciss, Schnaubert, Melis, & Goguadze, 2011, Eichelmann, Andrès, Schnaubert, Narciss, & Sosnovsky, 2012). We call this interface *Structured Templates for Exercise Progressive Solutions* (STEPS). Fig. 2 presents an example of using the STEPS interface for computing a sum of two fractions with unlike denominators.[1] When the exercise starts, the learner is presented with a single input field that is used to supply the final solution. On top of it, an interactive element allows for manipulation of intermediate steps. It consists of a drop-down menu, which the learner can use to specify the intention of the step. Its entries consist of operations that are relevant to fraction-addition problems: expand, reduce, find prime factors, find common denominator, compute inverse, transform, add, subtract, multiply, divide, find least common multiple, find greatest common divisor. When an operation is selected, a corresponding template is inserted into the next line which provides learners with answer fields for the respective operation. For instance, if learners choose to expand a fraction, they will be presented with the template "□ expanded with □ is □". This approach decouples learners' intentions (e.g., expanding a fraction), from the actual execution (computing the expansion), and eliminates the need to guess what learners are trying to achieve. Learners may insert as many steps as needed; they can delete steps, modify previous steps, and add new steps at any place in the solution path. The interaction ends once the learner submits the final answer.

3. Methods

This study is based on data from the interdisciplinary project, "Adaptive Tutoring Feedback" (AtuF), which is conducted by the Psychology of Learning and Instruction Research Group of the Technische Universität Dresden, and the Centre for e-learning Technology of the German Center of Artificial Intelligence. The following descriptions are limited to the part of the evaluation study that addresses the effects of feedback content and individual learner characteristics on achievement, motivation, and post-feedback behaviour of learners who solve fraction addition tasks with the ActiveMath environment described in Section 2.

3.1. Participants and design

This study was conducted with 186 sixth and seventh graders who were recruited by advertisements placed in schools, sports centres, and museums, as well as in a local newspaper in a large German city. The participants were between 10 and 14 years old (mean age = 12.01; SD .76), and were rewarded with 10 Euros for participation. The sixth graders had completed at least the basic level units of the school fraction curriculum (e.g., basic fraction concepts, ordering fractions, adding and subtracting fractions with common and unlike denominators); the seventh graders had completed the second level units (e.g., computations with all kinds of fractions, multiplication and division of fractions, etc.).

The study used a pre-test/treatment/post-test design. During the treatment phase, the participants were provided with three sets of 10 multi-trial fractions tasks. Each task was designed to contain a typical fraction error; we will refer to them hereafter as tasks with typical errors (TWTEs). For each of the TWTEs learners were asked to detect and correct the error. All learners received confirmatory or corrective feedback indicating the correct location of the error. The feedback that learners received with respect to their error correction step varied according to five different feedback strategies (four tutoring strategies and one control strategy).

Each of the four tutoring feedback strategies provides: (a) knowledge of result and a hint after the first failed correction attempt, (b) knowledge of result and an explanation after the second failed correction attempt, and (c) knowledge of the correct response in form of a worked-out example after the third failed correction attempt. The feedback content of the tutoring hints and explanations address either

[1] The experiment described in this paper was conducted with German schoolchildren. Therefore, the language of the system interface as well as the language of all learning tasks and feedback messages was German. Figs. 2, 4 and 5 present English translations of original system screenshots.

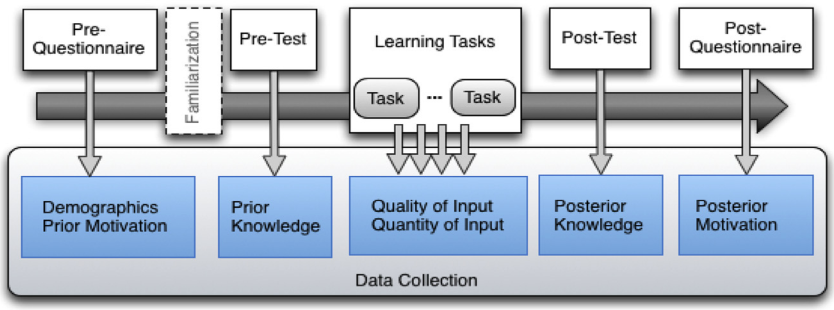

Fig. 3. Study workflow.

procedural or conceptual knowledge necessary to correct the typical error (see Section 3.4 for further details). The tutoring feedback strategies combine procedural and conceptual hints and explanations, which lead to the following four strategies:

- Conceptual hint – conceptual explanation – worked-out example (CH-CE-WE).
- Conceptual hint – procedural explanation – worked-out example (CH-PE-WE).
- Procedural hint – conceptual explanation – worked-out example (PH-CE-WE).
- Procedural hint – procedural explanation – worked-out example (PH-PE-WE).

Learners were either randomly assigned to one of the tutoring feedback conditions, or to the control condition. In the control condition learners received knowledge of result after the first correction attempt, and knowledge of the correct response in the form of a worked-out example after the second attempt:

- Knowledge-of-result – worked-out example (KR-WE).

It is important to emphasize that knowledge-of-result feedback was part of every hint and explanation.

3.2. Procedure

Participants worked on laptop computers connected to the Internet. All tests, questionnaires, and learning tasks were administered using the ActiveMath system. Instructions were given in written form, explained by trained experimenters, and supported by a projector presentation. The experiment consisted of six phases: pre-Questionnaire, familiarization, pre-test, learning phase, post-test, and post-questionnaire; Fig. 3 illustrates this workflow in detail. After the experimenters provided a short explanation about the purpose of the AtuF project, participants were asked to fill in the pre-questionnaire, which addressed demographic and motivational measures. Then, participants were given several trial exercises to help familiarize them with the STEPS-interface and the overall exercise environment of

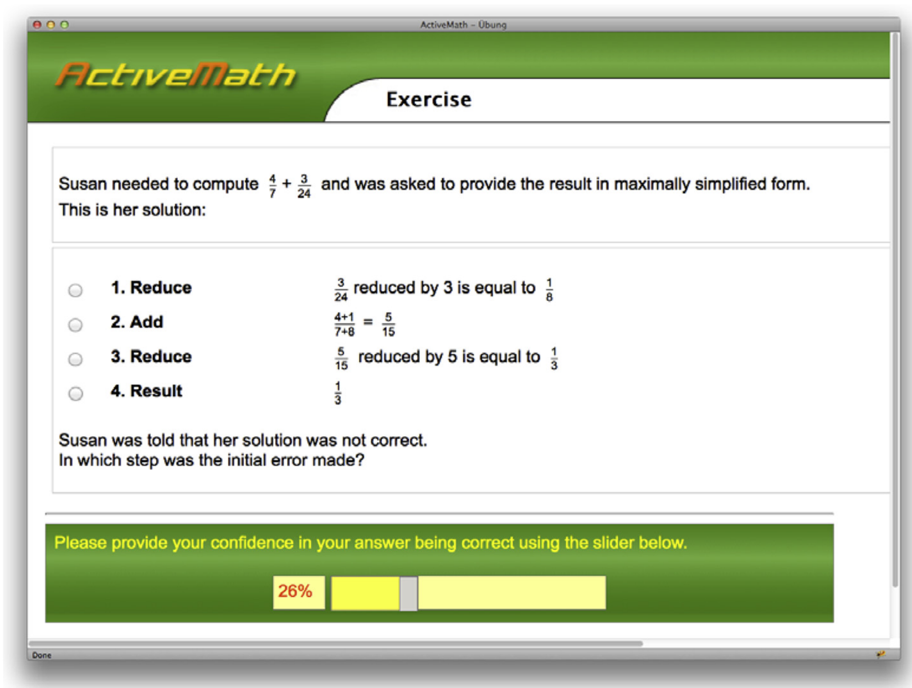

Fig. 4. Screenshot of a TWTE error-detection step (translated to English).

ActiveMath. Next, participants took a 15-min pre-test assessing their prior knowledge concerning fractions. After a short break of approximately 10 min, the participants worked on TWTEs for 45 min. During this phase, the participants were free to work at their own pace, and were provided with three sets of 10 TWTEs addressing various fraction competencies (i.e., represent fractions in various formats; add fractions with common denominators, order fractions, expand fractions, and add fractions with unlike denominators). Furthermore, the participants were supported by different feedback strategies, which they were randomly assigned to in the beginning of the experiment. After another 10-min break, the learners took a 15-min post-test, similar to the pre-test, and again, filled in a motivational questionnaire.

3.3. Learning tasks: tasks with typical errors

The TWTEs provided in the treatment phase are multi-trial exercises presenting step-by-step solutions to fraction problems that contain typical errors in the domain. These TWTEs were developed and evaluated on the basis of psychological and empirical task and error analyses (Eichelmann, Narciss, Faulhaber, & Melis, 2008, Melis, Faulhaber, & Eichelmann, 2008; Eichelmann, Narciss, Schnaubert, & Melis, 2012). Each TWTE consists of two distinct subtasks: An error-detection subtask in multiple-choice response format (i.e. learners mark the location of the error, cf. Fig. 4), and an error-correction subtask. The error-correction subtask had to be completed by using a restricted version of the STEPS interface. Participants had to choose the desired action from a drop-down menu (e.g., “expand” or “add”), and had to execute that action accordingly (cf. Fig. 5).

Participants had one trial to detect the error and up to three trials to correct it. Additionally, they were required to indicate their response confidence for error detection and the first trial of error correction. If participants failed to correct the error, they received tutoring feedback and were asked to try again. After the third (or a successful) trial, participants received a worked-out solution to the problem and moved on to the next task.

3.4. Feedback content and strategies for the error-correction subtask

For each of the addressed task requirements in the error-correction subtask of the learning exercises, procedural and conceptual tutoring feedback hints and explanations were developed on the basis of cognitive task- and error-analyses (Eichelmann et al., 2012).

- Procedural hints and explanations focused on the arithmetic operations that had to be used for solving a target exercise. The procedural hints were developed based on fundamental rules of fraction learning. While these hints merely pointed out the appropriate strategy to use, or mentioned one key aspect of the arithmetic operation (e.g., “When expanding a fraction, one alters the numerator and the denominator equally.”), explanations elaborated the hint information with more details about how to put this procedure into practice this procedure (e.g., “When expanding a fraction, one alters the numerator and denominator equally. To do this, multiply the numerator and denominator by the same number.”).
- Conceptual hints and explanations focused on the conceptual knowledge important for solving the target exercise. The conceptual hints and explanations were developed based on definitions of key concepts and core theorems in the domain of fractions. While the conceptual hints provided some information about the concepts addressed in the problem at hand (e.g., “When expanding a fraction, its value must not change.”), the explanations also contained the hint, but either provided additional information about the concepts that had to be applied in order to solve the task, or explained how a concept was relevant to the problem at hand (e.g., “When expanding a fraction, its value must not change. While expanding, the denominator increases, that means the partitioning becomes more fine-grained. But since the value of the whole fraction does not change, the numerator has to be altered in the same way.”).

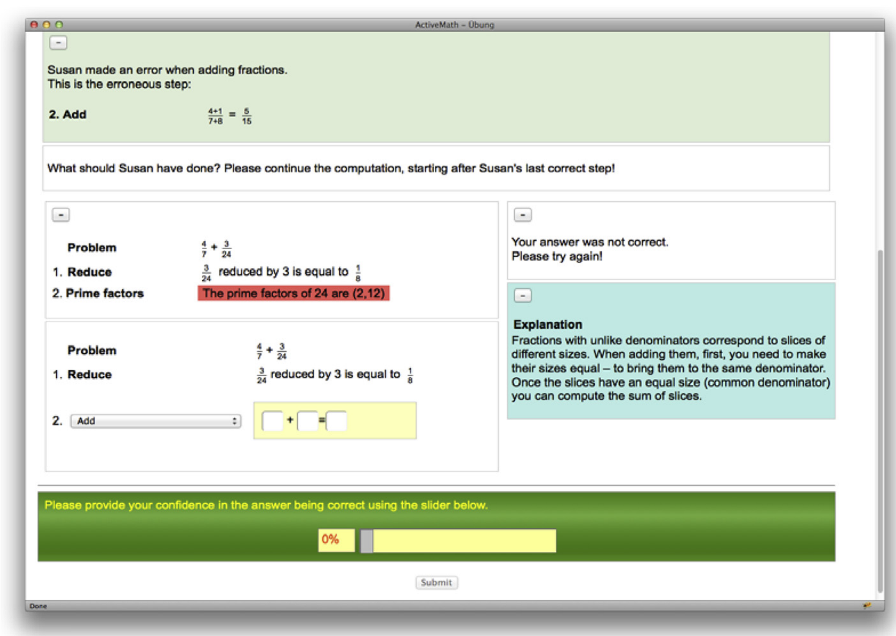

Fig. 5. Screenshot of a TWTE error-correction step (translated to English).

The tutoring feedback strategies for the error-correction tasks combine the procedural and conceptual hints as well as explanations, and present them to learners with a three-trial feedback strategy. If a learner succeeds in correcting an error, the system presents a confirmatory positive knowledge-of-result feedback message (e.g., "Well done"), followed by the presentation of a worked-out solution to the problem (knowledge-of-correct-result, KCR). The learner is then directed to the next task. If the first error-correction attempt is failed, the system provides negative knowledge-of-result feedback (e.g., "your response is not correct"), along with a tutoring hint, which contains either procedural or conceptual information depending on the feedback strategy, and prompts the learner to try again. If the second error-correction attempt is incorrect, the system provides negative knowledge-of-result feedback (KR-), and a tutoring explanation (procedural or conceptual), and prompts the student to try again. After the third failed error-correction attempt, the system provides KR- and KCR in the form of a worked-out solution, concludes the interaction, and directs the learner to the next exercise.

Learners in the control condition did not receive tutoring feedback and were only allowed one trial to correct the error, which was followed by KR as well as KCR in form of a worked-out solution.

Such a design, based on the three-trial tutoring feedback strategy with a stepwise increase of feedback specificity, has been chosen, because (a) several reviews and syntheses of feedback research recommend it (e.g., Narciss, 2008; 2012b; Hattie & Gan, 2011; Shute, 2008), and (b) it allows investigating how learners' post-feedback behaviour in each step relates to learner, task, and feedback characteristics.

The systematic between-subjects variation of the knowledge type addressed by feedback content (procedural vs. conceptual) was included in order to contribute empirical findings to the controversial debate on beneficial vs. detrimental effects of providing learners with procedural hints or explanations. Since some mathematics education researchers recommend providing learners with both, procedural and conceptual explanations (Bokhove & Drijvers, 2012), we included also the two tutoring feedback strategies that combine procedural and conceptual feedback messages.

3.5. Measures

3.5.1. Demographics

Demographic variables (e.g., gender, age, and school background) were assessed by the demographic questionnaire in the beginning of the study.

3.5.2. Motivation

Learners' motivation was assessed via the Expectancy-value-Form of domain-specific Learning Motivation (EWF-LM). The EWF-LM was developed on the basis of an integrative expectancy-value model of motivation (Narciss, 2006), which is rooted in Heckhausen's cognitive expectancy-value model of motivation (Heckhausen, 1989), Eccles's and Wiegfield's social-cognitive expectancy-value model (Eccles & Wigfield, 2002), and Pintrich's framework for motivation in education (Pintrich, 2003). Following the integrative expectancy-value model, the EWF-LM consists of four scales: intrinsic value (eight items; e.g., "I enjoy solving fraction exercises"); attainment value and perceived competence (six items; e.g., "Fraction tasks offer me an exciting occasion to demonstrate my abilities", "I am good at solving fraction tasks"); and fear of failure (three items; e.g., "I am worried about mistakes, even if nobody would see them") in relation to fraction tasks. Participants responded to the EWF-LM items on a 6-point rating scale (0 = "*Not true at all for me*"; 6 = "*Completely true for me*"). The development and iterative improvement of these scales was carried out in several phases. Former versions of the EWF-LM have been used in prior studies in other domains, e.g. subtraction tasks (Narciss & Huth, 2006) and scientific writing (Proske, Narciss, & McNamara, 2012). The EWF-LM was evaluated with a sample of 210 sixth and seventh graders. The psychometric values of the items and scales were satisfactory. Cronbach's alpha for the scales ranged from .79 to .89 (Puta, Schnaubert, & Narciss, 2013).

3.5.3. Fraction performance

To assess participants' levels of performance in the domain of adding fractions, we designed two parallel versions of a fraction test consisting of 24 items addressing five relevant task requirements: adding fractions with common denominators, adding fractions with unlike denominators, expanding fractions, representing fractions (e.g., transform a fraction visualized as a figure into a numerical form), and ordering fractions according to their size. The item difficulty of these types of items ranged between .11 and .89, as revealed by a pilot study conducted with sixth and seventh graders. Learners received up to four sets of exercises, depending on their working speed. All of these sets contained respectively one exercise for each task requirement presented in the following order:

1. Add two fractions with a common denominator.
2. Add two fractions with unlike denominators.
3. Expand a fraction with a given number.
4. Represent a fraction.
5. Order fractions according to size.

Participants were given 15 min to respond to the test items at their own pace using the STEPS-interface. Students were asked to complete as many exercises as they could within the 15 min. No feedback was provided. The number of correctly solved exercises was used as an indicator of the learners' performance on fraction arithmetic tasks.

3.6. Immediate behavioural reactions to task and feedback

To analyse the learners' behaviour within the multi-trial learning tasks, we differentiated between system-provided trials (= steps) and the learners' reaction (=behaviour). The error-correction subtask was implemented as a multi-trial procedure supported by tutoring feedback in order to focus on the specific learner behaviour within each step:

1. The initial step following the presentation of the error-correction subtask, which included information (KR, KCR) about the error-detection subtask (cf. Fig. 4);
2. the second step following the tutoring hint (cf. Fig. 5);
3. the third step following the tutoring explanation.

For each of the three steps we analysed two characteristics of the learners' inputs: (1) the quality of input (i.e., the correctness of the input with regard to the task requirements), and (2) the quantity of input (i.e., presence/absence of input).

3.6.1. Solving: quality of input

The system automatically classified the learners' input as correct or incorrect. A correct solution was defined as "any step that might have been helpful in solving the task and was executed correctly". For example, if the student was asked to correct a common denominator within a task of "adding fractions with unlike denominators", the student was able to approach the task in several ways: do a prime-factorization of the denominators to find a common denominator, state a correct common denominator, convert any of the two fractions to the correct denominator, or add previously converted fractions – under the condition that the student initially converted the fractions correctly. The key-question for the binary variable concerning the quality of input was: did the student provide a correct solution with regard to the task requirements or not?

3.6.2. Skipping: quantity of input

If a student submitted blank input fields, the behaviour was defined as a skip. If a learner provided any type of input, the behaviour was defined as a non-skip, regardless of elaboration and/or correctness of the input. For instance, it was not counted as a skip if learners chose which operation to apply (e.g., expand a fraction), even if they did not enter anything in the STEPS template associated with this operation. Only if a student did not provide the system with any information (ergo, if the "Submit"-button was clicked without using the STEPS interface), the behaviour was regarded as a skip. The key question for this binary variable was: did the student provide any possible solution to the problem at hand, or not?

If a student produced a continuous sequence of skips, only the first skip was included in further analysis. In other words, consecutively skipping all three trials did not contribute to increasing skipping counts for trials 2 and 3. We interpret the occurrence of skipping as a transition from trying, to giving-up. If a student made the decision to start skipping and did not change this decision throughout the following trials (either consciously, or unconsciously, without even looking at feedback messages), it would not make sense to associate repetitive skipping with the effect of more elaborate feedback.

4. Data sample and analyses

When working on TWTEs, learners were allowed to work at their own pace, and to attempt as many tasks as possible within the 45-treatment session. We filtered out the data logs of participants who worked too slowly to even complete the first TWTE set. We also excluded several incomplete logs (data was missing due to system failures). The reduced sample consists of data from 124 participants (57 females, 67 males) distributed per condition as follows:

- CH-CE-WE: 23 total = 12 male + 11 female;
- CH-PE-WE: 26 total = 11 male + 15 female;
- PH-CE-WE: 27 total = 14 male + 13 female;
- PH-PE-WE: 25 total = 14 male + 11 female;
- KR-WE: 23 total = 16 male + 7 female.

To ensure that the results that were derived from this reduced sample would also hold for the complete sample, first, we conducted a preliminary analysis with the complete sample (except for the analyses of post-feedback behaviour), and then, repeated it with the reduced sample. The patterns of findings for these samples were consistent. Thus, hereafter we will only report the analyses and results for the reduced sample.

When analysing the post-feedback behaviour, we excluded the data sets of the control group because the participants of this group were not provided with tutoring feedback components ($N = 101$). Furthermore, we excluded the logs for the two TWTEs that required the ordering of fractions because the implementation of these tasks did not follow the same algorithm as the implementation of the eight other tasks (i.e. instead of separate error-detection and error-correction tasks, learners were asked to identify a correct procedure and execute it in one single step). Consequently, the analyses of post-feedback behaviour were conducted with the total of 808 task completions. The resulting task sample was balanced according to task-requirements.

For each of the 808 task completions, we obtained the behavioural data for the initial step, because every learner had to answer the error-correction subtask at least once. However, for the second and third steps (following the hint and the explanation), the data samples were smaller, as some learners were able to solve the tasks without feedback and some gave up on the task in the first step. The final dataset included 808 task completions for the initial step, 310 for the second step (808–478 solve Step 1–20 skip Step 1), and 165 for the third step (310–58 solve Step 2–87 skip Step 2). Fig. 6 visualizes this dataset.

5. Results

This work examines the extent to which learner characteristics and feedback content impact learners' improvement of performance when practicing fraction arithmetic with TWTEs and various feedback strategies. Second, it explores whether learner characteristics and feedback content influence learners' post-feedback behaviour, particularly their skipping behaviour and their error-correction achievement after they had received feedback. To this end, we assessed cognitive as well as several motivational learner variables prior to, as well as

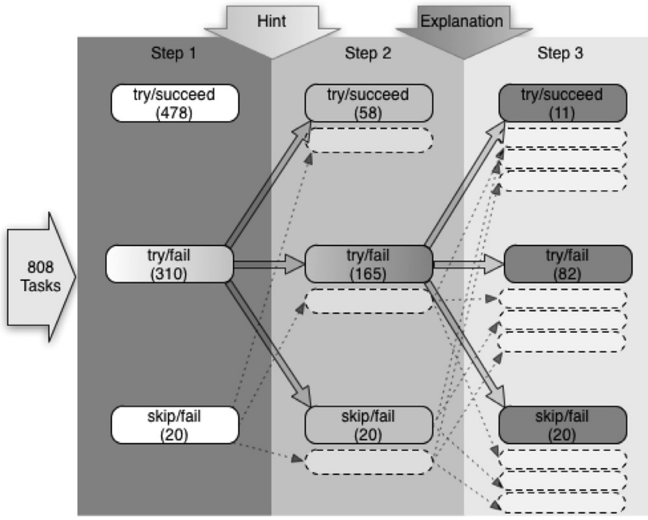

Fig. 6. Graph representing student behaviour.

following learners' work on fraction-TWTEs with various feedback strategies (Table 2 offers an overview of the means and standard deviations of these variables).

5.1. Learner characteristics and their relations to performance

To explore learner characteristics and their relations to performance, we conducted a correlation analysis with the achievement measures (i.e., pre- and post-test performance, learners' engagement with TWTEs in the treatment phase, and the number of correctly solved TWTEs), and the selected motivational variables (see Table 1). This analysis revealed that all learner characteristics assessed in the pre-test had significantly positive correlations with their respective post-test variables (r ranges between .63 and .71). The correlation for pre- and post-test performance was .63; pre- and post-measures of perceived competence correlated with $r = .63$; attainment values correlated with .67; intrinsic values with .71, and the fear-of-failure measures – with .70.

Furthermore, several motivational measures correlated with the pre- and/or post-test performance measures, and the achievement measures of the treatment: both perceived competence measures have significantly positive correlations with the performance measures of pre- and post-test, and the treatment measures (r range between .21 and .52). The intrinsic value measures correlated significantly positively with the number of TWTEs that were solved correctly in the treatment phase ($r_{pre} = .17$; $r_{post} = .28$). The same pattern emerged with regard to the pre-assessed attainment value ($r = .18$). The pre-assessed fear-of-failure measure had significant negative correlations with pre-test performance ($r = -.21$), and the number of TWTEs that were solved correctly in the treatment phase ($r = -.23$).

None of the motivational variables were significantly correlated to the learning gain measures ($r < .10$). However, gender was significantly correlated with the learning gain measures (Spearman's rho = −.27, and −.31; $p < .001$). This result indicates that boys achieved a

Table 1
Means, standard deviations and correlations among achievement and motivational variables ($N = 124$).

		Performance		Learning gain		TWTE		Intrinsic value		Attainment value		Perceived competence		Fear of failure	
		Pre	Post	Relative	Absolute	Engage	Correct	Pre	Post	Pre	Post	Pre	Post	Pre	Post
	Mean	8.12	11.59	.28	3.47	19.38	10.94	53.95	59.41	56.05	53.23	59.33	49.66	35.52	38.08
	SD	3.33	3.57	.27	2.99	6.12	7.55	22.99	22.88	22.27	20.58	23.81	22.59	26.26	29.61
Performance[a]	Post	.63**	–												
Learning gain[a]	Relative	−.10	.68**	–											
	Absolute	−.37**	.50**	.93**	–										
TWTE[a]	Engage	.24**	.32**	.17	.11	–									
TWTE[a]	Correct	.26**	.32**	.13	.09	.60**	–								
Intrinsic value[a]	Pre	.04	.16	.14	.14	.12	.17#	–							
	Post	.03	.12	.09	.12	.15	.28**	.71**	–						
Attainment value[a]	Pre	.11	.10	−.01	−.00	.10	.18*	.72**	.62**	–					
	Post	.09	.02	−.09	−.08	.06	.14	.48**	.55**	.67**	–				
Perceived competence[a]	Pre	.22*	.24**	.06	.03	.25**	.29**	.76**	.54**	.72**	.51**	–			
	Post	.12	.21*	.11	.11	.32**	.52**	.55**	.63**	.54**	.46**	.63**	–		
Fear of failure[a]	Pre	−.21*	−.14	−.02	.07	−.14	−.18	−.24**	−.25**	−.23**	−.05	−.35**	−.26**	–	
	Post	−.20*	−.16#	−.05	.03	−.10	−.23*	−.23**	−.29**	−.30**	−.10	−.33**	−.32**	.70**	–
Gender[b]		.07	−.14	−.27**	−.31**	.06	−.02	.10	.16	.18*	.10	.27**	.14	−.16	−.07

$^{**}p < .01$; $^{*}p < .05$; $\#p < .10$.
[a] Pearson's correlation coefficient.
[b] Spearman's Rho coefficient.

Table 2
Means and standard deviations for achievement and motivational variables by feedback strategy and gender.

			Feedback strategy									
			KR-WE		CH-CE-WE		CH-PE-WE		PH-CE-WE		PH-PE-WE	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Achievement												
Performance	Pre	Female	6.57	3.31	8.64	4.20	6.64	2.29	8.79	2.94	7.79	3.68
		Male	9.24	3.15	9.09	3.14	7.60	3.33	8.31	4.11	7.45	2.70
	Post	Female	11.00	2.89	13.36	4.48	11.73	2.10	11.86	3.30	12.64	4.14
		Male	11.94	3.65	12.64	3.64	10.20	3.45	10.08	2.72	10.45	4.11
Learning gain	Absolute	Female	4.43	2.76	4.73	3.17	5.09	1.38	3.07	2.79	4.86	3.23
		Male	2.71	2.82	3.55	3.01	2.60	2.16	1.77	3.19	3.00	3.85
TWTE	Engage	Female	21.00	5.60	19.00	5.48	16.18	3.97	19.07	7.84	20.79	7.23
		Male	23.35	5.54	20.55	5.61	16.07	5.24	19.62	4.89	17.45	6.01
TWTE	Correct	Female	8.14	6.91	11.42	8.45	8.73	4.41	13.71	8.94	12.36	9.97
		Male	10.59	8.56	9.45	5.47	8.73	6.30	13.31	7.73	11.36	5.63
Motivation												
Intrinsic value	Pre	Female	45.71	26.08	58.61	23.63	61.52	23.96	46.43	20.77	46.67	17.54
		Male	52.55	21.91	59.70	25.67	55.33	27.37	60.26	16.41	52.42	28.48
	Post	Female	45.71	15.48	64.72	25.16	65.15	21.15	54.29	19.89	52.14	17.57
		Male	63.53	18.43	56.36	31.21	58.44	29.49	68.97	16.96	59.09	27.08
Attainment value	Pre	Female	45.71	24.77	59.44	25.02	57.58	19.61	49.52	22.45	48.57	19.11
		Male	54.90	18.93	61.82	25.14	56.89	26.29	56.92	17.77	69.09	23.90
	Post	Female	40.00	20.73	60.00	19.90	55.76	24.27	54.76	16.57	43.33	18.81
		Male	50.59	15.29	58.18	16.89	54.22	23.21	57.95	20.80	54.55	28.72
Perceived competence	Pre	Female	44.57	24.49	59.67	23.04	52.00	28.73	49.71	25.04	55.14	18.31
		Male	64.71	27.32	68.36	16.92	59.47	26.05	69.85	17.41	63.27	23.45
	Post	Female	31.43	17.50	57.67	27.63	50.55	17.46	45.14	24.07	46.86	17.36
		Male	50.82	23.95	45.82	25.07	51.20	26.63	55.38	18.75	54.18	21.42
Fear of failure	Pre	Female	41.43	31.32	44.17	32.32	41.82	19.40	39.29	21.29	34.29	30.56
		Male	33.53	25.23	25.45	28.06	31.33	25.88	31.54	26.72	36.36	26.18
	Post	Female	51.43	23.40	41.67	32.15	46.36	33.25	32.14	25.77	35.71	33.45
		Male	39.41	26.80	25.45	34.75	45.33	32.70	33.85	27.25	33.64	25.80

lower learning gain than girls. We also found significant positive correlations between gender and pre-assessed perceived competence (Spearman's rho = .27, $p < .01$), and gender and pre-assessed attainment value (Spearman's rho = .18, $p < .05$). The respective correlations after the treatment were not statistically significant (Spearman's rho $r < .14$).

Given the latter finding, we provide the overview of means and standard deviation of all assessed variables separately for boys and girls (Table 2).

We also ran gender-specific correlation analyses in order to further explore the interrelations between achievement and motivational variables. Fig. 7 provides an overview of the significant correlations we found through these analyses.

Boys' and girls' correlation patterns share the following features. The number of correctly solved TWTEs in the treatment phase is significantly positively correlated to the performance and perceived competence measures of pre- and post-test. The higher participants scored on the pre-assessed performance and perceived competence measures, the more TWTEs they were able to solve correctly, and the higher they scored on the respective post test measures. Furthermore, the number of correctly solved TWTEs significantly positively correlated to intrinsic value after the treatment phase.

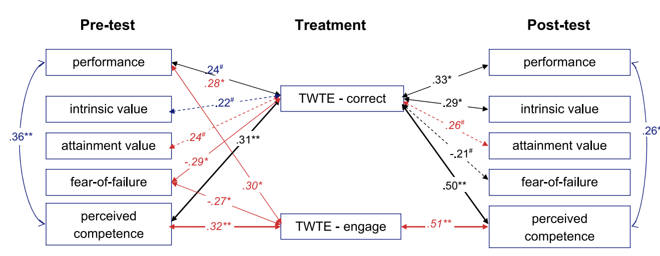

Fig. 7. Overview of significant correlations between pre- and post-assessed learner variables and treatment measures (red-italicized = female ($n = 57$); blue-non-italicized = boys ($n = 67$), black = both). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

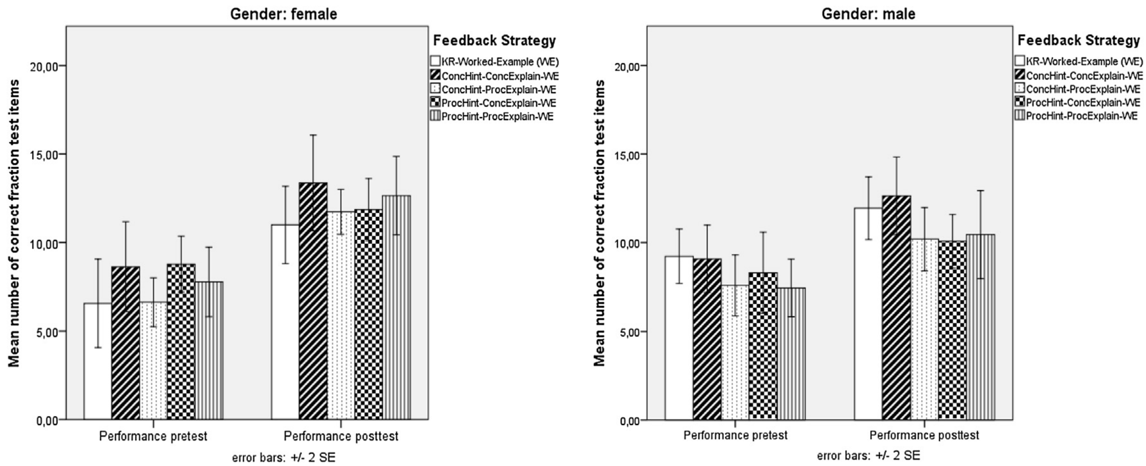

Fig. 8. Performance in pre- and post-test by gender and feedback strategy ($N = 124$) KR = knowledge of result; ConcHint = conceptual hint; ConcExplain = conceptual explanation; ProcHint = procedural hint; ProcExplain = procedural explanation, WE = worked-out example with correct solution.

For girls, we found significant negative correlations between pre-assessed fear of failure and the treatment measures. The higher their fear of failure, the less they engage in TWTEs, and the less successful they are in solving them correctly. Moreover, we found a significant positive correlation between the perceived competence measures for girls, but not for boys.

5.2. Feedback strategies, gender and knowledge gain

Our analyses so far have revealed that gender, which is a stable learner characteristic, relates to outcome variables. To investigate whether there are differential gender-based effects of the experimental feedback strategies on the knowledge gain from pre- to post-test, we ran a repeated-measures analyses of variance (ANOVA) with the between factors gender (2) and feedback strategy (5), and the within measure achievement in pre- and post-test. This ANOVA revealed a significant main effect of the within-factor, Wilk's Lambda = .39; $F(1/114) = 179.41$; $p < .001$; partial eta^3 = .61. Moreover, the ANOVA yielded a significant interaction between the within-factor and gender, Wilk's Lambda = .92; $F(1/114) = 10.25$; $p = .002$; eta^2 = .08. No other main effects or interactions were significant. Fig. 8 illustrates that under all feedback conditions, boys and girls achieved higher levels of performance in the post-test. Furthermore, it reveals that the average increase in performance from pre- to post-test is higher for girls than for boys even though there were no significant differences of the pre-test performance.

Follow-up analyses revealed that under all tutoring feedback conditions girls had a higher knowledge gain than boys $F(1/99) = 4.699$, $p = .03$, partial eta^2 = .05, while under the control condition the boys had a slightly higher, yet not significantly higher knowledge gain, $F(1/23) = 1.87$, $p = .19$, partial eta^2 = .08. Significant differences in learning gain of girls and boys occurred with the feedback-strategy providing a conceptual-hint, then a procedural explanation $F(1/25) = 11.18$, $p = .003$, partial eta^2 = .32.

5.3. Feedback strategies, gender, and motivational changes

Given the observed gender-specific correlation patterns, we also ran a repeated-measures multivariate analysis of variance (MANOVA) with the between factors gender (2), and feedback strategy (5), and the repeated measures intrinsic value, attainment value, perceived competence, and fear of failure. This MANOVA revealed a multivariate main effect of the within factor indicating that there are significant changes in motivation from the pre- to the post-test, Wilk's Lambda = .61; $F(4/112) = 17.81$; $p < .000$; partial eta^2 = .39.

Univariate follow-up ANOVAS showed the following effects for perceived competence: a significant main effect of the within factor $F(1/115) = 8.40$; $p = .004$, as well as of the between factor gender, $F(1/115) = 5.58$; $p < .02$; partial eta^2 = .05. In addition, we observed a significant interaction between the within factor and gender, $F(4/115) = 4.47$; $p = .04$; partial eta^2 = .04. Learners' perceived competence decreased from pre- to post-test, yet, the decline of perceived competence was significantly larger for boys than for girls. In the pre-test, boys rated their competence significantly higher than girls across all experimental conditions. In the post-test, this was still the case under the three conditions, KR-WE, CH-PE-WE, PH-CE-WE, and PH-PE-WE, even though boys' perceived competence decreased significantly more than girls' perceived competence (see Fig. 9). Under the CH-CE-WE condition, boys' perceived competence decreased the most, while girls perceived competence remained almost at the same level. A post-hoc contrast analysis shows that this interaction is statistically significant, $F(1/21) = 12.54$; $p = .002$; partial eta^2 = .37.

The follow-up ANOVAs further revealed that the increase in intrinsic value from pre- to post-test is statistically significant, $F(1/115) = 28.35$; $p < .001$; partial eta^2 = .20.

5.4. Feedback strategies, gender, and practice behaviour

To examine if learners' behaviour when practicing with TWTEs differs depending on (a) the feedback strategies they worked with, and (b) their gender, we ran two analyses of covariance (ANCOVA) with the between factors feedback strategy (5) and gender (2), and the covariate pre-test performance for the dependent variables number of TWTEs engaged in and number of correctly solved TWTEs.

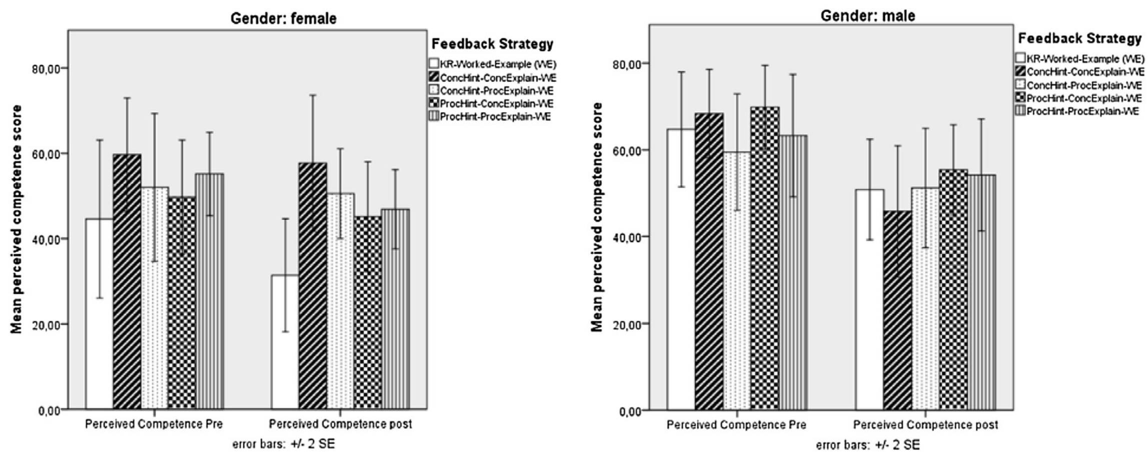

Fig. 9. Perceived competence by feedback strategy and gender ($N = 124$) KR = knowledge of result; ConcHint = conceptual hint; ConcExplain = conceptual explanation; ProcHint = procedural hint; ProcExplain = procedural explanation, WE = worked-out example with correct solution.

The ANCOVA for the number of TWTEs that learners engaged in revealed a significant main effect of the factor feedback strategy, $F(4/113) = 2.75$; $p = .03$; partial eta^2 = .09. Furthermore, it revealed a significant effect of the covariate pre-test performance $F(1/113) = 4.95$; $p = .03$; partial eta^2 = .04. The effect of gender, and the interaction between the between-factors was not statistically significant ($F < .07$). Hence, with regard to task engagement, female and male learners behaved similarly during practice. Under the conceptual-procedural feedback strategy, they engaged in fewer tasks than under all other feedback strategies (see Table 2 for means and SDs).

The ANCOVA for the number of correctly solved TWTEs provided neither significant main effects of the between factors gender and feedback strategy ($F < 1.4$), nor a significant interaction between these factors ($F < .18$). However, the effect of the covariate pre-test performance was statistically significant $F(1/113) = 6.99$; $p = .009$; partial eta^2 = .06. These findings reveal that learners' prior levels of knowledge influence their performance during practice, while gender and feedback do not.

5.5. Learner characteristics and student's post-feedback behaviour

To explore the impact of learners' characteristics on their post-feedback behaviour, we conducted detailed log-file analyses for the data sets of those participants who had worked on at least the complete first block of TWTE in the treatment phase. As mentioned above, the data sets of learners who worked too slowly to complete the first set of TWTEs, or whose log-files contained missing data due to system failures were excluded from these analyses. Since we were particularly interested in student behaviour after procedural and conceptual hints and explanations, we did not include the data of participants who received the KR-WE strategy. Consequently, the data set for this analysis included only logs of 50 male and 51 female participants.

5.5.1. Learner characteristics vs. post-feedback skipping rate

To explore if and how learner characteristics (gender, motivational variables, prior knowledge, etc.) relate to learners' post-feedback behaviour, we computed Spearman's correlation coefficient between the learner characteristics and behavioural variables.

Fig. 10 summarizes the results of the correlation analysis for the skipping rates after hints and explanations. It reveals that, for girls, pre-assessed fear-of-failure measures correlated marginally significantly with the skipping rate after explanations. All other pre-assessed learner characteristics did not correlate significantly with skipping behaviour. However, there are several significant correlations between skipping and post-treatment motivational learner characteristics: Most importantly, for boys and girls, a higher skipping rate after hints were significantly correlated to lower ratings of intrinsic value and perceived competence. A significant negative correlation between the skipping rate after explanations and perceived competence was also found for both boys and girls. For girls, we also found significant correlations between skipping after explanations and fear of failure, as well as intrinsic motivation.

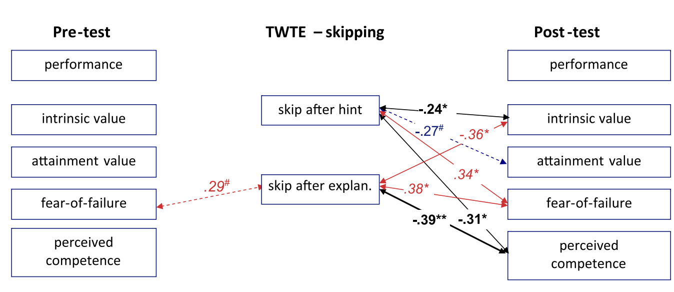

Fig. 10. Overview of significant correlations between pre- and post-assessed learner variables and post-feedback skipping rate (red-italicized = female ($n = 57$); blue-non-italicized = boys ($n = 67$), black = both. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

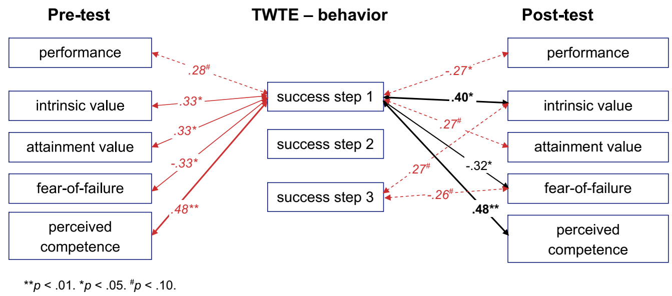

Fig. 11. Overview of significant correlations between pre- and post-assessed learner variables and post-feedback success-rate in correcting errors without tutoring feedback (step 1), after hints (step 2), and after explanations (step 3) (red-italicized = female ($n = 57$); blue-non-italicized = boys ($n = 67$), black = both). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.5.2. Learner characteristics vs. post-feedback success rate

A correlation analysis with regard to the post-feedback success rate in error correction and the pre-assessed learner variables revealed significant correlations for the girls only. The success rate of the first error-correction step was significantly correlated with all the pre-assessed motivational variables. Most of the motivational variables assessed after the treatment were also significantly correlated to the success-rate of the first step, but here for boys and girls similarly (see Fig. 11).

5.5.3. Feedback content and learner characteristics vs. skipping and success rate

To explore if feedback content has a gender-specific effect on learners' post-feedback behaviour, we ran univariate analyses of variance (ANOVA) with the between factors gender (2), and feedback content (2) for the following behavioural variables, which were obtained through the log-file analyses:

- Skipping rate after hints (the number of skips after hints divided by the number of occasions for skipping).
- Skipping rate after explanations (the number of skips after explanations divided by the number of occasions for skipping).
- Success rate after hints (the number of successful error corrections after hints divided by the number of occasions for error corrections).
- Success rate after explanations (the number of successful error corrections after explanations divided by the number of occasions for error corrections).

The ANOVA for the skipping rate after hints showed no significant main effects of feedback content and gender ($F < .24$). The interaction between feedback content and gender did not reach the significance level of .05; $F(1/95) = 3.16$; $p < .08$; partial eta$^2 = .03$. Boys tended to skip more often after a conceptual hint than after a procedural hint; while, for girls, feedback content did not affect skipping behaviour at this step.

However, the ANOVA for the skipping rate after explanations revealed a significant effect of feedback content: $F(1/85) = 3.8$; $p = .05$; partial eta$^2 = .04$. The skipping rate after conceptual explanations was higher than after procedural explanations. Fig. 12 illustrates these findings.

The ANOVAs for success rates after hints and explanations did not provide any significant main effects or interactions (all $F < .1$). Fig. 13 shows that the mean success-rate after a feedback hint ranged between .27 and .19, while the success-rate after an explanation varied between .02 and .08. Note that the explanation-level feedback was only provided to participants who failed twice in correcting the error in the TWTE.

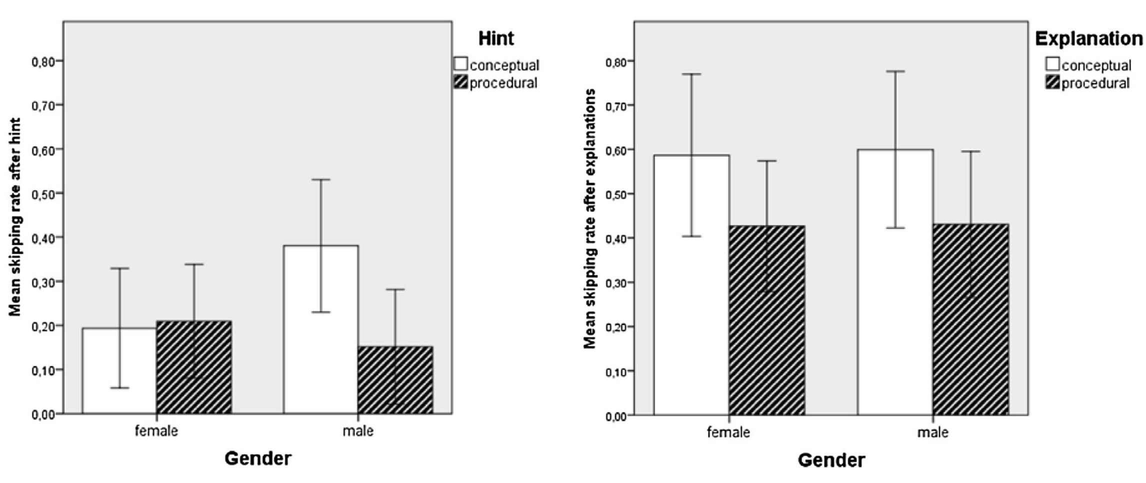

Fig. 12. Skipping rate by gender and feedback content (conceptual – procedural) after hints (left side) or explanations (right side).

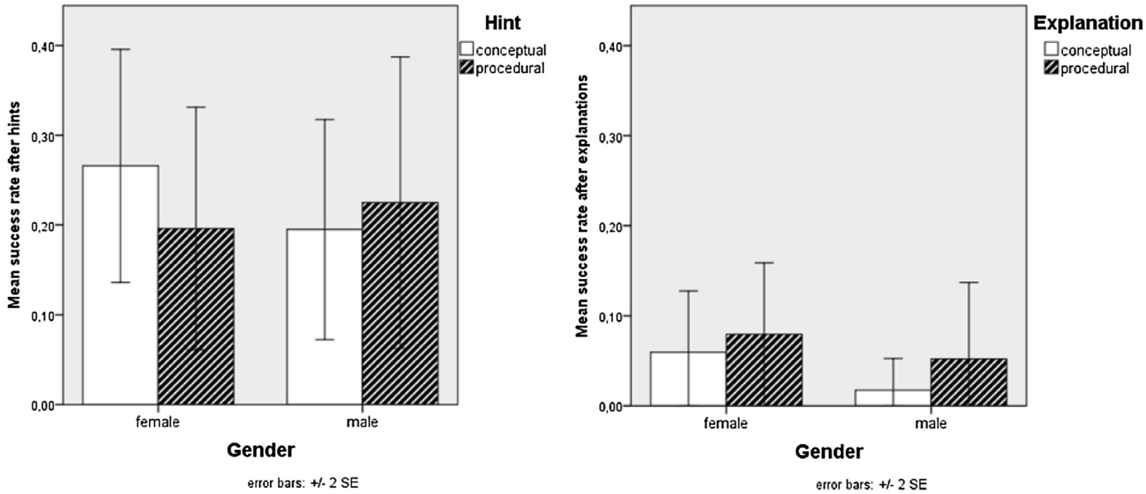

Fig. 13. Success rate by gender and feedback content (conceptual – procedural) after hints (left side) or explanations (right side).

6. Discussion

The goal of the presented work was to explore various factors that potentially influence the effectiveness of tutoring feedback and that may serve as a basis for implementing formal strategies for adapting feedback within a computer-based educational system. We focused on two sets of factors: feedback-related characteristics (including the type of knowledge communicated by the feedback-content: procedural or conceptual; and feedback specificity, that is, the level of elaboration of feedback messages: hint or explanation), and learner-related characteristics (including prior knowledge, gender, and several parameters describing learners' motivational state). The impact of these factors was investigated with regard to outcome measures (i.e., learning gain from pre- to post-test, motivational changes from pre- to post-test), and with regard to a set of behavioural measures collected in the AES's log-files during learners' practice, including number of TWTEs attempted, number of correctly solved TWTEs, the post-feedback skipping rate, and the post-feedback success rate. By including behavioural measures, we aimed at shedding light on how learner and feedback characteristics are related to learners' behaviour when practicing with an AES.

In summary, our results regarding the impact of feedback strategies and learner characteristics on the outcome variables revealed several significant gender-specific differences (research question 1) but no significant differences due to feedback strategies (research question 2). Boys benefitted less from practicing with feedback-equipped TWTEs than girls. Boys had a significantly lower level of knowledge gain, and their perceived competence declined significantly more from pre- to post-test. Interestingly, the decline in perceived competence (observed for both boys and girls) was not related to a decline in intrinsic motivation, as on average, intrinsic motivation increased significantly from pre- to post-test.

The finding that girls gained higher scores from pre- to post-test, is in line with prior research that showed gender-specific effects of score improvements across multiple assessments (e.g., Schleicher, Van Iddekinge, Morgeson, & Campion, 2010).

Since boys had a lower knowledge gain from pre- to post-test, the finding that boys' perceived competence declined more than girls' (research question 3) supports theoretical and empirical work on the role of mastery experiences for the development of perceived competence (Bandura, 1997). Even though learners did not receive immediate feedback for the test-items, boys may have had relatively fewer mastery experiences than the girls. They may have perceived that solving fraction test items in the post-test is not much easier for them compared to the pre-test. Since we did not explicitly measure ease-of-task completion, we cannot conclude to what extent this preliminary explanation accounts for declines in perceived competence. Further studies should thus include measures of ease-of-task completion in order to investigate this interesting issue for further research.

The overall decline in perceived competence may also be a methodological artefact. The motivational questionnaire was answered before the pre-test, but after the post-test. Hence, when learners responded to the pre-test questionnaire, they could not rely on concrete experiences with the fraction tasks; yet, they were able to do so after the post-test. Since there is empirical evidence that learners tend to overestimate their performance (see Bouffard & Narciss, 2011 for a review), learners may have simply corrected their perceived competence-bias from pre- to post-test using their experiences with the test-items. Investigating this issue was not within the scope of this paper, but certainly deserves further attention.

The results regarding the impact of feedback strategies and learner characteristics on learner's behaviour during practice with the AES, revealed that feedback strategies and pre-test performance (but not gender) had an impact on how many tasks learners attempted to solve (research question 4). Learners that were exposed to the conceptual-procedural feedback strategy engaged in fewer tasks than learners in all other feedback-strategy groups. Nevertheless, the knowledge gain under this condition was not significantly lower than for other conditions (and girls' learning gain was the highest among all groups, see Table 2). Furthermore, only learners' prior performance had a statistically significant impact on how many tasks they solved correctly in the treatment phase, and neither variable correlated with learning gain. Thus, one may conclude that there is no simple relation between the implemented feedback strategies, profiled learner characteristics, and computed outcome, or (rather global) behavioural variables.

In order to better understand how feedback strategies and learners' characteristics relate to learner behaviour (when learners practice with TWTEs that are equipped with tutoring feedback), we conducted a more fine-grained analysis using behavioural events which occurred immediately after learners were provided with feedback. We chose two post-feedback variables, which quantify two phenomena of student behaviour within problem-solving tasks: solving vs. failing, and skipping vs. trying. Both variables characterize an immediate learner-reaction to the learning task and the feedback messages, and are observable in learners' problem solving activities. Consequently, those variables can serve as indicators for feedback effectiveness. Post-feedback solving behaviour is an indicator for the feedback's capability to foster a learner's transition from not being able to solve an exercise to being able to solve it after attending to the feedback message. Post-feedback skipping behaviour is an indicator for the feedback's incapability to keep the learners on task even if they encounter difficulties.

From an instructional perspective, post-feedback skipping behaviour may be considered an even more important indicator of feedback effectiveness than post-feedback solving. If a feedback message convinces a learner to continue practicing, continue trying to solve an exercise, it creates an opportunity for learning, even if the attempt does not lead to the correct solution. A learner who does not solve an exercise might still learn something from the practice, while a learner who skips an exercise, most probably, does not learn anything. From a motivational perspective, a feedback message that succeeds in keeping learners on tasks, creates an opportunity for a mastery experience (in the case of a successful task completion). Mastery experiences that can be attributed to learner characteristics (i.e., learner's effort invested, or learner's competence) are considered a major factor that helps increase learners' confidence in their competence (Bandura, 1997). Even if a further attempt results in a failure, it offers an opportunity for testing if a selected error-correction strategy is successful, and for developing strategies and skills for coping with, and learning from errors. Hence, a high skipping rate indicates that learners tend to avoid testing their capabilities in correcting their errors and in doing so they miss chances to improve not only their cognitive, but also their motivational and metacognitive skills.

In summary, our findings reveal that learner characteristics had only a small impact on the post-feedback skipping-rate (research question 5), but no impact on the post-feedback success-rate (research question 6). Furthermore, the results indicate that the impact of feedback content may increase from trial to trial. After the first trial, only boys who received a conceptual hint were more likely to skip than boys who received a procedural hint, while there was no difference in girl's skipping-rate depending on the feedback content. After the second trial, boys and girls were more likely to skip after being exposed to a conceptual rather than a procedural explanation. These findings indicate that, in the domain of fraction learning, conceptual feedback content is less likely to be effective in keeping learners (boys in particular) on task.

These rather mixed results regarding the differential effects of feedback content and some of the explored learner variables, supports prior theoretical, meta-analytical, and empirical work on feedback (e.g., Arroyo et al., 2012; Narciss, 2008; Hattie & Timperley, 2007; Shute, 2008), suggesting that differential effects of various types of feedback may not occur consistently but rather under certain individual and situational conditions. For example, in line with the work of Arroyo and her colleagues (which revealed rather complex results with regard to gender differences of learners using tutoring systems), we also found interactive effects of student gender and hint content (Arroyo et al., 2001, 2006).

6.1. Limitations and implications for future research

The theoretical and methodological approaches of this work have several strengths and limitations. First, the design and evaluation of the feedback strategies were based on the Interactive-tutoring-feedback loops model (Narciss, 2008, 2012, 2013). The ITF-model suggests that tutorial feedback strategies are complex multidimensional and multifunctional instructional approaches, and that their effects may occur only under certain individual and/or situational conditions (see also Arroyo et al., 2001, 2006; Shute, 2008). If all feedback dimensions and all potential factors and effects are taken into account, the complexity becomes unmanageable within an empirical study. Therefore, we have limited our design by focussing on the between-subjects manipulation of one feedback dimension (i.e. feedback content: procedural vs. conceptual), while controlling the form and mode of feedback presentation (i.e., three trial feedback strategy with increasing feedback specificity after each trial), as well as the functional aspect of feedback. All participants were exposed to a feedback strategy providing step-by-step support (knowledge-of-result feedback – feedback hint – feedback explanation – worked-out example). Consequently, with each trial, the level of feedback specificity increased, while the feedback-content was varied systematically between experimental groups. Thus, all participants received feedback messages with the higher level of feedback specificity only if they had failed two times to correct the error. Therefore, the present data allow gaining insights on how feedback content and learner characteristics may affect learners' post-feedback behaviour after the first tutoring feedback message (the hint level). Yet, after the second tutoring feedback message (the explanation level) the effects of feedback content, feedback specificity, learner characteristics and learner's coping with their failures in previous attempts can be hardly disentangled. The inclusion of feedback strategies combining procedural and conceptual feedback messages contributes further to these qualifications of our findings after the explanation-level feedback. Students exposed to the combined CH-PE-WE and PH-CE-WE feedback strategies, received both, procedural and conceptual information. Since the text of explanation-level feedback messages included hint-messages, these students were potentially provided with more information (both types of a hint + complimenting part of an explanation) than the students of symmetric CH-CE-WE and PH-PE-WE conditions (same type of a hint twice + complimenting part of an explanation). Interestingly, we did not find a superiority of the combined feedback strategies with regard to their effects on knowledge gain and changes in motivation. This observation confirms prior findings, which showed that the mere amount of feedback information does not determine the efficiency of a feedback strategy (e.g., Kulhavy, White, Topp, Chan, & Adams, 1985; see also Narciss, 2008; Shute, 2008). Further studies are needed to investigate, if and to what extent differential effects occur for feedback strategies providing more specific feedback messages at an earlier trial.

Second, in order to investigate the effects of feedback content and individual learner variables, we used TWTEs providing learners with worked-out examples, in which typical errors for fraction learning were included. Consequently, the domain of the study was very specific with well-structured learning tasks that require students applying fraction addition rules but did not address their general problem solving ability or any higher order skills. With this methodological approach we were able to control the variety of error types, which had to be addressed by the feedback. Therefore, it was possible to tailor the feedback content to these typical errors. Based on cognitive task- and error analyses we developed procedural and conceptual hints and explanations for each of those typical errors the students had to correct when working with the TWTEs. The procedural feedback content addressed (rather concretely) knowledge on how to proceed in order to correct the typical error provided by the TWTE, but it did not explicitly address the concepts underlying the rationale for the procedure. In contrast, the conceptual feedback-content addressed these concepts, but did not explicitly provide concrete information on how to proceed. Thus, with the latter type of feedback content, students had to deduce the specific steps necessary for error-correction. In a preliminary study, which served as the basis for improving the readability of the designed feedback messages, students rated the conceptual feedback messages as more difficult to understand. Thus, the conceptual feedback messages were reworked and rechecked in order to improve their readability. After this iterative improvement process, the conceptual feedback messages, in particular, the conceptual explanations were longer than the procedural ones. Thus, the effects of feedback content on skipping behaviour might not be entirely due to the semantic content, but also due to further factors closely related to the semantic content (e.g., level of verbosity). For example, boys may have been less attentive to more verbose feedback messages. Future studies should thus examine the effects of the level of verbosity and the type of feedback content.

Third, the multi-trial learning tasks enabled us to collect process data on several behavioural events for each of the error-correction attempts. The focus of the presented study was on investigating skipping vs. trying behaviour, as well as successful vs. unsuccessful error correction. The frequencies of these distinct behavioural events could be reliably identified by analysing the log-files of the multi-trial TWTEs. Yet, there was a high level of variability in terms of how students interacted with these multi-trial tasks. They could have solved the task on any trial, they could have skipped any trial, and they were allowed to work on the tasks at their own pace. Thus, several methodological challenges had to be tackled. Because of the variability in students' pace, different students worked on different numbers of tasks. We coped with this challenge by including only the data of the students who had worked on at least one set of eight tasks. However, exclusion of the students who did not complete, at least, the first set of tasks makes our findings somewhat biased against very slow working students. Furthermore, since in each trial some students successfully corrected the error, the number of included events decreased with each trial. Therefore, the differences in frequencies of the selected behavioural events might also have been due in part to this variability across cases and tasks. A more fine-grained investigation of skipping and solving behaviour, also taking into account other behavioural variables (e.g., time-on-task before skipping or solving), or even patterns of several behavioural variables, is an open issue for further research.

Fourth, TWTEs confronted all students with typical errors, including students who would have been able to solve the task and would not have committed these errors themselves. Based on the present data, it is thus difficult to disentangle the effects of prior knowledge and the efficiency of flag-error feedback. Further studies should investigate, whether the efficiency of flag-error feedback is also strongly related to individual learner characteristics if learners are provided with flag-error feedback after errors they have committed themselves.

6.2. *Implications for instructional design and practice*

The present findings suggest several implications for the development of personalized feedback strategies: First, the findings suggest including not only students' prior knowledge, but also motivational variables, student's intrinsic motivation in particular, perception of competence, and fear of failure into the design of personalized feedback strategies. Second, the findings indicate that in mathematics education (and, perhaps, in other domains), gender also has to be taken into account. Third, for the development of dynamic personalized feedback strategies, the assessment components should not only focus on behavioural outcome variables (i.e., the solving rate, or other indicators of transitioning from failure to success), but should also focus on motivation-related process variables, skipping behaviour in particular. Investigating the added value of personalized feedback strategies based on these implications remains an open issue for further research.

Acknowledgements

This work was supported by the German National Science Foundation (DFG – Deutsche Forschungsgemeinschaft) under the project *AtuF: Adaptive Tutoring Feedback* (Grants ME 1136/8-1 and NA 738/10-1).

We want to express our gratitude to two anonymous reviewers for their thorough and constructive feedback to an earlier draft of this paper. Many thanks also to Marie and Meher Lippmann for editing and correcting language flaws.

References

Akbulut, Y., & Cardak, C. S. (2012). Adaptive educational hypermedia accommodating learning styles: a content analysis of publications from 2000 to 2011. *Computers & Education, 58*(2), 835–842.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned. *The Journal of the Learning Sciences, 4*(2), 167–207.

Andrès, E. (2012). Towards detailed diagnostics of unstructured learner's problem solutions. In *Presented at the doctoral consortium of the 7th European Conference on Technology Enhanced Learning (EC-TEL), Saarbrücken* (pp. 1–6).

Arroyo, I., Beck, J. E., Beal, C. R., Wing, R., & Woolf, B. P. (2001). Analyzing students' response to help provision in an elementary mathematics Intelligent Tutoring System. In R. Luckin (Ed.), *Papers of the AIED-2001 workshop on help provision and help seeking in interactive learning environments* (pp. 34–46).

Arroyo, I., Woolf, B. P., & Beal, C. R. (2006). Addressing cognitive differences and gender during problem solving. *International Journal of Technology, Instruction, Cognition and Learning, 4*, 31–63.

Arroyo, I., Woolf, B. P., Royer, J. M., Tai, M., Muldner, K., Burleson, W., et al. (2012). The impact of animated pedagogical agents on girls' and boys' emotions, attitudes, behaviors and learning. In *Proceedings of the 2011 IEEE 11th International Conference on Advanced Learning Technologies (ICALT)* (pp. 506–510). http://dx.doi.org/10.1109/ICALT.2011.157.

Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: Freeman.

Bokhove, C., & Drijvers, P. (2012). Effects of feedback in an online algebra intervention. *Technology, Knowledge and Learning, 17*(1–2), 43–59.

Bouffard, T., & Narciss, S. (2011). Benefits and risks of positive biases in self-evaluation of academic competence: introduction. *International Journal of Educational Research, 50*(4), 205–208.

Brusilovsky, P., & Henze, N. (2007). Open corpus adaptive educational hypermedia. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The adaptive web* (pp. 671–696). Heidelberg: Springer.

Brusilovsky, P., & Vassileva, J. (2003). Course sequencing techniques for large-scale web-based education. *International Journal of Continuing Engineering Education and Life Long Learning, 13*(1), 75–94.

Choe, H., Bae, Y., Kim, T., & Lee, T. (2004). *Work in progress-the study of Web-based adaptive feedback based on the analysis of individual differences*. In *Frontiers in education 2004 34th annual* (Vol. 1). T2C–T25–T2C–T26.

Chuang, S.-H., & O'Neil, H. F. (2006). *Role of task-specific adapted feedback on a computer-based collaborative problem-solving task (No. 684)*. Regents of the University of California.

Conati, C., & Manske, M. (2009). Evaluating adaptive feedback in an educational computer game. In Z. Ruttkay, M. Kipp, A. Nijholt, & H. H. Vilhjálmsson (Eds.), *Intelligent virtual agents* (pp. 146–158). Berlin Heidelberg: Springer.
Cook, D. A. (2012). Revisiting cognitive and learning styles in computer-assisted instruction: not so useful after all. *Academic Medicine, 87*(6), 778–784.
Dennis, M., Masthoff, J., & Mellish, C. (2012). Adapting performance feedback to a learner's conscientiousness. In J. Masthoff, B. Mobasher, M. C. Desmarais, & R. Nkambou (Eds.), *User modeling, adaptation, and personalization* (pp. 297–302). Berlin Heidelberg: Springer.
Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132.
Economides, A. A. (2006). Adaptive feedback characteristics in CAT. *International Journal of Instructional Technology and Distance Learning, 3*, 15–26.
Eichelmann, A., Narciss, S., Faulhaber, A., & Melis, E. (2008). Analyzing computer-based fraction tasks on the basis of a two-dimensional view of mathematics competences. In J. Zumbach, N. Schwartz, T. Seufert, & L. Kester (Eds.), *Beyond knowledge: The legacy of competence* (pp. 125–134). Netherlands: Springer.
Eichelmann, A., Narciss, S., Schnaubert, L., Melis, E., & Goguadze, G. (2011). Design und Evaluation von interaktiven webbasierten Bruchrechenaufgaben. In H. Rohland, A. Kienle, & S. Friedrich (Eds.), *DeLFI 2011 – Die 9. e-Learning Fachtagung Informatik* (pp. 31–42). Bonn: Köllen Verlag.
Eichelmann, A., Andrès, E., Schnaubert, L., Narciss, S., & Sosnovsky, S. (2012). Interaktive Fehler-Finde- und Korrektur-Aufgaben. In G. Csanyi, F. Reichl, & A. Steiner (Eds.), *Digitale Medien – Werkzeuge für Forschung und Lehre* (pp. 401–412). Münster: Waxmann.
Eichelmann, A., Narciss, S., Schnaubert, L., & Melis, E. (2012). Typische Fehler bei der Addition und Subtraktion von Brüchen – Ein Review zu empirischen Fehleranalysen. *Journal für Mathematik-Didaktik, 33*(1), 29–57.
Gallagher, A. M., & Kaufman, J. C. (Eds.). (2004). *Gender differences in mathematics: An integrative psychological approach*. Cambridge University Press.
Gertner, A. S., Conati, C., & VanLehn, K. (1998). Procedural help in Andes: generating hints using a Bayesian network student model. In J. Mostow, & C. Rich (Eds.), *Proceedings of the 15th national conference on artificial intelligence*. Palo Alto, CA: AAAI Press.
Gimeno Sanz, A., & De-Siqueira, J. M. (2009). Designing feedback to support language acquisition using the ingenio authoring tool. *Procedia – Social and Behavioral Sciences, 1*(1), 1239–1243.
Goguadze, G. (2009). Representation for interactive exercises. In J. Carette, L. Dixon, C. Sacerdoti Coen, & S. M. Watt (Eds.), *Intelligent computer mathematics* (pp. 294–309) (Berlin, Heidelberg).
Goldin, I. M., Koedinger, K. R., & Aleven, V. (2012). Learner differences in hint processing. In K. Yacef, O. Zaïane, H. Hershkovitz, M. Yudelson, & J. Stamper (Eds.), *Proceedings of the 5th international conference on educational data mining* (pp. 73–80). Worcester, MA: International Educational Data Mining Society.
Goodman, J. S., Wood, R. E., & Chen, Z. (2011). Feedback specificity, information processing, and transfer of training. *Organizational Behavior and Human Decision Processes, 115*(2), 253–267.
Hancock, T. E., Stock, W. A., & Kulhavy, R. W. (1992). Predicting feedback effects from response-certitude estimates. *Bulletin of the Psychonomic Society, 30*(2), 173–176.
Hancock, T. E., Thurman, R. A., & Hubbard, D. C. (1995). An expanded control model for the use of instructional feedback. *Contemporary Educational Psychology, 20*(4), 410–425.
Hattie, J., & Gan, M. (2011). Instruction based on feedback. In R. E. Mayer, & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). New York: Routledge.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.
Heckhausen, H. (1989). *Leistungsmotivation* (pp. 231–278). Berlin, Heidelberg: Springer.
van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education, 58*(1), 263–272.
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254.
Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology, 10*, 285–291.
Lin, L., Atkinson, R. K., Christopherson, R. M., Joseph, S. S., & Harrison, C. J. (2013). Animated agents and learning: does the type of verbal feedback they provide matter? *Computers & Education, 67(C)*, 239–249.
Mavroudi, A., & Hadzilacos, T. (2012). Broadening the use of e-learning standards for adaptive learning. In E. Popescu, Q. Li, R. Klamma, H. Leung, & M. Specht (Eds.), *Advances in web-based learning-ICWL 2012* (pp. 215–221). Berlin Heidelberg: Springer.
Melis, E., Goguadze, G., Homik, M., Libbrecht, P., Ullrich, C., & Winterstein, S. (2006). Semantic-aware components and services of ActiveMath. *British Journal of Educational Technology, 37*(3), 405–423.
Melis, E., Faulhaber, A., & Eichelmann, A. (2008). Interoperable competencies characterizing learning objects in mathematics. In B. P. Woolf, E. Aïmeur, R. Nkambou, & S. P. Lajoie (Eds.), *Proceedings of the 9th international conference on Intelligent Tutoring Systems (ITS)* (pp. 416–425). Berlin, Heidelberg: Springer.
Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: a comparison of human tutors and intelligent tutoring systems. *The Journal of Learning Sciences, 2*, 277–305.
Mitrovic, A. (2011). Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User Modeling and User-Adapted Interaction, 22*(1–2), 39–72.
Mory, E. H. (1991). *The effects of adaptive feedback on student performance, feedback study time, and lesson efficiency within computer-based instruction*. The Florida State University.
Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed.). (pp. 745–783) New York NY: Simon & Schuster Macmillan.
Narciss, S. (2006). *Informatives tutorielles feedback*. Münster: Waxmann.
Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. G. van Merrienboer, & M. P. Driscoll (Eds.) (pp. 125–143). Mahaw, NJ: Lawrence Erlbaum Associates.
Narciss, S. (2012). Feedback in instructional contexts. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 1285–1289). Berlin Heidelberg: Springer.
Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the interactive tutoring feedback model. *Digital Education Review*, (23). http://greav.ub.edu/der/.
Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. Niegemann, R. Brünken, & D.Leutner (Eds.), *How to design informative tutoring feedback for multi-media learning* (pp. 181–196). Münster: Waxmann.
Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction, 16*(4), 310–322.
Olney, A. M., Graesser, A. C., & Person, N. K. (2010). Tutorial dialog in natural language. In R. Nkambou, R. Mizoguchi, & J. Bourdeau (Eds.), *Advances in intelligent tutoring systems* (pp. 181–206). Berlin Heidelberg: Springer.
Parvez, S., & Blank, G. (2008). Individualizing tutoring with learning style based feedback. In *Intelligent tutoring systems* (pp. 291–301). Berlin Heidelberg: Springer.
Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: concepts and evidence. *Psychological Science in the Public Interest, 9*, 105–119.
Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667–686.
Popescu, E. (2009). Evaluating the impact of adaptation to learning styles in a Web-based educational system. In M. Spaniol, Q. Li, R. Klamma, & R. W. H. Lau (Eds.), *Advances in web based learning–ICWL 2009* (pp. 343–352). Berlin Heidelberg: Springer.
Proske, A., Narciss, S., & McNamara, D. S. (2012). Computer-based scaffolding to facilitate students' development of expertise in academic writing. *Journal of Research in Reading, 35*(2), 136–152.
Puta, M., Schnaubert, L., & Narciss, S. (2013). *Erwartungs-Wert Fragebogen zur bereichsspezifischen Lernmotivation (EWF-LM): psychometrische Eigenschaften und offene Fragen* (under revision).
Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Metacognitive practice makes perfect: improving students' self-assessment skills with an intelligent tutoring system. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial intelligence in education*. Berlin Heidelberg: Springer.
Schleicher, D. J., Van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2010). If at first you don't succeed, try, try again: understanding race, age, and gender differences in retesting score improvement. *Journal of Applied Psychology, 95*(4), 603–617.
Senko, C., & Harackiewicz, J. M. (2005). Regulation of achievement goals: the role of competence feedback. *Journal of Educational Psychology, 97*(3), 320–336.
Serge, S. R., Priest, H. A., Durlach, P. J., & Johnson, C. I. (2012). The effects of static and adaptive performance feedback in game-based training. *Computers in Human Behavior*, 1–9.
Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.
Shute, V. J., & Zapata-Rivera, D. (2008). Adaptive technologies. In J. M. Spector, M. D. Merrill, J. G. van Merrienboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 277–294). Mahaw, NJ: Lawrence Erlbaum Associates.

Smits, M. H., Boon, J., Sluijsmans, D. M., & 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*(2), 183–193.
Sosnovsky, S., Dietrich, M., Andrès, E., Goguadze, G., Winterstein, S., Libbrecht, P., et al. (2013). Math-bridge: bridging the gaps in European remedial mathematics with technology-enhanced learning. In P. Fischer, R. Hochmuth, D. Frischemeier, & T. Wassong (Eds.), *Using tools for learning mathematics and statistics* (pp. 437–451). Berlin Heidelberg: Springer-Verlag.
Spada, H., Plesch, C., Wiedmann, M., Kaendler, C., Deiglmayr, A., Mullins, D., et al. (2012). *D1.6: Final report on the STELLAR Delphi study: Future directions for TEL and TEL research: Areas of tension, core research areas, and grand challenge problems*. In *The STELLAR Network of Excellence*.
Thurlings, M., Vermeulen, M., Bastiaens, T., & Stijnen, S. (2012). Understanding feedback: a learning theory perspective. *Educational Research Review*, (9), 1–15.
Timmers, C. F., Broek den, J. B.-V., & van den Berg, S. M. (2012). Motivational beliefs, student effort, and feedback behaviour in computer-based formative assessment. *Computers & Education*, 1–19.
Tobias, S. (1989). Another look at research on the adaptation of instruction to students characteristics. *Educational Psychologist, 24*(3), 213–227.
Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64*(1), 37–54.
Tuckman, B. W., & Sexton, T. L. (1992). The effects of informational feedback and self-beliefs on the motivation to perform a self-regulated task. *Journal of Research in Personality, 26*(2), 121–127.
Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior, 27*(1), 118–130.
VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221.
Vasilyeva, E., De Bra, P., Pechenizkiy, M., & Puuronen, S. (2008). Tailoring feedback in online assessment: influence of learning styles on the feedback preferences and elaborated feedback effectiveness. In P. Diaz, Kinshuk, I. Aedo, & E. Mora (Eds.), *2008 Eighth IEEE international conference on advanced learning technologies* (pp. 834–838). Los Alamitos, CA: IEEE Computer Society.
Vasilyeva, E., Pechenizkiy, M., & De Bra, P. (2008). Adaptation of elaborated feedback in e-learning. In W. Nejdl, J. Kay, P. Pu, & E. Herder (Eds.), *Adaptive hypermedia and adaptive web-based systems* (pp. 235–244). Berlin Heidelberg: Springer.
Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: a framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*(1), 49–64.
Wolf, T. (2010). Assessing student learning in a virtual laboratory environment. *IEEE Transactions on Education, 53*(2), 216–222.
Wulfeck, W. (2009). Adapting instruction. In D. D. Schmorrow, I. V. Estabrooke, & M. Grootjen (Eds.), *Foundations of augmented cognition neuroergonomics and operational neuroscience. 5th international conference, FAC 2009* (pp. 687–695). Berlin Heidelberg: Springer.