Learning and Instruction 87 (2023) 101799

Contents lists available at ScienceDirect

Learning and Instruction

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

A sequence of learning processes in an intelligent tutoring system from topic-related appraisals to learning gains

Johann Chevalère[a,b,*], Hae Seon Yun[b,c], Anja Henke[a,b], Niels Pinkwart[b,c,d], Verena V. Hafner[b,c], Rebecca Lazarides[a,b]

[a] Department of Educational Sciences, University of Potsdam, Potsdam, Germany
[b] Science of Intelligence, Research Cluster of Excellence, Berlin, Germany
[c] Department of Computer Science, Humboldt-Universität zu Berlin, Berlin, Germany
[d] Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Berlin, Germany

ARTICLE INFO

Keywords:
Control-value theory
Intelligent tutoring system
Open-ended learning environment
Enjoyment
Boredom

ABSTRACT

Although intelligent tutoring systems (ITSs) are increasingly used, it is unclear which psychological processes precede students' learning gains. Using a pre- and posttest design, the present study examined a sequence of psychological processes informed by control value theory. We investigated (a) whether secondary school students' topic-related cognitive appraisals (value and control) affected their task-related affective (enjoyment and boredom) and cognitive (engagement and performance) outcomes while using the ITS and (b) whether task-related outcomes affected learning. Path analyses showed that students' topic-related interest, but not perceived utility, personal importance or self-efficacy, was associated with task-related enjoyment. In turn, enjoyment showed reciprocal effects on and of engagement and ongoing task performance, which predicted final performance and, ultimately, learning gains. The influence of boredom, in contrast, was minimal along this sequence. More generally, the findings highlight the difficulty of establishing a clear pattern of sequential causation derived from control value theory for the current ITS context, with evidence demonstrating the systematic influence of confounders accounting for the predicted relations among components. Despite these limitations, we identified key psychological processes involving the contribution of affective and cognitive processes to learning in the ITS context.

1. Introduction

Digital learning is being democratized worldwide and represents a major resource for the future of education. Advantages of digital learning are the flexibility of the learning settings and the enhancement of individualized and complex learning (Jacobson et al., 2017; Lazarides & Chevalère, 2021). However, it is unclear whether psychological processes that are known to influence learning progress in teacher-led, traditional large group settings also apply to digital learning settings (Loderer et al., 2020). An established theoretical framework that describes such psychological processes is Pekrun's control-value theory (CVT, Pekrun, 2006; Pekrun & Perry, 2014), which assumes a central role of human emotions specifically evoked during learning activities in driving the cognitive and motivational resources underpinning learning and achievement (Pekrun & Linnenbrink-Garcia, 2012; Pekrun & Perry, 2014). These theoretical assumptions have proven robust across a range of school subjects (Forsblom et al., 2022; Kögler & Göllner, 2018; Li & Wei, 2022; Mercan, 2020; Simonton et al., 2017) and student populations (Lichtenfeld et al., 2022; Mercan, 2020; Putwain et al., 2021). It is, however, of utmost importance to examine whether these theoretical tenets also apply to digital learning settings aimed at enabling students to learn autonomously, some of which are increasingly being developed, such as intelligent tutoring systems (ITSs) (Alkhatlan & Kalita, 2018). ITSs are computer programs that typically include an interface that communicates with the learner, a model of the learner (e.g. current knowledge level), a model of the domain knowledge (e.g., knowledge that needs to be conveyed and the difficulty level of specific tasks), and a model of the tutor (e.g. teaching methods tailored to specific needs), and tailor learning content and instructional feedback to the learner (Ma et al., 2014). Meta-analyses by Ma et al. (2014) and Kulik and Fletcher

* Corresponding author. Laboratoire de Psychologie Sociale et Cognitive (LAPSCO) UMR 6024, CNRS & Université Clermont-Auvergne, Clermont-Ferrand, France.
E-mail addresses: johann.chevalere@uca.fr (J. Chevalère), yunhaese@informatik.hu-berlin.de (H.S. Yun), anja.henke@uni-potsdam.de (A. Henke), pinkwart@hu-berlin.de (N. Pinkwart), hafner@informatik.hu-berlin.de (V.V. Hafner), rebecca.lazarides@uni-potsdam.de (R. Lazarides).

https://doi.org/10.1016/j.learninstruc.2023.101799
Received 9 August 2022; Received in revised form 28 February 2023; Accepted 31 May 2023
Available online 8 June 2023

(2016) on the effectiveness of ITSs showed that the use of an ITS was associated with greater achievement in comparison to teacher-led, large group instruction, non-ITS computer-based instruction, and textbooks. However, the benefits were generally lower for primary and secondary school students than for postsecondary school students. To better understand the suitability of ITSs for secondary education, we tested how the affective and cognitive processes described by CVT apply to an ITS context. In the following, we first introduce CVT and then review a set of studies that test its assumptions in ITSs.

1.1. Control value theory: achievement emotions and performance

Pekrun's CVT provides 'an integrative framework for analyzing the antecedents and effects of emotions experienced in achievement and academic contexts' (Pekrun, 2006). Achievement emotions encompass outcome-related emotions, referring to emotions experienced in relation to academic success and failure, and activity-related emotions, which are experienced during achievement activities (i.e., studying, attending classes, completing assignments, or taking exams) (Camacho-Morles et al., 2021; Pekrun, 2006). Activity-related emotions differ in terms of their valence (positive or negative, e.g., enjoyment vs boredom) and activation (activating vs. deactivating, e.g., anger vs. boredom; see Pekrun et al., 2007, for more details). CVT assumes that activity-related emotions directly predict achievement outcomes. Specifically, enjoyment and boredom play a pivotal role in increasing or reducing the allocation of cognitive resources available for tasks, which is why activity emotions relate to engagement and effort while completing tasks and the deployment of learning strategies ultimately affecting academic performance (Pekrun et al., 2007). Among the causal antecedents of activity emotions, the theory emphasizes learners' cognitive appraisals of the learning situation in terms of students' expectations and attributions of control and value (Pekrun & Perry, 2014). Appraisals of control refer to the perceived controllability of a given activity or topic (e.g., beliefs about one's ability to master a given topic), while appraisals of value refer to the subjective value ascribed to it (i.e., whether or not learners find a topic interesting, important or useful).

Within the sequence of CVT components from antecedents to achievement, mediational processes are assumed to take effect (Pekrun et al., 2002), some of which have been proven (Goetz et al., 2020; Tze et al., 2021). For example, Tze et al. (2021) showed that the relationship between math appraisals and math performance in Grade 4 students from 53 educational systems (i.e., the 2015 Trends in International Mathematics and Science Study International database) was mediated by enjoyment and boredom. This means that activity-related emotions at least partly account for the influence students' appraisals of the topic to be learned have on their performance while learning that topic.

In addition, it is assumed that CVT components are linked by reciprocal causation, that is, feedback loops involving mutual reinforcements of appraisals, emotions, and learning outcomes over time (Pekrun, 2006; Pekrun et al., 2014, 2017; Pekrun & Stephens, 2010; Putwain et al., 2018). For example, in a one-year longitudinal study among students from 25 primary schools, Putwain et al. (2018) showed that students' higher enjoyment and lower boredom with respect to math lessons predicted higher subsequent grades in math; in turn, greater achievement predicted subsequent enjoyment and low boredom.

1.2. CVT in digital learning and intelligent tutoring systems

Until now, it is still unclear whether tenets of CVT hold in digital contexts. Indeed, CVT posits that antecedents of achievement emotions also include the characteristics of the learning environment, such as task demands, the structure of goals, and the presence of feedback (Pekrun, 2006). This implies that the nature of the learning environment may influence the cognitive and motivational processes underpinning performance. In particular, ITSs often maximize learner-centered learning processes by presenting ill-defined problem-solving tasks embedded in open-ended environments. While granting students more flexibility and autonomy than traditional instruction, the complexity of ITS environments may result in high demands placed on students (Biswas et al., 2016; Daniels & Stupnisky, 2012; Land, 2000). Consequently, the characteristics of ITSs might potentially affect the sequence of CVT processes through the experience of activity-related emotions (Plass & Kaplan, 2016), calling for an examination of their relations to antecedents and learning outcomes.

Regarding the links between cognitive appraisals and activity emotions, it is unclear whether students' topic-related appraisals precede their emotions for students who learn with an ITS in a similar way as in traditional learning settings. Because ITSs place high demands on students (i.e., the interaction with ITSs is sometimes reported as "tedious" by students) (D'Mello, 2021) and the self-regulation skills required to overcome the challenging open-ended nature of ITSs are still in development in adolescence (Pintrich & Zusho, 2002), the question arises as to whether and which appraisals of value and control regarding a given topic may be predictive of a positive emotional experience while working with the ITS. In theory, students ascribing high value to and/or feeling self-efficacious on a particular topic are expected to enjoy learning activities and show grit while exhibiting low boredom, according to CVT and the situated expectancy value theory (SEVT, Eccles & Wigfield, 2020). The meta-analysis by Loderer et al. (2020) brought partial evidence confirming the applicability of this assumption in a variety of technology-rich environments including ITSs, among others, by showing overall positive relationships between a conjunction of perceived control measures (including topic-related self-efficacy and self-concept) and a conjunction of subjective value measures (including subject-related interest, importance, and utility, i.e., perceived usefulness for future goals) and activity-related enjoyment ($r = 0.50$ and $r = 0.56$, respectively). To our knowledge, however, no study has yet documented the independent contributions of topic-related interest, importance, utility, and self-efficacy to task-related enjoyment in the specific context of ITSs.

The effects of topic-related appraisals of value and control on activity boredom in the ITS context are also rarely investigated (Loderer et al., 2020). Studies that focus on similar educational technologies such as Moodle may be informative when aiming to develop hypotheses about the relations between appraisals and boredom in ITS, as these environments share some characteristics with ITSs including self-paced, flexible learning and the availability of feedback on student performance. Berweger et al. (2022) showed that individual differences in students' interest and utility value appraisals with regard to the subject of educational sciences predicted their activity boredom levels negatively while using Moodle. Partially confirming these results with a similar technology, Acosta-Gonzaga and Ramirez-Arellano (2021) found that a latent factor comprising subject-related appraisals of control and value predicted a negative-emotion latent factor comprising boredom, frustration, and anxiety for the school context in general.

Taken together, this previous work supports CVT in that higher subject-related appraisals of control and value are associated with more positive and less negative experiences while completing digital learning tasks. However, when disentangling the contribution of each subject-related appraisal of control and value to activity emotions, findings reveal more fragmentary support for CVT in digital contexts (Berweger et al., 2022), which might also apply to ITSs. Therefore, in the present study we sought to determine whether each of the subject-related appraisals of control and value could independently enhance emotional experience while learning with an ITS.

1.3. The role of activity emotions in performance in intelligent tutoring systems

CVT assumes that activity emotions influence achievement. According to Pekrun (2006), achievement encompasses students' engagement or motivation to learn, task performance, and learning. Studies using ITSs generally show that the tenets of CVT hold in these contexts, in that more enjoyment and less boredom are associated with better learning outcomes. Using Betty's Brain, an ITS for climate change, Munshi et al. (2018) found that high task-performers were significantly associated with higher task-related delight (a proxy for enjoyment), lower task-related boredom, and higher learning gains than low performers on a pencil and paper pre-posttest. These findings not only suggest that task-related activity emotions predict task-related performance, but also that higher task performance yields better learning gains after students study the materials with the ITS. Similarly, a study by Cloude et al. (2020) focusing on negative emotions in MetaTutor, an ITS for biology, found that students experiencing increasing levels of boredom during the task showed lower learning gains on an independent pre- and posttest knowledge test.

Task-related engagement in ITSs (i.e., how immersed in, focused on, and concentrated students are with the system, Baker et al., 2010) has also been shown to vary depending on activity emotions and to play a role in task performance and learning. Andres et al. (2019) analyzed the frequency of students' affective transitions during Betty's Brain sessions every 20 s and found that two of the three most prevalent transitions involved associations between delight and engaged concentration, suggesting that both states potentialize one another. Additionally, Graesser et al. (2022) showed that for students in their optimal zone of engagement, their performances in a version of AutoTutor for literacy predicted improvements in reading comprehension skills. Conversely, boredom has been shown to best predict a state of disengagement in AutoTutor (D'Mello & Graesser, 2010) and to correlate negatively with learning gains in Betty's Brain (Andres et al., 2019). For that reason, studies have employed techniques aimed at reducing boredom by offering encouragement (Arroyo et al., 2010; D'Mello & Graesser, 2013) or increasing learning support (Arroyo et al., 2010; Chevalère et al., 2023; Forbes-Riley & Litman, 2009). For example, a recent study by Chevalère et al. (2023) varied the hint delivery strategy in Betty's Brain, where students could receive increasingly detailed hints either according to their consecutive failures on the task or as they increasingly self-reported emotions incompatible with learning. Unfortunately, however, this manipulation yielded no overall benefit with respect to students' boredom.

To summarize, the reviewed research offers substantial evidence supporting the global predictions of CVT in terms of the presence and direction of effects among contiguous components of the model. Therefore, just as in non-digital learning environments, the existence of an overall pattern of sequential processes from distal antecedents to learning outcomes (e.g., Tze et al., 2021) can be expected in the ITS context (i.e., *topic-related cognitive appraisals → task-related activity emotions → task engagement → task performance → learning gains*), which we investigated in the present study.

1.4. *The present study*

The main contributions of this study to existing research are related to the fact that, on a theoretical level, we examine whether the predictions of CVT can be transferred to the context of learning with digital technologies in science classrooms. On an empirical level, studies using digital technologies have already shown the importance of topic-related cognitive appraisals for task-related activity emotions (Acosta-Gonzaga & Ramirez-Arellano, 2021; Berweger et al., 2022), of emotions for task engagement (Andres et al., 2019; D'Mello & Graesser, 2010; Munshi et al., 2018), of engagement for task performance (Graesser et al., 2022), and of task performance for overall learning gains (Graesser et al., 2022; Munshi et al., 2018). However, previous studies still show fragmentary findings that have never been considered together in an ITS context. We thus provide a unique examination of the transition from students' topic-related appraisals to digital task-related emotional experience and learning, and test the sequence of constructs described in CVT in a pre- and posttest design before and after students learn with an ITS. We operationalize activity emotions by assessing task-related enjoyment and boredom, as they represent two extremely valenced states yielding opposite effects on learning outcomes (Pekrun, 2006) and are among the emotions most frequently experienced by students in the classroom (Butz et al., 2015).

Here, we used Betty's Brain, an ITS that uses a learning-by-teaching paradigm. This effective method consisting of having students learn on behalf of (and teach) other students as opposed to learning for themselves (Bargh & Schul, 1980; Fiorella & Mayer, 2013, see also Koh et al., 2018), has been adapted and proven effective in Betty's Brain (Chase et al., 2009). More precisely, students teach a virtual peer-agent, Betty, as much as possible about the topic of climate change by reading a science book, building a concept map, and receiving hints from a virtual teacher agent (more details below) (Biswas et al., 2016). In regard to the literature, the present study asked the following research questions:

RQ1: What are the relations between topic-related cognitive appraisals and task-related outcomes in the ITS context?

RQ2: What are the relations between task-related outcomes and learning outcomes in the ITS context?

Our assumptions are depicted in Fig. 1.

Against the theoretical background of CVT, we formulated the following hypotheses to answer the research questions:

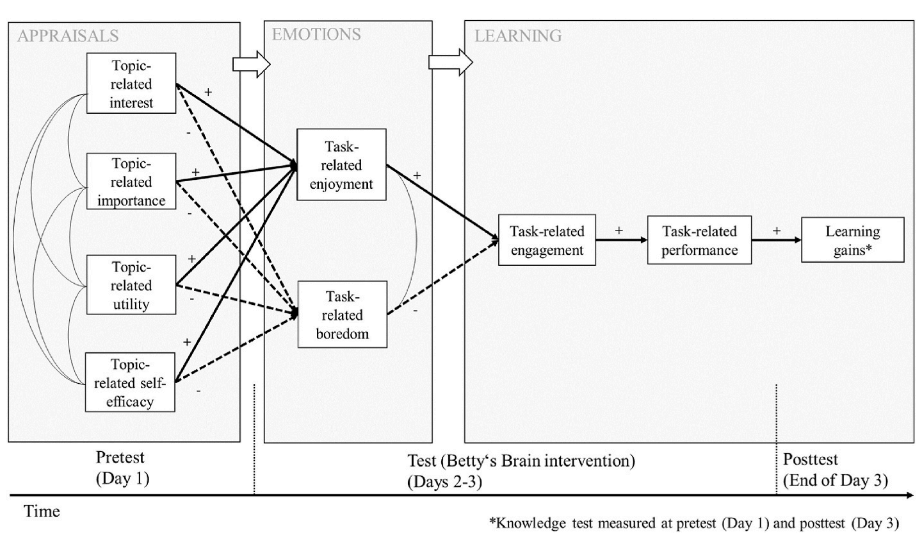

Fig. 1. Theoretical model of the relationships between topic-related appraisal, achievement emotions, and learning outcomes in 'Betty's brain'.

Hypothesis 1. Higher topic-related interest, importance, utility, and self-efficacy toward the topic of climate change (pretest phase) will be associated with higher levels of enjoyment and less boredom with Betty's Brain (test phase).

Hypothesis 2. Higher levels of enjoyment and lower levels of boredom (test phase) will be associated with higher levels of task engagement in Betty's Brain (test phase).

Hypothesis 3. Task engagement (test phase) will be positively associated with concept map performance in Betty's Brain (test phase).

Hypothesis 4. A higher performance on the concept map in Betty's Brain (test phase) will be associated with a higher increase in overall knowledge (learning gain from pretest to posttest) of climate change.

Hypothesis 5. There will be reciprocal relations between emotions (enjoyment and boredom) and learning-related outcomes (engagement and/or ongoing performance) during the Betty's Brain activity (test phase).

2. Methods

2.1. Sample

The study was conducted in five German secondary schools partially representative of the diversity of school types in Germany, including two 'Gymnasien' (academic secondary schools, 40%), one 'Gesamtschule' (comprehensive school, 20%), one 'Oberschule' (secondary school, 20%), and one 'Berufsschule' (vocational school, 20%). The ratio of school types in our study also roughly corresponds to the actual proportion of these school types at the country level when these school types in particular are compared to one another (42%, 29%, and 25% for Gymnasien, Gesamtschule, and Oberschule, respectively, with the exception of the Berufsschule, which currently only accounts for 4% of the ratio and is therefore slightly overrepresented here) (Statista, 2022). The sample used was composed of 140 participating students in the 7th to 10th grades (grades 7 and 8: 15.71%; grade 9: 37.14%; grade 10: 47.13%; M_{age} = 15.15, SD_{age} = 1.95, 48.29% girls, 78.57% native German speakers). Only students whose parents provided a signed informed consent form and who themselves consented to participate were included in the study.

2.2. Procedure

Data were collected over a maximum of four 1-h sessions embedded within a three-day protocol, which took place during each school's science week. Fig. 2 provides an illustration of the study design.

On pretest session 1 (day 1), students used an online platform to enter their demographic data and respond to an online questionnaire assessing CVT components with respect to the broader topic of climate change regardless of the Betty's Brain activity, as well as their prior knowledge by completing a knowledge test described below. On pretest session 2 (day 1), students received a general introduction to Betty's Brain and a training on concept maps by a student teacher of geography involved in the project, after which they were presented with an individual tutorial directly implemented in the software. During experimental test sessions 1 to 4 (i.e., the intervention spanned from day 2 to day 3), students worked independently on Betty's Brain for approximately 60 min each session with minimal assistance from the teacher, who would intervene in case the students had questions. During these sessions, the data of interest with regard to the digital learning task were collected through automatic 10-min state-like measures of activity emotions and engagement and assessments of performance implemented with the software. In-between the sessions and when students finished the task, they worked in collaborative groups on activities not related to climate change. Finally, at the end of the protocol (post-session on day 3), all participating students completed the knowledge test again and were presented with a general debriefing on the experiment.

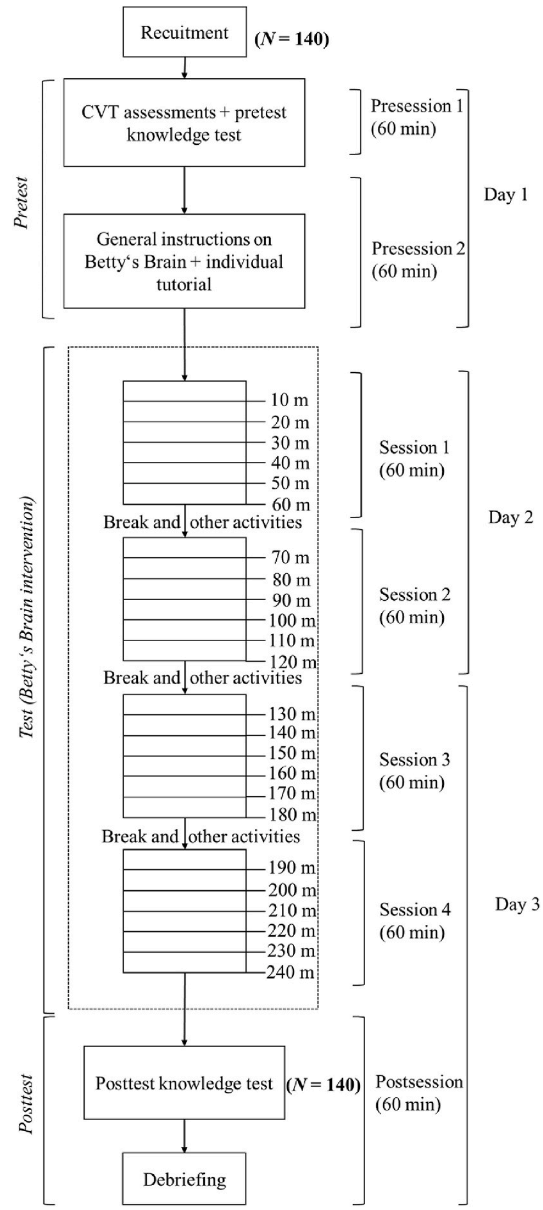

Fig. 2. The current study design.

2.3. Betty's Brain

Fig. 3 provides an illustration of the environment in Betty's Brain translated into the German language. When working with Betty's Brain, students learn the content of a "science book" that explains the phenomenon involved in climate change through text and illustrations and build a concept map, that is, a visual representation linking together via causal relations the scientific concepts just learned. The science book is divided into four chapters discussing different aspects of climate change and students are free to read the science book and build the concept map at their own pace (Biswas et al., 2016).

Upon request by the student and automatically every 10 min, the virtual teacher agent provides feedback on the accuracy of the concept map (an example of a concept map is depicted in Fig. 3) and delivers a hint with different levels of detail to either help the learner fix incorrect conceptual links or suggest missing links in the concept map (see Segedy, 2014, for more details on hint generation).

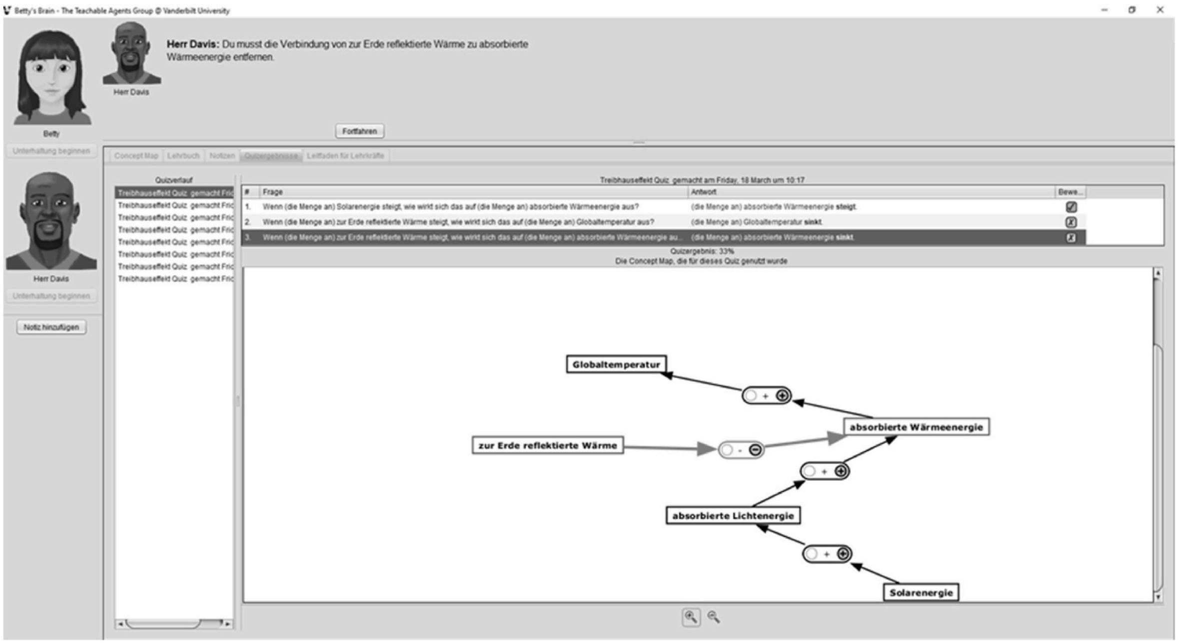

Fig. 3. An illustration of the Betty's brain environment.
Note. Mr. Davis, the mentor agent, gives a hint and performance feedback automatically every 10 min in the present study.

For research purposes pertaining to a larger research project (including Chevalère et al., 2023 publication), students were assigned randomly to two versions of the software that differed subtly in the way hints from the virtual teacher were delivered. In one version - the emotion-based strategy - students received hints according to their self-reports from a list of emotional categories derived from the cognitive disequilibrium framework (D'Mello & Graesser, 2012). This theoretical framework describes the dynamics of the cognitive-affective states that emerge during complex learning, with a focus on the transitions between states of engagement, confusion, frustration, and boredom. In the second version - the progressive hint strategy - students received hints according to their consecutive failures on the task (Segedy, 2014). The experimental condition was aimed at reducing students' boredom while interacting with the ITS, but showed no overall benefit (Chevalère et al., 2023). To further ensure that the present hypotheses related to CVT could be tested independently of the different Betty's Brain versions, we conducted a series of independent sample t-tests with Bonferroni corrections prior to all analyses (α level cutoff at $p < 0.05/2$ conditions $= 0.025$) on measures at pretest, test, and posttest (described below). Results of the t-tests revealed non-significant differences across the two conditions, $ts(138) \leq |1.33|$, $ps \geq 0.18$, $ds \leq 0.22$, in the variables included in this study, ruling out any significant influence of the software's versions on the findings. In all analyses of the present study, the group membership (G1 vs. G2) was used as a covariate. More information about the two versions is available in Chevalère et al. (2023).

2.4. Measures

2.4.1. Perceived topic-related appraisals

Topic-related values were assessed using the Scale Assessing Subjective Educational Task Values [Skala zur Erfassung subjektiver schulbezogener Werte, SESSW] by Steinmayr and Spinath (2010). The validated 9-item Likert-type scale includes three dimensions of perceived value (interest, importance, and utility) derived from SEVT (Eccles & Wigfield, 2020). We have adapted the questionnaire to the topic of climate change with 3 items for each dimension, such as '*It is important for me to be familiar with the topic of climate change*' (importance dimension) [in German], and responses ranged from 1 '*does not apply at all*' to 5 '*applies completely*' [in German]. The reliability of the three-item scale was good (α = 0.86) (Taber, 2017). More details about the reliability analyses are available in Appendix A.

Topic-related control was assessed using the German version of the self-efficacy questionnaire by Kunter et al. (2002) used in the 2000 PISA survey in Germany. The validated 4-item Likert-type scale was originally developed to assess self-efficacy in mathematics. In the present study, we have adapted the wording of the items to assess students' competence beliefs about the topic of climate change through items such as '*I am convinced that I can master the skills that are being taught on the topic of climate change*' [in German]. Reponses range from 1 '*does not apply*' to 4 '*applies*') [in German]. The reliability of the four-item scale was acceptable (α = 0.75).

2.4.2. Task-related activity emotions and engagement

Task-related activity emotions and engagement were directly assessed in the Betty's Brain environment through single-item adjectives presented automatically every 10 min. We focused on the two highly valenced activity emotions of enjoyment and boredom (Pekrun, 2006). Engagement was assessed in accordance with the taxonomy defined in previous studies using ITSs (Andres et al., 2019; D'Mello et al., 2007) based on the cognitive disequilibrium framework (D'Mello & Graesser, 2012). Items consisted of 5-point Likert-type items related to each of the three words '*Engaged/Concentrated,*' '*Enjoyed,*' and '*Bored*' (presented in that order) with response categories ranging from 1 '*not at all*' to 5 '*very strong*' [in German]. The item wording was '*How strong are you experiencing the following states right now? Please click on the button that best describes the intensity of each state*' [in German]. The use of single-item adjectives minimizes disruptions to the learning process and is in line with previous research that has measured emotional states over time (Duffy et al., 2020; Gogol et al., 2014). The reliability of the repeated single items computed using the maximum number of measurement occasions ($n = 6$) common to all participants ($N = 140$) was good ($ICC_{\text{Enjoyment}} = 0.85$, $ICC_{\text{Boredom}} = 0.89$, $ICC_{\text{Engagement}} = 0.89$, see Appendices A and B for more details).

2.4.3. Task performance

Task performance in Betty's Brain was assessed every 10 min following the emotion prompt. The scoring procedure builds upon the original scoring system in Betty's Brain (Segedy, 2014), which provides an accuracy score for the current concept map by calculating the proportion of correct links out of the total number of correct links derived from the fully correct concept map for a single chapter. As described in Appendix C, Text C.1, the original scoring system was expanded to account for the entire Betty's Brain task comprising the four chapters (/100%), referred to as 'raw performance.' The reliability of the repeated raw performance single item over 6 consecutive assessments was good ($ICC = 0.87$). Ultimately, the maximum score that a student j reached at the end of the task was adjusted by that student's frequency of progress along the task (i.e., the number of improvements (successes) i made on the causal map for a single student j), referred to as the 'final performance score.'

2.4.4. Overall knowledge test on climate change

The pretest knowledge test was administered as a self-developed online knowledge test. The knowledge test assesses general aspects of climate change addressed in Betty's Brain and consists of seven close-ended (/16.5) and one open-ended question (/13.5) for a maximum score of 30 points. More details about the knowledge test and the scoring system can be found in Appendix, Text C.2. The reliability of the 24-item test at pretest was good ($\alpha = 0.87$). The posttest knowledge test on climate change was identical to the one presented at pretest. The reliability of the 24-item test at posttest was good ($\alpha = 0.87$).

2.5. Statistical analyses

To answer the research questions, we used path analyses performed in R version 4.2.1 (R Core Team, 2022) using the 'lavaan' package version 0.6–12 (Rosseel, 2012). To test our hypotheses, a series of path analyses were conducted on the data from 140 participants (no missing or extreme values) to model the relations between perceived topic-related values, self-efficacy, task-related activity emotions, engagement, performance, and pretest and posttest knowledge. All model parameters were estimated using maximum likelihood estimation, and 95% confidence intervals were generated using bootstrapping (5,000 resamples) and percentile methods (Efron & Tibshirani, 1993). Following Kline (2016), we reported $\chi2$, CFI, RMSEA, and SRMR as fit indices. Values of CFI > 0.95, RMSEA < 0.07, and SRMR < 0.08 indicated good model fit (Hooper et al., 2008). The required sample size ($N \geq 127$) was determined a priori with a power analysis ($\beta = 0.80$, $\alpha = 0.05$) conducted with G*power software version 3.1 (Faul et al., 2009) in the context of multiple regression. In the absence of an available prior effect size for the present model, we expected a medium overall effect ($f^2 = 0.15$, Cohen, 1988) for a total of 12 predictors. Multilevel analyses were not performed due to the small number of participating schools ($N = 5$) and total classrooms ($N = 11$). However, to control for the variability in age and gender across schools, these two variables were used as covariates in all analyses.

We employed a three-step procedure to test the predictions of CVT, inspired by Rohrer et al.'s (2022) recommendations to investigate the causal validity of the theoretical model in Fig. 1. As the first step, an 'Average' model included the pretest, posttest, and final performance measures and the averages of all collected task-related measures of emotions and engagement (n measurement occasions over 4 1-h sessions). Endogenous and exogenous variables were mean-centered and entered as manifest variables to reflect the ordering of constructs in the CVT and the temporal ordering of events in the experiment (pretest, test, and posttest), as depicted in Fig. 1. The pretest knowledge test was entered as a covariate in the model. Consequently, *learning gains* corresponded to the residual score of the posttest knowledge test net of the influence of the pretest knowledge test (Xiao et al., 2019). To control for intervention effects, the two versions of Betty's Brain (the progressive hint strategy and the emotion-based strategy) were dummy-coded and used as a covariate in the model. An exhaustive representation of the Average model is available in Appendix D.

In the second step, a reverse causality analysis was conducted including a 'Forward' model and a 'Reverse' model. These models allowed us to manipulate the temporal precedence of task-related measures. To this end, we restricted the perimeter of analyses to the data common to all participants, that is, the first 1-h session working with the ITS (cf. Appendix B), and focused on task-related measures (i.e., enjoyment, boredom, engagement, and raw performance) assessed at T1 and T6, corresponding to the beginning ($t + 10$ min) and end ($t + 60$ min) of the first session. The Forward model reflected a similar temporal ordering of constructs as in the Average model, where activity emotions at T1 (enjoyment and boredom) were entered as predictors of learning-related measures at T6 (engagement and raw performance). The learning-related measures at T6 were entered as predictors of the final performance score assessed at the end of the task ($t + n$ minutes). The Reverse model inverted the temporal ordering of task-related measures, where learning-related measures at T1 (task-related engagement and raw performance) were entered as predictors of activity-related emotions at T6 (enjoyment and boredom). The activity emotions at T6 were then entered as predictors of the final performance score assessed at the end of the task. The Forward and Reverse models were compared to each other in light of the Average model to examine the causal reversibility of the theoretical assumptions. The two models included the same covariates as in the Average model.

Ultimately, in the third step, sensitivity analyses were conducted on the significant temporally-ordered paths identified in the Forward and Reverse models by considering the influence of observed confounders (e.g., the influence of engagement at T1 producing a spurious association between enjoyment at T1 and engagement at T6) (Elwert, 2013; Rohrer et al., 2022). The sensitivity to unobserved confounders was also examined, following MacKinnon and Pirlott's (2015) recommendations using the Average Causal Mediation Effect (ACME) technique (Imai et al., 2010) with 5,000 samples, bootstrapped 95% confidence intervals, and a percentile method, performed with the R package 'mediation' version 4.5.0 (Tingley et al., 2014). This method consists of assessing the extent to which an indirect effect, that is, the influence of a variable X (e.g., enjoyment T1) on an outcome Y (i.e., engagement T6) through a mediator M (here, engagement T1 as the confounder), is robust to the influence of an unobserved variable U. This is done by simulating a range of values for the magnitude of the correlation $\rho(U)$ between the residuals of the mediator M and outcome Y, to determine how influential the 'virtual' presence of the confounder U must be to invalidate conclusions about mediation (i.e., the cut-off $\rho(U)$ value at which the indirect effect equals zero, Imai et al., 2010). Using Hemphill's (2003) guidelines for the magnitude of correlation coefficients, an indirect effect was assumed to be either sensitive, $\rho(U) < 0.20$, robust, $0.20 \leq \rho(U) < 0.30$, or highly robust, $\rho(U) > 0.30$, to the unobserved confounder U.

3. Results

All coefficients of the models have been standardized, so they can be interpreted as effect sizes. Descriptive statistics and Pearson correlations among study variables can be found in Appendix E.

3.1. Average model

Based on current standards, the 'Average' model depicted in Fig. 4 showed a good fit, $\chi2\,(17) = 27.07$, CFI $= 0.981$, RMSEA $= 0.067$, SRMR $= 0.026$. An examination of the coefficients linking topic-related value and control appraisals to task-related activity emotions indicates that increases in students' interest in the broader topic of climate change resulted in more enjoyment when working with Betty's Brain, $\beta = 0.29$, $SE = 0.11$, $t = 2.64$, $p = 0.008$ [0.06; 0.51]. Conversely, predictions of the CVT were not supported in the data concerning the expected relation between topic-related interest and task-related boredom, and concerning relations between other value components (importance and utility) and control (self-efficacy) appraisals and activity emotions ($ps \geq 0.35$).

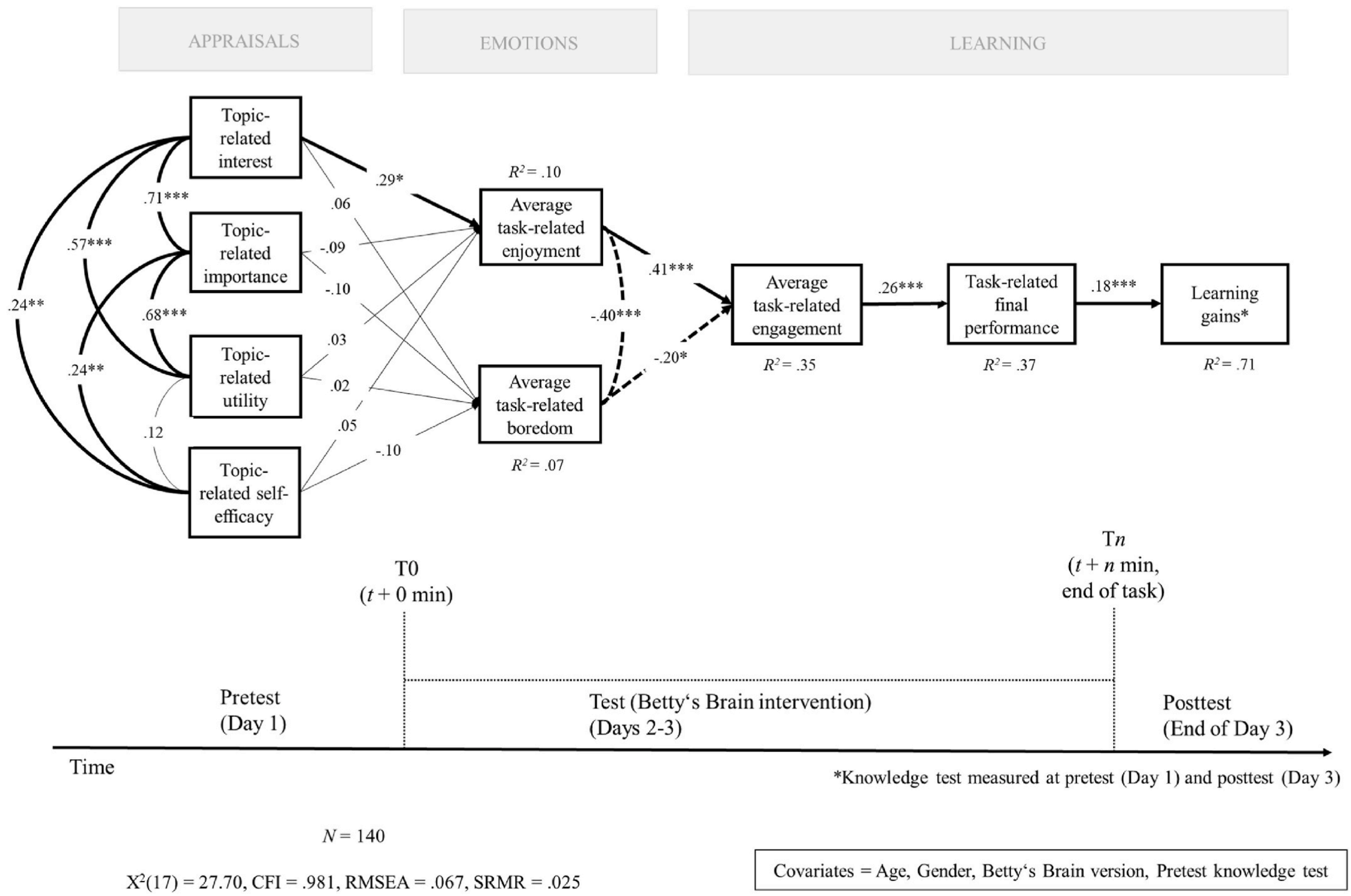

Fig. 4. Path diagram output with standardized estimates for the relations between cognitive appraisals, activity emotions, and learning outcomes in Betty's brain. *Note.* * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. Thick solid and dotted lines represent positive and negative significant relations.

The model remained unchanged when including only individual dimensions of topic-related value and control, ruling out explanations invoking collinearity among these variables. Regarding the relations between activity emotions and the learning components, task-related enjoyment during Betty's Brain was significantly and positively related to task-related engagement, $\beta = 0.41$, $SE = 0.09$, $t = 4.40$, $p < 0.001$ [0.20; 0.57], and negatively related to task-related boredom, $\beta = -0.20$, $SE = 0.10$, $t = -2.08$, $p = 0.038$ [-0.40; -0.02]. Among learning outcomes, more engagement in Betty's Brain was associated with a significantly better final performance score on the concept map, $\beta = 0.26$, $SE = 0.08$, $t = 3.43$, $p = 0.001$ [0.11; 0.40], which in turn yielded higher learning gains, $\beta = 0.18$, $SE = 0.06$, $t = 3.17$, $p = 0.002$ [0.070; 0.29] (a paired sample *t*-test revealed that learning gains were significant, $t(139) = 6.29$, $p < 0.001$). Between age and gender, only age had a significant effect on task-related final performance, $\beta = 0.20$, $SE = 0.07$, $t = 2.92$, $p = 0.004$ [0.06; 0.33], all other effects of age and gender being non-significant, all $ps \geq 0.057$. All reported effects in the path diagram were net of the significant effects of prior knowledge, all $\beta s \geq |0.22|$, $ps \leq 0.01$. Surprisingly, prior levels of knowledge had no significant relation to enjoyment during Betty's Brain, $p = 0.96$. The version of Betty's Brain (progressive hint delivery or emotionally adaptive) was unrelated to any other measure of the model, all $ps \geq 0.10$.

3.2. Reverse causality analysis

3.2.1. Forward model

The 'Forward' model depicted in Fig. 5, which assumes the temporal precedence of emotions, showed a good fit, $\chi2$ (22) = 26.81, CFI = 0.992, RMSEA = 0.039, SRMR = 0.028. In line with the Average model in Fig. 4, a similar association was found linking students' topic-related interest (assessed at pretest) to their task-related enjoyment after 10 min (T1) working with Betty's Brain, $\beta = 0.30$, $SE = 0.11$, $t = 2.78$, $p = 0.005$ [0.08; 0.51]. Students' initial enjoyment (T1) predicted their later engagement after 60 min (T6), $\beta = 0.26$, $SE = 0.09$, $t = 2.92$, $p = 0.003$ [0.08; 0.42], but the initial boredom (T1) did not predict later engagement (T6), $p = 0.45$, despite being significantly and negatively correlated with enjoyment (T1), $r = -0.26$, $SE = 0.09$, $t = -3.145$, $p = 0.002$ [-0.47; -0.09]. Interestingly, students' emotions (T1) were unrelated to their later raw performance (T6), $ps > 0.22$. The final performance score, assessed at the end of the task (T*n*), was predicted by students' engagement at T6, $\beta = 0.13$, $SE = 0.06$, $t = 2.42$, $p = 0.015$ [0.21; 0.24]. However, this relation was four times smaller than that involving the raw performance at T6, $\beta = 0.54$, $SE = 0.06$, $t = 9.09$, $p < 0.001$ [0.43; 0.66]. In sum, in line with the Average model, the Forward model revealed the following temporal sequence: *topic-related interest → task-related enjoyment T1 → task-related engagement T6 → task-related final performance → learning gains.*

3.2.2. Reverse model

The 'Reverse' model depicted in Fig. 6, which assumes the temporal precedence of learning-related task measures, showed a good fit, $\chi2$ (22) = 27.27, CFI = 0.990, RMSEA = 0.041, SRMR = 0.029. In contrast to the Average and Forward models in Figs. 4 and 5, no association was found between topic-related interest and learning-related measures at T1 (i.e., engagement and raw performance, $ps > 0.18$). Students' initial task engagement (T1) significantly predicted their later enjoyment after 60 min (T6), $\beta = 0.29$, $SE = 0.09$, $t = 3.28$, $p = 0.001$ [0.12; 0.46]. In parallel, students' raw performance at T1 significantly and negatively

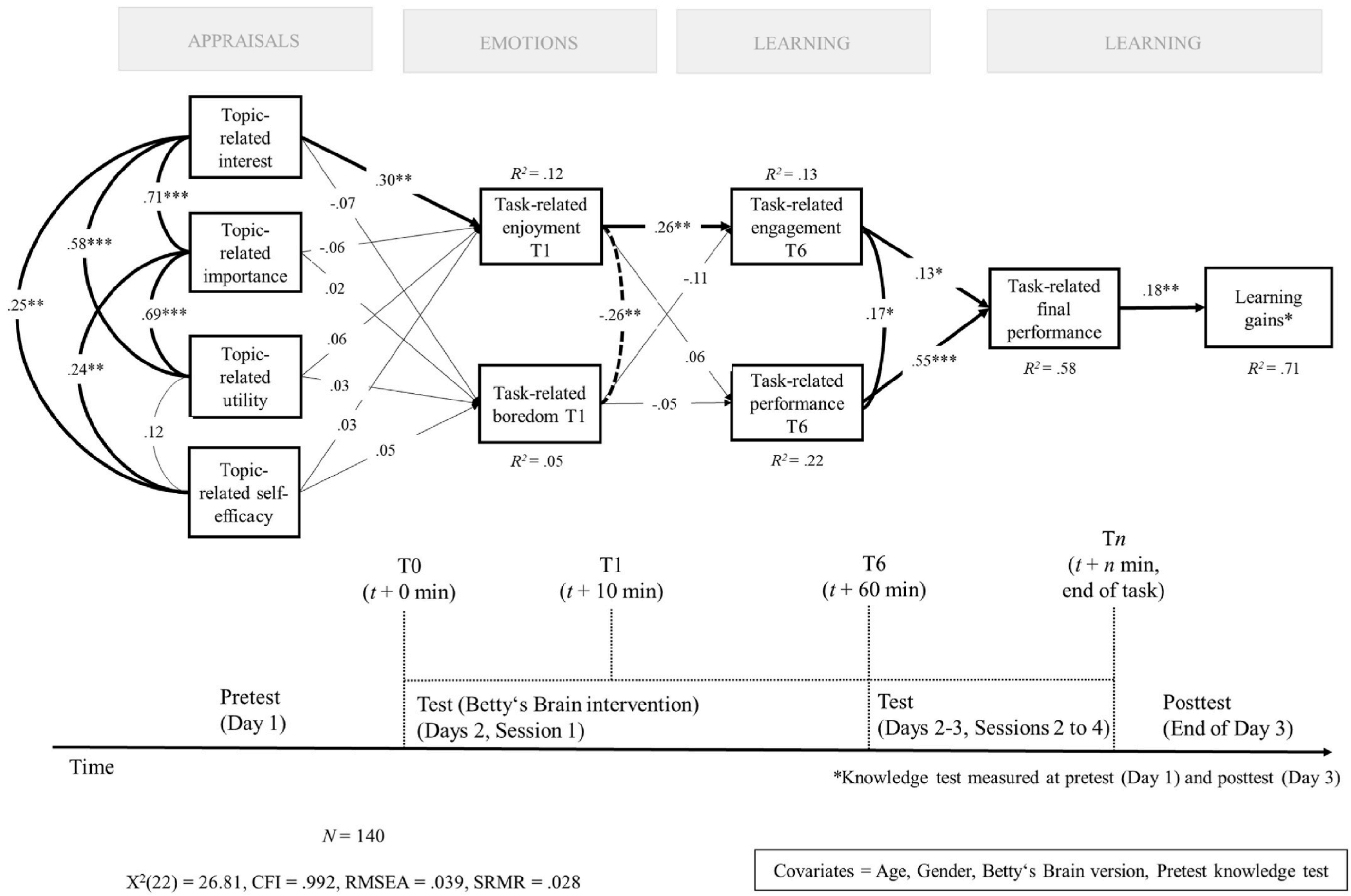

Fig. 5. Forward path diagram output with standardized estimates for activity emotions at T1 causing learning-related outcomes at T6 during the first Betty's brain session.
Note. Please refer to Fig. 4 for legend.

predicted their later enjoyment at T6, β = -0.18, *SE* = 0.08, *t* = 2.21, *p* = 0.027 [-0.29; -0.08]. However, only 4 participants out of 140 showed a T1 raw performance score >0 (*M* = 0.48, *SD* = 2.96), meaning that almost none of the students had shown any progress during the first 10 min on the task, which undermines the reliability of the coefficients involving performance at T1. Students' task-related enjoyment at T6 significantly predicted their final performance score at the end of the task (T*n*), β = 0.26, *SE* = 0.07, *t* = 3.82, *p* < 0.001 [0.13; 0.40]. Boredom at T6, in contrast, did not predict the final performance score, *p* = 0.53, despite being negatively correlated with enjoyment (T6), *r* = -0.38, *SE* = 0.08, *t* = −4.505, *p* < 0.001 [-0.54; -0.19]. In sum, the Reverse model showed that engagement, and possibly performance at an early stage of session 1 (T1), were associated with enjoyment at a later stage (T6), as shown by the following paths: *task-related engagement T1 → task-related enjoyment T6 → task-related final performance → learning gains* and *task-related performance T1 → task-related enjoyment T6 → task-related final performance → learning gains*.

3.3. *Sensitivity analyses with respect to observed and unobserved confounders*

The previous reverse causality analysis revealed three important points. First, after manipulating the temporal precedence (either T1 or T6) of activity emotions and learning-related measures, the paths involving boredom found in the Average model were no longer significant. Second, topic-related interest showed a unique association with task-related enjoyment at T1, which was not found for other task-related measures, consistent with the Average model. Third, reciprocal effects were found between task enjoyment and engagement.

As noted by Rohrer et al. (2022), 'temporal order cannot rule out confounding' (p. 3). Therefore, to further investigate the causality behind the previously identified sequences, we conducted sensitivity analyses with respect to observed and unobserved confounders where each sequence was broken down into simple mediation models. The mediation framework (Imai et al., 2010) allowed us to condition each predictor-outcome pair on their observed coevals (any other assessment made prior or concomitant to the predictor and outcome, used as mediators), in addition to systematically controlling for other covariates. Furthermore, the mediation framework allowed us to generate a value ρ(U) informing us about the sensitivity of each *predictor (X) → coeval (M) → outcome (Y)* triads to unobserved confounders. The detailed analyses presenting standardized regression coefficients and covariates are available in Appendix, Table F.1, and an intermediate summary of the resulting sequences is available in Table F.2.

Table 1 below summarizes the findings by contrasting the expected causal sequences found in the reverse causality analyses with the actual sequence found after the sensitivity analyses were performed. It revealed that nearly all paths identified in the reverse causality models (in bold) contained spurious associations due to the presence of observed confounders (in parentheses).

For the expected causal sequence of the Forward model involving the triads *topic-related interest → task-related enjoyment T1 → task-related engagement T6*, *task-related enjoyment T1 → task-related engagement T6 → task-related final performance*, and *task-related engagement T6 → task-related final performance → learning gains* (i.e., paths 1h, 3p, and 5b in Appendix, Table F.1), the analyses revealed that these paths were no

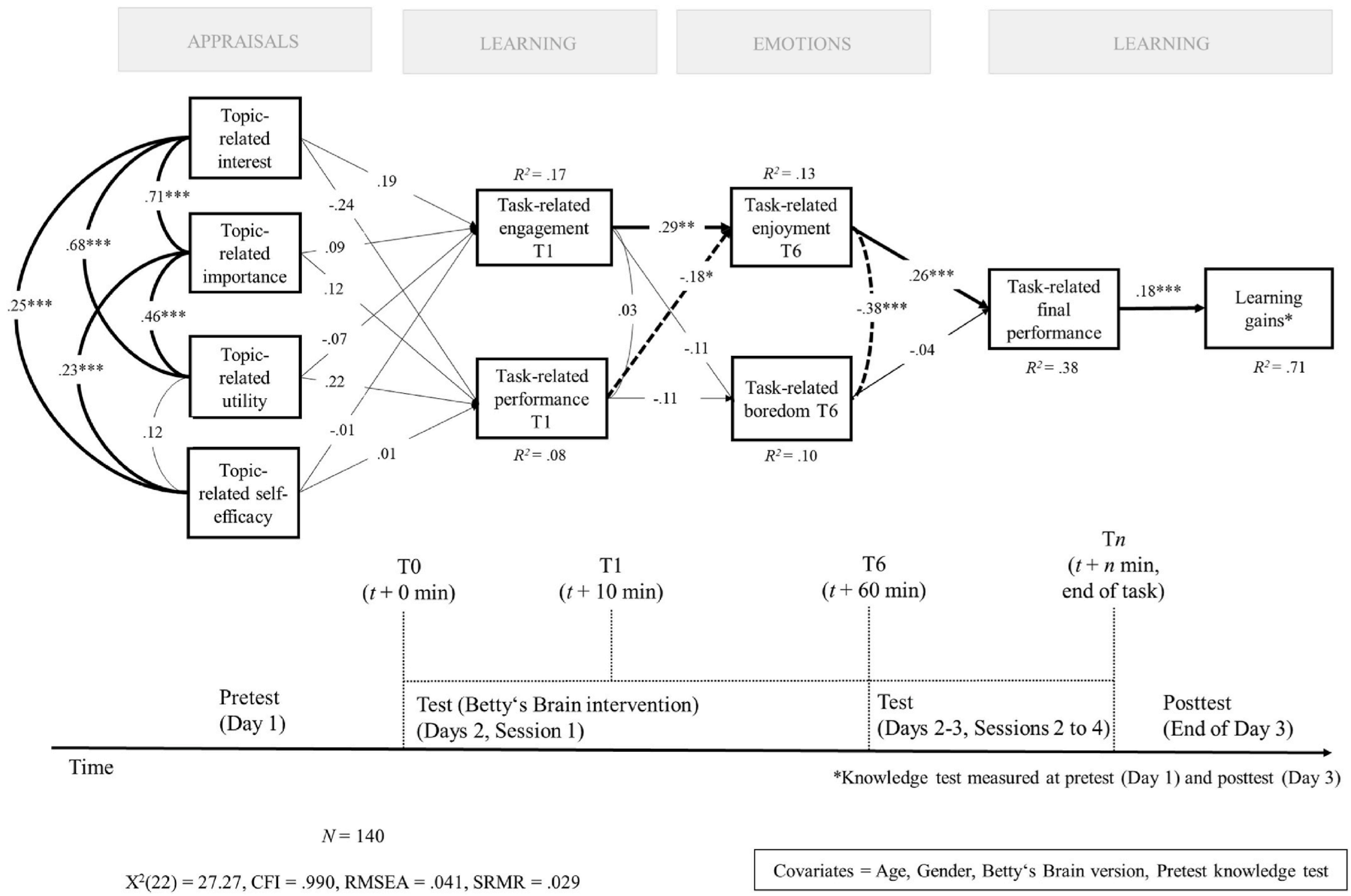

Fig. 6. Reverse path diagram output with standardized estimates for learning-related outcomes at T1 causing activity emotions at T6 during the first Betty's brain session.
Note. Please refer to Fig. 4 for legend.

Table 1
Comparison between expected causal sequences (in bold) and actual sequences after sensitivity analyses.

Reversed Causality Models	Time								
	Pretest		T1		T6		Tn		Posttest
Sequence 1 from Forward model	**Interest**	→	**Enjoyed T1**	→	**Engaged T6**	→	**Final Perf**	→	**L. Gains**
Actual sequence 1 (observed confounders)	Interest	→	Enjoyed T1 (Engaged T1)	→	Engaged T6 (Enjoyed T6, Performance T6)	→	Final Perf	→	L. Gains
Sensitivity to unobserved confounders			*Robust-to-highly robust*		*Highly robust*		*Sensitive*		
Sequence 2 from Forward model				X	**Performance T6**	→	**Final Perf**	→	**L. Gains**
Actual sequence 2 (observed confounders)				X	Performance T6	→	Final Perf	→	L. Gains
Sensitivity to unobserved confounders							*Sensitive*		
Sequence 3 from Reverse model		X	**Engaged T1**	→	**Enjoyed T6**	→	**Final Perf**	→	**L. Gains**
Actual sequence 3 (observed confounders)		X	Engaged T1	→	Enjoyed T6 (Engaged T6, Performance T6)	→	Final Perf	→	L. Gains
Sensitivity to unobserved confounders					*Highly robust*		*Sensitive*		
Sequence 4 from Reverse model		X	**Performance T1**	→	**Enjoyed 6**	→	**Final Perf**	→	**L. Gains**
Actual sequence 4 (observed confounders)		X	Performance T1	→	Enjoyed 6 (Performance T6)	→	Final Perf	→	L. Gains
Sensitivity to unobserved confounders					*Highly robust*		*Sensitive*		

Note. The bottom medium grey area represents paths involving the Performance T1 measure, for which only 4 participants out of 140 show a score > 0/100 (M = 0.48, SD = 2.96) and should thus be considered with caution.

longer significant when coevals and other covariates were controlled for, $ps \geq 0.11$. Despite that, the unique connection between topic-related interest and enjoyment was not explained by any observed confounder (i.e., paths 1c and 1e, $ps \geq 0.12$) and was highly robust to unobserved confounders, $\rho(U) = 0.31$. As Table 1 summarizes, from task enjoyment T1 onwards, the relations found in the Forward model existed due to the influence of confounders *engaged T1, enjoyed T6,* and *performance T6* (see also Summaries 1 and 2, Table F.2, and paths 1a, 1g, 2b, 3a, 3b, 4b, and 5c in Table F.1).

Similarly, as shown in Table 1, for the expected causal sequence of the Reverse model involving the triads *task-related engagement T1 → task-related enjoyment T6 → task-related final performance* and *task-related enjoyment T6 → task-related final performance → learning gains* (i.e., paths 3q and 5a in Appendix, Table F.1), the analyses revealed that these paths were no longer significant when coevals and other covariates were controlled for, $ps \geq 0.60$. Table 1 shows that these relations existed in the Reverse model due to the influence of confounders *engaged T6* and *performance T6* (see also Summary 1 and 2, Table F.2, and paths 2b, 3a, 3b, 4b, and 5c in Table F.1).

For the expected causal sequence of the Reverse model involving *task-related raw performance T1* → *task-related enjoyment T6* → *task-related final performance* and *task-related enjoyment T6* → *task-related final performance* → *learning gains* (i.e., paths 3r and 5a in Appendix, Table F.1), the analyses revealed that these paths were no longer significant when coevals and other covariates were controlled for, $ps = 0.55$. Table 1 shows that these relations existed in the Reverse model due to the influence of confounder *performance T6* (see also Summary 3, Table F.2, and paths 3l, 4b, and 5c in Table F.1). Note that the results of this particular sequence may not be reliable due to the very small number of participants showing a raw performance score above 0.

Finally, for the only path of the Forward model unbiased by observed confounders, *task-related raw performance 6* → *task-related final performance* → *learning gains*, meaning that the unique variance of *performance 6* predicted *learning gains* through the *final performance* score, $\beta = 0.08$, $SE = 0.04$, $t = 2.03$, $p = 0.030$ [0.001; 0.18] (path 5c in Appendix, Table F.1), the mediation model proved sensitive to the presence of an unobserved confounder *U*. In this case, the indirect effect of the final performance score was abolished at $\rho(U) = 0.18$.

4. Discussion

The present study was the first to test the direct assumptions regarding the sequence of processes described in the CVT in an ITS embedded in an open-ended learning environment. Using a pre- and posttest design, we assessed students' topic-related appraisals of value and control, task-related activity emotions (enjoyment and boredom), and learning outcomes (task-engagement, task-performance, and learning gains). We analyzed the data using a three-step procedure incrementally examining the causality of the CVT assumptions (i.e., a path analysis of the averaged task measures, a temporally-ordered reverse causality analysis, and a sensitivity analysis with respect to confounders). We expected a sequence of psychological processes starting with topic-related appraisals influencing task-related emotions (H1), and then emotions influencing task engagement (H2), engagement influencing task performance (H3), and, ultimately, task performance influencing learning gains (H4). Reciprocal relations among task-related measures were also expected (H5). Although the models brought partial support for the application of CVT to the ITS context on a correlational basis, causation was more difficult to establish due to the influence of multiple confounders challenging the expected unidirectional and reciprocal relations. In the following, we answer our two research questions in greater detail.

4.1. RQ1: what are the relations between topic-related cognitive appraisals and task-related outcomes in the ITS context?

Examining the applicability of CVT in ITSs is useful to determine the suitability of that technology to effectively support the learning of a given topic or subject. One important feature of CVT is the assumed sequential causation between the various components, from appraisals to learning via emotional experience. More precisely, CVT emphasizes that learners' value attribution and mastery expectations regarding a subject to be learned foster positive emotional experiences during the learning process (Pekrun et al., 2007; Pekrun & Perry, 2014). Regarding that specific relation, for which we hypothesized positive associations between topic-related interest, importance, utility, self-efficacy, and task-related enjoyment, as well as negative associations with task boredom (H1), we found puzzling results. Partially validating H1, the interest value component was the only bridge linking topic-related appraisals to task-related enjoyment, a finding quite robust to reverse causality and confounders, and no association was found between emotional experience and other appraisals of value (importance, utility) and control (self-efficacy). The lack of any relation between topic-related control and task-related enjoyment and boredom is at odds with CVT in general, but partially in line with other empirical findings based on Moodle (Berweger et al., 2022). Berweger et al. similarly found non-significant relations between expectancies of success and enjoyment and boredom on a between-person level. According to the authors, emotions evoked during digital learning activities may be more closely determined by the momentary sense of control students perceive over the task itself than by general expectations of competence on the topic to be learned. Our data suggest that similar conclusions may apply in the ITS context. This calls for more fine-grained considerations of students' perceived control over the various facets of the task (Butz et al., 2015), which may include technology-related features (i.e., navigational properties, amount of accessible resources), pedagogy-related features (degree of structuration of the learning support, clarity of learning goals), and content-related features (i.e., causal maps, clarity of the knowledge content). In addition, ill-defined goals in ITSs are known to place high demands on cognition and self-regulation (Biswas et al., 2016; Land, 2000; Munshi et al., 2018). Therefore, in the ITS context, topic-related perceived control and activity emotions may show weaker relations than those involving task demands and students' ability to cognitively and behaviorally cope with these demands (Butz et al., 2015; Ruthig et al., 2008; Zhao et al., 2021).

Regarding attributions of value, the predominant contribution of topic-related interest to task enjoyment at the detriment of utility and importance can be reconciled with the SEVT, highlighting the necessity of distinguishing between the three value components (Putwain et al., 2018). The interest component reflects intrinsic motivation, that is, the inherent satisfaction of pursuing an activity, while the utility and importance components entail extrinsic motivation, that is, when learning is instrumental to attaining outcome rewards (Eccles, 2005). Our findings align with Frenzel et al. (2007), who showed that only the intrinsic component was related to enjoyment, an activity emotion, whereas extrinsic value components were related to outcome-related emotions such as shame or pride. Simonton and Garn (2020) also confirmed these findings and concluded that enjoyment relates to intrinsic value because it directs attention to the content of the learning activity rather than to achievement outcomes such as grades. To better understand how similarly this applies to the ITS context, the intrinsic component has been shown to positively relate to academic buoyancy, while extrinsic components do not. This suggests that only interest may provide the grit necessary to deal with multiple academic challenges efficiently (Simonton & Garn, 2020).

Our finding that interest was unrelated to boredom was quite surprising (H1), given previous evidence in digital contexts (Acosta-Gonzaga & Ramirez-Arellano, 2021; Berweger et al., 2022), but it can be understood in light of different motivational processes. According to Pekrun et al. (2010), lack of interest entails a lack of approach motivation, while boredom triggers avoidance of the situation. In the present study, bored students might have been predominantly motivated to escape the situation, possibly due to task-related characteristics, rather than being uninterested in the topic in the first place. On the contrary, even for students initially uninterested in climate change, simply working with a new learning technology may have sufficed to elicit engagement and circumvent boredom, which may additionally explain the non-significant relations in the left portion of the reverse causality models. In fact, the disconnect between topic-related interest and task-related boredom and engagement echoes a conceptual distinction made between stable/dispositional components of interest and fluctuant/situational components of interest (Ainley, 2017; Hidi & Renninger, 2006; Krapp, 2007; Wild, 2022). For example, Wild (2022) discusses the relative independence between the content of interest (e.g., the topic of climate change) and the actions of interest (e.g., reading a book or working with a digital technology) that are carried out to engage with the object of interest. Likewise, Hidi and Renninger (2006) emphasize that interest development may be triggered by specific situations and be facilitated by the structure of the environment. When students encounter a new activity, they come into contact with stimuli that attract their attention (Krapp, 2007). The novelty triggering situational interest may also explain why students (n = 136 out of 140) showed no improvement during the first 10 min on task, as they might have initially been engaged in appreciating the stimuli of the environment before engaging in the learning task (136 students showing a score of 0 at T1 also rules out the involvement of performance as a meaningful factor at this early stage, disproving its negative associations with topic-related interest and task enjoyment at T6). The independent role of situational interest is not incompatible with the positive link between topic-related interest and task enjoyment, though. According to Hidi and Renninger (2006) and Wild (2022), the situational component of interest and the stable personal dispositions may also influence each other. Students assigning a personal value to the object of interest (i.e., the topic), who thus show eagerness to acquire new domain-specific knowledge (Ainley, 2017), might be more likely to develop positive emotions toward object-related actions when dealing with the object (Krapp, 2007).

4.2. RQ2: what are the relations between task-related outcomes and learning outcomes in the ITS context?

The CVT assumes that activity enjoyment and boredom respectively increase and decrease task engagement (H2), which then influences task performance (H3). At the end of the sequence, higher task performance increases learning gains (H4). The proposed Average model, which considered all measurement occasions of engagement, enjoyment, and boredom regardless of their temporal ordering, confirmed these predictions. Hence, from a correlational point of view, the data accredit this portion of the theory in the ITS context (*task-related activity emotions → task engagement → task performance → learning gains*).

However, the subsequent analyses qualify this interpretation. First, the Forward and Reverse models gave only partial credit to H2. In contrast to enjoyment, initial boredom was unrelated to later engagement, and vice versa. The boredom-engagement relation losing its significance after manipulating the temporal precedence suggests that these links might be more situation-sensitive than expected. Goetz et al. (2013) and Goetz and Frenzel (2006) proposed that boredom may vary qualitatively, from indifferent to reactant boredom, according to situational factors and possibly with time. Indifferent boredom pictures a less aversive, enjoyment-compatible profile relative to reactant boredom (Goetz et al., 2013). We found that the magnitude of the negative boredom-enjoyment relation was large at T6 while moderate at T1, which may impact how much boredom at T1 predicted engagement at T6 and vice versa. In other words, the boredom-engagement predictive relations might be sensitive to the extent of the momentary boredom-enjoyment compatibility.

Second, in line with Andres et al. (2019), enjoyment and engagement were found to reciprocally influence each other, providing partial support for the applicability of the expected reciprocal causation assumptions between emotions and learning-related outcomes in the ITS context (H5) (Pekrun et al., 2006). However, the reciprocal influence of enjoyment and engagement and their contributions to final task performance were largely explained by the concomitance of their respective coevals. For example, the association between engagement (T6) and final performance (Tn) was explained by enjoyment (T6), possibly reflecting the multidimensional state of flow (Csikszentmihalyi, 1990). This also supports Sinatra et al.'s (2015) claim that focusing on a single dimension of engagement (e.g., affective) disregards the fact that other dimensions (e.g., behavioral, cognitive) actually occur simultaneously. Moreover, while task performance did not play a major role in the enjoyment-engagement reciprocal relations at an early stage (from T1 to T6), its contribution at T6 became determinant in explaining the influence of enjoyment (T6) on final performance (Tn) and learning gains (posttest) (cf. sensitivity analyses in Table 1). To say the least, this unexpectedly deep and changing intertwinement of the CVT components is not easily reconcilable with the theory positing separate classes of components causing each other rather monolithically. This pattern of results seems more consistent with an understanding of learning as a dynamic and flexible schema of multidimensional components. Ainley (2017) and Hidi and Renninger (2006) emphasize the changing balance in the mental unit or schema among affective and cognitive components of learning to the extent of past interactions with the task, where the expansion of knowledge (e.g., the increasing task performance) progressively co-occurs with affective processes to define the interest experience. Future ITS studies might clarify this point by employing latent variable approaches that consider the trajectories of multiple dimensions activated simultaneously, as a way to reflect Ainley's (2017) time-varying schemas of interest/engagement.

4.3. Limitations

The present study has a number of limitations. One of them has to do with the lack of assessment of student appraisals of value and control with respect to the use of the ITS itself, and inversely, the lack of assessments with respect to activity emotions targeting the broader topic of climate change. This information might have shed important light on the disconnect between topic-related cognitive appraisals and task-related activity emotions found here. Furthermore, considering this distinction might be useful to examine potential reciprocal processes among and across components pertaining to these two domains (topic and technology). Second, we focused on a restricted set of emotions, enjoyment and boredom. Although these are important due to their ambivalent effects and are well documented (Camacho-Morles et al., 2021), allowing for study comparisons, other achievement emotions might by more informative or relevant with regard to ITSs, especially confusion, which is thought to be central for complex learning material (D'Mello & Graesser, 2012). Another limitation concerns the great difficulty of maintaining a large sample size beyond the first 1-h session due to the large variability in students' completion time (see Appendix B). Placing the focus on the first session has limited our analyses with respect to investigations of the temporal precedence of the CVT components that may benefit from more powerful methods at short time scales, such as automatic interaction-based affect detection techniques (Andres et al., 2019; Baker & Ocumpaugh, 2014).

4.4. Conclusions for theory and educational practice

Although CVT studies suggest an overall applicability of CVT to digital learning environments by demonstrating the existence of pathways connecting achievement to its determinants, the existing studies provide only partial and fragmentary support for the applicability of the theory in different contexts, when constructs are examined separately. For example, considering relations between cognitive appraisals and activity emotions, there is to date no systematic replicability for the equation that appraisal A predicts achievement emotion B, in context C, using digital learning technology D, in subject E, etc. (Bewerger et al., 2022). Our study contributes to theory development in the context of CVT with evidence that partially supports the application of CVT to ITSs. Our results indicate that the psychological functionality of a 'positive path' – in which students' intrinsic value predicts enjoyment, which in turn results in higher engagement and thus higher performance and learning gains – might be valid across contexts, but the findings also question the monolithic relations among classes of CVT components by revealing their deep intertwinement during ITS activity. Our results nevertheless suggest the relative importance of specific task value facets for learning in ITS environments. In particular, the findings make it clear that intrinsic value facets were most relevant for learning processes.

More speculatively, one possible explanation that needs further investigation might be that the situational interest elicited by the structure of the digital learning environment prevails over the instrumental (i.e., extrinsic) interest that students ascribe to the topic-related content of learning. Our results on the relative importance of single task-value facets emphasize that the complex psychological processes described in CVT warrant further research, especially favoring experimental and comparative approaches that disentangle the contribution of the above variety of factors, including the causal influence of the type of digital technology.

The present study has practical implications regarding the implementation of ITSs embedded in open-ended learning environments in classrooms. First, the findings clearly suggest that Betty's Brain is a suitable tool for learning about climate change (Biswas et al., 2016) when students perform reasonably well on the task. From pretest to posttest, the average learning gain was about 14.4%. Second, the task is appropriate for eliciting enjoyment and engagement sufficiently to potentialize performance and, consequently, learning. It is good news that students reported boredom levels below midrange on average – relative to enjoyment and engagement, which measured around and above midrange, respectively – indicating that the use of the ITS to learn about climate change did not lead to any major reluctance.

More speculatively, however, the disconnect between topic-related interest and boredom suggests ambivalent implications requiring teachers' attention. On the positive side, for students initially uninterested in the topic, the discovery of the new technology might suffice to trigger situational interest eventually leading to positive learning outcomes. This does not rule out a possible negative side, though: For some students initially interested in the topic, the ITS experience on its own right might deceive and thus interfere with learning. For this reason, we encourage teachers using ITS technologies to be vigilant regarding students' levels of boredom with regard to the technology as a means to achieve learning objectives. Short screening sessions while working with the ITS might be informative when identifying students who might or might not benefit from it.

Declaration of interest statement

Declaration of interest: none.

Funding statement

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence" - project number 390523135. The funding source was not involved in the study design; in the collection, analysis, or interpretation of the data; in the writing of the report; or in the decision to submit the article for publication.

Data statement

The data and code analysis that support the findings of this study are available on request from the corresponding author.

Ethical statement

This study received approval from the University of Potsdam Ethics Committee (number 82/2020) and from the Brandenburg Ministry of Education, Youth and Sports (number 82-E1-2020). The study complies with APA ethical standards in the treatment of the human sample.

Author statement

Johann Chevalère: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing - review & editing., Hae Seon Yun: Conceptualization, Investigation, Methodology, Software, Supervision, Validation, Writing - review & editing., Anja Henke: Data curation, Resources, Validation, Writing - review & editing., Niels Pinkwart: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing - review & editing., Verena V. Hafner: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing - review & editing., Rebecca Lazarides: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing - review & editing.

Acknowledgments

We would like to thank all students participating in this study as well as the teachers and school headmasters who made this study possible. We also thank Dustin Börner, Lea Trenkel, André Seidel, Volha Taliaronak, Jonas Frenkel, Jako Roder and Nina Brendel for their help and support. We sincerely thank two anonymous reviewers for their helpful comments to improve the quality of the paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2023.101799.

References

Acosta-Gonzaga, E., & Ramirez-Arellano, A. (2021). The influence of motivation, emotions, cognition, and metacognition on students' learning performance: A comparative study in higher education in blended and traditional contexts. *Sage Open, 11*(2). https://doi.org/10.1177/21582440211027561

Ainley, M. (2017). Interest: Knowns, unknowns, and basic processes. In P. A. O'Keefe, & J. M. Harackiewicz (Eds.), *The science of interest* (pp. 3–24). Springer International Publishing/Springer Nature. https://doi.org/10.1007/978-3-319-55509-6_1.

Alkhatlan, A., & Kalita, J. (2018). *Intelligent tutoring systems: A comprehensive historical survey with recent developments. ArXiv:1812.09628* http://arxiv.org/abs/1812.09628.

Andres, J., Ma, A. L., Ocumpaugh, J., Baker, R. S., Slater, S., Paquette, L., Jiang, Y., Karumbaiah, S., Bosch, N., Munshi, A., Moore, A., & Biswas, G. (2019). Affect sequences and learning in betty's brain. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 383–390. https://doi.org/10.1145/3303772.3303807

Arroyo, I., Mehranian, H., & Woolf, B. P. (2010). *Effort-based tutoring: An empirical approach to intelligent tutoring. Educational data mining 2010*. June 11-13. Pittsburgh, PA, USA: The 3rd International Conference on Educational Data Mining http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=F2AD57F10AE1E35BABADEDADB646A6B7?doi=10.1.1.174.6048.

Baker, R. S. J.d., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223–241. https://doi.org/10.1016/j.ijhcs.2009.12.003

Baker, R. S. J.d., & Ocumpaugh, J. (2014). Interaction-based affect detection in educational software. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), *The oxford handbook of affective computing*. Oxford University Press.

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72*(5), 593. https://doi.org/10.1037/0022-0663.72.5.593

Berweger, B., Born, S., & Dietrich, J. (2022). Expectancy-value appraisals and achievement emotions in an online learning environment: Within- and between-person relationships. *Learning and Instruction, 77*, Article 101546. https://doi.org/10.1016/j.learninstruc.2021.101546

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From design to implementation to practice a learning by teaching system: betty's brain. *International Journal of Artificial Intelligence in Education, 26*(1), 350–364. https://doi.org/10.1007/s40593-015-0057-9

Butz, N., Stupnisky, R., & Pekrun, R. (2015). Students' emotions for achievement and technology use in synchronous hybrid graduate programmes: A control-value approach. *Research in Learning Technology, 23*, 1–16. https://doi.org/10.3402/rlt.v23.26097

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review, 33*(3), 1051–1095. https://doi.org/10.1007/s10648-020-09585-3

Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology, 18*(4), 334–352. https://doi.org/10.1007/s10956-009-9180-4

Chevalère, J., Lazarides, R., Yun, H. S., Henke, A., Lazarides, C., Pinkwart, N., & Hafner, V. V. (2023). Do instructional strategies considering activity emotions reduce students' boredom in a computerized open-ended learning environment? *Computers & Education.* , Article 104741. https://doi.org/10.1016/j.compedu.2023.104741

Cloude, E. B., Wortha, F., Dever, D., & Azevedo, R. (2020). How do emotions change during learning with an intelligent tutoring system? Metacognitive monitoring and performance with MetaTutor. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.

Daniels, L., & Stupnisky, R. H. (2012). Not that different in theory: Discussing the control-value theory of emotions in online learning environments. *Internet and Higher Education*. https://doi.org/10.1016/J.IHEDUC.2012.04.002

D'Mello, S. K. (2021). Improving student engagement in and with digital learning technologies. *Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, 79–104. https://doi.org/10.1787/8a451974-en

D'Mello, S., & Graesser, A. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS), 2*(4), 1–39.

D'Mello, S., & Graesser, A. (2010). Modeling cognitive-affective dynamics with hidden markov models. *Proceedings of the Annual Meeting of the Cognitive Science Society, 32* (32). https://escholarship.org/uc/item/06b952g5.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

D'Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems, 22*(4), 53–61. https://doi.org/10.1109/MIS.2007.79

Duffy, M. C., Lajoie, S. P., Pekrun, R., & Lachapelle, K. (2020). Emotions in medical education: Examining the validity of the Medical Emotion Scale (MES) across authentic medical learning environments. *Learning and Instruction, 70*, Article 101150. https://doi.org/10.1016/j.learninstruc.2018.07.001

Eccles, J. S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. In A. J. Elliot, & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105–121). Guilford Publications.

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, Article 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall. https://doi.org/10.1007/978-1-4899-4541-9

Elwert, F. (2013). Graphical causal models. In S. Morgan (Ed.), *Handbook of causal analysis for social research. Handbooks of sociology and social research*. Springer. https://doi.org/10.1007/978-94-007-6094-3_13.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 38*(4), 281–288. https://doi.org/10.1016/j.cedpsych.2013.06.001

Forbes-Riley, K., & Litman, D. J. (2009). Adapting to student uncertainty improves tutoring dialogues. In V. Dimitrova, et al. (Eds.), *Artificial intelligence in education* (pp. 33–40). IOS Press. https://doi.org/10.3233/978-1-60750-028-5-33.

Forsblom, L., Pekrun, R., Loderer, K., & Peixoto, F. (2022). Cognitive appraisals, achievement emotions, and students' math achievement: A longitudinal analysis. *Journal of Educational Psychology, 114*(2), 346–367. https://doi.org/10.1037/edu0000671

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics —a "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education, 22*(4), 497. https://doi.org/10.1007/BF03173468

Goetz, T., & Frenzel, A. C. (2006). Phänomenologie schulischer Langeweile [Phenomenology of boredom at school]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie [Journal of Developmental and Educational Psychology], 38* (4), 149–153. https://doi.org/10.1026/0049-8637.38.4.149

Goetz, T., Frenzel, A. C., Hall, N. C., Nett, U. E., Pekrun, R., & Lipnevich, A. A. (2013). Types of boredom: An experience sampling approach. *Motivation and Emotion, 38*, 401–419. https://doi.org/10.1007/s11031-013-9385-y

Goetz, T., Keller, M. M., Lüdtke, O., Nett, U. E., & Lipnevich, A. A. (2020). The dynamics of real-time classroom emotions: Appraisals mediate the relation between students' perceptions of teaching and their emotions. *Journal of Educational Psychology, 112* (6), 1243–1260. https://doi.org/10.1037/edu0000415

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & Preckel, F. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology, 39*(3), 188–205. https://doi.org/10.1016/j.cedpsych.2014.04.002

Graesser, A. C., Greenberg, D., Frijters, J. C., & Talwar, A. (2022). Using AutoTutor to track performance and engagement in a reading comprehension intervention for adult literacy students. *Revista Signos. Estudios de Lingüística, 54*(107). http://www.revistasignos.cl/index.php/signos/article/view/776.

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58*(1), 78–79. https://doi.org/10.1037/0003-066X.58.1.78

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*, 53–60. https://doi.org/10.21427/D7CF7R

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2010). Causal mediation analysis using R. In H. Vinod (Ed.), *Advances in social science research using R. Lecture notes in statistics* (Vol. 196). Springer. https://doi.org/10.1007/978-1-4419-1764-5_8.

Jacobson, M. J., Markauskaite, L., Portolese, A., Kapur, M., Lai, P. K., & Roberts, G. (2017). Designs for learning about climate change as a complex system. *Learning and Instruction, 52*, 1–14. https://doi.org/10.1016/j.learninstruc.2017.03.007

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press. n.d. https://www.guilford.com/books/Principles-and-Practice-of-Structural-Equation-Modeling/Rex-Kline/9781462523344. (Accessed 25 July 2022)

Kögler, K., & Göllner, R. (2018). Control-value appraisals predicting students' boredom in accounting classes: A continuous-state-sampling approach. *Empirical Research in Vocational Education and Training, 10*(1), 4. https://doi.org/10.1186/s40461-018-0065-8

Koh, A. W. L., Lee, S. C., & Lim, S. W. H. (2018). The learning benefits of teaching: A retrieval practice hypothesis. *Applied Cognitive Psychology, 32*(3), 401–410. https://doi.org/10.1002/acp.3410

Krapp, A. (2007). An educational-psychological conceptualisation of interest. *International Journal for Educational and Vocational Guidance, 7*(1), 5–21. https://doi.org/10.1007/s10775-007-9113-9

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., & Stanat, P. (2002). *Dokumentation der Erhebungsinstrumente [PISA 2000: Documentation of assessment instruments*. Max Planck Institute für Bildungsforschung [Max Planck Institute for Human Development. PISA 2000.

Land, S. M. (2000). Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research & Development, 48*(3), 61–78. https://doi.org/10.1007/BF02319858

Lazarides, R., & Chevalère, J. (2021). Artificial intelligence and education: Addressing the variability in learners' emotion and motivation with adaptive teaching assistants. *Bildung und Erziehung, 74*(3), 264–279. https://doi.org/10.13109/buer.2021.74.3.264

Lichtenfeld, S., Pekrun, R., Marsh, H. W., Nett, U. E., & Reiss, K. (2022). Achievement emotions and elementary school children's academic performance: Longitudinal models of developmental ordering. In *Journal of educational psychology, No pagination specified-No pagination specified*. https://doi.org/10.1037/edu0000748

Li, C., & Wei, L. (2022). Anxiety, enjoyment, and boredom in language learning amongst junior secondary students in rural China: How do they contribute to L2 achievement?. In *Studies in second language acquisition* (Vols. 1–16). https://doi.org/10.1017/S0272263122000031

Loderer, K., Pekrun, R., & Lester, J. C. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction, 70*, Article 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918. https://doi.org/10.1037/a0037123

MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Personality and Social Psychology Review, 19*(1), 30–43. https://doi.org/10.1177/1088868314542878

Mercan, F. C. (2020). Control-value theory and enjoyment of science: A cross-national investigation with 15-year-olds using PISA 2006 data. *Learning and Individual Differences, 80*, Article 101889. https://doi.org/10.1016/j.lindif.2020.101889

Munshi, A., Rajendran, R., Penn, J. O., Biswas, G., Baker, R. S., & Paquette, L. (2018). Modeling learners' cognitive and affective states to scaffold srl in open-ended learning environments. In *UMAP 2018 - proceedings of the 26th conference on user modeling, adaptation and personalization* (Vols. 131–138). https://doi.org/10.1145/3209219.3209241

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz, & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). Academic Press. https://doi.org/10.1016/B978-012372545-5/50003-4.

Pekrun, R., Goetz, T., Daniels, L., Stupnisky, R., & Perry, R. (2010). Boredom in achievement settings: Exploring control-value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology, 102*(3), 531–549. https://doi.org/10.1037/a0019243

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., Hall, N. C., Goetz, T., & Perry, R. P. (2014). Boredom and academic achievement: Testing a model of reciprocal causation. *Journal of Educational Psychology, 106*(3), 696–710. https://doi.org/10.1037/a0036006

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development, 88*(5), 1653–1670. https://doi.org/10.1111/cdev.12704

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259–282). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-2018-7_12.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In *International handbook of emotions in education* (pp. 120–141). Routledge/Taylor & Francis Group.

Pekrun, R., & Stephens, E. J. (2010). Achievement emotions: A control-value approach. *Social and Personality Psychology Compass, 4*(4), 238–255. https://doi.org/10.1111/j.1751-9004.2010.00259.x

Pintrich, P. R., & Zusho, A. (2002). The development of academic self-regulation: The role of cognitive and motivational factors. In *Development of achievement motivation* (pp. 249–284). Academic Press. https://doi.org/10.1016/B978-012750053-9/50012-7.

Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In *Emotions, technology, design, and learning* (pp. 131–161). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-801856-9.00007-4.

Putwain, D. W., Pekrun, R., Nicholson, L. J., Symes, W., Becker, S., & Marsh, H. W. (2018). Control-value appraisals, enjoyment, and boredom in mathematics: A longitudinal latent interaction analysis. *American Educational Research Journal, 55*(6), 1339–1368. https://doi.org/10.3102/0002831218786689

Putwain, D. W., Schmitz, E. A., Wood, P., & Pekrun, R. (2021). The role of achievement emotions in primary school mathematics: Control–value antecedents and achievement outcomes. *British Journal of Educational Psychology, 91*(1), 347–367. https://doi.org/10.1111/bjep.12367

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rohrer, J. M., Hünermund, P., Arslan, R. C., & Elson, M. (2022). That's a lot to PROCESS! Pitfalls of popular path models. *Advances in Methods and Practices in Psychological Science, 5*(2). https://doi.org/10.1177/25152459221095827

Rosseel. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. http://www.jstatsoft.org/v48/i02/.

Ruthig, J., Perry, R., Hladkyj, S., Hall, N., Pekrun, R., & Chipperfield, J. (2008). Perceived control and emotions: Interactive effects on performance in achievement settings. *Social Psychology of Education, 11*, 161–180. https://doi.org/10.1007/s11218-007-9040-0

Segedy, J. R. (2014). *Adaptive scaffolds in open-ended computer-based learning environments [Doctoral dissertation]*. Vanderbilt University. https://ir.vanderbilt.edu/handle/1803/12486.

Simonton, K. L., & Garn, A. C. (2020). Control–value theory of achievement emotions: A closer look at student value appraisals and enjoyment. *Learning and Individual Differences, 81*, Article 101910. https://doi.org/10.1016/j.lindif.2020.101910

Simonton, K. L., Garn, A. C., & Solmon, M. A. (2017). Class-related emotions in secondary physical education: A control-value theory approach. *Journal of Teaching in Physical Education, 36*(4), 409–418. https://doi.org/10.1123/jtpF.2016-0131

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist, 50*(1), 1–13. https://doi.org/10.1080/00461520.2014.1002924

Statista. (2022). *Number of general schools in Germany in the 2020/2021 school year, by school type*. https://www.statista.com/statistics/1182435/number-of-general-schools-by-school-type-germany/?fbclid=IwAR2iZ4avER20D9nlfZV9pkjGuBJ8roXnB-bwmUKWJkZJI64KDgT2KH0pn4k. (Accessed 25 July 2022).

Steinmayr, R., & Spinath, B. (2010). Konstruktion und erste Validierung einer Skala zur Erfassung subjektiver schulischer Werte (SESSW) [Construction and first validation of a scale assessing subjective educational task values]. *Diagnostica, 56*(4), 195–211. https://doi.org/10.1026/0012-1924/a000023

Taber, K. S. (2017). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R Package for causal mediation analysis. *Journal of Statistical Software, 59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05

Tze, V. M. C., Li, J. C.-H., & Parker, P. C. (2021). A mediation analysis of emotions based on the control-value theory. In *Current psychology: A journal for diverse perspectives on diverse psychological issues*. https://doi.org/10.1007/s12144-021-01840-2

Wild, S. (2022). Trajectories of subject-interests development and influence factors in higher education. In *Current psychology: A journal for diverse perspectives on diverse psychological issues*. Advance online publication. https://doi.org/10.1007/s12144-021-02691-7.

Xiao, Z., Higgins, S., & Kasim, A. (2019). An empirical unraveling of lord's paradox. *The Journal of Experimental Education, 87*(1), 17–32. https://doi.org/10.1080/00220973.2017.1380591

Zhao, T., Fu, Z., Lian, X., Ye, L., & Huang, W. (2021). Exploring emotion regulation and perceived control as antecedents of anxiety and its consequences during covid-19 full remote learning. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.675910