Computers in Human Behavior 29 (2013) 845–857

Contents lists available at SciVerse ScienceDirect

Computers in Human Behavior

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

# Motivating agents in software tutorials

H. van der Meij *

University of Twente, Faculty of Behavioral Sciences, Department of Instructional Technology, P.O. Box 217, 7500 AE Enschede, Netherlands

ARTICLE INFO

Article history:
Available online 16 January 2013

Keywords:
Animated Pedagogical Agents
Tutorial
Software training
Motivation

ABSTRACT

Pedagogical agents can provide important support for the user in human–computer interaction systems. This paper examines whether a supplementary, motivating agent in a print tutorial can enhance student motivation and learning in software training. The agent served the role of motivator, attending the students to issues of task relevance and self-efficacy. The agent was presented in the tutorial by means of images and written messages. An experiment compared the agent condition with a no-agent (control) condition. Participants were 49 students (mean age 11.3 years) from the upper grades of elementary school. Data on motivation and learning were gathered before, during and after training. The findings revealed that students in the agent condition did significantly better on skills measures during and after training (i.e., performance indicators, posttest, and retention test). In addition, marginally significant differences favoring these students were found for flow experience during training and for motivational gains on task relevance and self-efficacy after training. The design strategies of the motivating agent are considered relevant for the creation of Animated Pedagogical Agents.



## 1. Introduction

Can an agent boost motivation and increase learning when students learn from a software guide? This study examines that question in a situation in which a print tutorial instructed elementary school students about Microsoft Word's formatting options. In the control condition students worked with a tutorial without an agent. In the experimental condition they received the same tutorial with a motivating agent added.

The agent was displayed as an annotated photograph of a young male (Max) who was the same age as the students. Max was introduced as a co-student. Max plays the role of motivator. Throughout the tutorial Max sides with the students, conveying written messages that highlight the relevance of an activity for students and stimulate their confidence in their abilities to overcome obstacles. The underlying idea of the study is that the agent raises students' motivation during training and thereby stimulates their active engagement with the tutorial, which increases learning.

Our inspiration for Max came from studies that have included Animated Pedagogical Agents (APAs) to motivate users in virtual (computer-based) environments. We therefore begin this paper with a summary on the designs and effects of such APAs. Next, we introduce the agents that we found used in software manuals. Thereafter, we describe Max's design and discuss the set-up, findings and conclusions from an experiment in which we examine whether an agent can enhance motivation and learning in a print tutorial.

## 2. Motivating Animated Pedagogical Agents (APAs)

APAs are fast becoming hugely popular in human–computer interaction systems. APAs come in different guises, including humans (e.g., AutoTutor – Graesser & McNamara, 2010), animals (e.g., Herman the Bug – Moreno, Mayer, Spires, & Lester, 2001), and inanimate objects (e.g., Microsoft's Clippy – Haake, 2009). Fig. 1 displays, from left to right, Clippy from Microsoft Office, the educational technology wizard, Chris, in the study by Baylor and Ryu (2003) and Joe, an AutoTutor in the study by Graesser and McNamara (2010).

An important stimulus for the development of APAs is that they can humanize the user experience. Some of the first APAs were introduced into virtual environments specifically for that purpose. To make users' interactions with the environment more life-like, considerable design efforts went into creating a likeable, intelligent, credible, and trustworthy APA.

Empirical studies have examined the claim that the presence of an APA primes a social interaction schema that positively influences the user's appraisal of the system, an effect variously labeled as social cue, social agency, social presence or persona effect (henceforth simply the persona effect) (e.g., André, Müller, & Rist, 1996; Atkinson, Mayer, & Merrill, 2005; Bates, 1994; Cassell, 2000; Choi & Clark, 2006; Lester et al., 1997; Moreno et al., 2001; Moundridou & Virvou, 2002; Paiva & Machado, 1998; Picard & Klein, 2002; Reeves & Nass, 1996). These studies have validated

* Tel.: +31 053 4893656; fax: +31 053 4892895.
E-mail address: H.vanderMeij@utwente.nl

0747-5632/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chb.2012.10.018

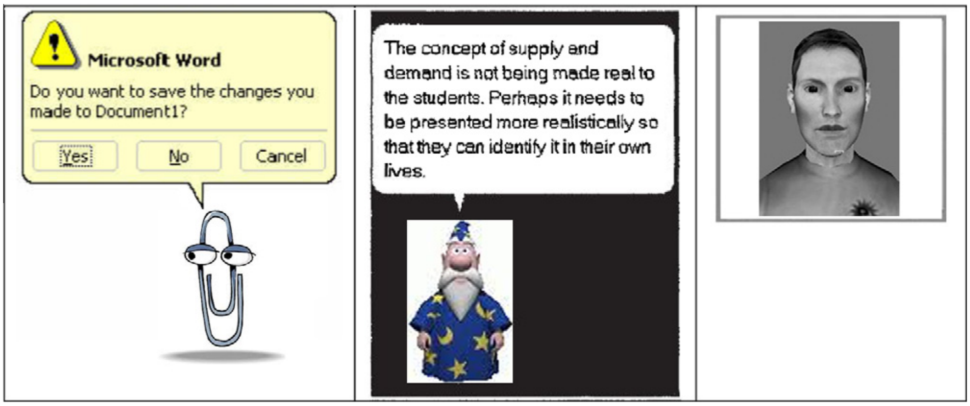



**Fig. 1.** Three APA-variants.

the persona effect, by and large. When an APA is present, the virtual environment is often seen as more entertaining and enjoyable. Granted that the presence of an APA can make the student's experience of the virtual environment more life-like, what more can be done to enhance student motivation (and learning)? Only a few empirical studies on APAs have addressed this issue. Below we discuss these studies, focusing on those manipulations that directly addressed the motivational qualities of the APA, and on the measures of student motivation and learning used.

Moreno and Mayer (2004) examined whether the delivery mode and communicative style of the APA (an insect called Herman the Bug) affected college students' motivation and learning of botany. Delivery mode was either low or high immersion. In low immersion mode participants worked with a desktop computer. In high immersion mode a head-mounted display provided visual information that changed each time the participant moved in any way. The idea was that a higher immersion would lead to a greater sense of physical and social presence, and hence to less perceived difficulty of the material and deeper learning. One communicative style was non-personalized formal speech. When using this style, the APA addressed the users in a neutral, third-person format. The other communicative style was personalized; the APA used informal, conversational messages incorporating "I" and "you" when communicating with the user. The idea was that the personalized APA would be more successful than the non-personalized APA in creating a sense of social presence and relationship that should affect motivation and learning. That is, the students should perceive the material as easier and requiring less cognitive effort, and learn more.

The study was set up as a factorial design with four conditions: delivery mode (low versus high immersion) and communicative style (non-personalized versus personalized speech). The 48 participants in the study were randomly distributed over conditions. Only students with low scores for experience with botany (self-ratings) were included. The dependent variable for student motivation was a difficulty scale. Students were asked to rate difficulty and cognitive effort ("How difficult was the material?" and "How much effort was required to learn the material?") on a 10-point Likert scale, with higher scores indicating greater difficulty/effort. Learning was assessed with a retention test and a transfer test. The retention test consisted of three questions about types of objects discussed in the program (e.g., "Please write down all the types of roots that you can remember from the lesson"). The transfer test presented seven problems for students to solve. These problems asked students either to design a plant that could live in a specific environment (e.g., low sunlight, high temperature), or to identify the type of environment in which a given plant would flourish. The final question for both problem types asked students to justify their answer ("Why do you think the plant will survive/flourish in this environment?").

There was no support for the hypothesis that a higher immersion would affect the scores for difficulty/effort and learning. Also, no interaction effects were found. With regard to communicative style, the prediction was confirmed that personalized messages would reduce perceived difficulty ($p = 0.05$, $d = 0.67$). In addition, there was a significant effect of personalization on learning, for both retention ($p = 0.002$, $d = 0.77$), and transfer ($p = 0.0001$, $d = 1.64$).

Baylor, Shen, and Warren (2004) examined the influence of motivational messages and emotional support from a male APA on 67 college students who were learning about word problems in math. The motivational messages were briefly described as consisting of five types: verbal suggestions (e.g., "Try to repeat the four concepts in your mind: principal, rate, time and interest"), affiliation (e.g., "I'll be with you all the way"), positive feedback (e.g., "So far you're doing great"), self-efficacy (e.g., "Just keep on trying. You can do it"), and emotional support (e.g., "Don't panic. Just take it easy and listen closely"). Emotional support was provided via facial expression and vocal tone quality, and was characterized as either positive or evasive.

The study was set up as a factorial design with four conditions: With or without motivational message, and with positive or evasive emotional support. The dependent variable for student motivation, self-efficacy, was measured with the domain-specific question, "How sure are you that you can correctly solve a percentage word problem?". To assess learning, an open-ended question asked participants to solve a (new) percentage word problem. The answers to both questions were scored on a 5-point Likert scale, with higher scores indicating stronger self-efficacy or a better solution.

Significant increases in students' self-efficacy were found when students received positive emotional expressions and when motivating messages were present, with all comparisons yielding a $p$ value < .001 and effect sizes of $d = 0.82$ or higher. Condition did not affect learning outcomes.

Arroyo and her colleagues have conducted several studies examining the influence of APAs on students' motivation and learning in an intelligent tutoring system on geometry problems. Arroyo, Woolf, Royer, and Tai (2009) designed a motivating APA that gave students feedback after they completed a math problem, and tested its effects in two consecutive studies. Following recommendations from Dweck (2007), the APA's messages disregarded success and valued effort. For instance, if a problem had been solved correctly but the student had put in low effort, the APA would react with a message like "That was good, however, I prefer harder questions so that we learn from the help that the computer gives, even if we get them wrong". And for a high-effort correct answer the APA might react with "Hey, congratulations. Your effort paid off, you got it right".

Study 1 included 38 high school students, with approximately equal numbers of male and female participants. There were three conditions: a male APA, a female APA, or no APA. Motivational measures came from repeated answers to a question about self-efficacy, task relevance, frustration or excitement. That is, every 5 min during training and after completing a problem the student would be asked the question "How confident, interested, frustrated, or excited do you feel right now?". Answers were scored on a 5-point Likert scale. Effort was assessed with a user model. That is, the degree of effort a student invested in solving a specific math problem was compared to what was expected for that math problem on the basis of thousands of past student interactions with it. Mainly the student's time per action was assessed (Arroyo et al., 2010). No details about the mathematics tests were provided. No significant effects for the presence of the APA on motivation or learning were found.

Study 2, which included 29 undergraduate female students, examined the impact of APA gender on student motivation and learning. Students were randomly assigned to Jake, the male APA, or Jane, the female APA. Students' self-efficacy, task relevance and math performance were measured before and after training (no details provided). Just as in study 1, there were also repeated motivational measures during training. For the measured during training, a significant difference was found only for excitement, with the male APA yielding higher scores ($p < 0.01$, $d = 0.68$). A significant effect of APA gender was found on self-efficacy gain ($p < 0.02$, $d = 0.99$). Female students who worked with the male APA had higher self-efficacy gains (post–pre). The male APA tended to lead to higher learning gains ($p = 0.058$).

In later work, Arroyo and her colleagues (Arroyo, Woolf, Cooper, Burleson, & Muldner, 2011) again studied the effects of an APA on student motivation and learning in an intelligent tutoring system on geometry problems. In this study the APA's messages were based on Weiner's attribution theory, and a distinction was made between "attribution training" and "effort-affirmation". The interventions were linked with the stage of problem solving. Attribution training messages were given only before students started on a task. Messages of this type tried to address students' beliefs about why they may succeed or fail at tasks (e.g., "Keep in mind that when we are struggling with a new skill we are learning and becoming smarter!", and "We will learn new skills only if we are persistent. If we are very stuck, let's call the teacher, or ask for a hint from Wayang!"). Effort-affirmation messages appeared only after students had correctly solved a math problem. These messages acknowledged and might praise a correct solution (e.g., "That was too easy for you. Let's hope the next one is more challenging so that we can learn something", and "Good job! See how taking your time to work through these questions can make you get the right answer?"). In addition, the APA might prompt students to reflect on their problem-solving strategies after they gave an answer (e.g., "Are we using the correct strategy to solve this? What are the different steps we have to carry out to solve this one?").

Participants were 108 high school students. There were three conditions: a male APA, a female APA, or no APA. Motivational measures came from a pre-training questionnaire with three questions about self-efficacy (e.g., students were asked to compare their math ability to that of other students, and to compare mathematics to other subjects) and three questions about task relevance that asked whether students liked or valued math. Every 5 min during training and after finishing a problem the student would be asked "How confident, interested, frustrated, or excited do you feel right now?". Answers were scored on a 5-point Likert scale. No details about the mathematics tests were provided.

The presence of an APA significantly affected the students' math interest during training ($p < 0.05$). In addition, there was a significant effect of the APA on math confidence during training which varied by the gender of the participants. That is, the APA improved the confidence of the female participants, but not the males ($p < 0.01$). The presence of the APA also significantly reduced the students' frustration during training ($p < 0.01$). This effect was especially strong for female students ($p < 0.005$). Significant effects of the female APA were found on task relevance gain (post–pre, $p < 0.05$) and self-efficacy gain (post–pre, $p < 0.05$). No effect of the APA on learning was found.

These studies reveal that the design of APAs that seek to address student self-appraisals of motivation, rather than appraisals of the APA or system, is oriented more toward what Moreno (2005) calls the internal properties of the APA. In her review on APAs, Moreno concluded that effects on student motivation and learning are more likely to be due to internal properties of the agents, such as their content, than to external properties such as their visual and auditory features. A further noteworthy feature of the studies by Arroyo et al. (2009, 2011) is that a motivational theory was used to design the APA.

## 3. Motivating agents in software tutorials

Software tutorials have rarely employed agents to motivate the reader. A longstanding exception is the series of manuals known as "For Dummies" books (see www.dummies.com). A unique feature of these manuals is the foregrounding of the author as agent. The author explicitly and regularly steps forward with meta-comments to the readers about their learning experience.

Humor is a key feature of this agent's talk. More importantly, the "For Dummies" writer regularly addresses reader motivation directly with comments such as "Congratulations, now let the champagne bubble", "This tip is wonderful", and "Have fun". In addition, the agent frequently acknowledges frustrations and uncertainties that may arise in software training. In combination with humor, this leads to comments such as "A tranquillizer is optional in this section", "Unfortunately WinWord has no easy command for this", "You don't like it? Click the Undo button to return to your original headache", and "To assist you in giving a text the right margins WinWord comes to save you" (Gookin, 1994). None of these comments is accompanied by an image of the agent, however. Only the author's voice comes through in these written messages.

Motivating agents also feature in Field's (2005) print tutorial on SPSS, a software program for conducting statistical analyses. This tutorial features six different agents. Besides functioning as entertainment, each agent fulfills a specific cognitive or affective role. Among others, the tutorial features Smart Alec – a know-it-all guy who helps when things get difficult, Curious Cat – whose main role is that of looking cute and making bad cat-related jokes, and Brian Haemorrhage – who asks questions constantly and who becomes more despondent as the book progresses (see Fig. 2). Humor is again a corollary feature in the attempted solution or mediation of obstacles for the reader by these agents.

Does the presence of these agents positively contribute to user motivation and learning from these tutorials? Although these are hugely successful user guides, little is known about the effects of the presence of agents on the user. More generally, little is known about the effects of motivating agents in software tutorials. Only three empirical studies have been published.

One study (van der Meij, 2008) investigated whether secondary school students' motivation and learning would improve with the presence of a female agent in a tutorial on Microsoft Word's formatting options. The agent's messages revolved around issues of

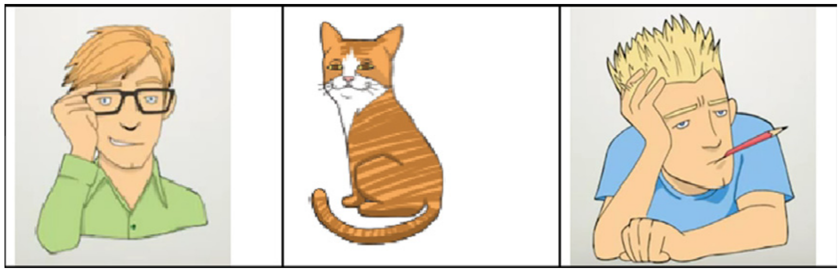


**Fig. 2.** Three Agents in Andy Fields' *Discovering Statistics Using SPSS* (2009). From left to right: Smart Alec, Curious Cat and Brian Haemorrhage.

motivation and cognition. The messages always addressed task relevance (e.g., "Oh yes. How funny. That's handy"), or user confidence (e.g., "I can remember this"). In addition, she expressed important thought processes (e.g., "Just one word? All one by one?"), feelings (e.g., "Cool", "Careful", "... awful ..."), or a combination of the two. No effects of the agent on motivation or learning were found.

In a follow-up study (van der Meij, Op de Weegh, & Weber, 2009), extensive efforts were made to improve the design of the motivational agent. An experiment then compared the influence of three agent conditions (i.e., cognitive, motivational or mixed) with that of a non-agent condition. Just as before, the study included secondary school students and the tutorial dealt with Microsoft Word's formatting options. The study revealed that, compared to the cognitive and non-agent condition, the motivational and mixed agent conditions led to significantly higher scores on motivation-related measures of mood and adoption of prescribed methods during training. In addition, students in these conditions also scored significantly higher on appraisals of task relevance and self-efficacy after training. Performance measures indicated that conditions did not differ on immediate posttest scores, but on a retention test the motivational and control conditions had significantly higher scores than the other conditions.

The third study (van der Meij, 2012) examined Clark and Choi's (2005) claim that an agent can tax the user beyond a level that is functional. Their argument is that an agent who supplements the instructions poses an additional burden on the user's processing capacities that can be detrimental. In other words, for some audiences the agent's messages may actually be harmful rather than beneficial. This issue was examined in a study in which elementary school students with extremely low reading skill were asked to process a tutorial with or without a motivating agent. The findings in this experiment were nearly the exact opposite of those of the earlier study by van der Meij et al. (2009). That is, student motivation increased significantly more in the control than in the agent condition. In addition, whereas no difference between conditions was found on the posttest, the control condition performed significantly better on the retention test.

## 4. Design of the motivating agent for the software tutorial

The present study follows along the lines of the study by van der Meij et al. (2009). In this case a regular elementary school audience forms the target population for a tutorial on Microsoft Word's formatting options. As with the agent created by Arroyo et al. (2011), the design of the agent was based on a motivational theory. However, the agent's design was based on expectancy-value rather than attribution theory, for two reasons.

One reason is that expectancy-value theory (Eccles & Wigfield, 2002) has "some of the strongest empirical support in educational settings" (Pintrich & Schunk, 2002, p. 89). Expectancy-value theory indicates that student motivation derives from a combination of expectancies for success and valuation of success. Self-efficacy and task relevance are seen as key constructs for the first and second component, respectively. Self-efficacy refers to students' belief in their capacity to succeed for specific tasks (Bandura, 1997). Task relevance has to do with incentives or reasons for task engagement.

Another reason is that an instructional theory with a proven track record exists that describes how to design the agent's messages for task relevance and self-efficacy. Keller's ARCS-model (1987, 2010) includes four conceptual categories, namely Attention, Relevance, Confidence and Satisfaction, that subsume many facets of motivation. The agent we designed concentrates on the relevance and confidence components of this model, because their design guidelines address appraisals of task relevance and self-efficacy.

The actual design of the agent in our study was based on the design guidelines from the ARCS-model (Keller, 1987, 1999, 2010; Keller & Kopp, 1987). This model offers extensive advice on how to enhance students' perceptions of task relevance and self-efficacy. Various empirical studies, none including agents, have substantiated its effectiveness (e.g., Feng & Tuan, 2005; Huett, Kalinowski, Moller, & Huett, 2008; Keller & Suzuki, 2004; Loorbach, Karreman, & Steehouder, 2007; Loorbach, Steehouder, & Taal, 2006; Song & Keller, 2001). Table 1 describes the strategies for the two motivational concepts on which the agent focuses and illustrates the messages the agent conveys to the students.

## 5. Research questions

Two conditions were compared. In the control condition students worked with a tutorial without an agent. In the agent condition, a motivational agent was added to the tutorial. Table 2 summarizes the dependent variables that are measured in the study. Below we detail our five research questions.

(1) Does condition affect *training time*? Training time may be influenced by how engaging each condition is, or the extent or difficulty of the learning material. If one condition leads to higher learning outcomes but takes longer, it is possible that this is due to time-on-task rather than to differences in instructional effectiveness. The tested prediction is that there are no significant differences between conditions for training time.

(2) Does condition affect motivational and functional states during training? According to the cognitive-motivational process model (Vollmeyer & Rheinberg, 1999, 2006), motivational and functional states are important mediators of the influence of initial motivation on learning (compare Lau & Roeser, 2002). A motivational state is a monitor of the fun, fear, frustration and similar feelings that students may experience during training. In this study we look at mood and worries as measures of motivational state. Mood

**Table 1**
An illustration of the agent's messages for the various strategies that the ARCS-model proposes to enhance students' appraisals of task relevance and self-efficacy.

| Motivational concept | Strategy | Illustrative agent message |
|---|---|---|
| Task relevance | *Goal orientation*<br>(How can the current task be made valuable and stimulating for the student?) | "Do I recognize this? Yes. This is a very annoying problem."<br>"If this works I can use it. Very handy for reports and the like." |
| | *Motive matching*<br>(How does this task fit the students' needs?) | "I dearly want to learn how to present a text nicely on paper."<br>"I always mess around with Tab. I know that that isn't right. But how is it done properly?" |
| | *Familiarity*<br>(How can the task be tied to the students' experiences?) | "This is also how I often see it."<br>"… the ruler? I always mess around with the TAB-key." |
| Self-efficacy | *Learning requirements*<br>(How can the agent build positive expectations for success?) | "Not a problem. I just have to choose the right buttons."<br>"This is easy." |
| | *Personal control*<br>(How can the agent convince students that their success is based on ability and effort?) | "Oops that's tough. I must pay attention."<br>"If I can do one, I can also do the other." |
| | *Success opportunities*<br>(How can the agent support or enhance students' belief in their competence?) | "Great. I can now adjust margins just as I want to."<br>"Gotcha. I did it. Now move on." |

**Table 2**
Overview of the measurement instruments and dependent variables in the study.

| When measured | Instrument | Dependent variable |
|---|---|---|
| Before training | *General Initial Motivation Questionnaire (GIMQ)*<br>Aim: To assess initial, general appraisals of motivation | *General motivational appraisals*<br>• Probability of success<br>• Anxiety<br>• Interest<br>• Challenge |
| | *Specific Initial Motivation Questionnaire (SIMQ)*<br>Aim: To assess initial, task-specific appraisals of experience and motivation for each of five main tasks | *Specific motivational appraisals*<br>• Task experience<br>• Task relevance<br>• Self-efficacy |
| | *Pretest*<br>Aim: To assess task-specific prior skill for each of five main tasks | *Domain knowledge*<br>• Prior skill |
| During training | *Mood indicator (MI)*<br>Aim: To assess mood states after completion of each of two chapters | *Motivational state appraisal (mood valuation)*<br>• Positive, neutral, negative |
| | *Flow and Worry Questionnaire (FWQ)*<br>Aim: To assess motivational and functional states after completing each of five main tasks | *Motivational state appraisal*<br>• Worries<br>*Functional state appraisal*<br>• Flow |
| | *Task Performance Indicator (TPI)*<br>Aim: To assess training time and success on instructed tasks (5) and exercises (2) | *Training performance*<br>• Success on task instructions<br>• Success on exercises |
| | *Time*<br>Aim: To assess training time | *Training duration*<br>• Time |
| After training | *Specific Final Motivation Questionnaire (SFMQ)*<br>Aim: To assess final, task-specific appraisals of motivation | *Specific motivational appraisals*<br>• Task relevance<br>• Self-efficacy |
| | *Agent Appraisal Questionnaire (AAQ)*<br>Aim: To assess the attention paid to the agent and the perceived credibility of the agent | *Agent appraisal*<br>• Agent reading frequency<br>• Agent credibility |
| | *Posttest (Post)*<br>Aim: To assess task-specific skills for five instructed tasks immediately after training | *Domain knowledge*<br>• Post skill |
| | *Retention test (Ret)*<br>Aim: To assess retention of task-specific skills 3 weeks after training | *Domain knowledge*<br>• Retained skill |

reflects the students' emotional state at a particular moment. Worries can be defined as their state of anxiety during training. Functional states refer to the students' concentration and task engagement during training. This study examines flow as a measure of functional state. Flow refers to completely focused motivation. It is a pleasant state of concentration and effort during task performance (Csikszentmihalyi, 1975, 1991). The tested prediction is that the agent condition yields more positive moods, less worries and more flow.

(3) Does condition affect task performance during training? Motivation influences goal setting, goal choice, task engagement, systematicity of effort and persistence (e.g., Britner, 2008; Caprara et al., 2008; Lau & Roeser, 2002; Vollmeyer

& Rheinberg, 2006). These activities contribute to the effort that students spend on processing the tutorial. Both conditions were expected to experience considerable success on the main tasks in the tutorial. The tested prediction is that the agent condition yields higher scores for success on task instructions and exercises.

(4) Does gender affect agent reading frequency and credibility? After training is completed students in the agent condition are asked to rate agent reading frequency and to give an appraisal of the credibility of the agent. Both boys and girls were expected to attend to a larger percentage of the agent's messages and to find these messages credible. The tested prediction is that boys give higher ratings than girls because of the greater model-target similarity (Bandura, 1997).

(5) Does condition affect *motivational gains*? Several researchers have argued that motivation is best measured in a task-specific way (e.g., Bandura, 2012; Boekaerts & Corno, 2005). This study assesses specific motivation before and after training. Each assessment revolved around the two key constructs from expectancy-value theory (Eccles & Wigfield, 2002), namely, perceptions of task relevance and self efficacy. Both conditions were expected to show substantial motivational gains on these constructs (pre–post). The tested prediction is that the agent condition yields higher motivational gains.

(6) Does condition affect *learning gains*? To assess learning, the students were asked to complete the five formatting tasks from the tutorial without having it available for reference. Skills measures were taken before and after training. Both conditions were expected to show substantial learning gains (pre–post, and pre-retention). The tested prediction is that the agent condition yields higher learning gains.

## 6. Method

### 6.1. Participants

The participants were 50 students from the fifth-grade and sixth-grade classrooms of an elementary school. The group consisted of 29 boys and 21 girls, with a mean age of 11 years and 3 months. Participants could perform basic tasks in Word, but generally failed in handling the formatting options discussed in the manual. Participants were randomly assigned to conditions, with stratification for gender and classroom. One student was dropped from the analyses because he fell ill during training.

### 6.2. Instruments

#### 6.2.1. Design of the tutorial

The control and experimental tutorial are identical apart from the presence of the agent. The tutorial is presented as a printed booklet with a leaflet binder. The tutorial discusses several text formatting options for Microsoft Word. Students from the target audience generally do not know how to use these options to improve the presentation of their school reports. Chapter one (7 pages, questionnaires excluded) revolves around adjusting the margins for a complete text. Chapter two (10 pages) discusses the formatting of paragraphs, citations and enumerations. Both chapters instruct students to use the ruler for these formatting options. Topics are presented in a simple-to-complex order. This design principle is most clearly evident in the positioning of the (simpler) instructions for handling the right margin before the (more complex) instructions for handling the left margin.

Pilot tests revealed that a training time of 70 min should enable all students to complete these chapters. The tutorial also contains a third chapter that deals with the construction of an automatically generated table of content. The chapter was added to accommodate fast students. It did not involve the use of the ruler and therefore was expected not to interfere with the findings for the first two chapters on which we restrict all discussion.

The general design of the manual is based on the minimalist approach, meaning that: (a) the manual is action-oriented, (b) tasks are anchored in the domain of formatting a school report, (c) there is information to prevent errors, and (d) all major information types (e.g., goals, actions, feedback, error) have their own, unique presentation format (see van der Meij & Carroll, 1998). Three specific design features should be further mentioned.

One, each chapter begins with a 1-page overview in which the main concept is defined, and a relevance organizer describes and depicts 'before' and 'after' states for the formatting options discussed in the chapter. The organizers have a motivating purpose; they aim to 'sell' the goal (van der Meij & Gellevij, 2004).

Two, in line with the findings from Moreno and Mayer (2004), a more personal style of communication is used in the tutorial than is usually the case. Actions are described in personalized form (''You click ...'') rather than in the more customary commanding and non-personalized style (''Click ....''). The personalization also aligns well with the messages from the agent in the experimental tutorial.

Three, each chapter is rounded off with an exercise. These exercises should help augment the learning effect of the task instructions. The exercises in the tutorial extend the one-shot experience of following task instructions. Each exercise includes the same tasks as in the instructions. That is, the exercise in chapter 1 asks the students to modify the left and right margins of a text, and the exercise in chapter 2 asks students to format the paragraphs, a citation and an enumeration in a text. Exercises employ different texts than used in the instruction, and they give information about related goals, but not about the specific means for achieving these. Thus, they stimulate the user to engage in further practice that is important for learning. Minimalism advocates the use of 'on your own' sections for this purpose, but empirical studies show that exercises induce more compliance and also have a stronger impact on learning (Glasbeek, 2004; Wiedenbeck, Zavala, & Nawyn, 2000).

#### 6.2.2. Design of the agent

Max first makes his appearance in the introduction, where he presents himself as someone who is in a comparable situation to the reader: a student from the upper grades of elementary school who needs to hand in nicely formatted reports, which he finds difficult to achieve (see Fig. 3).

Thereafter, Max always appears at every introduction of a new task, shows up about once after each three action steps, and in all exercises. In all, Max is displayed 20 times in the tutorial. Following Brave, Nass, and Hutchinson (2005), Max is always presented in writing as well as visually in the form of a photo. Except for the introduction, the visual is always a close-up of his face (see Fig. 4).

To make Max a convincing model, and also to strengthen positive or moderate negative motivational states of the students, his messages also convey a broad range of emotions and feelings (Baylor & Kim, 2004; Clore & Palmer, 2009; Dehn & Van Mulken, 2000). This was done through the additional inclusion of motivational words and comments. Thus, Max occasionally refers to his affect states and frames of minds with inserted adverbs and adjectives (e.g., ''Cool'', ''Careful'', ''... awful ...'') and in separate sentences (e.g., 'What a nuisance', 'I am curious') that are occasionally strengthened with exclamation marks ('Wonderful !', 'Did it right !!!').

To find the proper tone of voice, expression, and word choice for Max we conducted a pilot study in which students were asked to think-aloud. These protocols and an inspection of popular youth magazines assisted us in making Max a life-like agent who might

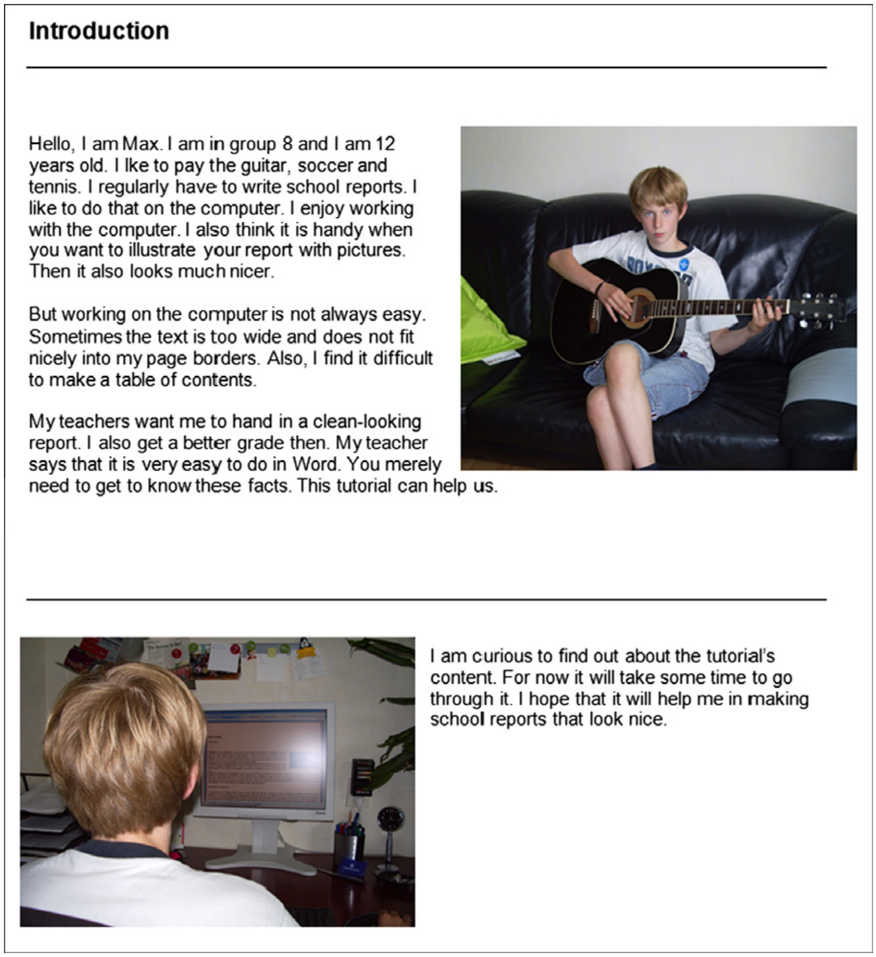



**Fig. 3.** The introduction of the motivating agent Max.

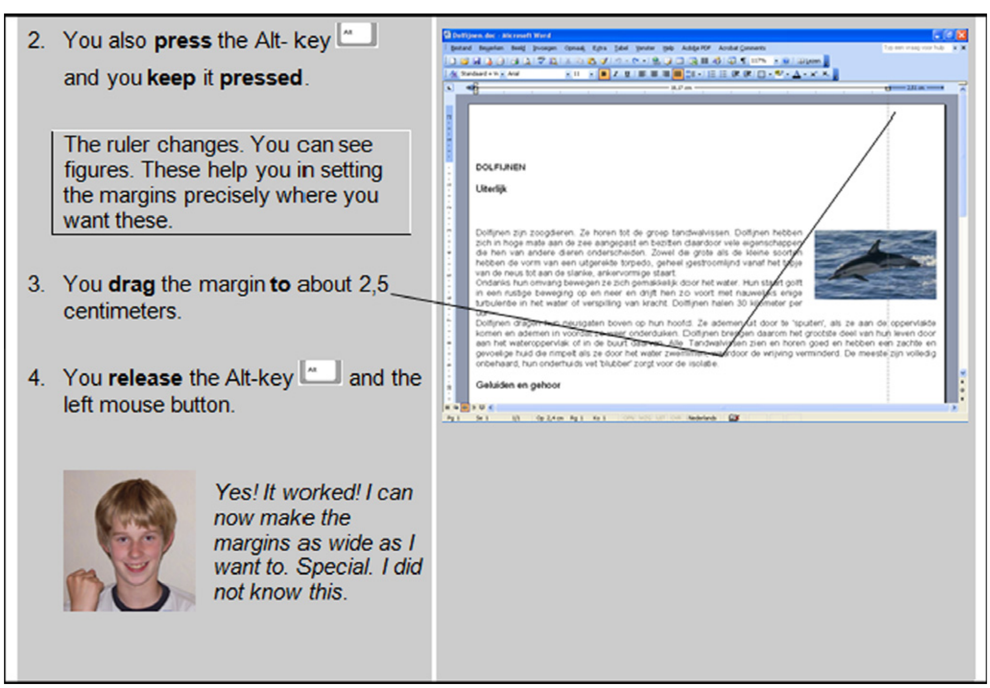
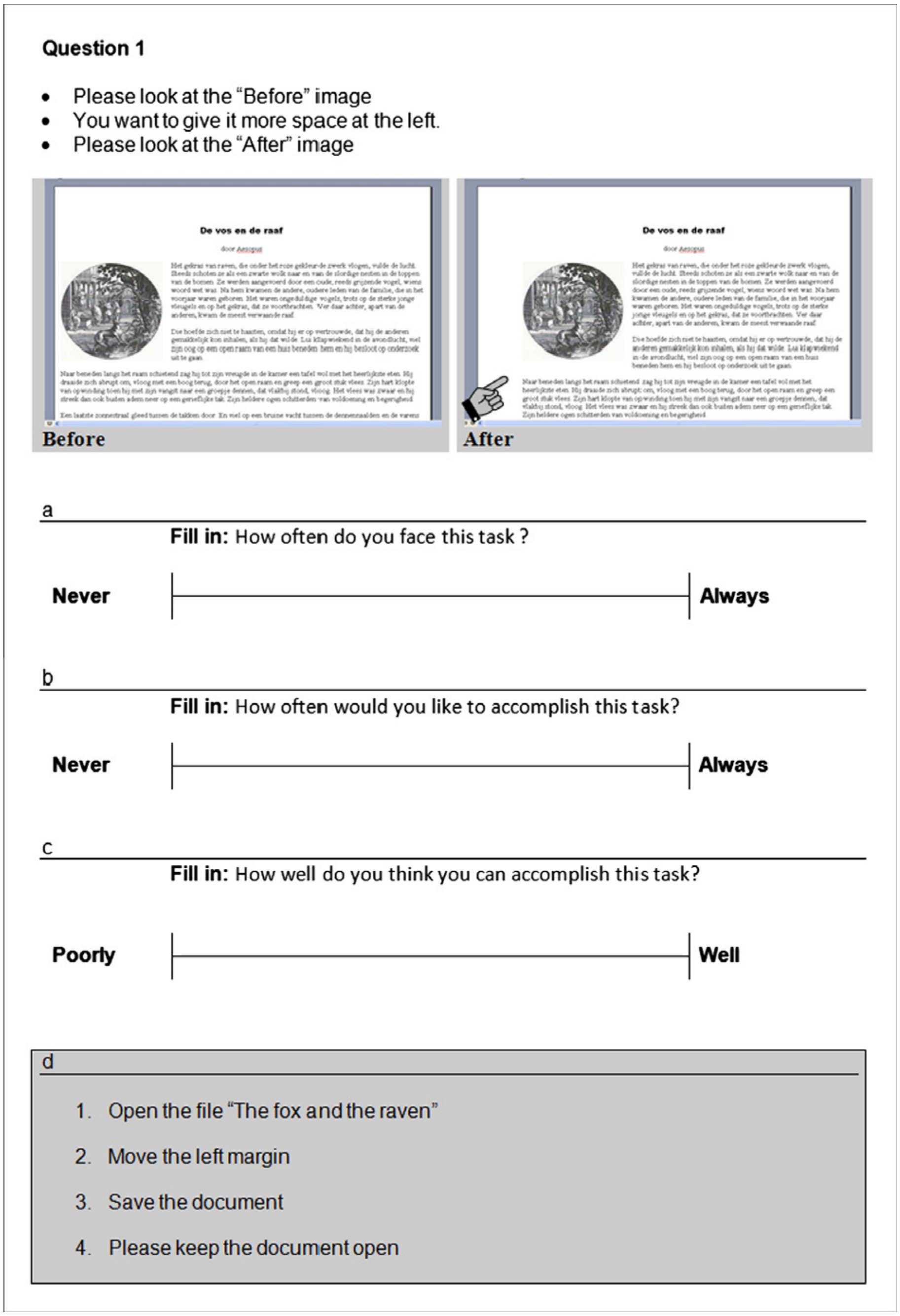



**Fig. 4.** Task instructions supplemented with an image of and message from Max.


**Question 1**

- Please look at the "Before" image
- You want to give it more space at the left.
- Please look at the "After" image

De vos en de raaf

door Aesopus

**Before**

De vos en de raaf

door Aesopus

**After**

a

**Fill in:** How often do you face this task ?

**Never** **Always**

b

**Fill in:** How often would you like to accomplish this task?

**Never** **Always**

c

**Fill in:** How well do you think you can accomplish this task?

**Poorly** **Well**

d

1. Open the file "The fox and the raven"
2. Move the left margin
3. Save the document
4. Please keep the document open

**Fig. 5.** An item from the Specific Initial Motivation Questionnaire and Pretest (section d).

appeal to the target audience. Fig. 5 illustrates how Max is presented alongside the task instructions in the tutorial.

The presence and nature of Max's task relevance messages vary depending on the task or situation. His self-efficacy messages show a growing belief in capacity within and across tasks; Max expresses stronger beliefs in competence and positive outcomes as task execution progresses.

*6.2.3. Measurement instruments before training*

*6.2.3.1. General Initial Motivation Questionnaire (GIMQ).* The GIMQ is an adapted version of the QCM from Rheinberg, Vollmeyer, and Burns (2001), which measures the students' appraisals of Probability of success, Anxiety, Interest, and Challenge for a future undertaking (also see Vollmeyer & Rheinberg, 1999). First, the student receives a broad description of the assignment they will be facing and the support therein. Thus, they are told that they are going to learn new formatting options from Microsoft Word, and that they will have to do this on their own using only a tutorial for support. Next, the GIMQ asks the students for their self-appraisals for this undertaking. Probability of success refers to the student's belief that he or she can succeed (e.g., "I think I can successfully complete this task"). Anxiety represents the negative incentive of failure (e.g., "When I think of this task I am worried"). Interest refers to positive affect and evaluations (e.g., "I like it that you learn new things with this task"). Challenge assesses whether the assignment is perceived as an achievement situation in which the student wants to have success (e.g., "I'm really going to try as hard as I can on this task"). All constructs are measured with five questions. [For Interest one question was dropped from the analyses after we found it negatively affected reliability.] Answers are given on a 7-point Likert-scale with one end of the extreme described as "Does not fit me" and the other as "Fits me". Reliabilities for the four constructs were satisfactory: Probability of success ($\alpha = 0.75$), Anxiety ($\alpha = 0.81$), Interest ($\alpha = 0.78$), and Challenge ($\alpha = 0.71$).

*6.2.3.2. Specific Initial Motivation Questionnaire (SIMQ).* The SIMQ contains a set of three questions on Experience, Task relevance, and Self-efficacy (see questions a, b and c in Fig. 5) to which students respond in connection with each of the five main tasks in the tutorial. Students see the task, respond to the initial motivation questions, and then attempt to complete the task (pretest, question d). The Experience question asks, "How often do you face this problem?" The Task relevance question asks, "How often would you like to solve this problem?" The Self-efficacy question asks, "How well do you think you can solve this problem?" Answers are given on an open line in which only the end points are labeled. Students answer each question by making a cross at what they decide is the right place on this line. The student's score is determined by using a transparent template that divides the line into equal sections, by which the number (0–7) corresponding to the position of the student's cross can be gauged.

*6.2.3.3. Pretest.* The pretest is presented as a separate question in the SIMQ (see the grey box in Fig. 5). The question invites students to open a file on the computer and to try to complete the given task in Word. The Pretest awards a score of 0 points for each task that the student cannot accomplish or for which an incorrect method is used. A good solution (and method) yields a score of 1. The maximum score for the Pretest is 5. Scores are presented as a percentage.

*6.2.4. Measurement instruments during training*

*6.2.4.1. Mood indicator (MI).* Mood states are measured with a set of five pictograms plus descriptor (see Fig. 6) from which the student is asked to select the one that best fits his or her current emotional state (see Read, 2008). Pictograms (a smiley) and text represent the following moods: happy, certain, neutral, uncertain and angry. Mood is measured immediately after students have completed an exercise, which occurs at the end of each chapter (hence: twice). The MI was presented together with the FWQ on a separate page that was inserted into the tutorial booklet.

In line with the research of Vollmeyer and Rheinberg (1999, 2006) the analysis of mood concentrated on its valence (i.e., positive, neutral or negative). Happy and certain are scored as signals of a positive mood; uncertain and angry are signals of a negative mood. Scores are presented as a percentage. Thus, a score of, say 50% for Positive mood indicates that the student selected the happy or certain smiley at one of the two measurement points for mood.

*6.2.4.2. Flow and Worry Questionnaire (FWQ).* Flow is measured with an adapted version of the Flow Short Scale (FKS) from Rheinberg, Vollmeyer, and Engeser (2003). Eleven items represent the six characteristics of Flow: (1) challenge-skill balance, (2) merging of actions and awareness, (3) unambiguous feedback, (4) concentration on the task at hand, (5) time transformation, and (6) fluency of action. Examples are "I felt pleasantly challenged", "I had the feeling that I had everything under control", "I was completely lost in thought", and "The right thoughts came without effort". Three items measure Worries, an anxiety-evoking, negative motivational state (e.g., "I am worried about failing"). Answers are given on a 7-point Likert-scale. Flow and Worries are measured immediately after a student has completed the instructions for one of the main tasks in the tutorial. Thus, these constructs were measured five times during training. Reliability scores for Flow and Worries were satisfactory at all points (Flow: $\alpha = 0.74, 0.72, 0.83, 0.78, 0.81$; Worries: $\alpha = 0.68, 0.66, 0.67, 0.60, 0.65$).

*6.2.4.3. Task Performance Indicator (TPI).* The TPI indicates how successful students were during training. A distinction is made between success in following task instructions (5) and on exercises (2). Students' performances on the training tasks are analyzed by consulting their practice files. The analysis is the same as for the pretest. Scores are presented as a percentage.

*6.2.4.4. Time.* Training time is the total time for processing the first two chapters from the tutorial. Time records are provided by the students. Each time the tutorial presents a Questionnaire, the last question on that page asks students to fill in the time indicated on their computer clock.

*6.2.5. Measurement instruments after training*

*6.2.5.1. Specific Final Motivation Questionnaire (SFMQ).* The SFMQ contains four task-specific questions about task relevance ("I think a nice margin is important") and four task-specific questions about self-efficacy (e.g., "I can now indent the first line of a paragraph"). Answers are given on a 7-point Likert-scale. Reliability for each construct was good (Relevance: $\alpha = 0.86$; Confidence: $\alpha = 0.88$). [Note: the SFMQ also included questions about attributions for success. Because the reliability for this construct was too low, these outcomes are not reported.]

*6.2.5.2. Agent Appraisal Questionnaire (AAQ).* The first question in the AAQ assesses the student's frequency of reading the agent's messages (i.e., "How often did you read Max's comments?"). Then there are 11 questions about the credibility of the agent (e.g., "Max said what I was thinking of", and "I felt just as Max did"). Answers are scored by drawing a cross somewhere on a continuous line for which the two extreme positions are described (i.e., 'completely agree' and 'completely disagree'). The student's score is measured by looking at the corresponding number on a transparent template that divides the line into ten equal sections. The maximum score is 10. Reliability was good (Cronbach $\alpha = 0.86$).

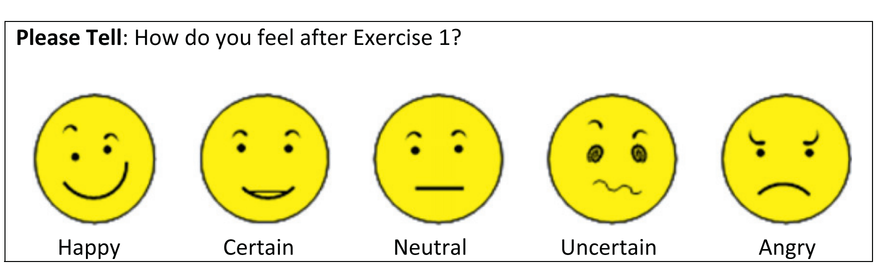



**Fig. 6.** The questionnaire for Mood.

*6.2.5.3. Posttest (Post) and Retention test (Ret).* These tests measure the students' skill in again accomplishing the five tasks discussed in the tutorial. They are measured hands-on and without the tutorial available. Students receive a test file whose format they must modify. Test instructions describe the specific changes that need to be made and a screen shot shows the intended end result. The Posttest and Retention test have a different surface appearance, but share the same underlying structure. Scoring is the same as for the Pretest.

### 6.3. Procedure

Two weeks before training students completed the General Initial Motivation Questionnaire. One week before training they completed the Specific Initial Motivation Questionnaire and Pretest. Training sessions took place in a separate room with small groups of about 10 students at a time. Students were each given a copy of the tutorial booklet and each worked at a computer on which to complete the tasks and exercises. Students were instructed to "read through the entire tutorial booklet, which helps you to format reports". They were to work on their own and call on the experimenter for help only when stuck. While students engaged in the training, they answered the questions on Mood, Flow, Worry and Time on the inserted pages in the tutorial booklet. Students were given a maximum of 70 min to work with the tutorial. Then a short break of 10 min followed. After the break, all students answered the Specific Final Motivation Questionnaire, and students in the experimental group also answered the Agent Appraisal Questionnaire. Next, students were tested for skills development with the Posttest. Three weeks later they took the Retention test. Students were allowed a maximum time of 25 min for completion of each test.

### 6.4. Data analyses

The study uses a pretest–posttest design with a control and experimental (agent) condition. Gain scores for task relevance and self-efficacy are computed by subtracting the SIMQ scores from the SFMQ values (after–before). Gain scores for posttest skill and retention test skill are computed by subtracting the pretest score from the posttest score (post–pre) or retention test score (ret–pre), respectively.

Before the main research questions were examined, a check on differences between conditions in initial motivation was done. This revealed that despite the random distribution there were statistically significant differences between conditions on Interest and Challenge from the GIMQ and on Experience from the SIMQ (with the control group scoring higher). These measures were therefore entered as covariates in the analyses of variance conducted, and we report adjusted means and standard error. As the measure for training time for chapter 1 produced a skewed distribution, the variance was corrected by applying a logarithmic transformation. In all analyses, the significance level was set at an $\alpha$ of 0.05 (two-tailed). Trends $(0.10 < p > 0.05)$ in the predicted direction are mentioned. Cohen's (1988) $d$-statistic is reported for effect size. These tend to be qualified as small for $d = 0.2$, medium for $d = 0.5$ and large for $d = 0.8$.

## 7. Results

### 7.1. Does condition affect training time?

Students in the agent condition completed training slightly faster than students in the control condition (see Table 3). However, conditions did not differ significantly for total training time, $F(1,44) = 1.50$, n.s. Detailed analyses revealed that the students in the agent condition took less time to complete the first chapter than students in the control condition, $F(1,44) = 4.29$, $p = 0.04$, $d = 0.64$. There was no difference between conditions for the second chapter, $F < 1$.

### 7.2. Does condition affect motivational states during training?

The data show that positive moods were found most frequently, followed by neutral moods. Negative moods were rather scarce (see Table 4). Students in the control condition more often indicated being in a positive mood, but this difference was not statistically significant, $F(1,44) = 1.03$, n.s. There was also no statistically significant difference between conditions for neutral mood, $F < 1$, or for negative mood, $F(1,44) = 1.25$, n.s.

The scores for flow were well above the scale midpoint of 3.5, indicating that the students were absorbed in their activities (see Table 5). Conditions differed marginally. Students who worked with the agent tutorial tended to give higher appraisals for flow than students in the control condition, $F(1,44) = 3.63$, $p = 0.06$. Exploratory analyses of the relationships between the students' flow experience and their final performances showed that there was a significant correlation with the posttest scores ($r = 0.41$, $p < 0.00$) as well as with the scores on the retention test ($r = 0.52$, $p < 0.00$).

The scores for worries were well above the scale midpoint of 3.5, which indicates the presence of relatively high anxiety levels (see Table 5). Conditions did not differ for worries, $F < 1$.

### 7.3. Does condition affect task performance during training?

The scores for training tasks indicate that students completed the majority of training tasks successfully when they had the task instructions available (see Table 6). There is a considerable drop in performance success for the exercises. The students must try to solve these on their own, but they can look back at the instructions in the tutorial. There is a statistically significant decline in performance in both the control condition, $t(24) = 7.40$, $p = 0.00$, and in the agent condition, $t(23) = 6.11$, $p = 0.00$.

Statistically significant differences were found between conditions on task performance during training. Students who worked with the agent tutorial were more successful on tasks completed with instructions than students in the control condition, $F(1,44) = 5.00$, $p = 0.03$, $d = 0.68$. These students also did better on the exercises, $F(1,44) = 5.31$, $p = 0.03$, $d = 0.71$.

### 7.4. Does gender affect agent reading frequency and credibility?

The ratings for reading frequency indicated that students said they nearly always read the messages from the agent (see Table 7).

**Table 3**
Adjusted means and standard error for training times by condition.

| Condition | Training time chapter 1 | | Training time chapter 2 | | Total training time | |
|---|---|---|---|---|---|---|
| | *M* | SE | *M* | SE | *M* | SE |
| Control ($n = 25$) | 25.60 | 1.75 | 27.73 | 1.85 | 53.33 | 2.61 |
| Agent ($n = 24$) | 19.96 | 1.79 | 28.49 | 1.89 | 48.45 | 2.67 |

**Table 4**
Adjusted means (percentages) and standard error for moods by condition.

| Condition | Mood | | | | | |
|---|---|---|---|---|---|---|
| | Positive | | Neutral | | Negative | |
| | *M* | SE | *M* | SE | *M* | SE |
| Control ($n = 25$) | 60.7 | 8.8 | 30.1 | 7.6 | 9.2 | 6.2 |
| Agent ($n = 24$) | 47.1 | 9.0 | 33.2 | 7.8 | 19.6 | 6.3 |

**Table 5**
Adjusted means and standard error for Flow and Worries by Condition.

| Condition | Flow | | Worries | |
|---|---|---|---|---|
| | *M* | SE | *M* | SE |
| Control ($n = 25$) | 4.89 | 0.15 | 4.91 | 0.24 |
| Agent ($n = 24$) | 5.31 | 0.15 | 5.23 | 0.24 |

Scale maximum is 7. A higher score indicates higher appraisal.

**Table 6**
Adjusted means (percentages) and standard error for tasks with instructions and exercises by condition.

| Condition | Tasks with instructions | | Exercises | | Cohen's *d*[*] |
|---|---|---|---|---|---|
| | *M* | SE | *M* | SE | |
| Control ($n = 25$) | 67.2 | 0.5 | 37.6 | 0.5 | 1.18 |
| Agent ($n = 24$) | 84.1 | 0.5 | 55.2 | 0.5 | 1.45 |

[*] Effect sizes within condition were computed using unadjusted scores.

**Table 7**
Adjusted means and standard error for reading frequency and credibility rating of the agent by gender.

| Condition | Reading frequency | | Credibility | |
|---|---|---|---|---|
| | *M* | SE | *M* | SE |
| Girls ($n = 11$) | 9.42 | 0.72 | 4.92 | 0.46 |
| Boys ($n = 13$) | 8.18 | 0.65 | 5.73 | 0.44 |

Scale maximum is 10. A higher score indicates higher appraisal.

**Table 8**
Adjusted means and standard error for task relevance by condition.

| Condition | Task relevance before | | Task relevance after | | Cohen's *d*[*] |
|---|---|---|---|---|---|
| | *M* | SE | *M* | SE | |
| Control ($n = 24$) | 3.21 | 0.36 | 5.55 | 0.17 | 2.71 |
| Agent ($n = 24$) | 2.84 | 0.37 | 5.94 | 0.17 | 3.11 |

Scale maximum is 7. A higher score indicates higher appraisal.
[*] Effect sizes within condition were computed using unadjusted scores.

There was no difference for gender, $F(1,19) = 1.44$, n.s. One boy who indicated that he never read any of Max's messages was excluded in the assessments for the agent's credibility ratings. Boys gave higher credibility ratings for the agent than girls, but the difference was not statistically significant, $F(1,18) = 2.36$, n.s.

### 7.5. Does condition affect motivational gain?

The gain scores for task relevance indicated that students in both conditions raised their valuation of the training tasks substantially $t_{control}$ (23) = 4.91, $p = 0.00$; $t_{agent}$ (23) = 7.15, $p = 0.00$ (see Table 8). Conditions differed marginally for this gain. Students who worked with the agent tutorial tended to increase their task relevance appraisals more than students in the control condition, $F(1,43) = 3.17$, $p = 0.08$.

**Table 9**
Adjusted means and standard error for self-efficacy by condition.

| Condition | Self-efficacy before | | Self-efficacy after | | Cohen's *d*[*] |
|---|---|---|---|---|---|
| | *M* | SE | *M* | SE | |
| Control ($n = 24$) | 4.64 | 0.36 | 5.84 | 0.20 | 2.75 |
| Agent ($n = 24$) | 4.37 | 0.37 | 6.27 | 0.20 | 2.60 |

Scale maximum is 7. A higher score indicates higher appraisal.
[*] Effect sizes within condition were computed using unadjusted scores.

The gain scores for self-efficacy also indicated that students raised their confidence ratings substantially $t_{control}$ (23) = 3.88, $p = 0.00$; $t_{agent}$ (23) = 4.06, $p = 0.00$ (see Table 9). Conditions differed marginally for this gain too. Students who worked with the agent tutorial tended to increase their self-efficacy more than students in the control condition, $F(1,43) = 2.99$, $p = 0.09$.

### 7.6. Does condition affect learning gain?

The gain scores for both the pretest–posttest and the pretest–retention test comparisons indicated that students in both conditions became substantially more skilled $t_{control;\ post\text{-}pre}$ (24) = 6.96, $p = 0.00$; $t_{agent;\ post\text{-}pre}$ (24) = 8.83, $p = 0.00$; $t_{control;\ ret\text{-}pre}$ (24) = 6.95, $p = 0.00$; $t_{agent;\ ret\text{-}pre}$ (24) = 9.13, $p = 0.00$ (see Table 10). In addition, only a small difference was found between the scores on the posttest and retention test within conditions (i.e., for the control condition: $t(24) = 1.05$, n.s.; for the agent condition: $t(23) < 1$, n.s.).

Conditions differed significantly in learning gains. Students who worked with the agent tutorial raised their skill level (post–pre) more than students in the control condition, $F(1,44) = 5.39$, $p = 0.03$, $d = 0.71$. These students did also significantly better on gain score for the pretest–retention test difference, $F(1,44) = 12.89$, $p = 0.00$, $d = 1.09$.

Exploratory correlational analyses of the relations between perceptions of task relevance and self-efficacy on the one hand and performance scores on the other revealed the following outcomes. Before the training the correlations between initial perceptions of task relevance and self-efficacy and initial skill were $r = 0.29$, $p = 0.04$, and $r = 0.26$, $p = 0.07$, respectively. After training the correlations between the students' final appraisals for these motivational constructs and their posttest scores were $r = 0.26$, $p = 0.08$, and $r = 0.31$, $p = 0.03$, respectively. For the retention test these correlations were $r = 0.44$, $p < 0.00$, and $r = 0.45$, $p < 0.00$.

## 8. Discussion and conclusion

The data indicated that both tutorials substantially influenced student motivation and learning. Significant and considerable gains in perceptions of task relevance and self-efficacy were found in the control as well as the agent conditions. Because these motivational constructs can strongly influence students' activities during instruction and the outcomes thereof (e.g., Bandura, 2012; Eccles & Wigfield, 2002), this is an important finding. In this respect it is interesting to note that this study also revealed that the relations between these motivational constructs and the students' skills were (near) significant both at the start and end of the experiment. Students in both conditions also made significant and considerable progress in their skills development on the main tasks of the tutorial.

A further noteworthy finding is that results for both conditions indicated a significant difference between success on tasks completed with instructions and exercises during training. Performance success on the former was much higher than on the

**Table 10**
Adjusted means (in percentages) and standard error for pretest, posttest and retention by condition.

| Condition | Pretest | | Posttest | | Retention test | |
|---|---|---|---|---|---|---|
| | *M* | SE | *M* | SE | *M* | SE |
| Control (*n* = 25) | 10.1 | 3.1 | 45.3 | 6.9 | 38.6 | 5.7 |
| Agent (*n* = 24) | 3.6 | 3.1 | 60.6 | 7.1 | 61.6 | 5.8 |

* Effect sizes for learning gains (unadjusted scores) within the control condition were: *d* (pre–post) = 1.50; *d* (pre–ret) = 1.50.
* Effect sizes for learning gains (unadjusted scores) within the agent condition were: *d* (pre–post) = 2.31; *d* (pre–ret) = 2.53.

latter. Task instructions prescribe all of the step-by-step actions that need to be performed to achieve task completion. In contrast, exercises merely provide a problem description and illustrate the end goal that students should try to achieve. Even with such a difference in support, the 30% drop is remarkable, because the exercises dealt with problems similar to those discussed in the task instructions and students could look back to these instructions for help with a solution.

The experiment supported the general prediction that students would benefit from the added presence of a motivating agent. The study gives a positive answer to the question whether an agent can raise motivation and increase learning when students study a software tutorial. But not all predictions were confirmed and not all findings were equally strong.

On several motivational measures the expected positive influence of the agent emerged, but differences with the non-agent condition were only marginally statistically significant. Students who worked with the agent tutorial tended to have somewhat higher scores for flow during training and they tended to realize somewhat higher motivational gains for appraisals of task relevance and self-efficacy. These finding are important signals of the influence of the agent on students' motivational development during their learning experience.

The findings for flow signal that the added presence of the agent did not cause cognitive overload, but rather that it positively affected the students' motivational state. Flow refers to complete absorption in an activity; it reflects the level of concentrated effort that the student experiences. One could also say that students who are in a flow experience optimal cognitive load. There is a balance between challenge level and capacity. Students are neither bored by task requirements that are too low, nor are they taxed by task demands that are too high for them.

The cognitive-motivational process model from Vollmeyer and Rheinberg (1999, 2000, 2006) assumes that initial motivation affects strategies and motivation during training which then influence performance. The students' functional state during training is considered one of the mediators for this influence. Vollmeyer and Imhof (2007) assessed flow as a measure of functional state in an empirical study where university students learned to use SPSS, a software program for conducting statistical analyses, and found a significant relation between flow and performance. The correlational outcomes found in the present study, even with the more distant performance on the retention test, align well with this finding.

While students raised their appraisals for task relevance and self-efficacy more in the agent than in the control condition, the data showed that these differences were not statistically significant. An explanation for the fact that only a trend was found is that it may not have been easy to achieve *more* motivational gains given that the control tutorial already included several important motivational measures. The inclusion of exercises fits the recommendation from Keller's ARCS-model (1987, 2010) to offer varied learning experiences that provide students with success opportunities and enhance their feelings of personal control. Another design feature that seems particularly relevant for students' self-efficacy beliefs is the simple-to-complex sequencing of topics. Two general design strategies that seem especially relevant for task relevance are the presence of relevance organizers and the personal form of address. Relevance organizers inform students about the initial and final state of the main tasks in the tutorial. Besides being brief, the organizers also visualize the before and after states. These concrete displays were expected to strongly appeal to the students' sense of task value for the formatting option that were discussed. The research from Moreno and Mayer (2004) indicated that students' perceived task relevance might benefit from the more personal form of address that was adopted in the tutorial.

On the skills measures the positive influence of the agent was unequivocal. In the agent condition students improved their skill levels significantly more than in the control condition. Student in both conditions started with (very) low skill levels, but students in the agent condition did much better even for task performance during training. These students scored almost 20% higher for the tasks completed with instructions as well as the exercises. A striking difference between conditions was also observed for the retention test, on which students in the agent condition scored almost 25% higher than students in the control condition. We have no ready explanation for why students in the agent condition were better able to maintain their skill level over time (i.e., from posttest to retention test) while that of the students in the control conditions dropped. It is possible that the higher scores on the exercises signal better learning during training and therefore partly account for this difference.

In conclusion, the question arises whether research on APAs can benefit from the insights offered in this study. We believe it can. In our view, the motivational design strategies that were employed are just as meaningful for designing the voice of an APA as they were for crafting the written messages of our agent.

## Acknowledgements

The author wishes to thank Martijn van de Lagemaat for his help in conducting this study. I also would like to thank Emily Fox for her valuable editorial assistance.

## References

André, E., Müller, J., & Rist, T. (1996). The PPP persona: A multipurpose animated presentation agent. In *Paper presented at the workshop on advanced visual interfaces*, Gubbio, Italy.

Arroyo, I., Woolf, B. P., Royer, J. M., & Tai, M. (2009). Affective gendered learning companions. In *Paper presented at the proceedings of the 2009 conference on artificial intelligence in education: Building learning systems that care: From knowledge representation to affective modelling*.

Arroyo, I., Woolf, B. P., Cooper, D. G., Burleson, W., & Muldner, K. (2011, 6–8 July). The impact of Animated Pedagogical Agents on girls' and boys' emotions, attitudes, behaviors and learning. In *Paper presented at the 11th IEEE International Conference on Advanced, Learning Technologies* (ICALT).

Arroyo, I., Woolf, B. P., Royer, J. M., Tai, M., Muldner, K., Burleson, W., et al. (2010). *Gender matters: The impact of animated agents on students'affect, behavior and learning*. Amherst: Computer Science Department, UMASS.

Atkinson, R., Mayer, R., & Merrill, M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent?s voice. *Contemporary Educational Psychology, 30*(1), 117–139. http://dx.doi.org/10.1016/j.cedpsych.2004.07.001.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman and Company.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38*(1), 9–44. http://dx.doi.org/10.1177/0149206311410606.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM, 37*(7), 122–125.

Baylor, A. L., Shen, E., & Warren, D. (2004). Supporting learners with math anxiety: The impact of pedagogical agent emotional and motivational support. In: *Paper presented at the Workshop on Social and Emotional Intelligence in Learning Environments.* International Conference on Intelligent Tutoring Systems (ITS), Maceió, Brazil.

Baylor, A. L, & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Proceedings of 7th International Conference on Intelligent Tutoring Systems (ITS)*, Lecture Notes in Computer Science (pp. 592–603), Berlin: Springer.

Baylor, A. L., & Ryu, J. (2003). The effects of image and animation in enhancing pedagogical agent persona. *Journal of Educational Computing Research, 28*(4), 373–394.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*(2), 199–231. http://dx.doi.org/10.1111/j.1464-0597.2005.00205.x.

Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies, 62*(2), 161–178. http://dx.doi.org/10.1016/j.ijhcs.2004.11.002.

Britner, S. L. (2008). Motivation in high school science students: A comparison of gender differences in life, physical, and earth science classes. *Journal of Research in Science Teaching, 45*(8), 955–970. http://dx.doi.org/10.1002/tea.20249.

Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., et al. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology, 100*(3), 525–534. http://dx.doi.org/10.1037/0022-0663.100.3.525.

Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM, 43*(4), 70–78.

Choi, S., & Clark, R. E. (2006). Cognitive and affective benefits of an Animated Pedagogical Agent for learning english as a second language. *Journal of Educational Computing Research, 34*(4), 441–466.

Clark, R. E., & Choi, S. (2005). Five design principles for experiments on the effects of Animated Pedagogical Agents. *Journal of Educational Computing Research, 32*(3), 209–225.

Clore, G. L., & Palmer, J. (2009). Affective guidance of intelligent agents: How emotion controls cognition. *Cognitive Systems Research, 10*(1), 21–30.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience*. NY: Harper Perennial.

Dehn, D. M., & Van Mulken, S. (2000). Impact of animated interface agents: a review of empirical research. *International Journal of Human Computer Studies, 52*(1), 1–22.

Dweck, C. S. (2007). The perils and promises of praise. *Educational Leadership, 65*(2), 32–39.

Eccles, J. S., & Wigfield, A. (2002). Motivation beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132. http://dx.doi.org/10.1146/annurev.psych.53.100901.135153.

Feng, S. L., & Tuan, H. L. (2005). Using ARCS model to promote 11th graders' motivation and achievement in learning about acids and bases. *International Journal of Science and Mathematics Education, 3*(3), 463–484.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.

Glasbeek, H. (2004). Solving problems on your own: how do exercises in tutorials interact with software learners' level of goal-orientedness? *Professional Communication, IEEE Transactions on, 47*(1), 44–53. http://dx.doi.org/10.1109/tpc.2004.824285.

Gookin, D. (1994). *Word for Windows 6 for Dummies*. San Mateo, CA: IDG books Worldwide.

Graesser, A., & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist, 45*(4), 234–244. http://dx.doi.org/10(1080/00461520), 2010, 515933.

Haake, M. (2009). *Embodied pedagogical agents: From visual impact to pedagogical implications*. Department of Design Sciences, Lund University, Sweden.

Huett, J. B., Kalinowski, K. E., Moller, L., & Huett, K. C. (2008). Improving the motivation and retention of online students through the use of ARCS-based e-mails. *American Journal of Distance Education, 22*(3), 159–176. http://dx.doi.org/10.1080/08923640802224451.

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development, 10*(3), 2–10.

Keller, J. M. (1999). Using the ARCS motivational process in computer-based instruction and distance education. *New Directions for Teaching and Learning, 78*, 37–47.

Keller, J. M. (2010). *Motivational design for learning and performance. The ARCS model approach*. New York, NY: Springer Verlag.

Keller, J. M., & Kopp, T. W. (1987). An application of the ARCS model of motivational design. In C. M. Reigeluth (Ed.), *Instructional theories in action: Lessons illustrating selected theories and models* (pp. 289–320). New York, NY: Erlbaum.

Keller, J. M., & Suzuki, K. (2004). Learner motivation and E-learning design: A multinationally validated process. *Journal of Educational Media, 29*(3), 229–239. http://dx.doi.org/10.1080/1358165042000283084.

Lau, S., & Roeser, R. W. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science. *Educational Assessment, 8*(2), 139–162. http://dx.doi.org/10.1207/s15326977ea0802_04.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: Affective impact of Animated Pedagogical Agents. In *Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems*.

Loorbach, N., Karreman, J., & Steehouder, M. (2007). Adding motivational elements to an instruction manual for seniors: Effects on usability and motivation. *Technical Communication, 54*(3), 343–358.

Loorbach, N., Steehouder, M., & Taal, E. (2006). The effects of motivational elements in user instructions. *Journal of Business and Technical Communication, 20*(2), 177–199.

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*(1), 165–173. http://dx.doi.org/10.1037/0022-0663.96.1.165.

Moreno, R. (2005). Multimedia learning with Animated Pedagogical Agents. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 507–523). New York, NY: Cambridge University Press.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with Animated Pedagogical Agents? *Cognition and Instruction, 192*(2), 177–213.

Moundridou, M., & Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning, 18*(3), 253–261.

Paiva, A., & Machado, I. (1998). Vincent, an autonomous pedagogical agent for on-the-job training. In *Paper presented at the Intelligent tutoring systems*.

Picard, R. W., & Klein, J. (2002). Computers that recognise and respond to user emotion: Theoretical and practical implications. *Interacting with Computers, 14*(2), 141–169.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education. Theory, research, and applications* (2nd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Read, J. C. (2008). Validating the fun toolkit: An instrument for measuring children's opinions of technology. *Cognition, Technology & Work, 10*, 119–128. http://dx.doi.org/10.1007/s10111-007-0069-9.

Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge University Press.

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktuller Motivation in Lern- und Leistungssituationen. *Diagnostica, 47*(2), 57–66. http://dx.doi.org/10.1026//0012-1924.47.2.57.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Die Erfassung des Flow-Erlebens [The assessment of flow]. In J. Stiensmeier-Pelster & F. Rheinberg (Eds.), *Diagnostik von Motivation and Selbstkonzept [Diagnosis of motivation and self-concept]* (pp. 261–279). Gottingen, Germany: Hogrefe.

Song, S. H., & Keller, J. M. (2001). Effectiveness of motivationally adaptive computer-assisted instruction on the dynamic aspects of motivation. *Educational Technology Research and Development, 49*(2), 5–22.

van der Meij, H. (2008). Designing for user cognition and affect in software instructions. *Learning and Instruction, 18*(1), 18–29. http://dx.doi.org/10.1016/j.learninstruc.2006.08.002.

van der Meij, H. (2012). Supporting children in improving their presentation of school reports. In M. Torrance, D. Alamargot, M. Castelló, R. Llull, F. Ganier, O. Kruse, A. Mangen, L. Tolchinsky & L. v. Waes (Eds.), *European research network on learning to write effectively* (pp. 145–148). Bingley, UK: Emerald Group Publishing.

van der Meij, H., & Carroll, J. M. (1998). Principles and heuristics for designing minimalist instruction. In J. M. Carroll (Ed.), *Minimalism beyond the Nurnberg funnel* (pp. 19–53). Cambridge, Mass: MIT Press.

van der Meij, H., & Gellevij, M. (2004). The four components of a procedure. *Professional Communication, IEEE Transactions on, 47*(1), 5–14. http://dx.doi.org/10.1109/tpc.2004.824292.

van der Meij, H., Op de Weegh, M. J., & Weber, I. H. M. (2009). Heeft een (papieren) medeleerling een meerwaarde in software handleidingen? *Pedagogische Studiën, 86*, 296–312.

Vollmeyer, R., & Imhof, M. (2007). Are there gender differences in computer performance? If so, can motivation explain them? *Zeitschrift für Pädagogische Psychologie, 21*(3), 251–261. http://dx.doi.org/10.1024/1010-0652.21.3.251.

Vollmeyer, R., & Rheinberg, F. (1999). Motivation and metacognition when learning a complex system. *European Journal of Psychology of Education, 14*(4), 541–554. http://dx.doi.org/10.1007/bf03172978.

Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction, 10*(4), 293–309. http://dx.doi.org/10.1016/s0959-4752(99)00031-6.

Vollmeyer, R., & Rheinberg, F. (2006). Motivational effects on self-regulated learning with different tasks. *Educational Psychology Review, 18*(3), 239–253.

Wiedenbeck, S., Zavala, J. A., & Nawyn, J. (2000). An activity-based analysis of hands-on practice methods. *Journal of Computer Assisted Learning, 16*(4), 358–365. http://dx.doi.org/10.1046/j.1365-2729.2000.00148.x.