Adapting Tutoring Feedback Strategies to Motivation

Susanne Narciss\textsuperscript{1}, Sergey Sosnovsky\textsuperscript{2}, Eric Andres\textsuperscript{2}

\textsuperscript{1} University Passau, Lehrstuhl fuer Psychologie, Dr. Hans-Kapfinger Str. 14b, 94030 Passau
Susanne.Narciss@uni-passau.de

\textsuperscript{2} German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3, 66123 Saarbruecken, Germany
{sergey.sosnovsky, eric.andres}@dfki.de

Abstract. This paper investigates tutoring feedback strategies adaptive to student motivation. Several static feedback adaptation strategies have been designed based on the interactive tutoring feedback (ITF) model, implemented within the Adaptive Educational System (AES) ActiveMath and evaluated in a study with 6th and 7th graders. Student motivation profile (high vs. low perceived competence and intrinsic motivation) has been used to assign them to either conceptual or procedural feedback condition. The data analysis shows that, for low motivated students, the benefits of adapting feedback to motivational profiles are visible in both, performance during treatment and knowledge gain from pre- to post-test. For highly motivated students, no significant effects have been registered. These findings shed light on the role of motivation in tutoring feedback processing and have important methodological implications for designing feedback strategies in AESs.

1 Introduction

Effective feedback is one of the most powerful factors influencing learning in various instructional contexts including technology-enhanced learning (TEL) environments (e.g., \cite{1, 2}). The term “feedback” refers to all post-response information helping students regulate further learning \cite{3, 4}. This notion of feedback can be traced back to early cybernetic views of feedback and implies that the core aim of feedback is to eliminate the gap between the current and the desired states of learning \cite{5, 6}.

The interactive potential of modern TEL environments makes it possible to provide students not only with basic feedback types merely communicating the outcome of a student response, but also with tutoring feedback guided by strategies. Tutoring feedback strategies combine formative elaborated feedback (such as hints, explanations, attribute-isolation examples) with instructional design methods (e.g., mastery learning or scaffolding) to support students in acquiring target competencies, detecting errors, overcoming learning obstacles, etc. In doing so, tutoring feedback strategies offer
"strategically" useful information for task completion without immediately providing the correct solution [3, 12].

There is a fast-growing body of empirical research on feedback adaptation. Corbalan, Paas, and Cuypers investigated the effect of different feedback levels (after each step vs. on the final step) in basic linear algebra tasks on perception, learning and motivation [7]. They found that students in general tend to prefer fine-grained feedback on each step and suggested the need for future studies on adapting the feedback level to individual student expertise. Mitrovic, Ohlsson, and Barrow evaluated the impact of providing positive feedback in a constraint-based tutor that previously delivered only post-error feedback [8]. Their findings indicate that the inclusion of positive feedback can lead to faster learning. The research of Dennis, Masthoff, and Mellish shows that students themselves consider that adapting feedback may be not needed for successful students, but can be very important when a student is under-achieving [9]. Conati and Manske explored the effectiveness of adaptive feedback in a serious learning game. Their results suggest that when feedback is too frequent and more informative than the students need, its effectiveness decreases [10]. Goldin, Koedinger, and Aleven data-mined several models measuring individual feedback value and found that, for different students, the value of feedback messages is different; this allowed them to hypothesise about the existence of a feedback processing skill that defines the individual effectiveness of feedback depending on the level of provided details [11].

Most of the existing research, however, focuses on adapting feedback to student cognitive and meta-cognitive characteristics. The purpose of this work is to investigate how tutoring feedback strategies can be adapted to student motivation. In the reminder of this paper, we present an experimental study, which explores if and under what conditions students with different motivational profiles benefit from different tutoring feedback strategies.

2 The Context of this Work

2.1 Interactive Tutoring Feedback Model

The ITF-model considers feedback as a multidimensional instructional activity helping students to regulate their learning process and assisting them in acquiring knowledge and competencies needed to perform learning tasks. Such conceptualization of feedback facilitates analysis of possible factors and effects of (tutoring) feedback strategies [2, 3, 12]. Rooted in regulatory paradigms from the systems theory, the ITF-model views feedback as one of the several basic components of a generic feedback loop and suggests to differentiate between the two interacting feedback loops: the student feedback loop and the loop of the external feedback source. The controlled process for the two interacting feedback loops consists of the acquisition of competencies necessary to master the requirements of learning tasks. Building on models of self-regulated learning (e.g., [13]) and on current multidimensional notions of competencies (e.g., [14]), the ITF-framework suggests to determine the controlled
variables for a learning process with regard to the cognitive, motivational, and meta-cognitive variables that are relevant for acquiring the desired competencies.

To regulate and control the acquisition of competencies, both feedback loops need to determine a standard or reference level for each of the controlled variables. Furthermore, it is assumed that the actual state of competencies is (continuously) assessed by both the student (internal assessment) and the external feedback source (external assessment). These assessments result in the internal and external feedback on the actual state of the student’s competencies. In both feedback loops, the feedback on the current state of competencies has to be compared to the respective desired level of competencies in order to evaluate to what extent the desired level of competencies has been achieved. Based on this comparison, the external controller (e.g., an AES or a teacher) generates an external feedback message. If a gap between the desired and the current level of competencies has been detected, this external feedback message may provide evaluative information pertaining to this gap, as well as tutoring information (i.e., a suggestion of an action that would help eliminating the gap).

The student has to process the external feedback along with the internal feedback in order to generate ideas or suggestions regarding one or several control actions that may help to proceed in the direction of the desired level of competency. Finally, the student has to implement the selected control action.

Given the complexity of this interaction between the feedback loops, the ITF-model suggests that the effects of tutoring feedback strategies depend not only on the properties of the feedback strategy, but also on various conditions and characteristics of the feedback receiver (i.e., a student), the parameters of the feedback source, and the instructional context [12]. An essential precondition for external feedback being efficient is the mindful processing of this feedback by the student.

2.2 AtuF Project

This work was part of the project “Adaptive tutoring Feedback” (AtuF) that has the overall goal to systematically develop and evaluate adaptation strategies that allow tailoring tutoring feedback to students’ needs. In the first project phase, the methodological, conceptual and technological framework for investigating tutoring feedback adaptation has been developed. Additionally, potential factors that should be taken into account when designing feedback adaptation strategies were investigated with an experimental study [15]. This work presents the results of the 2nd project phase which focused on the design and evaluation of static feedback adaptation strategies tailoring the content and sequence of feedback messages to student motivational profiles.

3 Experiment Description

3.1 The domain of fraction arithmetic

The target domain of this research is the subset of basic fraction arithmetic taught in 6th and 7th grades of the German school program. We have employed the two-
dimensional framework of fraction competencies [16] to represent target knowledge components and cognitive operations in the domain, as well as the typical errors students make while solving fraction exercises. All assessment and training exercises were modelled in terms of the elements of this framework. On the top level, exercises were grouped into the following five large categories based on the requirements addressed by the encoded typical error:

- **Represent**: Representation of fractions in various formats (e.g., candy bar, pie)
- **Transform**: Expansion and cancelling of fractions
- **Add (common)**: Addition of fractions with common denominators
- **Add (uncommon)**: Addition of fractions with uncommon denominators
- **Order**: Ordering and comparison of fractions

### 3.2 The exercise environment of ActiveMath

ActiveMath is a Web-based intelligent AES for mathematical domains [17]. It allows students to work on learning material at their own pace with various degrees of system support, ranging from intelligent problem solving support to personalized courses and adaptive navigation [18].

Interactive exercises in ActiveMath can have multiple steps and combine various types of tasks from single choice to formula input. On the representational level, ActiveMath exercises are finite state machines (FSM) consisting of nodes corresponding to (sub-)tasks and system’s feedback, and transitions encoding correct and typical incorrect student responses [19]. Representations of an exercise can be automatically enriched by a tutoring strategy that defines a general pattern for reacting to the actions and requests of the student solving the exercise [20]. Technically speaking, a tutoring strategy is a transformation of a given exercise automaton into a new exercise automaton, which expresses the needed tutoring behaviour with respect to the instructional goals of the strategy.

Based on this framework, the AtuF exercise strategies can express different tutoring behaviour by varying the sequence and the types of feedback messages presented in a concrete instance of the exercise depending on various information stored in the student model. For example, the same exercise can offer a conceptual hint to a student with a high level of motivation and a procedural hint to a low motivated student.

### 3.3 Assessment tests

In order to test target student competencies before and after training, two parallel sets of fraction assessment items were created. The items were designed by systematically combining conceptual and/or procedural knowledge components with one of the cognitive operations. The sets covered all of the suggested processes and knowledge types and their combinations with items of various response formats (e.g., multiple-choice, short-answer, constructed-response, and graphical input) [21].

To examine the psychometric properties of these tests, a dedicated pen-and-paper experiment was conducted with 293 6th and 7th graders. The analysis showed that the item difficulties ranged from .03 to .88. The difficulties were normally distributed for
each set with a mean of .45 (pre-test) and .41 (post-test). For each set of tasks the mean discriminatory index was .31 (ranging from -.25 to .61). Reliability-analysis showed a Cronbach's alpha of .73 (pre-test) and .80 (post-test), containing only results of students who worked on every task.

Finally, we have selected two isomorphic subsets of these exercises for implementing in ActiveMath. The resulting pre- and post-tests consisted of 29 items each. While taking the tests within ActiveMath, students worked with series of exercises grouped according to the target operators (the number of exercises within groups ranged from 4 for Represent to 8 for Order). Within each group, students were limited in time. If a student did not answer all the exercises from a group, ActiveMath interrupted her/him and transitioned to the next group.

### 3.4 Training fraction competencies with tasks with typical errors

In AtuF, tasks with typical errors (TWTE) were used for training fraction competencies. The TWTE were designed around typical errors in the domain identified on the basis of psychological and empirical task and error analyses [22].

We used TWTE instead of conventional fraction tasks to investigate more precisely the question of how efficient the developed feedback strategies are for various typical errors. Addressing this issue with conventional fraction tasks is a methodological challenge due to a high variability of students' error rates and of the types of errors occurring when students work on conventional tasks. Using TWTE for evaluating the effects of the developed feedback components helped us to design and deliver feedback messages that directly target the known error.

Interaction with a TWTE differs from a conventional exercise. In a TWTE students do not commit their own errors by calculating an answer to an exercise; instead, they are provided with a worked step-by-step solution of an exercise that already contains a typical error. Then, as the first task, they are asked to detect the error. If they respond correctly, the system acknowledges it; otherwise, they are provided with a simple knowledge-of-result feedback that indicates the correct option. Figure 1 demonstrates an example of a student interaction with the error-detection phase of a TWTE.

In the second phase, students have to correct the erroneous exercise step (see Figure 2). If they are not able to correct the error on their own, they receive tutoring feedback. Students have three trials to accomplish this task, after each incorrect trial, the feedback details increase. After the third incorrect trial, ActiveMath presents the final feedback in terms of a worked-out solution and moves to the next exercise.

Overall, the treatment session included 30 TWTEs. Students received them in blocks of ten exercises. Within each block, exercises followed the same sequence (in terms of the covered operator): Represent, Represent, Transform, Transform, Add (common), Add (common), Add (uncommon), Add (uncommon), Order, Order. When analysing the data (see Section 4 below), we had to exclude the logs for all TWTEs covering ordering of fractions, because the implementation of these tasks did not follow the same algorithm as for the other eight (instead of separate error-detection and error-correction tasks, students were asked to identify a correct procedure and execute it in one single step). The final set of treatment exercises consists of 24 items.
Figure 1. Interface of an AtuF training exercise (error detection task)

Figure 2. Interface of an AtuF training exercise (error correction task)
3.5 Motivation questionnaire

To assess students’ motivation we used the Expectancy-value-Form of domain-specific Learning Motivation (EWF-LM). The EWF-LM has been developed on the basis of an integrative expectancy-value model of motivation (see [15] for more detail), and consists of the four scales: intrinsic value (eight items; e.g., “I enjoy solving fraction exercises”; Cronbach’s alpha = .87), attainment value (five items; e.g., “Fraction tasks offer me an exiting occasion for demonstrating my abilities”; Cronbach’s alpha = .73), perceived competence (three items; e.g., “I am good at solving fraction tasks”; Cronbach’s alpha = .81), and fear of failure (three items; e.g., “I am worried about mistakes, even if nobody would see them”; Cronbach’s alpha = .81) in relation to fraction tasks. Students had to respond to the EWF-LM items on a 6-point rating scale (0 = “Not true at all for me”; 6 = “Completely true for me”). Based on the findings from the prior AtuF studies [15, 23], the scales for intrinsic value and perceived competence were selected to guide the feedback adaptation.

3.6 Feedback strategies

Feedback messages in AtuF can vary along two dimensions. From the specificity standpoint, a feedback message can be a hint pointing out an appropriate procedure or a key concept, or it can be a more elaborate explanation describing how the procedure should be applied or why the particular concept is relevant to this task. From the point of type of knowledge communicated by the hint, the feedback can be procedurally or conceptually oriented. For example, in a TWTE on fraction expansion one of the following four messages could be delivered to a student depending on the strategy and the student progress with the task:

- Procedural hint: “When expanding a fraction, one alters the numerator and the denominator equally.”
- Procedural explanation: “When expanding a fraction, one alters the numerator and denominator equally. To do this, multiply the numerator and denominator by the same number.”
- Conceptual hint: “When expanding a fraction, its value must not change.”
- Conceptual explanation: “When expanding a fraction, its value must not change. While the fraction is expanded, 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.”

Using the ActiveMath exercise strategy framework described in Section 3.2, several static adaptation strategies for TWTE were developed. They do not alter the error detection phase in any way, but vary presentation of tutoring feedback in the error correction phase of TWTEs. Informed by the results of previous AtuF studies, we have made feedback sequences sensitive to both, student and content characteristics. Table 1 summarizes all feedback conditions.

---

2 The messages are translated from German
Table 1. Feedback strategies

<table>
<thead>
<tr>
<th>Operator</th>
<th>Strategy A</th>
<th>Strategy B</th>
<th>Strategy C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add (common)</td>
<td>KR+P-Hi; KR+P-Ex; KCR</td>
<td>KR+C-Hi; KR+C-Ex; KCR</td>
<td>KR; KR; KCR</td>
</tr>
<tr>
<td>Add (uncommon)</td>
<td>KR+P-Ex; KR+P-Ex; KCR</td>
<td>KR+C-Ex; KR+C-Ex; KCR</td>
<td>KR; KR; KCR</td>
</tr>
<tr>
<td>Order</td>
<td>KR+P-Hi; KR+P-Ex; KCR</td>
<td>KR+C-Hi; KR+C-Ex; KCR</td>
<td>KR; KR; KCR</td>
</tr>
</tbody>
</table>

Note: KR = knowledge of result (correct/incorrect); P-Hi = procedural hint; C-Hi = conceptual hint; P-Ex = procedural explanation; C-Ex = conceptual explanation; KCR = knowledge of the correct response (worked-out solution).

Three feedback strategies were developed: Strategy A predominantly uses procedural feedback; Strategy B is symmetric to A, but employs conceptual feedback components instead of procedural; Strategy C implements the control condition, it does not use conceptual and procedural hints and explanations, instead it presents simple knowledge-of-result feedback followed by the elaborated worked-out solution. We hypothesized that the procedural strategy A would be more beneficial for students with low motivation, while for highly motivated students either type of feedback strategy might be helpful.

Within strategies A and B, depending on the operator covered by the TWTE, the specificity of feedback differed. Our previous experiments have shown that for some operators higher feedback specificity does not bring any value. For example, exercises related to the Represent operator have very high solving rates (>96%) and, thus, do not require elaborated feedback components for the target population.

3.7 Design and procedure

The study consisted of six phases (Figure 3). The pre-questionnaire was used to record demographic data of the students (such as age and gender), as well as to assess their motivational parameters. A short familiarization session was followed by the pre-test assessing students’ prior knowledge of fractions. During the training session, students were exposed to different experimental (and control) feedback conditions. Finally, the post-test and post-questionnaires were used to obtain estimations of their posterior knowledge and motivation.

Figure 3. Overview of the experiment procedure

All interactions with the systems were stored in the log files of ActiveMath; this included student answers to questioners, assessment and treatment exercises.

Based on the results of the motivation pre-questioner, for each student, a cumulative motivation metric was computed as the average score of her answers to the ques-
tions about intrinsic motivation and perceived competence. If the metric value was below the threshold = 45 (experimentally identified during the previous AtuF study), the student was categorized as low motivated, otherwise she/he was considered a highly motivated student. After that, ActiveMath automatically assigned the student to one of the three feedback conditions trying to balance the groups with regard to motivation and gender.

4 Data Analyses and Results

During the treatment session, students were allowed to work at their own pace with as many tasks as possible within the given time limit. We inspected ActiveMath log files to identify the data of students who (a) worked too slowly to complete even the first block of 8 TWTEs, (b) succeeded in correcting all errors without any support (i.e., were not provided with tutoring feedback at all). We excluded these cases from further data analyses, since tutoring feedback messages can only be beneficial if students observe them at least several times [12]. We also had to exclude a few incomplete logs (data was missing due to system failures). The resulting sample used in this study consists of data from 128 students (57 female, 66 male; five preferred not to answer).

Table 2 shows how the students with different motivation profiles (i.e., low perceived competence and intrinsic value vs. high perceived competence and intrinsic value) have been assigned to the feedback conditions.

Table 2. Cross tabulation of feedback strategies and students’ motivation

<table>
<thead>
<tr>
<th>Feedback Strategy</th>
<th>perceived competence &amp; intrinsic value</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>procedural</td>
<td>17</td>
<td>36</td>
</tr>
<tr>
<td>conceptual</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>control</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>85</td>
</tr>
</tbody>
</table>

4.1 Preliminary analyses

Prior to performing statistical data analyses, we screened the data for outliers and normality. While performing the analyses, assumptions of the procedures were tested and met, unless otherwise noted. Furthermore, since the assignment to the feedback strategies was done based on students’ level of perceived competence and intrinsic value, we checked if there were any differences in the level of fraction competency prior to the treatment (performance on the pre-test) and for the first TWTE of each type during the treatment (baseline of successful error corrections in the treatment; see Table 3 for descriptive data). The analysis revealed no significant differences for these variables across different feedback strategies (all F<1), yet the low motivated students under the procedural feedback condition started with a significantly lower level of fraction competency as demonstrated by the significant interaction among the factors feedback and motivation ($F(2,122) = 3.54; p=.03; \eta^2=.06$). These students
had also the lowest baseline regarding their success-rate in correcting the TWTE, yet the difference there was not statistically significant.

### Table 3. Overview of means and standard deviations for competency levels, number of TWTE worked on, baseline of successful error correction in the treatment and feedback efficiency by feedback conditions and motivation

<table>
<thead>
<tr>
<th>Feedback strategy</th>
<th>Motivation</th>
<th>procedural</th>
<th>conceptual</th>
<th>control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>low</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Competency level</td>
<td>low</td>
<td>14.61</td>
<td>5.38</td>
<td>12.65</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>16.38</td>
<td>4.40</td>
<td>17.28</td>
</tr>
<tr>
<td>Post-test</td>
<td>low</td>
<td>16.59</td>
<td>4.75</td>
<td>16.59</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>18.32</td>
<td>3.97</td>
<td>18.89</td>
</tr>
<tr>
<td>Number of TWTE</td>
<td>low</td>
<td>19.00</td>
<td>4.46</td>
<td>18.47</td>
</tr>
<tr>
<td>worked on</td>
<td>high</td>
<td>16.59</td>
<td>3.95</td>
<td>16.47</td>
</tr>
<tr>
<td>Baseline</td>
<td>low</td>
<td>2.09</td>
<td>1.09</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>2.21</td>
<td>1.05</td>
<td>2.31</td>
</tr>
<tr>
<td>Treatment</td>
<td>low</td>
<td>.25</td>
<td>.28</td>
<td>.40</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>.27</td>
<td>.24</td>
<td>.27</td>
</tr>
<tr>
<td>Feedback</td>
<td>low</td>
<td>.33</td>
<td>.30</td>
<td>.45</td>
</tr>
<tr>
<td>efficiency</td>
<td>high</td>
<td>.37</td>
<td>.28</td>
<td>.42</td>
</tr>
</tbody>
</table>

### 4.2 Differential Effects of Feedback during Treatment

During the treatment, the participants had up to three attempts per error-correction task. If a student failed on the first attempt, she received a feedback message addressing this error according to the experimental condition. To verify how effective these messages were in terms of helping students correct the typical error in the next attempt with the task, we have computed two measures characterizing feedback efficiency: initial feedback efficiency and overall feedback efficiency (see Table 3). The first measure quantifies the efficiency of feedback provided after the first unsuccessful error correction attempt, and is computed as the number of successful error corrections after the first feedback divided by the difference between the number of all attempted tasks and the number of tasks solved correctly on the first attempt (i.e., tasks solved without any feedback). Additionally, the overall feedback efficiency is computed as the total number of successful error correction attempts after feedback of any level divided by the difference between the number of all attempted tasks and the number of tasks solved correctly on the first attempt (before seeing any feedback).

A feedback (3) by motivation (2) ANOVA for initial feedback efficiency revealed a significant main effect of feedback, $F(2,122) = 3.34; p = .04; \eta^2 = .05$, and a two-way interaction $F(2,122) = 3.05; p = .05; \eta^2 = .05$. For students with low motivation, procedural feedback strategy was significantly more efficient than the conceptual feedback strategy or the control condition strategy that provided a worked-out solution after the third incorrect attempt. For students with high motivation, there were no statistically significant differences in feedback efficiency after the first incorrect attempt (see Figure 4).
Regarding the overall feedback efficiency, we have also found a significant main effect of feedback $F(2, 122) = 3.14; p = .05; \eta^2 = .05$. The procedural feedback strategy was more efficient than the conceptual and the control strategy (see Figure 4).

**Figure 4.** Feedback efficiency after receiving feedback on the first unsuccessful error-correction attempt (left), and over all attempts (right) for students with low and high motivation by feedback condition

**Differential Effects of Feedback on Learning Gain**

The role of feedback as a teaching intervention is not only to facilitate problem solving, but, first and foremost, to improve students’ learning. To investigate if, how and under which individual conditions, the feedback strategies affected knowledge gain from pre- to post-test, we run a repeated-measure ANOVA with the between factors feedback (3), motivation (2), and the within factor (2) for the fraction competency measures of the pre- and post-test.

This analysis revealed a significant three-way interaction with Wilks’ Lambda = .94, $F(2, 122) = 4.07$, $p = .02$, $\eta^2 = .06$. Students with low perceived competence and intrinsic value improved their level of fraction competency significantly more with the procedural feedback strategy than with the conceptual or the control strategy. It is noteworthy, that under conceptual feedback condition, they had the smallest knowledge gain. Students with high perceived competence and intrinsic value improved their fraction competency level comparably under all feedback strategies. Figure 5 illustrates these findings.

It should be noted that low motivated students with the procedural feedback strategies had a significantly lower level of prior competency in the pre-test than all other groups. Yet, for all groups there was room for improvement since the fraction tests consisted of 29 items. Moreover, it seems interesting that irrespective of the feedback strategy, the low motivated students worked on significantly more TWTEs during treatment than the highly motivated students ($F(2, 122) = 6.17; p = .01; \eta^2 = .05$), (see Table 3 for the descriptive data).
Figure 5. Learning gain from pre- to post-test for students with low vs. high perceived competence and intrinsic value by feedback condition

5. Summary and Conclusions

This paper investigates the issue of how tutoring feedback within an AES can be adapted to the pre-existing level of student motivation. The domain of basic fraction arithmetic was used as an example; and the study focused on the two motivation parameters: perceived competence and intrinsic motivation. The findings support the hypothesis that students with low motivation would benefit from a feedback strategy providing procedural tutoring feedback messages more than from a feedback strategy with conceptual feedback messages. The findings also show that, for students with high motivation, either type of feedback strategy works equally well.

This experiment reveals some challenges of applying the ITF-model to the evaluation of tutoring feedback strategies. In order to provide students with occasions to use the feedback information for correcting the errors of the TWTE, we implemented two static multi-trial feedback strategies. These feedback strategies provided students with up to three trials to correct errors, and students were free to work at their own pace. The next level of feedback was offered only when a student did not succeed on the previous trial with the less detailed level of feedback. Through this approach, we obtained a rich dataset for investigating how feedback conditions, feedback properties and student feedback processing contribute to feedback efficiency. Yet, we were faced with a number of methodological challenges, including high student variability in terms of working pace, and the number of trials needed to complete tasks. The findings indicate, for example, that students with low motivation worked on more TWTEs than students with high motivation, but they had a lower success rate in completing the error correction step of TWTEs. Further analysis is needed to more accu-
rately account for the actual feedback processing. We plan to take a deeper look at the
timestamp data of the log files in order to investigate if and how the time students
spend on answering the task after a feedback message is connected to successful error
correction.

Another challenge relates to the issue of how to assign students to the feedback
strategies. Based on our prior findings, the students of the present study were assigned
to the two feedback strategies according to their pre-existing level of motivation.
Their initial level of competency was controlled, but not used as an assignment crite-
rion. To classify students’ initial levels of motivation, the assignment procedure used
the median of the motivational variables obtained in a prior study with 207 students
from the same population (6th and 7th graders from Dresden schools) [15]. This as-
signment procedure proved to have limitations, since the median was different for the
present sample compared to the sample of the prior study. Furthermore, the comp-
etency level of the three low motivation groups was rather dissimilar
on the pre-
test.

Further development and evaluation of the classification and assignment procedure
integrating motivational, cognitive, and metacognitive variables into a feedback adap-
tation strategy is an interesting practical and methodological challenge that we plan to
address in our future work.

References