How teachable agents influence students responses to critical constructive feedback

Annika Silvervarg, Rachel Wolf, Kristen Pilner Blair, Magnus Haake and Agneta Gulz

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-168780

N.B.: When citing this work, cite the original publication.
This is an electronic version of an article published in:
https://doi.org/10.1080/15391523.2020.1784812

Original publication available at:
https://doi.org/10.1080/15391523.2020.1784812
Copyright: Taylor and Francis
http://www.tandf.co.uk/journals/default.asp
How Teachable Agents Influence Students’ Responses to Critical Constructive Feedback

Does a teachable agent influence the uptake or neglect of ‘critical constructive feedback’ and learning within a digital environment? 285 middle-school students engaged with a history learning game in a 2x2 study design. One dimension was inclusion of a teachable agent. Orthogonal was whether critical constructive feedback was presented automatically or only when students chose. Analyses showed that a teachable agent positively affected students’ responses to feedback and mitigated feedback neglect; the results were especially strong for lower-achieving students. Additionally, presence of a teachable agent improved post-test performance for students overall, and this effect was mediated by lower feedback neglect.

Keywords. critical constructive feedback; teachable agent; choice; game; low- and high-achieving students; learning

1. Introduction

Teachable Agents (TAs) are pedagogical agents based on the instructional approach of ‘learning by teaching’ (Bargh & Schul, 1980; Chase, Chin, Oppezzo, & Schwartz, 2009; Blair, Schwartz, Biswas, & Leelawong, 2007) and built on the premise that a powerful way to learn something is to try to teach it to someone else. With TAs, students take on the role of teachers and teach a pedagogical agent – or digital tutee – who then solves challenges and receives feedback.

Research has found benefits of TAs both in terms of learning outcomes and motivation. Students make more effort to learn in order to teach their TA than to learn for themselves, such as engaging in more reading and revising (Chase et al., 2009). They take responsibility for their TA and show engagement in instructing and interacting with them (Chase et al., 2009; Lindström, Gulz, Haake, & Sjödén, 2011). Performance is often higher in instructional conditions that involve TAs (Chin, Dohmen, & Schwartz, 2013; Pareto, Haake, Lindström, Sjödén, & Gulz, 2012). In particular, several studies show that lower-achieving students tend to benefit by interacting with TAs, which can help reduce the achievement gap (Chase et al., 2009; Sjödén & Gulz, 2015; Pareto et al., 2012; Tärning, Haake, & Gulz, 2017).
A number of explanations for the pedagogical power of TA-based learning environments have been proposed (see, for example, Blair et al., 2007; Kinnebrew & Biswas, 2011). There are likely several contributing mechanisms that lead to increased learning. In this paper, we focus specifically on the mechanisms of feedback. How do TAs influence the uptake or neglect of feedback within a digital environment, and how does this influence learning?

It is well established that feedback, defined as information regarding performance outcomes and learning processes, is important for learning (Mory, 2003; Hattie & Timperley, 2007; Shute, 2008; Van der Kleij, Feskens, & Eggen, 2015). In particular, this applies to feedback that is critical, in that it helps students identify their mistakes, and constructive, in that it helps them improve and make progress (Kluger & DeNisi, 1998). For this kind of feedback, we use the term critical constructive feedback (CCF 1).

However, many studies show high propensity for students to not make use of CCF – but avoid or in some way neglect it (Segedy, Kinnebrew, and Biswas, 2013; Wotjas, 1998; Clarebout & Elen, 2008; Duncan, 2007; Winter & Dye, 2004; Conati, Jaques, & Muir, 2013; Hounsell, 1987; Tärning, Lee, Andersson, Månsson, Gulz, & Haake, 2020).

One reason for neglect may be that the presence of critical feedback indicates that one has failed at solving a task, or at least could have performed it better. If a student feels uneasy when confronted with failure – or even anything that hints at failure – she may consciously or unconsciously avoid critical feedback (Chase et al., 2009). Indeed, it has been shown that students may avoid critical feedback because they interpret it as evaluative punishment (Hattie & Timperley, 2007).

Could TAs counteract this? Research suggest that they can (Chase et al., 2009). The increases seen in motivation and effort when learning on behalf of a TA may lead students to engage in behaviors they are otherwise prone to avoid, such as reading and acting on CCF. More specifically, the ego-threat of CCF may be mitigated by a mechanism that Chase et al. (2009) term an ego-protective buffer. In a TA-based learning environment it is the TA that is tested for its knowledge. When the TA fails at

---

1 It is important to note that by critical feedback, we do not mean that the feedback places criticisms on the learner. We simply mean that the feedback indicates a discrepancy between the current state and a goal state.
a test, the failure does not as directly bear on the student as when she takes a test herself. Even if students are aware that the TA’s knowledge reflects their teaching, the responsibility for failing is not only theirs. Instead of bearing the full burden, the responsibility of failure can be shared between the TA and student. Although this may benefit higher-achieving students as well, Chase et al. (2009) hypothesize that it may be particularly strong for lower-achieving students, who are more used to failing at school and may have more ego-threat. Thus, the ego-protective buffer mechanism may provide one explanation for why in particular lower-achieving students perform better when working with a TA.

In this study, we examine whether students’ inclinations to neglect CCF may be counteracted by including a TA in a digital learning environment on a social science subject. We also examine potential differences in the response patterns and performance between students grouped according to their achievement level – low, mid or high – based on classroom teachers’ assessment of their ability to process text-based information, an ability that according to discussions with the teachers, strongly impacts overall achievement in social science subjects for this age group.

Our research questions were: (i) Will there be differences in response patterns and/or performance between a condition with TA and a condition without TA? (ii) Will response patterns and/or performance differ depending on students’ achievement levels? We explored the questions under two instructional scenarios. In the first, CCF is automatically provided to the learner without his or her control, which is typical of instructional games and classrooms. In the second scenario, students have control over whether they receive CCF (they can accept it or decline it). While this is not typical of classroom situations, there is reason to believe increased agency might improve learning (Reeve, Nix, & Hamm, 2003; Deci & Ryan, 1985). This may, thus, be an important factor to explore with respect to design of educational software.

In a 2x2 design, 285 middle school students were randomly assigned to one of four conditions: (1) automatically receive CCF with TA, (2) automatically receive CCF without TA, (3) accept-or-decline-CCF with TA, and (4) accept-or-decline-CCF without TA. Click-stream log data provide process measures of feedback uptake, while an out-of-game post-test provides a measure of learning. We are particularly interested in potential differences between the two groups of low-achieving and high-achieving students, since differences are most likely to appear between those (rather than between one of these and the group of mid-achieving students) and because previous studies
have found differences between low- and high-achieving students with respect to TA-based learning environments.

The article begins by reviewing the literature relevant to this study. Then, it describes the digital learning environment and the four customizations that correspond to each of the four conditions. Next the study procedure and measures are presented, followed by the results. Finally, it concludes with a discussion of educational implications, limitations, and future research directions.

2. Literature Review

This section briefly reviews the relevant literature on teachable agents, methods for studying feedback and learning, and choice to receive feedback.

2.1. Teachable Agents

Software environments that include TAs have been shown to have a number of positive effects on students’ learning when compared to equivalent software without TA, Examples include positive effects on reasoning abilities and conceptual understanding (Pareto et al., 2012; Chin, Dohmen, & Schwartz, 2013) and on metacognitive processing (Chin et al., 2010; Lindström et al., 2011; Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010).

The effect of increased student effort and time spent by students (Chase et al., 2009), sometimes referred to as the protégée effect, is likely an underlying factor for some of the positive effects on learning outcomes. In studies reported by Chase et al. (2009), 10-11 as well as 13- to 14-year-olds spent significantly more time on learning activities when their task was to teach a TA compared to when learning for themselves.

An observation from several studies is, that the metaphor of the computer character as a tutee is readily accepted by students. They engage in teaching it although it is in fact nothing but a computer artifact. In effect, students seem to attribute mental states and responsibility to the character (Chase et al., 2009; Lindström et al., 2011). They approach the TA as an entity that can learn (respond to being taught by them) and ascribe it traits such as ‘brave’, ‘slow’, ‘smart’, ‘forgetful’, etc.

In turn, approaching a TA this way allows the student to share the burden of failures with it. For the present study addressing inclinations to (not) avoid critical information regarding failures, this is highly relevant. A related finding in the studies by
Chase et al. (2009), is that students who learned in order to teach their TA were significantly more inclined to talk about the errors and mistakes they had done on tasks. Talking about a mistake requires acknowledging it. Acknowledging a mistake, in turn, is a precondition for being open to critical, constructive feedback regarding the mistake.

2.2. Methods for Studying Feedback and Learning

Most studies of feedback and learning tend to focus on performance outcome measures (e.g. Hounsell, 1987; Wotjas, 1998; Clarebout & Elen, 2008). That is, the effect of feedback is measured by an examination of students’ subsequent performance: progress or not on the task for which feedback was provided. This leaves everything that happens from the moment of presentation of feedback till the final measurement of performance inside a black box.

Those studies that do try to examine students’ learning process behaviors as they engage with feedback are often based on self-report measures representing the learners’ views on their use of feedback, and are collected via surveys and interviews (Mahfoodh, 2017; Mulliner & Tucker, 2017; Narciss, 2013; Sargeant et al., 2011).

However recently, capitalizing on techniques such as eye-tracking, and interactive, data-logging possibilities of learning technologies, it has become possible to capture process data (Conati, Jaques, & Muir, 2013; D’Mello, 2019; Cutumisu, Blair, Chin, & Schwartz, 2017). For example, Tärning et al. (2020) found that feedback neglect happened at different rates at various stages of feedback processing (noticing the feedback, reading it, acting upon it, progressing from it). Building on this research, the present study uses process data collected via an educational digital game. It attempts to catch a series of stages relating to learners’ handling of critical feedback, rather than only examining an ‘output’ of student performance in relation to an ‘input’ of critical feedback.

2.3. Choice to Receive Feedback

Until recently most research on feedback has focused on situations where the feedback is provided for a learner – by teacher or software – whether she asks for it or not (Evans, 2013; Geitz, Joosten-Ten Brinke, & Kirschner, 2016). That is, the student has no control over the feedback as it arrives without choice. Some exceptions are the following: In a study by Aleven, Roll, McLaren, & Koedinger (2016), help-seeking students could ask
for different levels of feedback and hints in an intelligent tutoring environment that included meta-cognitive mentoring. D’Mello, Olney, Williams, and Hays (2013) used eye-tracking to investigate if and how long the learner looked at feedback when it was chosen or received. Cutumisu and collaborators (Cutumisu, Blair, Chin, & Schwartz, 2015; Cutumisu et al., 2017; Cutumisu, Chin, & Schwartz, 2019) conducted a series of studies where the participating students had a choice of receiving critical constructive feedback or positive feedback. The studies made use of a digital game about graphical design principles, and found that students’ game performance as well as their performance on a post-test on graphical design correlated significantly with both their tendency to choose CCF and their tendency to revise their tasks (Cutumisu et al., 2015; Cutumisu et al., 2017). Also, both behavioral tendencies correlated with broader academic performance – grades and national test scores in reading, math and science. These are findings that hook into educators’ intuitions that asking for and using CCF are productive learning behaviors associated with high achievements.

3. The TA-based Game ‘Guardian of History’

Guardian of History is a digital learning environment that employs aspects of artificial intelligence to track students’ learning processes and performance, as well as to provide them with customized feedback as they play the game. The learning domain is history, more specifically scientific discoveries and inventions during the 15th-18th centuries, and the instructional goals range from matters of historical fact (e.g., in 17th Century Europe, only boys were allowed to go to school) to conceptual knowledge requiring active comparisons and drawing of conclusions (e.g., the invention of the printing press led to increased literacy by enabling the cheap mass production of books, including the Bible, which made books available also to people who were not well off).

We first describe the standard Guardian of History game that includes a TA. The background narrative is as follows: Professor Chronos, the Guardian of History who watches over the passage of time, has a team of apprentices. When the student enters the learning environment for the first time, she meets the time elf Timy (teachable agent) who tells the student that s/he would very much like to be on this team. To join the team, one must pass a series of history exams given by Professor Chronos himself. Unfortunately, Timy suffers from temporary time-travel sickness and therefore cannot use the time machine to travel through time to learn about the past and about history.
But – Timy suggests – *the student* could do the time traveling for her/him, then return to teach Timy so that s/he can pass the exams. The goal of the game is to have Timy pass the exams and the student’s role is to make this happen by helping (teaching) Timy.

In this study, Professor Chronos provides six missions, each corresponding to an exam. They are provided in increasing order of difficulty. To collect information the student travels with a time machine to different historical scenarios and explores the surroundings via interactive objects and engages in text conversations with historical figures (Figures 1a and 1b).

![Figure 1 a). The student visits Gutenberg’s workshop in Mainz.](image)
Figure 1 b). Gutenberg tells about his Bible project.

Thereafter the student returns to the castle to teach Timy (Figures 2a and 2b). Teaching activities are carried out on a blackboard in the classroom, and use different formats, such as *conceptual maps*, *sorting tasks* or, as in figures 2a and 2b, a format centered around a historical turning point. Here the task is to place a correct series of states and consequences to represent “before the turning point” and “after the turning point”. The interaction makes use of drag-and-drop of pictures. Hovering over the picture brings out the corresponding statement in text. A fully completed task means a correct description on the black-board of the “printing revolution” – Gutenberg’s invention of the printing press – with its “before” and “after”.
Figure 2 a) Upon return the student goes to the classroom to instruct Timy at the black board.
Figure 2 b) Student instructs Timy by showing how to solve the task at the black board.

The game is used as a research instrument, customized in different ways for different studies (Silvervarg, Kirkegaard, Nirme, Haake, & Gulz, 2014; Silvervarg, Gulz, & Haake, 2018).

In the present study the standard game version that includes a TA is compared with a version without a TA (NoTA) where the student learns (only) for herself. In the NoTA condition, the background narrative is modified and the student herself is vying for a position as apprentice on the team and must pass Chronos’ tests herself.

Orthogonally there are two different conditions regarding delivery of CCF on non-passed tasks (in NoTA conditions, tasks the student herself does not pass; in TA conditions, tasks the TA does not pass). In the Choice conditions, the student gets the choice of whether or not to receive CCF. In the Automatic (Auto) conditions, CCF is provided automatically. Together, this creates four conditions: TA-Auto, TA-Choice, NoTA-Auto, NoTA-Choice.
3.1. Differences Between the Four Conditions

We describe the differences between the four conditions by focusing on what happens in the classroom when the student (or TA) answers Chronos’ questions and receives CCF.

3.1.1. TA-Auto condition

In figure (3a) the student has completed a task on the blackboard, and Timy has watched and learned. Next, the student hides behind a curtain – since s/he is secretly helping Timy – and it is now Timy’s turn to show the solution on the task to Professor Chronos in order to pass. (Timy is programmed to present a solution identical to the student’s – but not mirroring each action the student made in the same temporal order). Chronos goes through the presentation on the blackboard, evaluating it. Upon this, he provides a brief overall assessment, for example “Some of the answers are correct, but you still have some way to go, continue work on the task!” (Figure 3b).
Figure 3b. Chronos provides an overall assessment to the time elf (and indirectly to the student) after Timy has demonstrated his/her knowledge.

In the case of a no-pass, Chronos thereafter automatically provides critical, constructive feedback in the text box (Figure 3c). The feedback regards two – randomly chosen – mistakes.
When Chronos’ evaluation is finished, the student has the option – in all conditions – to time-travel to find information to complete her knowledge or understanding of things, or to make revisions without time travel. When the student returns to the classroom for another round on a task, she may use or not use received CCF for the revision.

3.1.2. TA-Choice condition

In TA-Choice, where the student can say yes or no to receive CCF, the interaction is identical to TA-Auto until after Chronos has provided the overall assessment (as in Figure 3b). Here the TA whispers to the student (Figure 4), “Shall we ask Chronos to say more about what was incorrect in the answer?” The student says “yes” or “no” to this by pressing a button.
Figure 4. The TA asks the student if s/he wants to ask for CCF.

If the student says “yes” to receive CCF, Chronos will provide CCF like in the TA-Auto condition (Figure 3c). If the student says “no”, no CCF is provided.

3.1.3. NoTA-Auto condition

In the NoTA conditions, a virtual character in the form of a clock is present in the classroom while the student solves the task. It is “sleeping” while the student is working on the blackboard (Figure 5a), and “wakes up” when the student needs Chronos to come to examine the result on the blackboard (Figure 5b). The main reason for including the clock is to control for the mere presence of an on-screen character across TA and NoTA conditions.
Figure 5a. The student works on the task on the blackboard, accompanied by a sleeping virtual clock character.
Figure 5b. The virtual clock character wakes up and fetches Chronos when the student has completed the task.

Professor Chronos evaluates what the student has presented on the blackboard. When Chronos has completed his evaluation, he provides an overall assessment and then goes on to provide CCF as in TA conditions (Figure 3b & 3c).

3.1.4. NoTA-Choice condition

In this condition Chronos, after having provided the overall assessment, asks if the student wants information about what she had not done well in the task. The student says ‘yes’ or ‘no’ by pressing a button (Figure 6). If ‘yes’, Chronos provides CCF, as in the TA conditions (Figure 3b & 3c).
In its function as research instrument the game logs data related to the students’ potential processing and use of critical, constructive feedback: whether they (in Choice conditions) accept the offer of CCF; whether they spend sufficient time to be able to read it; whether they use the feedback to collect more information for the task they or their TA did not pass; whether they use the feedback in order to revise the task they or their TA did not pass. The game also collects data on how many of the six missions – in increasing order of difficulty – that are passed. (In the TA conditions it is the TA that gets pass or not, but the TA performs according to what the student has taught; thus, the TA’s performance reflects the student’s performance).

4. Method – Description of the Study

The study compares the effects of ‘learning in order to teach a TA’ or ‘learning for oneself’ on behavioral responses to the CCF and on performance on a post-test. Based on previous research, we were particularly interested in potential differences between low- and high-achieving students.
4.1. Participants and Procedure

Participants were $N = 285$ grade 5 students aged 11-12 from six public middle schools in southern Sweden, with mixed (low to high) SES backgrounds. Before the interventions, teachers provided assessments (low, mid, high) of each student’s ability to process text-based information proficiency. The rationale for collecting teacher assessments of reading proficiency was that the educational game used in the study is inherently text-based, and that such proficiency according to the teachers strongly impacts overall achievement in social science subjects for this age group.

The students, from 11 classes, participated during three full lessons of 60 minutes. The first lesson started with a 15 minutes introduction, and the third lesson ended with a 15 minutes paper-&-pencil post-test. Class teacher(s) and at least one researcher per class of 25 students were present, with the class teacher(s) being in charge, and researchers assisting with technical issues. Due to technical problems with logging, 27 students were excluded from the analyses together with 15 students who didn’t finish the post-test. Another eight students with special needs and six students who didn’t complete the initial training round were also excluded. Thus, the final dataset for the analyses included 229 students (111 females).

Data collection took place Spring 2019. Students in each class were randomly assigned to one of the four conditions. The introduction was held in two separate (half class) groups with the students in the TA conditions in one room and the students in the ‘playing for oneself’ conditions in another room. A video presenting the central features and the respective game narrative was shown in front of class. At the end of the third lesson, students received a paper-&-pencil post-test to evaluate their knowledge and understanding of content processed during game-play.

4.2. Measures and Data Sources

Accepting offers of CCF (CCF-Accept). For the participants in the Choice conditions, this measure stands for the proportion of ‘yes’-answers when students are asked if they want to know more about the errors and mistakes in a just completed ‘no pass’ task. This measure is not applicable for the Auto conditions.
**Reading CCF-texts (CCF-Read).** This measure addresses the proportion of texts that (i) are exposed on the screen because they are either automatically exposed (in Auto conditions) or said ‘yes’ to (in Choice conditions) and (ii) are exposed on the screen sufficiently long to be read before clicked away. Based on literature on 11- to 12-year-olds’ reading speed and analyses of the time interval data, the cut-off was set to 4.25 seconds. (The system logs the time that passes from when the feedback textbox is presented until the student clicks it down.) Without eye tracking, it is not possible to know if students are actually reading the text, so this is a measure of the proportion of feedback texts potentially read (i.e., not dismissed without opportunity to read). CCF-Read is a proportion calculated against the total number of initial CCF opportunities for each student.

**Using CCF to revise ‘non-passed’ tasks (CCF-Use).** This measure addresses whether a student makes use of CCF to revise a task they did not pass. This information was retrieved from the game logs by means of a script, which identified instances in which students made changes related to the CCF offered by the system. CCF-Use is a proportion calculated against the total number of initial CCF opportunities for each student.

**Learning (out-of-game) performance.** The learning outcome was measured with a post-test with six question in multiple-choice format. An example: "What did Galilei discover: i) how a glass prism can split white light into its different color elements, ii) that there were moons around the planet Jupiter, iii) that the earth revolved around the sun, iv) the energy objects have when they move. One to four alternatives are correct.” The post-test was completed at the end of the third lesson using paper and pencil, with performance ranging from 0 to 6.

**5. Data Analyses and Results**

For analyses, we evaluated post-test performance scores and the three kinds of CCF-related behaviors that correspond to the measures ‘accepting offers of CCF’, ‘reading CCF-texts’, and ‘using CCF for revising non-passed tasks’. The treatment conditions were two agent conditions – *learning in order to teach a TA* vs. *learning for oneself* (Agent[TA/NoTA]) – and two orthogonal feedback conditions – *choice of feedback* vs.
automatic feedback (Feedback[Choice/Auto]). Participants were beforehand assessed by their teachers into three levels: low-, mid-, and high-achieving students (Achievement[Low/Mid/High]).

Post-test performance and the three behavioral CCF-related measures were analyzed with the statistical software environment R (R Core Team, 2019) using a linear model approach as follows: (1) choosing a best (simplest) model for each of the four measures by an analysis of variance for fitted models (R-base function: anova, Chambers & Hastie, 1992) and the Akaike information criterion (AIC), (2) multiple regression analysis of the chosen best model using low-achieving students as reference level against mid- and high-achieving students, (3) simultaneous test of linear models to evaluate indicated interaction effects (R-package: multcomp, Hothorn, Bretz, & Westfall, 2020). All reported p-values were evaluated at an alpha-level of .05.

229 participants (N = 229) were distributed between the experimental groups (Table 1) with no significant differences as to group sizes (Chi-square test: $\chi^2 = 8.747$, df = 11, $p = .65$).

Table 1. Distribution of participants (N = 229) over Agent[TA/NoTA] × Feedback[Choice(Ch)/Auto] × Achievement[Low/Mid/High]).

<table>
<thead>
<tr>
<th>Achievement level</th>
<th>Low</th>
<th>Mid</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TA</td>
<td>NoTA</td>
<td>TA</td>
</tr>
<tr>
<td>Agent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>Ch Auto</td>
<td>Ch Auto</td>
<td>Ch Auto</td>
</tr>
<tr>
<td>Group size (n)</td>
<td>21 14</td>
<td>21 15</td>
<td>21 18</td>
</tr>
</tbody>
</table>

5.1. Post-test Performance

Figure 7 presents the average post-test scores for the two Agent conditions against the two Feedback conditions separated on the three student achievement levels.
Figure 7. Post-test performance (means and standard error) for Agent[TA/NoTA] × Feedback[Choice/Auto] separated on Achievement[Low/Mid/High].

An evaluation of the full model with Post-Test Scores as the outcome variable and Agent conditions, Feedback conditions, and Achievement levels as predictor variables, resulted in an exclusion of Feedback from the model since it had no significant contribution. The evaluation procedure ended in a best (simplest) model of 
Post-Test Score ~ Agent * Achievement (interactions included). The resulting multiple regression test for post-test performance is presented in Table 2.

Table 2. Multiple regression test of the chosen model: Post-Test Scores [0-6] ~ Agent (Agnt) [TA/NoTA] * Achievement (Achv) [Low/Mid/High]; reference levels: Agent[TA], Achievement[Low].

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.000</td>
<td>0.209</td>
<td>19.12</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Agnt[NoTA]</td>
<td>-1.417</td>
<td>0.294</td>
<td>-4.82</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Achv[Mid]</td>
<td>0.397</td>
<td>0.288</td>
<td>1.38</td>
<td>0.169</td>
</tr>
<tr>
<td>Achv[High]</td>
<td>1.269</td>
<td>0.288</td>
<td>4.40</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Agnt[NoTA] : Achv[Mid]</td>
<td>0.644</td>
<td>0.397</td>
<td>1.62</td>
<td>0.106</td>
</tr>
<tr>
<td>Agnt[NoTA] : Achv[High]</td>
<td>0.866</td>
<td>0.416</td>
<td>2.08</td>
<td>0.039 *</td>
</tr>
</tbody>
</table>

Multiple $R^2$: 0.32, Adjusted $R^2$: 0.31

Referring to Table 2 and Figure 7, there was a significant effect of Agent, with higher post-test scores in the TA compared to the NoTA condition. Likewise, there was a strong significant effect of Achievement, with high-achieving students scoring higher on the post-test than low-achieving students. There was also a strong significant interaction of Agent by Achievement. A simultaneous test of linear models (Table 3) showed a strong significant effect in that low-achieving students performed better in the TA condition than in the NoTA condition. For high-achieving students there was no
effect of Agent. Mid-achieving students were in-between, showing a weak significant effect of Agent.

Table 3. Simultaneous tests of linear models to evaluate interactions on the model: Post-Test Score [0-6] ~ Agent [TA/NoTA] * Achievement [Low/Mid/High].

<table>
<thead>
<tr>
<th></th>
<th>TA (M (SD))</th>
<th>NoTA (M (SD))</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA.Low – NoTa.Low</td>
<td>4.00 (1.42)</td>
<td>2.58 (1.48)</td>
<td>1.417</td>
<td>0.294</td>
<td>4.823</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>TA.Mid – NoTA.Mid</td>
<td>4.40 (1.10)</td>
<td>3.62 (1.34)</td>
<td>0.772</td>
<td>0.267</td>
<td>2.896</td>
<td>0.012 *</td>
</tr>
<tr>
<td>TA.High – NoTA.High</td>
<td>5.27 (0.79)</td>
<td>4.72 (1.17)</td>
<td>0.550</td>
<td>0.295</td>
<td>1.865</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Significance codes (adjusted p-values): . 0.1 * 0.05 ** 0.01 *** 0.001

5.2. Accepting Offers of CCF

The option of saying “yes” or “no” when offered CCF only relates to the Feedback[Choice] condition (N = 116). Figure 8 presents the proportion of students’ accepting (saying yes to) CCF for Agent[TA/NoTA] separated on student achievement levels (Achievement[Low/Mid/High]).

Figure 8. Average proportion (means with standard error bars) of Accepted CCF-texts for Agent[TA/NoTA] separated on Achievement[Low/Mid/High].

A multiple regression analysis on the model with Accepted CCF-texts as the outcome variable and Agent conditions and Achievement levels as predictor variables (CCF Accept ~ Agent * Achievement) (Table 4) showed a significant effect of Agent with a higher proportion of students accepting offer of CCF in the TA compared to the
NoTA condition. The test also indicated a marginally significant interaction between Agent and Achievement.

Table 4. Multiple regression test of the model: CCF Accept [0-100] ~ Agent (Agnt) [TA/NoTA] * Achievement (Achv) [Low/Mid/High]; reference levels: Agent[TA], Achievement[Low].

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>81.088</td>
<td>4.084</td>
<td>19.853</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Agnt[NoTA]</td>
<td>-19.271</td>
<td>5.776</td>
<td>-3.336</td>
<td>0.0012 **</td>
</tr>
<tr>
<td>Achv[Mid]</td>
<td>1.685</td>
<td>5.776</td>
<td>0.292</td>
<td>0.77</td>
</tr>
<tr>
<td>Achv[High]</td>
<td>9.250</td>
<td>6.107</td>
<td>1.515</td>
<td>0.13</td>
</tr>
<tr>
<td>Agnt[NoTA] : Achv[High]</td>
<td>16.702</td>
<td>8.710</td>
<td>1.917</td>
<td>0.058</td>
</tr>
</tbody>
</table>

Multiple $R^2$: 0.21, Adjusted $R^2$: 0.17

A simultaneous test of linear models (Table 5) showed a significant effect of Agent in that low-achieving students were more likely to accept offer of CCF in the TA condition than in the NoTA condition. For mid- and high-achieving students there was no effect of Agent. Interestingly, low-achieving students in the TA condition are not significantly different from high-achieving students when it comes to accepting offered CCF, while low-achieving students in the NoTA condition are.

Table 5. Simultaneous tests of linear models to evaluate interactions on the model: CCF Accept [0-100] ~ Agent [TA/NoTA] * Achievement [Low/Mid/High].

<table>
<thead>
<tr>
<th>Contrasts</th>
<th>TA</th>
<th>NoTA</th>
<th>Simultaneous test of linear models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$ ($SD$)</td>
<td>$M$ ($SD$)</td>
<td>Estimate</td>
</tr>
<tr>
<td>TA.Low – NoTa.Low</td>
<td>81.1 (15.9)</td>
<td>61.8 (26.8)</td>
<td>19.271</td>
</tr>
<tr>
<td>TA.Mid – NoTA.Mid</td>
<td>82.8 (14.4)</td>
<td>78.4 (22.9)</td>
<td>4.410</td>
</tr>
<tr>
<td>TA.High – NoTA.High</td>
<td>90.3 (14.1)</td>
<td>87.8 (10.7)</td>
<td>2.570</td>
</tr>
<tr>
<td>TA.Low – TA.High</td>
<td>81.1 (15.9)</td>
<td>90.3 (14.1)</td>
<td>-9.250</td>
</tr>
<tr>
<td>NoTA.Low – NoTA.High</td>
<td>61.8 (26.8)</td>
<td>87.8 (10.7)</td>
<td>-25.951</td>
</tr>
</tbody>
</table>

Significance codes (adjusted $p$-values): . 0.1 * 0.05 ** 0.01 *** 0.001

5.3 CCF-reads

Figure 9 presents the average proportions of CCF-reads (out of possible opportunities)
for the two Agent conditions against the two Feedback conditions, separated on the three achievement levels.

![Graph showing average proportions of CCF-reads for Agent conditions, Feedback conditions, and Achievement levels.]

Figure 9. Average proportions (means with standard error bars) of CCF-reads for Agent[TA/NoTA] × Feedback[Choice/Auto] separated on Achievement[Low/Mid/High].

An evaluation of the full model with CCF-reads as the outcome variable and Agent conditions, Feedback conditions, and Achievement levels as predictor variables, resulted in a model without interactions (\( \text{CCF Read} \sim \text{Agent} + \text{Feedback} + \text{Achievement} \)). The resulting multiple regression test for CCF-reads is presented in Table 6.

Table 6. Multiple regression test of the chosen model: CCF Read [0-100] \sim \text{Agent (Agnt) [TA/NoTA]} + \text{Feedback (Fbck) [Choice/Auto]} + \text{Achievement (Achv) [Low/Mid/High]}; reference levels: Agent[TA], Feedback[Choice], Achievement[Low].

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>77.41</td>
<td>2.86</td>
<td>27.10</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Agnt[NoTA]</td>
<td>-18.57</td>
<td>2.60</td>
<td>-7.16</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Fbck[Auto]</td>
<td>6.70</td>
<td>2.60</td>
<td>2.57</td>
<td>0.011 *</td>
</tr>
<tr>
<td>Achv[Mid]</td>
<td>4.16</td>
<td>3.14</td>
<td>1.32</td>
<td>0.19</td>
</tr>
<tr>
<td>Achv[High]</td>
<td>11.61</td>
<td>3.30</td>
<td>3.52</td>
<td>&lt; 0.001 ***</td>
</tr>
</tbody>
</table>

Multiple \( R^2: 0.25 \), Adjusted \( R^2: 0.24 \)

Referring to Figure 9 and Table 6, there was a strong significant effect Agent and a weak significant effect of Feedback. For Achievement—there was a strong significant difference between low- and high-achieving students.

Taken together, the inclination to read CCF-texts was considerably higher in the TA conditions, and high-achievers were least prone to decline or click away CCF-texts.
In the Choice conditions, students could decline offers of CCF. Since a declined CCF has no chance to be read, this can partly explain the difference in the proportion of CCF-reads between the Auto and Choice conditions.

5.4. Using CCF

To examine CCF-use we reviewed the number of CCFs that students acted upon for revising a non-passed task divided with the number of CCFs they ‘could have’ acted upon (Figure 10).

![Bar chart showing average proportions of CCF-use for different conditions](image)

Figure 10. Average proportions (means with standard error bars) of CCF-use for Agent[TA/NoTA] x Feedback[Choice/Auto] separated on Achievement[Low/Mid/High].

An evaluation of the full model with proportion of used CCF as the outcome variable and Agent conditions, Feedback conditions, and Achievement levels as predictor variables resulted in a best (simplest) model of $CCF \text{ Use} \sim Agent + Achievement$ (Feedback and interactions excluded). The resulting multiple regression test for CCF-reads is presented in Table 7.
Table 7. Multiple regression test of the model: CCF Use [0-100] ~ Agent (Agnt) [TA/NoTA] + Achievement (Achv) [Low/Mid/High]; reference levels: Agent[TA], Achievement[High].

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>29.42</td>
<td>1.99</td>
<td>14.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Agnt[NoTA]</td>
<td>-3.57</td>
<td>1.98</td>
<td>-1.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.073 .</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0020 **</td>
</tr>
<tr>
<td>Achv[Low]</td>
<td>-8.21</td>
<td>2.51</td>
<td>-3.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0013 **</td>
</tr>
</tbody>
</table>

Multiple $R^2$: 0.073, Adjusted $R^2$: 0.061

Referring to Figure 10 and Table 7 and 8, the measured number of CCF-revised tasks was low and the standard deviation relatively high, why the analyses should be interpreted with some caution.

As for the results, there was a significant effect of Achievement in that high-achieving students used CCF to revise a task to a significantly higher degree than low- as well as mid-achieving students. It can also be seen, in Figure 10, that for low- and mid-achieving students the average proportion of CCF-use was higher in the TA than in the NoTA conditions (Figure 10).

5.5. Effects of Responses to CCF on Post-test Performances

The regression analyses so far have found an effect of Agent[TA/NoTA] on post-test performance. They have also found differences in CCF processing stages with respect to behavioral differences with and without TA. Separately, these analyses show that the Agent conditions affected post-test performance and CCF-related behaviors. Using mediation analyses, we examined whether the effect of Agent condition on post-test performance was mediated by (was affected by an underlying mechanism of) feedback-related behaviors.

Conditional process analyses using the SPSS statistical software implementation of PROCESS v.3.4 (Hayes, 2017) was used to determine the significance of direct and indirect effects by means of 5,000 bootstrapped samples. When used in the analyses, the Agent condition of NoTA was used as a reference. Figure 11 presents the simplest conceptual diagram of the mediation effects in question (Model A). Results, summarized in Table 8, included a significant direct effect of Agent condition on post-test score/performance. In other words, part of the effect of Agent on post-test
performance was independent of the intermittent Feedback-related behaviors, e.g., there may be general motivational factors. There was also a significant indirect effect through Reading CCF, suggesting that part of the effect of Agent on post-test was mediated by differences in feedback-reading behaviors between conditions. Interestingly, the effect of using CCF to revise tasks did not significantly mediate effects on post-test performance. Using CCF to revise occurred at relatively low rates across conditions, which may affect statistical results/power.

Figure 11. Conceptual diagram of the mediation analysis model.

We continued to probe this mediation effect by including student achievement as a moderating factor (Model B), as we hypothesized that achievement may be underlying the feedback-behavior mechanisms. As such, we included Achievement as a moderator of the direct link between Agent and post-test performance, and as a moderator of Agent on students’ reading of CCF. Results are presented in Table 8. We found the direct effect of Agent was significant for the low- and medium-achieving students, but not the high-achieving students. We also found that this indirect effect was not significantly different between achievement levels. Finally, although the indirect effect of Agent on post-test performance mediated through reading CCF was significant for all three achievement levels, the index of moderated mediation was not significant; i.e., the mediation effect was independent of Achievement.
Table 8. Summary of direct and indirect effects of probing CCF on post-test score computed via conditional process analysis. Significant effects (Sig.) are indicated with an asterisk (*).

**Model A**

<table>
<thead>
<tr>
<th>Effect Type</th>
<th>Effect Description</th>
<th>Effect Coef.</th>
<th>95% CI (bootstrap)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>Learner condition on posttest score</td>
<td>0.56</td>
<td>[0.18, 0.94]</td>
<td>*</td>
</tr>
<tr>
<td>Indirect</td>
<td>Learner condition on posttest score through reading CCF</td>
<td>0.37</td>
<td>[0.18, 0.62]</td>
<td>*</td>
</tr>
<tr>
<td>Indirect</td>
<td>Learner condition on posttest score through acting on CCF</td>
<td>0.00</td>
<td>[-0.06, 0.05]</td>
<td></td>
</tr>
<tr>
<td>Indirect</td>
<td>Learner condition on posttest score through reading CCF then acting on CCF</td>
<td>0.04</td>
<td>[-0.02, 0.11]</td>
<td></td>
</tr>
</tbody>
</table>

**Model B**

<table>
<thead>
<tr>
<th>Effect Type</th>
<th>Effect Description</th>
<th>Effect Coef.</th>
<th>95% CI (bootstrap)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>Learner condition on posttest score</td>
<td>1.06</td>
<td>[0.47, 1.65]</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>low performers</td>
<td>0.55</td>
<td>[0.03, 1.07]</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>mid performers</td>
<td>0.32</td>
<td>[-0.26, 0.90]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>high performers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect</td>
<td>Learner condition on posttest score through reading CCF</td>
<td>0.34</td>
<td>[0.11, 0.66]</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>low performers</td>
<td>0.21</td>
<td>[0.06, 0.42]</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>mid performers</td>
<td>0.23</td>
<td>[0.06, 0.46]</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>high performers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect</td>
<td>Learner condition on posttest score through acting on CCF</td>
<td>0.00</td>
<td>[-0.04, 0.03]</td>
<td></td>
</tr>
</tbody>
</table>

6. Discussion

In this study, we examined how students interacted with critical constructive feedback (CCF) in a digital learning environment. We focused on differences in response patterns and performance between conditions that included a teachable agent (TA) and conditions that did not, taking students’ achievement levels into consideration. Overall analysis showed that including a TA affected how students responded to CCF, and
mitigated inclinations to avoid or neglect CCF. These results were especially strong for lower-achieving students.

6.1. Effects of Conditions on Feedback Response Patterns

We examined three stages of students’ uptake or neglect of CCF. First, in the Choice conditions, students had the choice of accepting an offer of CCF or not. High-achieving students were highly inclined to say ‘yes’ to CCF regardless of whether there was a TA present. Low-achieving students without the presence of a TA were significantly less likely to say ‘yes’ to CCF than high-achieving students. However, in an environment that included a TA, there was no longer any significant difference between low- and high-achieving students in this respect. In other words, the response pattern of low-achieving students, when in TA conditions, became similar to that of high-achieving students.

Next, examined – for all conditions – the proportion of cases where students may have read the CCF-texts (i.e., did not dismiss them by saying ‘no’ or quickly clicking them away quickly). Students in the TA conditions dismissed feedback texts to a significantly lower degree. This was an effect that was present for students of all achievement levels. In the cases most similar to a typical digital learning environment – where learners themselves perform and feedback is provided automatically – CCF-text dismissal occurred in approximately 29% of the cases. When learners were responsible for teaching a TA, this decreased to approximately 6% of the cases.

Finally, we looked at the extent to which students made use of CCF for revising a non-passed task. Comparing low- and high-achieving students, we see (Table 8) that high-achieving students in the TA and NoTA conditions equal each other, making use of CCF in, 28% and 27,7% of the cases respectively. For low-achieving students in NoTA conditions the number is 15,8% while students in the TA conditions reached 23,1%. This pattern is similar to other patterns, in this and other studies, where the gap between students rated as lower- and higher-achieving is reduced by the presence of a TA, both in terms of behavior and performance.

On the whole, we see an overall tendency of low use of CCF, across conditions and achievement levels, in line with previous studies (e.g. Segedy, Kinnebrew, and Biswas, 2013; Clarebout & Elen, 2008; Wotjas 1998; Tärning et al., 2020). There are many possible reasons for this. First, students may not have actually read the feedback
text; we only estimate whether it was possible for them to read given the time slot the
text was on screen. Second, students may not remember what the feedback was, once
they get to revising the task. Third, students may find feedback texts difficult to
understand – even though the texts used in this study had been evaluated both by several
teachers and many students in previous studies – as reading comprehension abilities
vary considerably in the student population. Fourth, students may not think that reading
and/or acting upon feedback will pay off compared to other strategies, e.g. combinations
of reasoning and using trial-&-error.

6.2. Effects of Conditions and Feedback Response Patterns on Post-test Scores

Turning to the post-test scores, overall the TA conditions had a significant, positive
effect on learning outcomes for low- and mid-achieving students. In addition, the gap
between low- and high-achieving students was smaller in TA conditions (4.0 vs 5.3)
than in NoTA conditions (2.6 vs 4.7).

To look more specifically at relationships between TA condition, CCF
behaviors, and post-test, we conducted a moderated-mediation analysis that included
agent condition, feedback behaviors (reading and using CCF), and achievement level.
Results indicated both significant direct and indirect effects of Agent condition on post-
test performance. The direct effect, which was found for low- and mid-achieving
students, tells us that the inclusion of a TA may have had positive effects on learning
not related to feedback response patterns. These might include a general stronger
motivation to learn and perform, i.e. the protégé-effect. Importantly, the indirect effect
through reading CCF, which was found across achievement levels, suggests that part of
the TA condition effects on post-test performance can be explained by decreased
feedback neglect in the presence of a TA. This indirect effect was found for reading
feedback, but not for using CCF to revise. However, CCF-use happened at relatively
low rates across conditions, which may have contributed to it not being a significant
mediator.

Finally, we note that the interactions between TA/NoTA and level of
achievement that show up in the individual analyses are not reproduced in the mediation
analysis.
6.3. Instructional Implications

The results of this study highlight the importance of considering the processes of feedback uptake or neglect, in addition to its outcomes on post-test performance. This would represent an important shift in the literature on feedback. Studies on the benefits of feedback on learning are extensive (Hattie & Timperley, 2007), but how particular factors influence the uptake or neglect of CCF are not well known. This study makes two contributions. First, it shows that lower achieving students are less likely to read and make use of the feedback texts than higher achieving students. They are less likely to choose CCF, to read it when presented, and to act on it. This is problematic and highlights an area in which more research is needed to understand and reduce this gap.

A second contribution of the study is to demonstrate that one way to mitigate this feedback neglect effect involves including a TA in a learning environment. In focusing on how TAs affect students’ approaches to critical-constructive feedback the analysis found dramatically decreased neglect of CCF, particularly for students rated by their teachers as lower-achieving. The NoTA conditions corresponds to a ’standard’ situation for a learner in that what one does and accomplishes has a value primarily for oneself. The TA conditions may be described as a situation where ’someone’ else, as well, is directly affected by the student’s choices and accomplishments, 'someone' who can share both success and failure. This seems to have benefits for learning as a general outcome, as well as for the propensity to not dismiss CCF and, for lower-achieving students, the propensity to accept an offer of CCF and the propensity to make use of CCF.

In a TA-based environment that 'someone’ is a computer character. Other learning-by-teaching or collaborative educational scenarios may have similar effects. This research does not yet tease apart which elements of TAs increase productive feedback behaviors, and to what degree. Prior research has found that the social aspects of TAs increase general motivation, which may influence feedback uptake. Additionally, research has shown that TAs can provide an ego-protective buffer that makes CCF less fraught with negative feelings of failure, which could decrease neglect. Other learning environments that include similar features can be examined to determine if they also increase feedback uptake and decrease neglect, ultimately leading to positive effects on learning.
Research shows that, overall, neglect of feedback is prevalent. The take-home message is that we need different ways to encourage students to respond to feedback in productive ways.

### 6.4. Limitations and future Work

Our analyses do not discriminate between different tasks and changing levels of difficulty, nor do they consider how many attempts a student has made on a task, and they do not consider potential wearing-out effects, etc. To better understand students’ trajectories, more detailed and qualitative analyses of the logs of student interactions are required. Potentially, such analyses could be a basis for a follow-up large scale study. Similarly, more in-depth analysis could help explain why the prevalence of acting on feedback was relatively low (though consistent with other studies) and inform changes to the environment to increase action-on-feedback.

Our result suggests that the increased potential reading of CCF in the TA conditions had a direct positive effect on learning, measured by the post-test, for all students. However, an alternative explanation is that an inclination to dismiss CCF-texts – saying no when offered CCF or clicking away CCF-texts – correlates with (is a proxy for) a lower inclination to read all kinds of texts in the game. Since the game is heavily text-based with respect to learning content, less overall reading likely corresponds to less learning and lower post-test scores. More studies are needed to discriminate between these two possible explanations.

Finally, our study is limited to studying students’ responses to CCF when presented via text. We look forward to similar studies where CCF is provided in other formats and modalities, not the least since our results suggest that non-reading of CCF (declining offers of CCF-text or clicking away CCF-texts) is associated with less learning.

### References


https://doi.org/10.1111/bjet.12796


