Young Children's Experience and Preference of Feedback – Sense and Sensibility

Betty Tärning[1], Björn Sjödén[3], Agneta Gulz[1,2], Magnus Haake[1]

[1] Div. of Cognitive Science, Lund University, Helgonavägen 3,
223 62 Lund, Sweden
{betty.tarning, agneta gulz, magnus.haake}@lucs.lu.se
[2] Dept. of Computer Science, Linköping University, Campus Valla,
581 83 Linköping, Sweden
agneta gulz@liu.se
[3] Dept. for Teacher Education, Halmstad University, Kristian IV:s väg 3,
301 18 Halmstad, Sweden
bjorn.sjoden@hh.se

Abstract. This study explored the effects of adding visual continuous feedback in the form of feedback bars to a teachable-agent based learning game in mathematics. Forty-five (45) children, 8- to 12-years-old, from three Swedish school classes used the game during four math lessons. The focus was on how feedback to the students regarding their teachable agents learning progression – and different detailedness of such feedback – affects how the students (in a teacher role) experience the learning game. The results suggest that students were positive towards receiving immediate and continuous feedback, but their preferences with respect to the detailedness of the feedback differed according to their age. We found a divergence as to the preferred number of bars, where the 3rd and 5th graders preferred 1 or 3 bars but where the 2nd graders preferred the more detailed version (6 bars) despite their lack of understanding of what the different bars represented.

Keywords: Educational Technology; Recursive Feedback; Teachable Agents; Mathematics

1 Introduction

When a student encounters a new situation or task, the feedback provided to her can be crucial for the outcome of her performance. The scientific literature on feedback with respect to learning is vast. Most studies focus on written or spoken feedback provided either from a human or a computer to the student, and it is well established that feedback in general has a positive effect on students learning progression (e.g. [1, 2, 3]).

The growth of digital learning technologies enables the development of novel pedagogical approaches as well as the use of different kinds of feedback. One of these novel digital pedagogical approaches is the *teachable agent* paradigm [4] built upon the pedagogical principle of *learning-by-teaching*, i.e. the student takes the role of a teacher and teaches a digital tutee (see below). As the student here takes a teacher role, the teachable agent paradigm activates the use of *recursive feedback* [5], which is a feedback form normally reserved for the teacher. More specifically, recursive feedback is feedback to the teacher (in this case a student taking a teacher role) on her own knowledge and her own teaching gained through the observation of how her tutee (in this case a digital tutee) performs. Observing her tutees responses to what she has taught her or him, indirectly provides information on the quality of her own teaching.

Given the substantial evidence that feedback in general, and recursive feedback specifically, can have a positive effect on learning, this study focuses on students' experiences and preferences with respect to *recursive feedback* within the teachable agent paradigm.

Teachable agent-based learning environments [4] – the digital pedagogical context of our study – are built around the well-known pedagogical principle of *learning-by-teaching*. This principle states that teaching someone else is generally a powerful strategy to learn for oneself [6, 7, 8, 9]. In a teachable agent based educational game, the teachable agent (hence called *digital tutee*), initially exhibits no knowledge about the specific learning domain, but gradually learns from the student who in various ways takes a teacher role by explaining, showing and correcting the digital tutee – as well as answering questions. This role shifting in teachable agent-based learning environments thus promotes *recursive feedback*. Here, the student in her teaching role (henceforth also called student-teacher), gets feedback via her tutee's behavior. This is in contrast to the regular case of getting direct feedback based on the results of one's actions.

In sum, the present study focuses on how recursive feedback provided by a digital tutee's learning progression – and different level of details of such feedback – affects how students in their teacher roles experience the use of the learning game.

1.1 Related studies

Betty's brain [4, 10] is considered the first digital educational game using a teachable agent (TA). In the TA-game, students teach their tutee different topics on eco-systems by building causal models. In the original version of Betty's brain, the recursive feedback was implemented using the opportunity to pose quizzes and queries in order to experience the learning progress of the digital tutee. Three implementations of recursive feedback were evaluated in a quantitative study [10] with 50 fifth graders assigned to one of four experimental conditions (teach only, quiz, query, quiz & query). The results showed that students who could quiz Betty were helped in selecting what to teach Betty, whereas students who could query Betty improved their understanding of relations between concepts. In another, later, study [5] further confirmed that recursive feedback in TA-games have positive effects on students' performance and knowledge transfer. Ninety-four high school students were randomly assigned to one of four conditions, including 15 students assigned to a control condition. Forty students used a TA-version (*Moby*) of a learning game targeting logical reasoning (representing the recursive feedback condition). The two other conditions used different non-TA versions of the same learning game (regular feedback conditions). The students in the recursive feedback condition exhibited superior abilities to use logic and solve novel problems compared to the other groups.

However, in the two TA-games discussed above, students receive recursive feedback only when they enter a specific test mode, i.e. they have to take explicit actions such as assigning a quiz to their digital tutee. In contrast, in the present study students use an educational teachable agent game for math [11], where recursive feedback of the digital tutee's status is provided continuously during the entire game sessions. More specifically, in the teachable agent game (*Betty's Brain*) used in the studies by [10], the recursive feedback was provided after the learning activity by observing the agent's behavior when it is queried or quizzed. Likewise, in the teachable agent game (*The Moby Environment*) used by [5] the recursive feedback was provided after the learning activity by observing Moby playing a game. In contrast, the teachable agent game used in our study provided us with the possibility to provide continuous recursive feedback during the actual learning activities, as well as observing the teachable agent playing on her own.

On a more general level there are research fields targeting how young people handle the visual dynamic richness encountered in digital resources of different kinds, e.g. in the domains of Vision Research (e.g. [12]) and of Visual Literacy (e.g. [13, 14, 15]). However, studies on how school children handle visual information of different degrees of detailedness in the context of digital math games are to our knowledge absent. This includes studies that would focus on visual continuous feedback on development of knowledge in an area of math knowledge.

1.2 The original vs. the feedback extended versions of the game

The math game used in the study (*The Square Family* [11], figure 1) targets math and – more specifically – the base-ten concept. In the regular version of the math game, the recursive feedback is not readily accessible and the tutee's level of knowledge (reflecting the student's success as teacher) can only be inferred through the behavior of the digital tutee. A well taught tutee is, for example, likely to win over another student's less well taught tutee or to win over an in-game computer player. As for the actual teaching process there is only implicit information provided through the gradual increase in difficulty of the questions asked by the digital tutee. Students using this math game in previous studies [16] have also asked for more information on their digital tutees progress, not being satisfied with the intermittent indirect feedback.

In order to study continuous recursive feedback, we decided to explicitly visualize the gradual development of the digital tutee's "knowledge". In practice, we decided to address the rule-based game algorithms governing the base-ten concept, the game rules handling carry-overs and the rules managing the choice of cards. The rational for our choice was that these rule-based game algorithms together defines the knowledge domain that the digital tutee needs to master in order to play the game well. This approach is also in line with well-established design heuristics like: A system's status (in this case the tutee's knowledge and learning process) "*[...] should always keep the user informed about what is going on, through appropriate feedback within reasonable time*." [17].

For the actual visualization of the recursive feedback, we used a graphical format with progress bars presenting immediate, explicit and continuous information allowing students at all times to supervise and perceive the gradual development of their digital tutee's learning process.

For this recursive feedback to be genuinely useful, the student-teacher needs a declarative understanding of the material to be taught to the digital tutee. This poses a delicate problem as especially the younger students hardly can be expected to explicitly formulate the mathematics behind the base-ten concept. Thus, in order to evaluate if and how different levels of detail would affect the students' learning experience, three versions of the feedback format were developed with 1, 3, or 6 progress bars. The bars corresponded to a *low*, *middle* and *high level of detail* with regard to the underlying rule-based game mechanics.

1.3 Research questions

With these three feedback conditions (1, 3, or 6 bars) plus a control condition with no feedback bars, our objectives were to examine the effects of continuous recursive feedback and how different levels of detail would affect students' *preference*, *feelings of confidence*, *engagement* and *understanding* when playing the game. Furthermore, we wanted to examine how these objectives above relate to age comparing: 8 to 9, 9 to 10, and 11- to 12-year-olds, corresponding to Swedish school grades: 2, 3 and 5. Our research questions were the following:

1. Given the different feedback configurations (level of detail), in what ways will students:
   a) rate their preferences for the feedback configurations?
   b) experience confidence in what their tutee knows and their own teaching?
   c) experience engagement in their role as a teacher to their digital tutee?
   d) experience an understanding of the different feedback configurations and an understanding in their tutee's knowledge
   e) experience help/support versus distraction?
2. How will age affect the answers to questions 1 a) – 1 e)?

2 Method

2.1 Participants

A total of 45 students from three Swedish school classes from the same school participated: one 2nd grade class (n = 15), one 3rd grade class (n = 15) and one 5th grade class (n = 15). In the Swedish school system, 2nd grade corresponds to 8- to 9-year-olds, 3rd grade to 9- to 10-year-olds, and 5th grade to 11- to 12-year-olds.

In order to balance the groups, at least to some extent, the students were categorized and selected according to their mathematical skill levels as judged by their teacher. Of the 45 students (21 male and 24 female) participating in the study, 13 were categorized as high-achieving, 18 as mid-achieving and 14 as low-achieving. (See table 1.)

Table 1. The participants participating in the study displayed by grade, math skill level and gender.

	Mathematical skill level						
	High		*Middle*		*Low*		
Grade	Boys	Girls	Boys	Girls	Boys	Girls	Σ
2nd	4	1	1	4	3	2	15
3rd	0	3	4	4	3	1	15
5th	2	3	2	3	3	2	15
Σ	6	7	7	11	9	5	

2.2 Materials

The math game. For a detailed description of the educational math game used we refer to [11]. In short, the game trains basic arithmetic skills in a card game environment built around a graphical model of the base-ten concept (figure 1). The base-ten concept, which is trained already from grade 1, is an important basis for math all through school, and also a well-known bottleneck.

The regular version of the math game without feedback bars has been used in several previous studies, e.g. [18] explored the effects of the digital tutee on test performance whereas [19] provided other measures of math comprehension and motivation. In these studies, the game has been shown to have a positive effect on learning. In the present study we decided not to investigate this further and instead focus on the study of feedback. If we had also included pre- and post-tests targeting performance, this would probably have affected our primary focus of interest and increased the burden on the participating students. Furthermore, a reliable set of learning outcome results would require a dedicated "between group" design, considerably increasing the required number of participating students.

The game features two "players", where a player can be a student (in the role of a teacher), a digital tutee, or an in-game computer player. Figure 1 illustrates the game with a student playing to the left, with her digital tutee watching, against a "computer player" at skill level 3 to the right. Each player automatically gets a hand of cards that depict colored squares. A successful game move on the part of the student (in the situation illustrated in figure 1) involves choosing the card that yields a maximum number of carry-overs when added to the common game board in the middle. At the same time the card should ideally minimize the opponent's opportunities for carry-overs with his/her own cards.

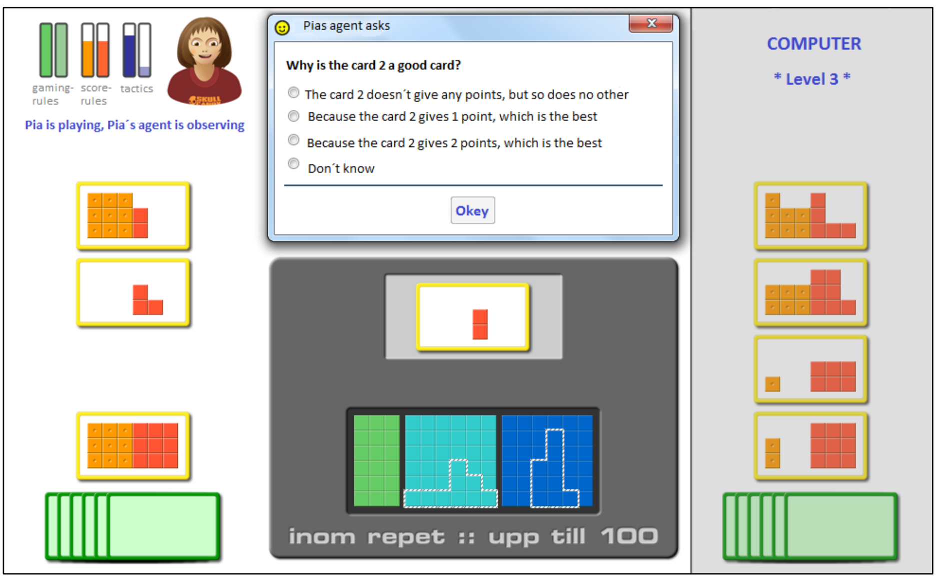


Fig. 1. Screenshots of the math game in the *Student Shows* mode with the digital tutee posing a question in response to the latest card played by the student. The feedback configuration (6 feedback bars) is found to the left of the digital tutee (upper left corner).

The game has four different playing modes: *Student Plays*, *Student Shows*, *Agent Tries* and *Agent Plays*. The four modes correspond to four subsequent steps of teaching and learning but can also be played in optional order.

1. *Student Plays:* The student can play the game by herself without the digital tutee.
2. *Student Shows:* The student shows the digital tutee that, in turn, learns from watching the student play. The tutee then poses questions to the student, for instance asking the student to explain why a choice of card is a good choice, whereupon the student answers the questions via a multiple-choice format.
3. *Agent Tries:* The digital tutee proposes cards based upon what s/he has learnt during *Student Shows*. The student can either accept or reject the proposed card. If the card is rejected, a supposedly better card must be chosen. Multiple-choice questions are also included in this mode to prompt the student to explain why a certain card was better than the card proposed by the tutee.
4. *Agent Plays:* The digital tutee plays on her or his own and the student can watch and reflect upon how well the tutee plays (which mirrors the student's teaching during the previous modes).

As mentioned above, feedback associated with the progress of the digital tutee was scarce in the regular game design. Only in the last mode (*Agent Plays*) would the performance of the digital tutee indicate whether the tutee had been well taught or not.

The feedback extended version of the math game. As stated above, the feedback extended versions of the game used for this study featured three different graphical configurations with 1, 3, or 6 bar(s), dynamically visualizing aspects of the digital tutee's progress and knowledge domain. The graphical dynamic feedback bars were situated next to the visually static digital tutee with explanatory text labels beneath them (figure 1 and 2). The bars did not change in response to every move made by the student-teacher, but according to the gradual fulfilment of the rules governing the game. In the game, this corresponds to the continuous rule-based evaluation of the cards picked by the student-teacher as well as the student-teacher's response to the questions proposed by the digital tutee. Thus, while teaching the tutee, the student-teacher could continuously observe changes in the bars directly mapped to changes in the game algorithms handling the digital tutee's knowledge representation.

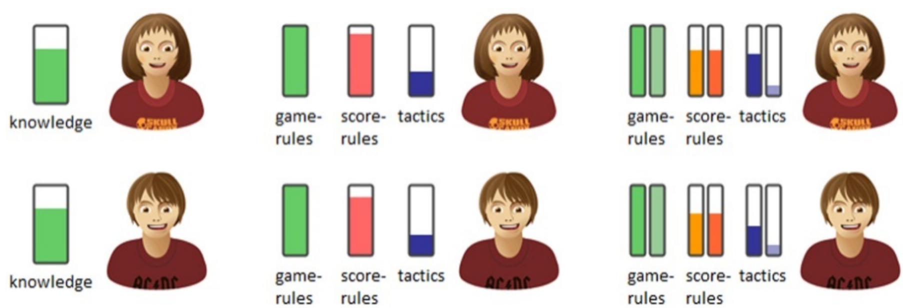


Fig. 2. The three different bar constellations with the male and female digital tutee. *Left*: least detailed configuration; *Middle*: mid-detailed configuration; *Right*: most detailed configuration.

The most detailed feedback configuration presented six bars in three pairs, each pair representing three aspects of the digital tutee's knowledge domain: (i) basic game rules, (ii) scoring rules, and (iii) tactics, correspondingly labelled "game rules," "score rules," and "tactics". The two bars in each pair separated the three knowledge domains with regard to tens and hundreds, respectively. The mid-detailed configuration with 3 bars displayed the same main parameters (game rules, score rules and tactics), but did not separate tens and hundreds. The least detailed configuration displayed only one feedback bar labelled "Knowledge" representing the digital tutee's overall knowledge with basic, scoring and tactics knowledge merged into one parameter. (See figure 2.)

The "game rules" bar responded to the digital tutee's questions about the basic ideas of the game as such, e.g. "*When do you get a star?*" The bar for "score rules" responded to more specific knowledge of how points were gained, e.g. "*Will the card 92 yield points?*" The bar for "tactics" responded to tactically choices of cards, e.g. "*Why is the card 78 better than 67?*" The general "knowledge" bar in the "1 bar configuration" responded to a combination of the previously mentioned parameters.

Questionnaires. In the study, questionnaires targeting the students' experiences (see Appendix) were distributed at four different times during the study period. A first questionnaire (*Questionnaire A*) targeted the regular game version without the added feedback bar(s) and included four questions about students' confidence and engagement: "*How confident are you as to what your tutee knows about the game?*," "*How confident are you in teaching your tutee?*," "*How eager are you to teach your tutee how to play the game?*," and "*What do you think about playing the game?*"

A second questionnaire (*Questionnaire B*) addressed the three feedback conditions (6, 3 and 1 bar(s)). This questionnaire one, two and three repeated the four questions from the first questionnaire, but also included three additional questions concerning the students understanding of the bars: "*Where you helped by the (n) bar(s) in understanding how much your tutee knows about the game?*" "*Did you find the (n) bar(s) easy or difficult to understand?*" and "*Was it disturbing or helpful to have the (n) bar(s) in the game?*"; "n" representing the number of feedback bars the student had encountered. At the last occasion when the students had experienced all three feedback conditions, Questionnaire B was used (with two questions added): "*How many bars did you find to be the best when playing?*" and "*Why?*" These two questions were added in order to record the students' preferences regarding level of detail in the recursive feedback.

A third questionnaire (*Questionnaire C*) was distributed after a final session where the students had been told to play with their own chosen number of feedback bars. In this questionnaire the students were asked two questions about their preferences regarding the level of detail in the recursive feedback: "*How many bars did you play with?*" and "*Why did you choose this one?*" In addition, the students were also asked to draw or write possible suggestions on how their tutees knowledge could be displayed in other ways than in the shape of bars.

The response formats of Questionnaire A and B differed between the grades. The 5th graders responded using visual analogue scale (VAS) format for all questions. For the younger students (2nd and 3rd graders) question 1 to 5 used five smileys ranging from "sad" to "happy" (figure 3) equivalent to a regular five-level Likert scale format. For question 6 and 7 in Questionnaire B, however, a visual analogue scale (VAS) format was used also for the 2nd and 3rd graders since these questions did not fit with the "smiley-format". Both the phrasing and scale formats for the questions were based on Bandura's design guidelines for self-efficacy scales [20]. The students were also familiar with this kind of questionnaires, having used similar ones regularly in the school's evaluation forms. Similar questionnaires have also been used and validated in previous studies (e.g. [21]. For this study, the questionnaire had been evaluated by teachers and pilot tested by external students of the same age groups.

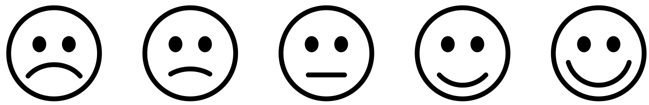


Fig. 3. The five smiley faces used in the questionnaire for the younger students (2nd and 3rd graders).

Preference choices. The students' preferences regarding the level of detail in the recursive feedback (i.e. the number of feedback bars) together with their arguments were recorded at three different times during the study, either directly by the principal investigator or as part of a questionnaire, see Questionnaire B and C above.

Focus group interviews. As a complement to the questionnaires, we followed up with semi-structured focus group interviews at the end of the study period; one for each grade. Each focus group included four students with at least one from each of the categories low, mid and high math-achievers. The interviews were based upon the questionnaires with some questions directly mapped, for example, question 7 in the interview and question 6 in Questionnaire B. Other interview questions were more targeted to probe for complementary information; for example, question 4 in the interview was asked as a complement to question 5, 6 and 7 in Questionnaire B. As a whole, interview questions 1-12 (except question 6 and 9) were in these ways directly or more complementary related to the questionnaires. Interview questions 13-16 were added to gain further information regarding the students' general impression of the game and feedback in general, for example: "*How can we make the bar(s) better?*" (see Appendix).

The interviews were digitally recorded, each lasting about 25 minutes. The interviews were later transcribed and categorized according to our research questions (preference, confidence, engagement and understanding). During the focus group interviews, the range of answers to a particular question shifted from no students answering to all four in the focus group answering. For some of the interview questions, the answers pointed in all different directions, for some there was more of consensus. All relevant results from the focus group interviews are presented in the paper.

2.3 Procedure

Throughout the study each student had their own digital tutee, although the students were playing in pairs, sharing a laptop. That is, each student answered the questions posed by their own digital tutee, while taking turns in front of the shared laptop. The students could not change partners during a session, but they were allowed to switch their partner between sessions. In case of an uneven number of students, one student would play against the computer. The reason why we had the students play in pairs was partly to incite motivation, but foremost that the number of laptops available was limited.

Since we used *three* feedback-extended versions of the math game, we discussed in which order students should be presented with and use them, given our purpose of comparing the effects of the different configurations representing different levels of detail. Due to the limited number of participants possible to recruit, we had no means to control for possible effects of order. Thus, in order to maintain consistency for a between group design, we decided to start with the 6 bars configuration, followed by the 3 bars configuration and lastly the 1 bar configuration. The order of the three experimental configurations (level of detail) could be debated. For the experimental design we made the choice to use a decreasing level of detail so that students would start off with the most detailed version and not have to interpret and process new elements of the feedback presentation format. We reasoned that most students would ascribe a meaningful interpretation of the 1 bar configuration and that this could interfere with and confound the later "meaning-making" of the more detailed configurations. Starting with the most detailed version, on the other hand, means that the gradual necessity for an expanding "meaning-making" is neutralized.

As for the procedure, the students were first introduced to the regular version of the math game without feedback bars in an introductory session. In the following three experimental sessions, all students were presented with the three different configurations of feedback bars. The students were told that the bars represented the digital tutees knowledge regarding the game but were never told in detail what the different bars represented. At the end of each experimental session, the students individually filled out a questionnaire. In a final game playing session, the students were offered a choice of which of the three feedback configurations they wanted to use during this session. Following the last session, there was a focus group interview with selected students. Two experimenters supervised all sessions and the whole study period lasted about 4 weeks. (See figure 4.)

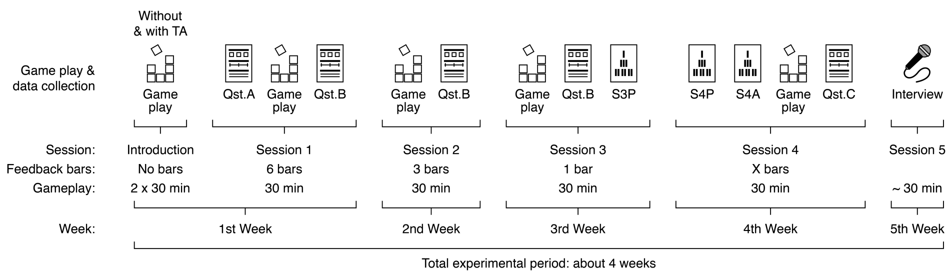


Fig. 4. Flow chart of the study procedure: The three questionnaires used in the study are abbreviated as: *Qst.A*, *Qst.B* and *Qst.C*. The three preference choices recorded in the study are abbreviated *S3P*, *S4P* and *S4A*.

In the following, the different study sessions are presented more in detail. (See also figure 4.)

Introductory session. All students received a total of 60 minutes introduction to the game during two subsequent sessions of 30 minutes each. During this introduction session, they played the regular game version without feedback bars. The students first played without the digital tutee in the mode *Student Plays*, where after they acquainted themselves with their digital tutee in the three modes: *Student Shows*, *Agent Tries* and *Agent Plays*. In order to balance out possible confounding agent gender effects, half of the students were assigned a female digital tutee and the other half a male digital tutee. This assignment was randomized and equally distributed across gender. All students played the same sub-game targeting basic addition with carry-overs.

First game session. Students began by filling out the 4 questions of Questionnaire A with respect to their experiences from the introduction session. After finishing the questionnaire, the students played the game with the 6 bars configuration, whereupon they filled out the 7 questions of Questionnaire B. At the start of this session the students were also told that there would be a competition held at the last session where we would see which of the students had done the best job in training their digital tutee. This way we hoped to motivate the students to focus on playing and encourage them to keep an eye on the feedback bars. For the rest of the study, we never instructed or prompted the students to specifically study the bars while playing.

Second and third game sessions. During the following two game sessions, students proceeded to play the game with 3 and 1 feedback bar, respectively. Both sessions ended with filling out the same 7 questions (Questionnaire B) as in the end of the first game session. In addition, at the end of the third game session, Questionnaire B was extended with a section presenting the students with three feedback constellations and asking which constellation they preferred and why. (We refer to this as the Session 3 Preference choice: S3P).

Fourth game session. In the beginning of the fourth (and last) game session, each student got to choose how many bars they wanted to have during this last session after which they proceeded to play with their chosen configuration. As the students played in pairs and it was not possible to have different number of feedback bars for their respective agents, the students had to first make an individual preference choice and after that a mutual choice upon the actual number of bars they were to use while playing the game. All student pairs smoothly agreed upon the number of bars to play with and no argumentations between the students were observed. The individual preferences were annotated as their second preference choice (Session 4 Preference choice: S4P) by the principal investigator whereas the actual mutual agreements of the student pairs (Session 4 Actual choice: S4A) were recorded as part of Questionnaire C distributed at the end of the session. While playing, the students could choose between challenging each other's digital tutees (*Agent Plays* for both sides) or to keep on training their tutees (*Student Shows* or *Agent Tries*).

Interviewing and closing session. One week after the fourth session, the main experimenter conducted focus group interviews with each class. Following the interview, she announced the winners from the game sessions in each class, that is, the student whose digital tutee had been best trained. A second and a third-place winner were also selected, and all three winners received a golden chocolate coin and all remaining students received a small piece of candy as thanks for their participation.

3 Results

Statistical analyses were performed with R version 3.2.4 [23]. In case of post-hoc analyses, Bonferroni corrections were used to compensate for multiple comparisons.

The analyzed data set derived from the questionnaire included a total of 35 students, representing the students who took part during all sessions of the study (figure 4).

3.1 Overall Results

The overall results showed: (i) a positive response to having feedback bars compared to no feedback bars and (ii) a divergence as to the preferred number of feedback bars in that 2nd graders preferred the configuration with 6 feedback bars whereas 5th graders favored 1 or 3 feedback bars. (As for the 3rd graders, their results wavered between the results of the 2nd and the 5th graders).

In the following, we will evaluate the preference choices in more detail as well as aspects of the students' learning experience (confidence, engagement, understanding and distraction).

3.2 Preference Choices

At two occasions, the students were asked about their preferred number of feedback bars, namely at the end of the 3rd session (S3P) and during the 4th session (S4P). Additionally, at the 4th session when the students were to play in pairs with an agreed upon number of feedback bars, their "actual" mutual choice was also recorded (S4A). For S3P and S4A, the students also commented upon their choices in the questionnaire.

Overall preference choices. For all three occasions where the students were asked about their preferred number of feedback bars (S3P and S4P) or instructed to make a mutual actual choice (S4A) in pairs, the configuration with 3 feedback bars was chosen by one third of the students. As for the two other configurations, the preference for 1 feedback bar gradually declined while the preference for 6 feedback bars gradually increased (cf. figure 5).

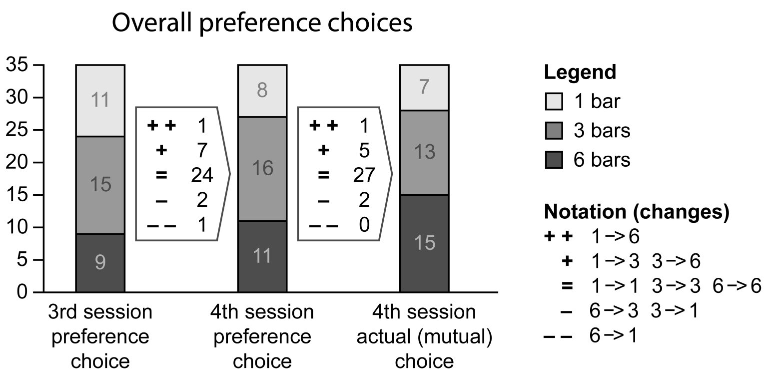


Fig. 5. The overall choices of feedback bars for S3P, S4P and S4A. Individual transitions (changes) between the occasions are notated between the bars. *Notation:* "+/–" denotes a single step change to the nearest higher/lower configuration of feedback bars; "+ +/– –" denotes a double step change from "1 to 6 / 6 to 1" feedback bars; and "=" denotes consistency (no change in the preferred number of feedback bars between occasions).

At a more detailed level (see figure 5), there were 19 transitions (changes) in total between the three occasions, with only three "double step" changes between 1 and 6 bars (or vice versa). Regarding the 16 "single step" changes (transitions between two adjacent levels of detail), 12 were towards a higher level of detail. Furthermore, half of the students (18 out of 35) did not change their preferences or actual choice over all three occasions.

Preferences of number of bars with respect to grade (age). The preferences for 1, 3 or 6 feedback bars varied with respect to which grade the children were in (figure 6). At S3P, there was a tendency (Spearman's rho: $\varrho(35) = .32$, $p = .064$) that higher grade correlated with a preference for fewer feedback bars and vice versa – a lower grade correlated with a preference for more feedback bars. For the two subsequent preference checks during the fourth session, this tendency was more evident (S4P: $\varrho(35) = .69$, $p < .001$; S4A: $\varrho(35) = .65$, $p = .001$).

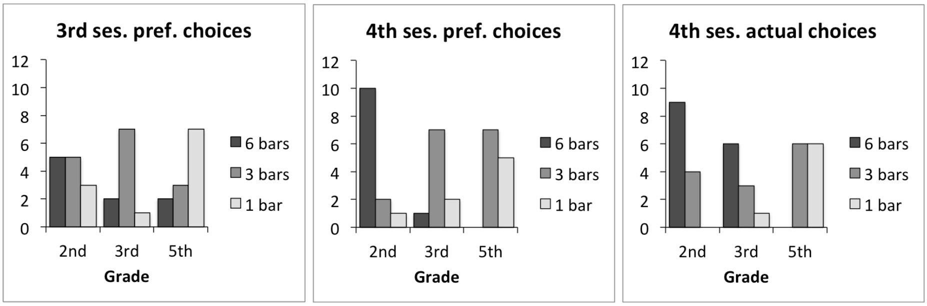


Fig. 6. Distribution (absolute numbers) of students' choices of feedback bars for each grade: (left) 3rd session preference choices, (mid) 4th session preference choices, and (right) 4th session actual choices.

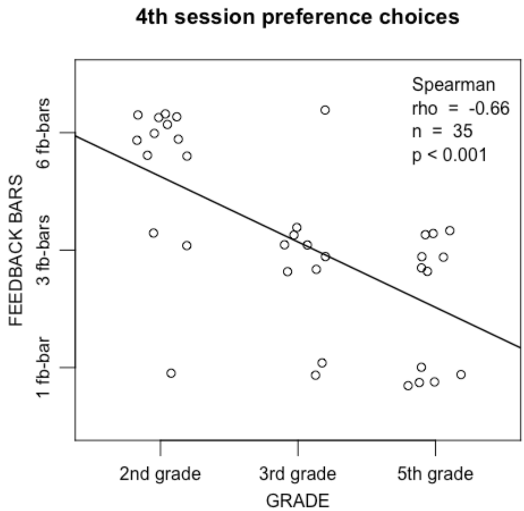


Fig. 7. Scatterplot for the 4th session preference choices (S4P) displaying the distribution for *Grades* x *No. of Feedback Bars*, including a fit line (linear regression) and results for Spearman correlation (with *Grade* and *No. of Feedback Bars* as 3 level ordinal data).

For a complementary analysis, the 24 students that were consistent in their two first preference choices (S3P and S4P) were evaluated. The rationale behind this more restricted selection was the assumption that these participants might represent more robust and valid preference data. The results, in line with the previous findings, now also put forth a strong significant correlation between grade and number of preferred feedback bars also for the 3rd session preference choices (S3P: $\varrho(24) = .65$, $p < .001$).

Taken together, the data suggest an inverse relation between grade and the number of preferred feedback bars, i.e. the 2nd graders preferred 6 feedback bars (but not 1), whereas the 5th graders preferred 1 or 3 feedback bars (but not 6).

3.3 Student Arguments on Preference Choices

For S3P and S4A, the students commented on their choice of feedback configurations in the questionnaires. The comments were by three coders sorted into the following categories: *Understanding*, *Don't know* and *Other* with one additional sub-category, *Social*, for S4A. The coders agreed on 74 out of the in total 79 available comments; after discussion, the coders reached an agreement for remaining comments.

Out of the 35 students who took part in all sessions, a clear majority (24 and 19 respectively for S3P and S4A, see table 1) referred to understanding or knowledge as arguments for their choices. For example, one 3rd grader said: "*Well, because with three bars, one can know more.*" and a 5th grader expressed: "*Then you can much better see what the agent is good at.*"

It should be noted, however, that during the focus group interviews, none of the students (including the 5th graders) could accurately explain what the individual feedback bars represented in the 6 bars configuration with paired bars. A 3rd grader asked: "*What does this light and dark mean?*"; referring to the different color shades of the paired bars and a 2nd grader wondered, referring to the paired bars: "*Why are there two bars?*"

Table 2. Categorized arguments (*Don't know*, *Understanding*, *Other* and NA/*Social*) for preference choices of S3P and S4A, separated on *Grade* (2nd, 3rd and 5th).

3rd session preference choice					4th session actual choice				
Grade	*Don't know*	*Under-standing*	*Other*	*NA*	Grade	*Don't know*	*Under-standing*	*Other*	*Social*
2nd	5	7	1	0	2nd	0	9	2	2
3rd	0	9	0	1	3rd	0	4	4	2
5th	1	8	2	1	5th	3	6	1	2
Σ	6	24	3	2		3	19	7	6

When the students were to make an actual choice of a feedback bar configuration (S4A; figure 6 & table 2), they were first asked to reflect on the three previous sessions with the three feedback configurations and also instructed to agree upon a choice after mutual discussions. Potentially this could result in more well-considered choices.

In their actual choice (S4A; table 2), most 2nd graders (9 out of 13) chose 6 bars and seven of these argued that 6 bars provided the best help in understanding how much knowledge their tutee had and/or gave them a better feeling of understanding the game, for example: "*It was easier to understand*," "*Because they were the best at explaining what the agent knew about the game*," and "*Because you got to know more in detail what it [the tutee] knows.*" However, as noted above, no student in the focus group interviews could accurately describe what the individual bars in the 6 feedback bar configuration represented. Two students (one 2nd and one 3rd grader) tried to explain the pairwise design of the 6 bar configuration by dividing the labels of the feedback bars: "*It is two, so the first might mean game and the other rules*," referring to the label "Game rules" and "*One of them means game and the other rules, points, and the other is rules and this one is just tactics, this is the easiest tactics and this is the hardest tactics.*"

Table 3. Categorized arguments (*Don't know*, *Understanding*, *Other* and *Social*) for preference choices from S4A (4th session gameplay with actual (mutual) choice) separated on *Grade* and *No. of Feedback Bars*.

		Don't know	*Under-standing*	*Other*	*Social*
2nd Grade	6 bars	–	7	2	–
	3 bars	–	2	–	2
	1 bar	–	–	–	–
3rd Grade	6 bars	–	3	1	2
	3 bars	–	–	3	–
	1 bar	–	1	–	–
5th Grade	6 bars	–	–	–	–
	3 bars	1	3	1	1
	1 bar	2	3	–	1

Also, among the 5th graders (table 3), arguments referring to better understanding were the most frequent (6 of 12), but in contrast to the 2nd graders the 5th graders preferred 1 or 3 bars. In view of the focus group interviews, the 5th graders seemed to have some understanding of the configurations with 1 or 3 feedback bars and reasoned about them, for example: "*When it was just one bar, it was much easier to understand because then everything was summed up and then you only needed to keep track of one bar [...] otherwise it can get so messy*" and "*You feel much more confident with what your agent knows when you have three bars*."

3.4 Focus Group Interviews

Next, we turn to the focus group interviews to further probe into the 2nd graders contradictory behavior regarding their preferences for 6 feedback bars without understanding what they stood for. In the interviews, 2nd graders reported an interest in their tutee's knowledge, but were at the same time unconcerned with, or oblivious to the fact that they did not understand the meaning of the 6 bars. When asked why they liked 6 bars better, they argued it was because they understood more and could better distinguish what types of knowledge their tutee had. As an example, one student answered the question why she did not like to have only one bar as follows: "*You didn't know if, if it is really bad at tactics you might think it is really good at this just because it has much in knowledge*," referring to the 1 bar configuration. As mentioned before, 2nd graders expressed an uncertainty regarding why there were pairs of bars in the 6-bar configuration. One student said, for example: "*On some [...] it is two bars on every [referring to the paired bars] and then it is tactics on one but which of them is tactics then, I don't quite get it, is the dark blue also tactics?*"

The 3rd and 5th graders also reported an interest in gaining more specific knowledge on how much their tutee knew – but they argued that 6 feedback bars were too many whereas 3 bars were just enough: "*I also think that 3 bars were better, that is 6 bars was to many and 3 was just the right amount*." When asked how many bars they preferred, one 3rd grader answered: "*Three. [Interviewer: "Why?"] Because if there was only one you couldn't understand anything, because everything was put together and then it was hard to understand, but when you had three, you understood the knowledge of scoring rules and knowledge and such things*." Another 3rd grader explained: "*I thought three was good because you could see things like tactics, even though everything is put together in this one [referring to 1 bar]. I liked three bars* *better because then you could see different things it was good at.*" One 5th grader expressed his preference for three bars with: "*You feel more certain about what you know about your tutee*" and another 5th grader presented a similar argument: "*I also think it was better with 3 bars because then you could see different things it was good at.*"

3.5 Questionnaire results

Overall, on Questions 1 to 7 in the questionnaire, the relative answering patterns were similar for students in all three grades, but with the 5th graders systematically scoring lower on Questions 1 to 5 (figure 8), possibly reflecting age related factors. These differences between the 5th graders and the two lower grades were significant for Questions 1 to 5 (post-hoc pairwise Wilcoxon: $p_{(5:2/3)} < .0.5$).

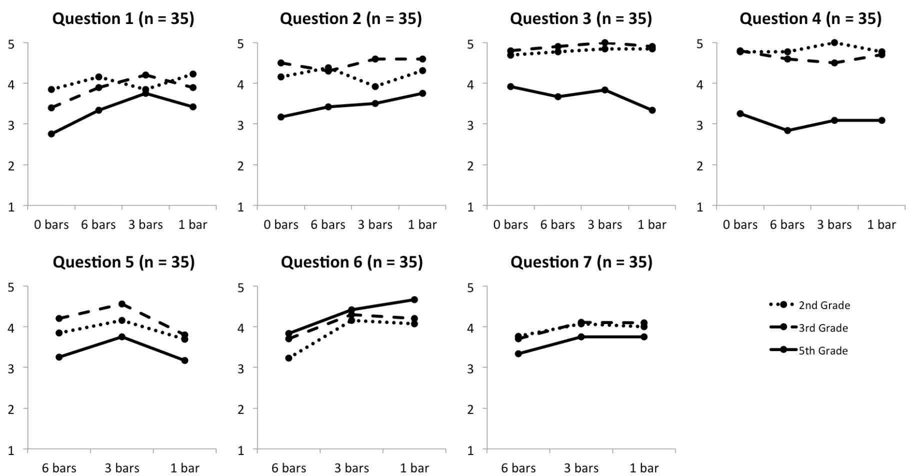


Fig. 8. Means for the three grades (age groups) over the four feedback configurations (no. of feedback bars) for Questions 1 to 7 in Questionnaire A and B.

Confidence. Students' confidence in their digital tutee's knowledge as well as their confidence in their own teaching was probed through Question 1: "*How confident are you as to what your tutee knows about the game?*" and Question 2: "*How confident are you in teaching your tutee?*" The means for the three grades over the four feedback conditions are plotted in figure 8.

The results showed a significant difference (post-hoc pairwise Wilcoxon: $p_{(0:6/3/1)} < .01$) regarding confidence on agent knowledge (Question 1) between the introductory session with no feedback bars, and each and one of the three game sessions of 6, 3 and 1 bars (figure 8). For Question 2, there were no differences between the introductory session with no feedback bar and the three subsequent game sessions with feedback bars.

In the focus group interviews, the students were once again asked how confident they were in what their tutee knew (Question 1). The students' answers supported the finding that the presence of feedback bars made the students feel more confident in what their tutee knew. In particular, they expressed insecurity about their agent's knowledge when the feedback bars were not present. A 2nd grader commented: "*It* *was hard without the bars*" and when asked what made them feel like they were in control, one 2nd grader said: "*I looked at the bars*."

As to Question 2 about how confident they were with respect to their own teaching, the focus group interview answers were somewhat mixed. The 2nd and 5th graders expressed that they felt more, or much more, confident in their teacher role when the feedback bars were present. A 2nd grader answered: "*Yes, of course you could know some when they chose cards, but that was not so much*" to the question if they felt more confident with what they were teaching with the feedback bars present. When asked if they felt more confident in their teacher role in the presence of feedback bars, all 5th graders answered "*Yes*"; none of them elaborated on their answers. Among the 3rd graders, however, two students said that the bars didn't make any difference: "*No, I didn't feel confident at all, I thought I had answered everything incorrect*" and "*I felt about 2% more confident*."

Thus, it appears that the presence of feedback bars made a difference with respect to the students' confidence in their agent's knowledge, but it is questionable whether they affected the students' confidence in their teacher role.

Engagement. To get an indication of how the feedback bars affected the students' overall engagement in the game, we compared their answers on Question 3: "*How eager are you to teach you tutee how to play the game?*" and Question 4: "*What do you think about playing the game?*" However, none of the two questions displayed any differences with regard to the feedback conditions (0, 1, 3, or 6 bars) they had been using. Notable was the 5th graders significantly lower scores on both questions, see figure 8.

No questions in the focus group interview explicitly targeted whether students felt more engaged when there were feedback bars compared to when there were no feedback bars, but they did report that they felt more encouraged to keep on playing when they could look at the bars and see what progress their digital tutee made. For example, one 3rd grader testified: "*It is important that things are happening all the time – otherwise one does not want to play*."

Students' understanding. Students' understanding of their tutee's knowledge as well as their understanding of the feedback bars, was evaluated through Question 5: "*Were you helped by the [n] bar(s) in understanding how much your tutee knows about the game?*" and Question 6: "*Did you find the [n] bar(s) easy or difficult to understand?*" The means for each of the three grades over the four feedback conditions are plotted in figure 8.

Answers on Question 5 indicated (tendency) that the configuration with 3 feedback bars is better in supporting students' understanding of their tutees knowledge compared to both 6 and 1 bars, however the results are not significant (Kruskal-Wallis: $X^2(2) = 4.11$, $p = .13$; post-hoc pairwise Wilcoxon: $p_{(3:1)} = .11$; $p_{(3:6)} = .72$).

Related answers provided in the interviews, however, suggested that the feedback bars did provide a feeling of understanding about how much their tutee knew about the game. For example, one 3rd grader said that he looked at the bars in order to know how much knowledge his tutee had, and a 2nd grader expressed: "*At first the tutee was really bad and then it became better and better and then you could also look at the bars*." Another 2nd grader expressed: "*I thought my agent was the worst out of all the agents*," when explaining how he felt about how much knowledge his tutee had without the feedback bars.

Furthermore, when asked during the interviews what they liked most with the bars, four out of five students said that the best thing with the bars was that one could see how much knowledge one's tutee had gained, for example: "*The good thing was that* *you could see how high he [the tutee] got.*" and "*I think it was the fact that you could see how much the agent knew.*"

Questionnaire answers on Question 6, on the other hand, put forth the configuration with 6 feedback bars as significantly harder to understand compared to the configurations with 3 and 1 bars (Kruskal-Wallis: X^2 (2) = 10.3, p = .006; post-hoc pairwise Wilcoxon: $p_{(6:3)}$ = .029; $p_{(6:1)}$ = .014). Taken together, 6 feedback bars seemed detrimental for the understanding of the feedback, whereas the configuration with 3 bars possibly could support students' understanding better than 1 bar only.

This finding was partly supported in the focus group interviews, where the 3rd and 5th graders expressed that 6 feedback bars were hard to understand, for example: "*Six were a bit too many*" and "*When there were 6 bars it was kind of messy, they were so many.*" The 2nd grades, to the contrary, argued that "*You understand more then,*" when asked why 6 bars were best, even though they were unable to explain the meaning of 6 bars.

Helpful or distractive. Question 7: "*Was it disturbing or helpful to have the (n) bar(s) in the game?*" targeted potential experiences of distraction or disturbance related to the feedback bars. Deeming from the results (figure 8), the feedback bars were perceived as helpful rather than distractive, and experiences of helpful versus disturbing were quite equally distributed over the three feedback configurations.

The focus group interview contained no explicit questions that used the word "distraction", but when asked about the most negative thing with the bar(s), no student mentioned or suggested that the bars were distracting or disturbing. Instead, these answers focused on not understanding what all bar(s) meant or the time it took to get the bar(s) to increase: "*The worst thing with the bars was that they took so long to fill up.*"

4 Discussion

The focus of this study was not whether the continuous recursive feedback would increase students' performance; something already proven in several studies [4, 10, 5] as well as for the math game used in this study [19]. Instead, we explored effects of recursive feedback – with regard to level of detail – on preference and experiences of confidence, engagement and understanding in a learning and teaching context.

More specifically, how should recursive feedback on the tutee's learning progress be designed with respect to the level of detail, if the goal is that students should experience the feedback as both understandable and engaging – and also through this feedback gain confidence in the task at hand?

4.1 Preference and understanding

The results suggested that the students responded positively to receiving immediate, continuous feedback on their digital tutee's learning process, but that the preferences with respect to the level of detail of the feedback differed according to the students' age. Not surprisingly, all students judged the configuration with 6 feedback bars to be the hardest to understand. More surprising, though, was the 2nd graders preference for 6 bars when given a choice. At the preference choice during the 4th game session when the students had played with all three configurations and then paused for one week, 77% of the 2nd graders chose 6 bars whereas only 10% of the 3rd graders and none of the 5th graders did so. In conjunction to their choice of 6 bars, 80% of these

2nd graders argued – sometimes at length – in terms of increased understanding to legitimate their choice. In the later focus group interviews, however, all 2nd graders had a hard time when asked to explain the meaning of the feedback bars.

On the whole, the majority of students argued that their choice of feedback bars added to their understanding about their digital tutee's knowledge. In alignment with this, the older students appear to be more mature in their selection of bars, selecting the configuration that matches their level of understanding (i.e. 1 or 3 bars). In contrast – and quite intriguingly – the younger 2nd graders seemed to choose a configuration that would make the game experience more enjoyable and fun – the more happening in the game, the more fun – rather than choosing a learning experience. Precisely as the older students, the 2nd graders in both the questionnaire comments and in the focus group interviews argued that their choice of 6 bars helped them know how much knowledge their digital tutee has. None of them, however, could explain what the six different bars actually meant. Furthermore, in contrast to the older students, the 2nd graders did not appear to be bothered by this. Rather, they seemed content with the fact that the bars had different labels, while the actual meaning of these labels seemed of less concern. For example, one 2nd grader said: "*It says under the bars what they mean*," but could not provide an answer when asked if she could explain further what the names of the bars meant.

To elaborate on the notion of "fun", this seemed to be a determinant factor for their choice. Also, when these young 2nd grade students in the focus group interview were asked for new ideas on how to design the feedback, all answers included elements of "fun" (together with "motion"). These concepts were presented by the older students too, but less frequently – and did not seem to influence older students' choice of feedback bars as strongly as the 2nd graders.

To conclude the findings on understanding and preference, it doesn't seem too far-fetched to paraphrase the well-known novel by Jane Austen; for the younger 2nd graders, the choice of feedback bars seems to be a question of "sense versus sensibility".

4.2 Experience of confidence and engagement

Another goal of the study was to evaluate whether the presence of (recursive) feedback bars would add to the students' experience of confidence and/or engagement while playing the game. According to [23], it is a challenge to engage students when they are to use technology in a learning situation. However – whereas engagement is central when designing games for entertainment – it is not the focus of an educational game where the primary goal is "learning".

Regarding confidence, presence of the feedback bars seemed to support the students' confidence in their digital tutee's knowledge. This feeling of confidence is hardly surprising, since the tutee's knowledge is directly visualized in the bars.

Presumably the bars would also make them feel more confident as teachers, since they continuously can monitor their tutee's development through the bars. However, for the students' confidence in their teacher role, the presence – or non-presence – of feedback bars did not seem to matter. One possible explanation is that the teaching experience the students could gain in the study was too limited.

In conclusion, just adding visual continuous (recursive) feedback is not enough to strengthen the student in their teacher role. This has to be catered for in other ways and needs to be further explored within the teachable agent paradigm.

As for engagement, neither the number nor the presence of feedback bars seemed to make a difference as to students' willingness to teach their digital tutee. Probably, the questionnaire didn't target the issue of feedback, but rather reflected a more general willingness to interact with digital learning games in the classroom. This could explain why the younger children (2nd and 3rd graders) "hit the roof" in their evaluation scores, while the older 5th graders were not as excited, having more prior experience with digital learning materials.

Also, all respondents from the focus group interviews agreed that as long as the bars were changing, they felt encouraged to keep on playing. As one student put it: "*It is important that things are happening all the time – otherwise you do not want to play*", again simply indicating a general positive effect from interaction with digital media. Another student declared that the feedback bars contributed to engagement in the sense that the students enjoyed seeing their digital tutee develop and learn: "*It was fun to see when it [the digital tutee] got better.*"

Lastly, we did not find any indications that the feedback bars were at any point disturbing to the students – rather the opposite. However, during the focus group interviews when asking the students about the worst parts with the feedback bars, we sometimes identified expressions of frustration. A couple of students mentioned that the bars didn't fill up "*fast enough*" and that this was frustrating. We do not interpret this frustration as a frustration regarding the feedback bars per se, but rather as a more general frustration regarding the fact that their digital tutee (and thus themselves) didn't learn quick enough.

4.3 Design implications

When designing an educational game with recursive feedback of this kind, the age span of the target group is important. Young children's attitudes as well as experience of understanding change with age. A priori, it seems reasonable that a game developed for 8-year-olds will work just as well for nine-year-olds. However, as our result indicates, even one year's difference may change the balance between experience of understanding with respect to level of detail and attitudes (fun factor).

With respect to the central content of the actual math game, the base-ten concept, it is not a simple task to provide accurate and meaningful feedback on the underlying mathematical rules. (Actually, a deeper conceptual understanding of these fundamental mathematical theorems is at University level.) Regarding the specific recursive feedback visualized in the math game used in the study, it is evident that this ought to be redesigned. It should balance the underlying mathematical concepts with a simplified but constructive and meaningful concretization adapted to the target group. From a more general design perspective, the results of our study suggest that the older students may benefit from meaningful and constructive feedback relating to the underlying learning objectives, in this case the mathematical base-of-ten concept, – while it is questionable for the younger children. Instead, it may be more beneficial to implement implicit scaffolding strategies for these younger children [24] – something that, on the other hand, would demand extensive design resources as one would need to implement individual adaptivity based on domain specific pedagogical strategies.

Lastly, it is all too easy to give in to a "fun factor" on the expenses of the "learning objectives" when developing educational digital games.

4.4 Limitations of the study

A limitation of the study was that the three feedback configurations (6, 3 and 1 bar) were presented in the same order to all participants. A randomized order for the encounter of the different configurations and/or only one of the three configurations for each class/grade would be needed for more general claims and to ascertain that there was no bias towards any configuration of bar(s). With the present experimental design, we cannot clearly answer whether the students would prefer a certain configuration "as it is". However, we are still able to evaluate possible grade dependent differences as all students follow the same experimental procedure. Furthermore, an alternative randomized design would require both more participants (school classes) as well as additional experimental resources – something we were not able to bring in for this study.

Also, the order with decreasing level of detail (i.e. decreasing number of feedback bars) should be mentioned. As discussed earlier, our rational to let the students encounter the feedback configurations in decreasing level of detail (6, 3 and 1 bar) instead of increasing (1, 3 and 6 bars) was to avoid influencing their interpretation of the more complex configurations. We believe, however, that additional studies can here provide a richer picture.

The generation of questions for the focus group interviews was based on the results from the preceding questionnaires. As the questions targeting engagement did not reveal any differences for the three feedback configurations, there were no questions directly targeting "engagement" (such as whether the students found themselves more or less *engaged* when the feedback bars were present or not); afterwards, we considered this a mistake. However, as the focus group interviews were semi-structured, several answers still indicated that the presence of feedback bars had a positive effect on engagement.

5 Conclusions

In sum, the results of this study: (i) displayed an age-related difference in the students' preferences as to the number of feedback bars as well as their experience of these different configurations; (ii) indicated that the feedback bars contributed to more positive game experience in that the students experienced an improved understanding of their digital tutee's knowledge and (iii) suggested that the feedback bars added extra value in terms of confidence and (possibly) engagement.

Especially, the results pointed at: (i) the youngest students (2^{nd} graders) being unbothered by their own trading of "understanding" for "enjoyment" and (ii) the students' eagerness to comment of dynamic aspects of game experience (several students mentioned the importance of "seeing things move" in order to keep up their interest).

An important lesson learned from this study is that the students' potential to understand or make sense of recursive feedback targeting teaching of fundamental math concepts, such as carry-overs, is a delicate issue that has to take into consideration the age and general performance level of a target group. Thus, in a case like the present when feedback should be used by students of a broad age span, one needs to put emphasis on what is potentially possible for different age groups to understand – or not understand. It is apparent that even with as little as a one-year difference in age, the understanding for different feedback configurations may diverge.

We also demonstrated in this study that design evaluations with children in these ages can be especially tricky. In their questionnaire responses, the 2^{nd} graders favored the highest level of detail in terms of understanding, while the questions during the focus group interview directly asking for explanations revealed an actual (and profound) unawareness of the meaning of their favored feedback configuration. In fact, their preference choice could possibly even be detrimental to their learning process. As for the 5^{th} graders – and to some extent the 3^{rd} grades – it seemed as if they had a better understanding of their own learning and were able to make a more sensible choice.

Thus, we want to emphasize a warning for easy, superficial feedback solutions for the younger children, as it seems quite easy to boost their positive attitude towards colorful and animated feedback on the expense of providing actual meaningful information.

References

1. Hattie J., Timperley H.: The power of feedback, Review of Educational Research, 77(1), pp. 81--112 (2007)
2. Shute V. J.: Focus on formative feedback, Review of Educational Research, 78(1), pp. 1531--89 (2008)
3. Black P., Wiliam D.: Assessment and classroom learning, Assessment in Education, 5(1), pp. 7--74 (1998)
4. Biswas G., Leelawong K., Schwartz D., Vye, N.: Learning by teaching: A new agent paradigm for educational software, Applied Artificial Intelligence, 19, pp. 363--392 (2005)
5. Okita S. Y., Schwartz D. L.: Learning by teaching human pupils and teachable agents: the importance of recursive feedback, Journal of the Learning Sciences, 22(3), pp. 375--412 (2013)
6. Annis L. F.: The processes and effects of peer tutoring, Journal of Educational Psychology, 2(1), pp. 39--47 (1983)
7. Bargh J. A., Schul Y.: On the cognitive benefits of teaching, Journal of Educational Psychology, 72(5), pp.593--604 (1980)
8. Papert S.: The children's machine: Rethinking school in the age of the computer. New York: Basic Books, (1993)
9. Renkl A.: Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results, Learning and Instruction, 5(1), pp.21--36 (1995)
10. Leelawong K., Davis J., Vye N., Biswas G., Schwartz D., Belynne K., Bransford J.: The effects of feedback in supporting learning by teaching in a teachable agent environment, in The Fifth International Conference of the Learning Sciences, pp 245--252 (2002)
11. Pareto L.: A Teachable Agent Game Engaging Primary School Children to Learn Arithmetic Concepts and Reasoning, International Journal of Artificial Intelligence in Education, 24(3), pp. 251--283 (2014)
12. Dye M., Bavelier, D.: Differential development of visual attention skills in school-age children. Vision Research, 50(4), pp. 452--459 (2010)
13. Miller E. B., Warschauer M.: Young children and e-reading: research to date and questions for the future, Learning, Media and Technology, 39(3), pp. 283--305 (2014)
14. Salmon L. G.: Factors that affect emergent literacy development when engaging with electronic books, Early Childhood Education Journal, 42(2), pp. 85--92 (2014)
15. Zucker T. A., Moody A. K., McKenna M. C.: The effects of electronic books on prekindergarten-to-Grade 5 students´ literacy and language outcomes: a research synthesis, Journal of Educational Computing Research,40(1), pp. 47--87 (2009)
16. Lindström P., Gulz A., Haake M., Sjödén B.: Matching and mismatching between the pedagogical design principles of a math game and the actual practices of play, Journal of Computer Assisted Learning, 27(1), pp. 90--102 (2011)
17. Nielsen J. 10 usability heuristics for user interface design. Retrieved from http://www.nngroup.com/articles/ten-usability-heuristics/. (1995)
18. Sjödén B., Tärning B., Pareto L., Gulz A.: Transferring teaching to testing – an unexplored aspect of teachable agents, in AIED 2011, LNCS, vol. 6738, pp. 337--344, Springer, Heidelberg, (2011)
19. Pareto L., Haake M., Lindström P., Sjödén B., Gulz A.: A teachable-agent-based game affording collaboration and competition: evaluating math comprehension and motivation, Educational Technology Research and Development, 60(5), pp. 723--751 (2012)
20. Bandura A.: Guide for constructing self-efficacy scales, in Self-efficacy Beliefs of Adolescents, pp. 307--337,Information Age Publishing, Greenwich, (2006)
21. Brännström J., Holm L., Lyberg-Åhlander V., Haake M., Kastberg T., Sahlén B.: Children's subjective ratings and opinions of typical and dysphonic voice after performing

a language comprehension task in background noise, Journal of Voice, 29(5), pp. 624--630 (2015)
22. R Core Team: R: A language and environment for statistical computing [Software], R Foundation for Statistical Computing (2016)
23. Kiili K.: Digital game-based learning: Towards an experiential gaming model, The Internet and Higher Education, 8(1), pp. 13--24 (2005)
24. Husain L., Haake M., Gulz A.: Supporting early math – rationales and requirements for high quality software, Journal of Computers in Mathematics and Science Teaching, 34(4), pp. 409--429 (2015)

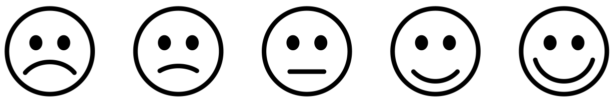
Appendix

Questionnaire A and B

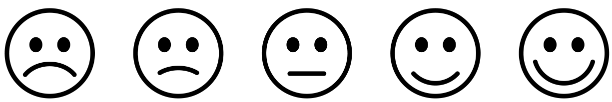
Question 1–4 constituted the pre-questionnaire (Questionnaire A) and Question 1–7 constituted the questionnaires after session 1-3 (Questionnaire B).

1. How confident are you as to what your tutee knows about the game?

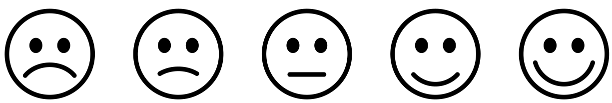
2. How confident are you in teaching your tutee?



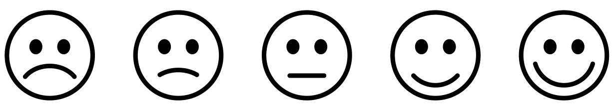
3. How eager are you to teach your tutee how to play the game?



4. What do you think about playing the game?



5. Were you helped by the (N) bar(s) in understanding how much your tutee knows about the game?



6. Did you find the (N) bar(s) easy or difficult to understand?

very hard					very easy

7. Was it disturbing or helpful to have the (N) bar(s) in the game?

distracted a lot					helped a lot

Focus group questions

1. What do think about playing the game?
   – What has been fun/not so fun?
2. How did you feel about the fact that you were teachers to your agent?
   – What was fun/not so fun about it?
3. Did it feel like you knew how much you agent knew about the game?
   – How could you know/not know?
4. Was it important to get information about how much knowledge you agent had?
5. Would you like to have had more information about how good your agent was
   – What would you like this information to look like?
6. These bars (pointing to bars), did you look at them often?
   – Why did you look at them/not look at them?
7. Did you understand the difference between 6, 3, 1 bar(s)?
   a) What was the difference?
   b) What did the bar(s) mean?
8. Does the number of bars matter?
   – Why/why not?
9. Did you bother to understand what the bars meant?
   – Why/why not?
10. Did you feel more confident in you teacher role when you knew how much knowledge your agent had?
    – Why/why not?
11. What was good about the bar(s)?
12. What was not so good about the bar(s)?
13. How can we make the bar(s) better?
14. What made you decide on how many bars you choose to play with the last time?
    – Why was this best?
15. If you yourselves were to choose a way to get this information presented on, what would it look like?
16. Do you often play other games at home?
    a) What types of games/how often?
    b) What types of information do you usually get in these games? / What does it look like/what do you think about it?
17. Did you understand how the agent got its knowledge?