

RESEARCH ARTICLE

Open Access



Compared to what? Effects of social and temporal comparison standards of feedback in an e-learning context

Marc P. Janson^{1*} Jan Siebert² and Oliver Dickhäuser¹

*Correspondence:
mjanson@mail.uni-mannheim.de

¹School of Social Sciences,
University of Mannheim, A5 6,
68131 Mannheim, Germany

²Hasso-Plattner-Institute,
University of Potsdam,
14482 Potsdam, Germany

Abstract

Performance evaluation is based on comparison standards. Results can either be contrasted to former results (temporal comparison) or results of others (social comparison). Existing literature analyzed potential effects of teachers' stable preferences for comparison standards on students' learning outcomes. The present experiments investigated effects of learners' own preferences for comparison standards on learning persistence and performance. Based on research and findings on person-environment-fit, we postulated a fit hypothesis for learners' preferences for comparison standards and framed feedback on learning persistence and performance. We tested our hypotheses in two separate experiments ($N=203$ and $N=132$) using different manipulations of framed feedback (temporal vs. social) in an e-learning environment, thus establishing high ecological validity and allowing objective data to be collected. We found first evidence for beneficial effects of receiving framed feedback towards own preferences on learning persistence and performance in our experiments. We tested fluency as a possible underlying psychological mechanism in our second experiment and observed a larger fit effect on learning persistence under disfluency. The results are discussed regarding a new theoretical perspective on the concept of preferences for comparison standards as well as opportunities for adaptive e-learning.

Keywords: Self-regulated learning, e-Learning, Reference norm orientation, Frame of reference, Feedback

Introduction

Feedback is considered to be one of the most powerful influences on learning outcomes (Hattie, 2009) that play a key role in many learning situations. However, some feedback has a more beneficial impact than others (Butler, 1987) and it has been pointed out that this impact is not necessarily always positive or directly related to behavior (Balcazar et al., 1985; Hattie & Timperley, 2007; Ilgen et al., 1979; Latham & Locke, 1991). The latter conclusion is based on a large amount of empirical support reviewed by Kluger and DeNisi (1996). Their meta-analysis of 608 studies on feedback interventions revealed a mean positive effect of feedback on performance ($d=0.41$), but feedback decreased performance in about one third of the studies. Literature on feedback interventions (Kluger & DeNisi, 1996,

1998) distinguishes different feedback forms like informative feedback, solely stating what is correct and what is wrong and elaborative feedback, which gives information about opportunities to improve learning. We are aware of research concluding that elaborative feedback is more effective than solely providing informative feedback (Bangert-Drowns et al., 1991). However, elaborative feedback is harder to provide, especially in an automated way in digital learning environments. Therefore, we aim to provide insights into how informative feedback—as the most basal form of feedback—can be modified to be more beneficial for learners' persistence and performance.

We use a differential perspective on feedback perception and propose certain feedback types to be more effective for specific subgroups. Within their feedback intervention theory, Kluger and DeNisi (1996, 1998) already state that behavior regulations depend on a comparison of the provided feedback and individuals' goals or standards and we will focus on these comparison standards. Educational scholars have discussed the role of different comparison standards on students' motivation and performance. The construct of *reference norm orientation* was suggested by Rheinberg (see Rheinberg, 2001, for a summary) in order to describe stable preferences towards different comparison standards of performance evaluation. Research revealed that preferences for comparison standards of teachers have a significant impact on learners' motivation (Dickhäuser et al., 2017; Lüdtke et al., 2005; Retelsdorf & Günther, 2011; Rheinberg, 2001). However, no existing literature focuses on the orientation of students and their preferences for different comparison standards in real learning contexts and potential effects on learning outcomes.

For the very first time, we investigated potential effects of those personal preferences for comparison standards of learners and provided feedback on actual learning outcomes, i.e., learning persistence and learning performance. In line with (a) the differential approach of feedback interventions theory (Kluger & DeNisi, 1996, 1998), (b) the general idea of *aptitude-treatment-interaction* (ATI, Cronbach & Snow, 1969) that students with different abilities benefit more or less from different interventions, treatments or instructional techniques and c) a large corpus of fit effects in psychological research (e.g., Edwards, 1991; Higgins, 2000; Porter & Umbach, 2006), we will propose interactions between personal preferences and different feedback types. Based on the relevance of designing good e-learning environments (Al-Fraihi et al., 2020; Mohammadhassan et al., 2022), the fact that learning success in digital environments is dependent on interindividual differences in self-regulation (Aparicio et al., 2017) and given that the field of e-learning opens up new possibilities to adapt the learning context to learners' specific needs (Seo et al., 2021; Shute & Towle, 2003), we investigated whether such postulated fit effects of different comparison standards affect persistence and performance on an e-learning platform for exam preparation. The present manuscript provides new theoretical perspectives on learners' preferences for comparison standards, their potential fit with the comparison frames of the feedback, and their importance for self-regulated learning. For overview of the constructs investigated in this manuscript see Table 1.

Theory

Comparison standards for feedback and individual preferences

What is a good performance? The answer to this question is not as easy as it seems at first glance. There might be situations in which a good performance can be derived

Table 1 Central constructs investigated in this manuscript

Feedback	Feedback is information about performance and it is one of the most powerful factors impacting individuals' learning outcomes (Hattie, 2009; Hattie & Timperley, 2007). It should be noted that feedback effects are in general beneficial, but can also be negative (Hattie & Timperley, 2007; Kluger & DeNisi, 1996, 1998). Within the feedback-intervention-theory it is stated that feedback needs to be meaningful for individuals to be effective (Kluger & DeNisi, 1998).
Comparison standards	Comparison standards are used to describe performance in different means (Rheinberg, 2001). On a criterial level performance can be compared if absolute standards exist (e.g., to achieve a driver's license you must provide skills and abilities to participate in public traffic safely and responsibly). Besides those absolute standards performance can also be compared with own former performance (temporal comparison standard) or performance of others (social comparison standard).
Preferences for comparison standards/frame of reference/reference norm orientation	Individuals differ in their preferences for temporal and social comparison standards (Dickhäuser et al., 2017; Lüdtke et al., 2005; Retelsdorf & Günther, 2011; Rheinberg, 2001). Individuals may have low or high preferences for one or both comparison standards as those are independent. Research for preferences for comparison (often called frame of reference or reference norm orientation) has focused on the effects of preferences of comparison standards of teachers on students' motivation and performance, stating a superiority of preferences for temporal comparisons standards.
Fit effects	Fit effects emerge if situational cues meet individual preferences. They are a common construct in psychological research (e.g., regulatory fit; Higgins, 2000) and are associated with positive effects on persistence and performance.
Fluency	Fluency (Alter & Oppenheimer, 2009) is the perceived ease of processing. This ease of processing is a mediating mechanism for many phenomena in social psychology and fluency can be affected by different facets. One facet of fluency is perceptual fluency emerging from easy to read information compared to hard to identify information (e.g., Eitel et al., 2014; Hansen et al., 2008).

from a certain objective criterion. A good high jump performance, for example, can be evaluated by whether the jumper dislodged the bar or not. But what is a good learning performance? In situations where no absolute criterion as comparison standard can be identified, performance can be either evaluated in contrast to the performance of others or in contrast to own former performance. These comparison standards have a long research tradition as, for example, the *social comparison theory* (Festinger, 1954) and *temporal comparison theory* (Albert, 1977) have addressed these two types of comparisons. Former research has focused on the effects of different feedback types on motivational and performance outcomes by comparing effects of task-oriented feedback and competitive feedback conditions (Butler, 1987; Covington & Omelich, 1984; Shih & Alexander, 2000).

The *achievement goal theory* (Dweck & Leggett, 1988; Elliot & McGregor, 2001) provides a good starting point for a closer look at learners' goals and preferences for comparison standards. The achievement goal theory states that two motivational systems guide learners' motivation in achievement-related situations. Mastery goals describe tendencies to improve knowledge while performance goals reflect tendencies to demonstrate competence. The theory has been further developed and researchers such as Nicholls (Nicholls, 1984, 1989) claim that mastery goals are task-oriented and performance goals are ego-orientated. This implies different reference points for learners to compare actions. The task-orientation refers solely to improvement in the task and therefore needs information of former and current performance (temporal comparisons). The ego-orientation compares current performance

to the performance of others for the sake of evaluation (social comparison). Educational psychology researchers have also solely focused on these preferences for comparison standards, especially on teachers and their respective effects on learners. They distinguish teachers' stable preferences for social and temporal comparisons as teachers' *reference norm orientation* (Retelsdorf & Günther, 2011; Rheinberg, 2001) or teachers' *frame of reference* (Dickhäuser et al., 2017; Lüdtke et al., 2005). In previous research, a stronger teacher preference for temporal comparison standards was associated with more beneficial outcomes, like more adaptive instructional styles promoting students' comprehensive learning (Retelsdorf & Günther, 2011), higher self-concept and more adaptive mindsets of students (Dickhäuser et al., 2017; Lüdtke et al., 2005). It is important to note that preferences for temporal and social comparison standards are not endpoints of one continuum, but rather independent constructs that moderately correlate (Dickhäuser et al., 2017; Retelsdorf & Günther, 2011).

Benefits of fit: when preferences meet standards

Instead of teachers' frame of reference, we focus on reference norm orientation of learners as the preference for a temporal or social comparison standard for the evaluation of own performance. Positive effects of a higher preference for temporal comparison standards (and negative effects of a higher preference for social comparison standards) on learners' persistence and performance can be derived from the analogous evidence from teachers' preferences for comparison standards (Dickhäuser et al., 2017; Lüdtke et al., 2005; Retelsdorf & Günther, 2011; Rheinberg, 2001).

However, in line with the feedback intervention theory suggested by Kluger and DeNisi (1998), one could argue that individuals with different preferences towards standards for performance evaluation benefit differently from different types of feedback. For example, individuals with a high preference for social comparison standards should prefer feedback that gives information about individuals' performance contrasted to the performance of others. If such preferences are met by the framed feedback given by a teacher or a learning system, i.e., if such students actually receive feedback based on social comparison, this should result in beneficial outcomes. To the best of our knowledge, we are the first to propose such a fit effect of preferences for comparison standards and framed feedback. However, fit effects in general are a well-established phenomenon in psychological literature: e.g., person-job fit in industrial and organizational psychology (Edwards, 1991), aptitude-treatment-interactions or choice of major subjects in college in educational psychology (Cronbach & Snow, 1969; Porter & Umbach, 2006) or regulatory fit in social psychology (Higgins, 2000). Indeed, the regulatory focus theory (Higgins, 1998) also refers to the perception of feedback. Positive effects of a fit between situational regulatory cues and chronic regulatory orientations have been identified in various disciplines, e.g., beneficial effects on health behavior (Hong & Lee, 2008). A recent meta-analysis underlines the robustness of the phenomena (Motyka et al., 2014) and Keller and Bless (2006) found evidence that induced fit during tests enhances test performance

of high school students. Therefore, it is plausible that in an educational context, fit effects on learning outcomes are observable.

Motivated by fitting feedback

Which outcomes are affected by fit? For regulatory fit it has been shown that many fit effects occur due to a motivational or volitional process (Motyka et al., 2014). In one of their studies, Spiegel and colleagues (2004) asked students to write a non-compulsory report in their leisure time. Students first reported their chronic regulatory orientations and were then told to first think about time slots, places and techniques, which were either favorable for fulfilling the task (promotion) or hindered fulfilling the task (prevention). More participants under fit (where the instruction met their chronic regulatory orientation) handed in reports than participants under misfit (where instruction did not meet their chronic regulatory orientation). This can be explained through higher persistence caused by matching instructions and personal preferences. Freitas and Higgins (2002) postulated “that another determinant of action enjoyment is the action’s fit with one’s phenomenological state, such as one’s mood, mind-set, or regulatory focus” (p. 1) and directly assessed participants’ feelings about goal pursuit under fit. Under regulatory fit, participants reported higher task enjoyment and higher willingness to repeat the task than under regulatory misfit. We propose that such processes are triggered by the joint operation of preferences for comparison standards and framed feedback: For learners who understand a good performance in terms of improvement, a feedback intervention providing information about former performance and current performance should be more motivating and foster more persistence than for learners who do not prefer comparisons of current performance with former performance. Indeed, the latter should yield lower self-regulated learning activities as it may be non-informative or even aversive.

Persistence in e-learning environments is also dependent on individuals’ self-regulation (Aparicio et al., 2017; Lee & Lee, 2008). Consequently, based on the general framework of fit effects (Kluger & DeNisi, 1996, 1998) and transferring the evidence and mechanisms of regulatory fit (Freitas & Higgins, 2002; Higgins, 2000) to learners’ preferences for comparison standards and the context of feedback, we propose similar effects on learning persistence in an e-learning environment. To respect the bi-dimensionality of preferences for comparison standards (Dickhäuser et al., 2017; Retelsdorf & Günther, 2011), we propose separate hypotheses for both orientations in the present work:

H1a: Feedback with a temporal (social) comparison standard will enhance (decrease) learning persistence for learners with higher preferences for temporal comparison standards.

H1b: Feedback with a social (temporal) comparison standard will enhance (decrease) learning persistence for learners with higher preferences for social comparison standards.

Cognitive benefits of fit

Beside those motivational effects of induced fit between preferences for comparison standards and presented type of feedback, one could speculate about additional processes. Keller and Bless (2006) discussed whether cognitive processes could explain the observed fit effects in students' test performance as well. Received instructions (or feedback) which are framed in the reader's preferred way, should be easier to process than instructions which do not meet readers' strategic orientations. Therefore, fit should foster information processing while misfit might bind cognitive resources. This assumption is also indirectly supported by findings on effects of (dis-)congruency. Several studies have identified reduced cognitive performance, when congruency of information with stereotypes or expectancies is not given (Macrae et al., 1993; Stangor & McMillan, 1992). Transferring those results to the context of feedback framed with different comparison standards, we propose higher performance for individuals who receive feedback framed in correspondence to their preferences. We aim to investigate those effects on learning performance in an e-learning environment as well as previous research outlining the importance of enhanced motivation on e-learning performance (Castillo-Merino & Serradell-López, 2014). On the other side, misfitting feedback should reduce learning performance as cognitive resources are limited. This is also in line with *cognitive load theory* stating that learning performance is dependent on the amount of extraneous cognitive load, information that is hard to process or irrelevant (Sweller, 2010). Taken together, learners with a high preference for temporal comparison standards should process temporal feedback with ease while for learners with low preference for those comparisons, the feedback should lead to misfit and reduce task performance as additional cognitive resources are spent on feedback processing.

It is plausible to assume that the effects are only observable in the presence of the feedback, as cognitive resources should only be strained when difficult-to-process information is presented. Therefore, we predict positive fit effects for those exercises introduced with feedback simultaneously:

H2a: Feedback with a temporal (social) comparison standard will enhance (decrease) learning performance for learners with higher preferences for temporal comparison standards.

H2b: Feedback with a social (temporal) comparison standard will enhance (decrease) learning performance for learners with higher preferences for social comparison standards.

Fit and fluency

With our present research, we also want to investigate possible psychological mechanisms underlying the proposed fit effects. As a foundation, we use research on regulatory focus as "It feels right" is Higgins's (2000) main description of regulatory fit. Freitas and Higgins (2002) identified perceived task enjoyment, willingness to repeat the task or perceived success at a task as an outcome of regulatory fit. It is plausible to assume that the metacognitive experience of perceived ease of processing could explain such outcomes and fit effects in general. The concept of *fluency* (Alter

& Oppenheimer, 2009) refers to this ease of processing, which impacts numerous outcomes like confidence, truth judgments or liking (Hertzog et al., 2003; Reber & Schwarz, 1999; Reber et al., 1998). Perceptual fluency derives from visual ease and impacts affective outcomes as well. Stimuli on highly contrasted backgrounds (i.e., fluent) are judged as prettier than disfluent ones (those on low contrasted backgrounds; Hansen et al., 2008; Reber & Schwarz, 1999; Reber et al., 1998). We believe that effects of feedback on learning outcomes go along with ease of processing as well and propose enhanced learning outcomes or at least enhanced learning persistence via higher ease of processing under fitting feedback. Despite the fact that misfitting feedback should bind additional cognitive resources, fitting feedback should be more easily processed and enhance performance. Fluency (Alter & Oppenheimer, 2009) by definition relies on the same psychological principle as effects of ease of processing. Therefore, we hypothesize that fitting feedback leads to higher ease of processing as well as induced fluency. Because experimental manipulation of processes has more leverage regarding causality, we want to test this by additionally manipulating the perceived ease of processing (Spencer et al., 2005). Also, the manipulation of perceptual fluency is a common method of manipulating the perceived ease of processing in recent research (Eitel et al., 2014; Flavell et al., 2020; Godinho & Garrido, 2021). We propose that an additional manipulation of (perceptual) fluency should alter the strength of fit effects. If fitting feedback has an impact on learning outcomes via higher ease of processing, an additional disfluent cue should diminish those effects. Adding additional fluent cues should strengthen the fit effects, respectively:

- H3a: Fit effects of preferences for temporal comparison standards and framed feedback on learning persistence are stronger under fluent than disfluent conditions.
- H3b: Fit effects of preferences for social comparison standards and framed feedback on learning persistence are stronger under fluent than disfluent conditions.
- H3c: Fit effects of preferences for temporal comparison standards and framed feedback on learning performance are stronger under fluent than disfluent conditions.
- H3d: Fit effects of preferences for social comparison standards and framed feedback on learning performance are stronger under fluent than disfluent conditions.

To sum up, we propose two separate fit effects of preferences for social and temporal comparison standards and framed feedback for two different learning outcome variables and an additional underlying mechanism. We investigated these in two studies. Experiment 1 focused on the fit effects of both orientations and framed feedback (social vs. temporal) on learning time (H1) and performance (H2). Experiment 2 was a conceptual replication (Schwarz & Strack, 2014; Stroebe & Strack, 2014) testing H1 and H2 with an additional manipulation of fluency to test the proposed mechanism behind fit effects (H3). Former research on preferences for comparison standards relies on self-report data (Dickhäuser et al., 2017; Retelsdorf & Günther, 2011), which can be inaccurate or even biased (Pintrich, 2004; Winne et al., 2002). For our studies, we investigated our proposed fit effects in a digital environment using e-learning software, which provides users with exercises. We used available objective learning data instead of relying on self-reports. Furthermore, providing evidence for a fit-effect in an e-learning environment underlines

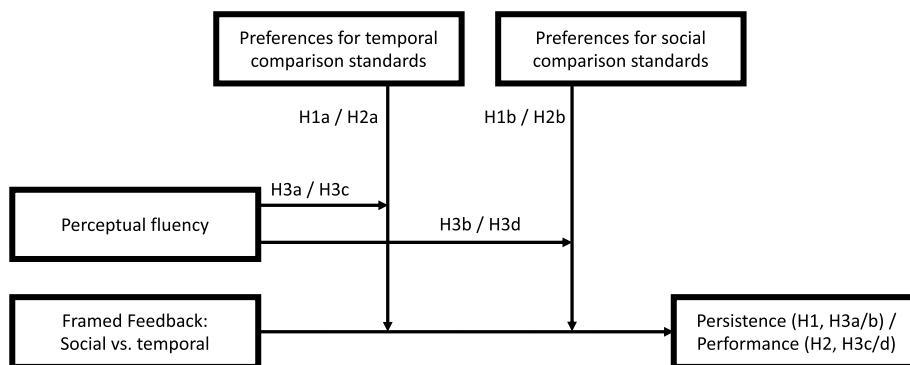


Fig. 1 Conceptual model of fit effects of preferences for comparison standards and feedback with temporal or social comparison standards in dependency of perceptual fluency. The conceptual model is analogous for both outcomes (persistence and performance) investigated in this manuscript

the tremendous opportunities of adaptive e-learning (Shute & Towle, 2003) and leverages research on ATI as the limitations of classroom-settings are mitigated (Corno & Snow, 1986; Cronbach & Snow, 1969). The complete conceptual model of our proposed fit effects of preferences for comparison standards, framed feedback and perceptual fluency on learning persistence and performance is represented in Fig. 1.

Experiment 1

Methods

Participants

203 participants (178 female, 23 male, 2 other) were willing to participate in the experiment in exchange for course credit. Participants' mean age was 21.84 years ($SD=3.76$). We collected data from four different courses and observed users preparing for final exams.¹

Design and overview

We invited users of the e-learning software (Siebert and Janson, 2018) to participate in the experiment during the fall semester 2018. The learning software provides multiple choice questions and statistical problems for psychology students at a German university. We conducted a within-subjects field experiment investigating the effects of individuals' preferences for comparison standards and feedback providing different comparison standards. It is important to note, that participants were not differentiated into groups based on their preferences for comparison standards. Instead, we used their respective orientations as interacting personality trait for the evaluation of two feedback-types either representing temporal or social comparison standards. We chose a within-design for the feedback variation two reasons. First, a within analysis of user behavior in the software provides a much larger dataset controlling for interindividual differences in user behavior. Therefore, the power of this experimental design was higher compared to a between-participant design. Second, as manipulations between users could influence

¹ The courses were "quantitative methods I", "empiric scientific working", "multivariate methods" and "testing & deciding". Note that users could have used the software to prepare for more than one package.

learning outcomes (and exam performance, respectively) a within analysis was more appropriate considering ethical concerns about studies on exam preparation.

Users of the e-learning software were automatically redirected to our web-survey at their first login. After giving informed consent to participate in this experiment, sociodemographic variables were collected. Following, we measured preferences for social and temporal comparison standards of the users in a randomized order. Users were able to start learning after finishing the questionnaire or declining participation in this experiment. Until the exams, participants used the learning software individually. For each single learning session, it was randomly selected whether the social or temporal framed feedback was presented during this particular session and participants got the respective feedback after every tenth item.

Materials

Learning Software. The learning-software (Siebert and Janson, 2018) was designed to overcome possible material effects during learning (remembering items not concepts) by providing exercises, where numbers are generated randomly and cover stories vary about different topics, maximizing the positive effects of testing on learning (Carpenter, 2009; Roediger & Karpicke, 2006). For multiple choice items, alternating answer options are implemented. Those features are combined with an adaptive selection algorithm, which selects items solved correctly in the past less often to make use of the spacing effect (Rawson & Dunlosky, 2011; Son & Simon, 2012).

Preferences for Comparison Standards Assessment. To assess learners' preferences for comparison standards, we used a questionnaire provided by Schöne and colleagues (2004), which provides two different scales for preferences for social and temporal comparison standards. Items like "It would be a good performance if it is better than a previous performance" measure preferences for temporal comparison standards and items like "It would be a good performance if it is better than others' performance" measure preferences for social ones. As the questionnaire only provides three items per scale and internal reliabilities of the temporal scale were only sufficient in some studies ($\alpha=0.55-0.84$), we included two additional items per scale. These new items were "It would be a good performance if you develop positively" and "It would be a good performance if you do more over time" for the temporal orientation and "It would be a good performance if you surpass others' performance" and "It would be a good performance if you do better than others". Additionally, we entered every item as inverted versions of the original ones. These inverted items were worded in terms of the perception of bad performance. Thus, the assessment includes ten items per scale in total. Answers were collected on a five-point Likert-scale using (1) "strongly disagree" and (5) "completely agree" as end-points. All items are included in the Additional file 1.

Social/Temporal Feedback. We implemented feedback providing information either with a social or a temporal comparison standard into the software. The social-framed feedback provided information about the current *elaboration score* (a parameter reflecting total learning performance within the software on a 0–100 scale) of a user compared to the mean elaboration score of all other users of the same package in the last 50 days (in this way, only the performance of peers preparing for the same exam at the same time was presented). For the temporal feedback, we compared the current elaboration-score

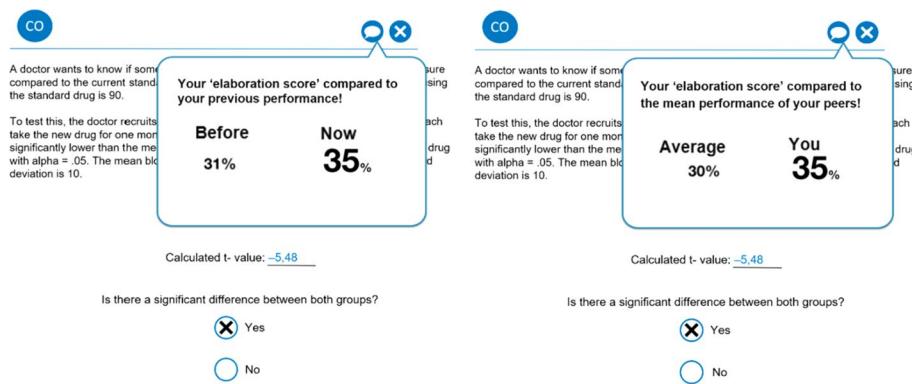


Fig. 2 Feedback types used in experiment 1. Temporal feedback on the left; social feedback on the right

of a user with the last presented elaboration score. For the temporal feedback the presented information was introduced with: “Your ‘elaboration score’ compared to your previous performance!” and the current as well as the elaboration score of the last feedback were displayed. The presented social feedback information was introduced with: “Your ‘elaboration score’ compared to the mean performance of your peers!” and the current as well as the mean elaboration score of the peers were provided. It should be noted that in both conditions the feedback about the current elaboration score was identical and only the comparison standard varied. The different types of feedback are illustrated in Fig. 2.

We pretested the different feedback-types. We asked 20 pretesters to rate illustrations of the feedback, whether “this type of feedback represents a comparison of performance with...” on a seven-point Likert-scale with the endpoints “the performance of others” (1) and “own former performance” (7). Pretesters perceived the temporal feedback as expressing more of a temporal comparison standard than a social comparison standard compared to the social feedback, $t(19)=8.93, p<0.01$.

Learning Outcomes. Learning time as a measurement for learning persistence and learning performance was provided by the logfiles of the learning software. For each learning session the time stamps of the login and the last activity before (automatic) logout were collected.² Learning time was operationalized as the difference score between those two time stamps. Learning performance was operationalized as the number of correct answers right after receiving feedback divided by the number of total received feedback per sessions.

Results

Preliminary analyses and descriptives

We investigated the items for preferences for comparison standards a) to ensure psychometric properties of the used scales and b) to validate the assumption that preferences for temporal and social comparison standards are distinct constructs. Reliability

² This way the data is not blurred by sessions where users forgot to logout. After ten minutes the learning software automatically logs off. The time differences used for analysis are based on login and last activity. Only short absences from the keyboard (e.g., going to the toilet) cannot be distinguished from actual learning time. Such noise in the data should be distributed equally in both conditions and should be negligible.

analysis for both scales using Cronbach's alpha provided evidence for very good internal consistency (temporal: $\alpha=0.89$; social: $\alpha=0.95$). Both scales correlated weakly, $r=0.27$, $p<0.01$, underlining the importance of testing fit effects separately. On average, participants rated medium preferences for temporal comparison standards ($M=3.11$, $SD=0.43$) as well as for social comparison standards ($M=3.05$, $SD=0.37$). Additional descriptive analyses regarding age and gender are included in the Additional file 1. We also inspected the 4108 learning sessions with more than 10 answered exercises (participants in sessions with less did not receive our manipulated feedback). Mean learning time per session was 41.09 min ($SD=38.38$). As learning time was extremely skewed ($sk=2.69$) and to obtain a normal distribution of residuals, we logarithmized the learning time. Regarding session performance, on average 61% of those exercises presented together with the feedback (every tenth item) were answered correctly.

Main analyses

We used the R- 'lme4' package (Bates et al., 2014) to conduct stepwise multi-level regression analyses on session learning time and session performance. We first entered main effects of the preferences for comparison standards and feedback-type (effect-coded with -1 for temporal and 1 for social feedback) in step one and the interactions in step two. The scores for the preferences for temporal and social comparison standards were standardized. For both dependent variables we entered the time till the final exam (standardized) as well as the uninvestigated outcome as another control variable.

Learning persistence. The learning times of the 4108 observed learning session were calculated by the difference of last activity of the learning session and session-onset as a measurement of learning persistence. We conducted a multi-level regression analysis with logarithmized learning time (in minutes) as the dependent variable and type of feedback and preferences for comparison standards as predictors as well as the time till exam as a covariate. Fit effects were entered as interactions between type of feedback and preferences in step two. Sessions were clustered within individuals and subjects to control for different intercepts. All regression coefficients are presented in Table 2. In step one, we observed no significant effects on learning time, despite the control variables indicating longer learning sessions closer to the exam date, $b=0.071$, $p<0.001$, and higher persistence with medium success probability, $b=2.529$, $p<0.001$. Regarding the preferences for comparison standards of learners, we observed only tendencies of higher learning time with higher preference for temporal comparisons, $b=0.048$, $p=0.149$, and lower preferences for social comparisons, $b=-0.038$, $p=0.247$ as well as a small tendency towards longer learning times in sessions with temporal framed feedback, $b=-0.016$, $p=0.146$. Out of the proposed fit effects, the interaction between feedback type and preferences for temporal comparison standards was not significant, $b_{\text{temporal} \times \text{feedback}}=0.002$, $p=0.881$,³ but we found a significant interaction between feedback type and preferences for social comparison standards, $b_{\text{social} \times \text{feedback}}=0.025$, $p=0.034$. Therefore, we can only support hypothesis H1b, stating higher learning time for participants with

³ P-values for our proposed effects can be divided by two, if the interactions are in the stated direction as we explicitly formulated directed hypotheses. However, all reported p-values are two-tailed.

Table 2 Multilevel regression on learning persistence and performance in experiment 1

	Persistence		Performance	
	Model 1 B(SE)	Model 2 B(SE)	Model 1 B(SE)	Model 2 B(SE)
Intercept	7.279*** (0.091)	7.278*** (0.091)	0.531*** (0.046)	0.531*** (0.045)
Session level (L1)				
Time to exam ^a	0.071*** (0.014)	0.071*** (0.014)	0.012* (0.007)	0.012* (0.007)
Session performance ^b	2.529*** (0.110)	2.532*** (0.011)		
Learning time ^a			– 0.048*** (0.007)	– 0.048*** (0.007)
Feedback	– 0.016 (0.011)	– 0.016 (0.011)	– 0.006 (0.006)	– 0.006 (0.006)
Learner level (L2)				
Temporal RNO ^a	0.048 (0.033)	0.047 (0.033)	– 0.014 (0.012)	– 0.014 (0.012)
Social RNO ^a	– 0.038 (0.033)	– 0.039 (0.033)	0.009 (0.012)	0.009 (0.012)
Cross-level-interactions				
Temporal RNO × Feedback		0.002 (0.012)		0.005 (0.006)
Social RNO × Feedback		0.025** (0.012)		– 0.006 (0.006)

N=4108. RNO = reference norm orientation (preferences for comparison standards). Persistence measured as learning time per session (L1, logarithmized), performance as mean performance on items presented with feedback (L1, logarithmized). Effect coding of feedback type (– 1 temporal, 1: social). ^astandardized variables, ^bsession performance entered as $p \times (p-1)$ to investigate higher persistence under medium difficultly exercises

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

higher preferences for social comparison standards when a social comparison standard is presented compared to a temporal comparison standard.

Learning Performance. We chose the same analytical procedure for the effects on the success rate on items presented together with the feedback. This mean exercise performance per session was nested within participants and session. First, we entered the control variables as well as type of feedback and reference norm orientations as predictors. All regression coefficients are presented in Table 2. There were no significant main effects of either feedback type, individuals' social or temporal preferences for comparison standards. However, exercise success on items after feedback was more likely closer to the exam, $b=0.012$, $p=0.086$, and in shorter learning sessions, $b=-0.048$, $p<0.001$. We found no support for our proposed fit effects as we revealed no higher success probability with higher preferences for temporal comparison standards under temporal feedback conditions compared to social feedback conditions, $b_{\text{temporal} \times \text{feedback}}=0.005$, $p=0.444$, nor a significant interaction of feedback-type and preferences for temporal comparison standards remained insignificant, $b_{\text{social} \times \text{feedback}}=-0.006$, $p=0.306$.

Discussion

Experiment 1 tested the effects of different types of feedback on actual learning. The types were framed according to different comparison standards and we tested whether the effects of feedback depended on learners' preferences for temporal and social comparison standards. We were only partially able to support our proposed fit effects. In particular, we observed a first fit effect of preferences for social comparison standards and the presented feedback-type on learning persistence. Participants with higher preferences for social comparison standards learned longer in sessions with feedback which compares their learning score to the score of others. This calls for a closer look

at interindividual differences regarding established constructs. Based on the existing literature (Dickhäuser et al., 2017; Retelsdorf & Günther, 2011) higher motivation to learn could have been expected for individuals with higher preferences for temporal comparison standards, as those individuals have higher beliefs in change of abilities. Also, due to its nature, social feedback is more stable compared to temporal feedback as when all learners improve (or decline) in same similar manners, social feedback will be rather constant. However, we only observe small tendencies in the main effects of the preferences for standards in our first experiment, not replicating the proposed dominance of temporal comparison standards. This might be explained as those main effects might be too small to explain the large variance in learning persistence and performance in our e-learning environment. However, we see that—as proposed—differences in preferences for social comparison standards alter the association between feedback-type and learning time. Even though we did not observe any fit effects on learning performance, these results provide first empirical support for the central proposition of our fit-assumptions. We want to highlight that in the present experiment the amount of variance in learning sessions (even within individuals) was enormous and this might explain insignificant results. It seems obvious that learning times ranging from minutes to several hours are not only dependent on fitting or misfitting feedback, but on external factors. A student entering the university library after a lunch break to start learning for the whole afternoon will stay longer in the e-learning environment compared to a student arriving at home and using some spare time for a little more practicing before dinner. Obviously, those learning times are only marginally influenced by fitting or misfitting feedback. This also holds true for the learning performance. Performance on items is mainly dependent on the difficulty of the exercise. With a large variety of exercises and interindividual as well as intraindividual (over-time) variance, it is hard to detect effects on aggregated learning performance. Another limitation of the experimental manipulation was the use of the learning index as feedback. The overall learning index increases (or decreases) rather slowly and feedback on such a semi-volatile score might not be as informative as intended. Taken together, the first experiment can be seen as (a) first evidence for a more differentiating approach to the question of whether temporal or social comparison standards should be used and (b) as a starting point to dive deeper into the underlying processes of fit effects, which we will address in experiment 2 by investigating ease of processing as an underlying mechanism.

Experiment 2

The second experiment had three main goals. First, we conducted a conceptional replication (Schwarz & Strack, 2014; Stroebe & Strack, 2014) of experiment 1 by varying the manipulation of feedback with different comparison standards. As we discussed, the feedback of the first experiment only provided feedback on aggregated learning scores. We consider feedback on item level as more concrete and therefore more relevant for individuals and aimed to adapt the feedback towards the exercise level. Second, we changed the design to vary the feedback within learning sessions to better investigate fit effects on learning persistence by observing session abort after fitting or non-fitting

feedback instead of investigating the effects of constant fitting or non-fitting feedback on learning time. Finally, we want to concentrate more on potential underlying processes of fit effects by additionally manipulating perceptual fluency.

Methods

Participants

We collected data from 132 participants (20 male, 112 female) with a mean age of 20.8 years ($SD=2.87$) in exchange for course credit.⁴

Design and overview

Similar to experiment 1, we used the e-learning software and a within subject design and invited users to participate in this study in the spring semester 2020. Compared to experiment 1 we switched from a *between-session* to *within-session* design. While in experiment 1 the feedback-type presented during a session was constant until session abort, resulting in learning time per session as dependent variable, now the feedback-type was randomized at each presentation. This way we were able to analyze, after which feedback-type learners abort sessions and after which learners continue. By still manipulating within subjects but now on every instance of feedback instead of every session we were able to reduce noise and still respect ethical concerns. We gathered additional parameters like timestamps, session length, and processing time of exercises of the learning software to further reduce error variance. Furthermore, we manipulated fluency in addition to comparison standards. Hence, every instance of feedback was randomly selected from a two (social vs. temporal) by two (fluent vs. disfluent) set. The procedure was analogous to the procedure in experiment 1. Participants answered scales for preferences for comparison standards at their first log in and used the software for self-regulated exam preparation. After every seventh item the feedback was presented. It was randomly selected whether social or temporal comparisons were presented and whether the next question was displayed fluently or disfluently on every single instance of feedback.

Materials

Learning Software. The learning software was the software used in experiment 1 (Siebert and Janson, 2018).

Preferences for Comparison Standards Assessment. We used the same items for the assessment of preferences for comparison standards as in experiment 1.

Social/Temporal Feedback. As we intended experiment 2 to be a conceptual replication, and periodical feedback on a total learning index might not be very informative, we implemented new feedback to the software. After every seventh exercise the software provided feedback on the performance of the last seven items in a pop-up on the next exercise, which is illustrated in Fig. 3. A green checkmark or a red cross for each of the seven items indicating success or failure. Additionally, this exercise's specific feedback

⁴ We had to exclude 58 participants who did not answer attention checks correctly. Participants were psychology students at a German university using the software as exam preparation for "quantitative methods II" and "diagnostical psychology".

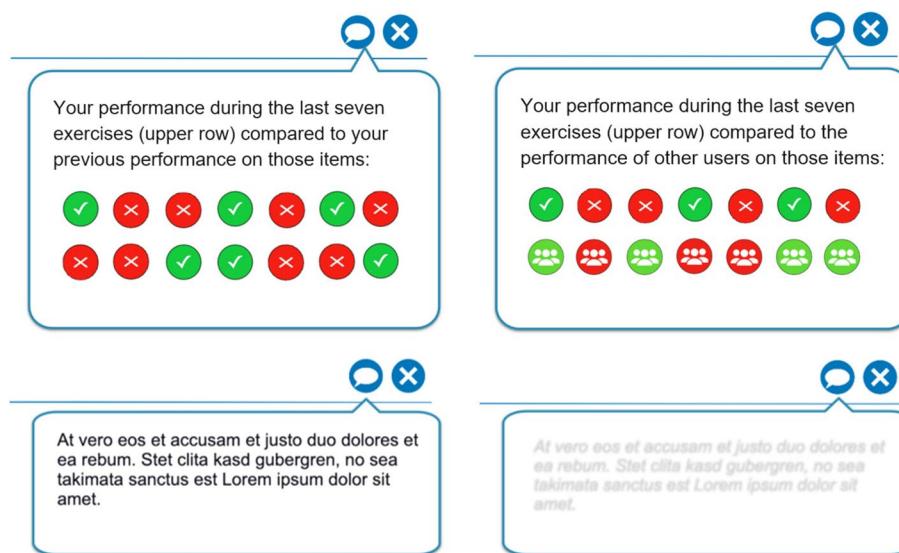


Fig. 3 Feedback types and fluency manipulations used in experiment 2. Top left = temporal feedback, top right = social feedback, bottom left = fluent learning advice, bottom right = disfluent learning advice

was either compared to the users' last performance on those seven exercises (temporal condition) or the mean performance of other users on those items (social condition) with an additional green or red symbol below. The pop-up had to be clicked away to continue with the next exercise. Again, the feedback was pretested. 20 pretesters perceived the temporal feedback to express more of a temporal comparison standard than a social comparison standard compared to the social feedback, $t(19)=6.96$, $p<0.01$, using the same item like in the first experiment.

Fluency. We added a perceptual fluency manipulation using contrast manipulations similar to those in existing studies (Hansen et al., 2008; Reber et al., 1998, 2004). In both conditions, the feedback occurred on the screen as a pop-up window together with the next exercise greyed out in the background. After clicking away the feedback, the text color of the next exercise changed back to black in the fluent condition, resulting in easier to read text with a higher contrast. In the disfluent condition, after the feedback was clicked away, the exercise remained greyed out and font color switched back to black with the next exercise. We presented sample exercises with fluent and disfluent font types to our pretesters and asked "how readable is the text in this exercise" on a seven-point Likert-scale with the endpoints "very bad" (1) and "very good" (7). The pretesters rated the fluent font type as easier to read than the disfluent, $t(19)=11.07$, $p<0.01$.

Learning outcomes. Again, we used logfiles from the e-learning software (Siebert and Janson, 2018). This time, we logged whether learning activities were terminated after feedback, i.e., no learning activities within one hour after the last seen feedback, or not as an indicator of learning persistence and performance on exercises presented after the feedback as an indicator of learning performance. Additionally, we only logged data from learning sessions on non-mobile devices (feedback might not be presented correctly on mobile devices). In total, 4690 instances of feedback and the performance on the consecutive exercises were analyzed in this experiment.

Results

Preliminary analyses

Again, the scales assessing preferences for comparison standards provided good internal consistency (temporal: $\alpha=0.85$; social: $\alpha=0.95$) and correlated moderately, $r=0.29$, $p<0.01$. However, in this sample we observed a higher preference on average for temporal comparison standards ($M=4.19$, $SD=0.59$) than for social comparison standards ($M=2.73$, $SD=0.86$). Additional descriptive analyses regarding age and gender are included in the Additional file 1. On average, terminal session abortion probability in the learning software was 21% (mean abortion after 5 feedback). The mean proportion of correct exercises presented together with feedback was 71%.

Main analyses

We used generalized linear regressions to predict the binary outcomes session abortion and item performance by preferences for comparison standards of the user, type of feedback (effect coded: $-1 = \text{temporal feedback}$; $1 = \text{social feedback}$) presented and fluency (effect coded: $-1 = \text{fluent}$; $1 = \text{disfluent}$). To respect the hierarchical data structure, we computed multi-level regressions with each instance of feedback on L1 and users and learning packages as L2 variables. Additionally, we added standardized time until exam, number of already presented feedback during this session, proportion of right answers of the last seven exercises and processing time (only for the analyses on learning performance) as control variables.

Learning persistence. Again, we entered predictors stepwise to our regression to test our hypotheses separately. First, we entered the main effects of the preferences for comparison standards, feedback-type and fluency, as well as the control variables standardized time to exam, number of already presented feedback and proportion of right answers. Instead of the proportion of right answers p we used proportion of right answers times the proportion of wrong answers $p \times (1 - p)$ as exercises with medium difficulty are more motivating than solving everything correctly or incorrectly. In this first step, only control variables predict session abortion significantly: session abortion was less likely with more time until the exam, $b=-0.110$, $p=0.015$, more medium solution probability, $b=-1.241$, $p=0.007$ and less feedback presented previously, $b=0.065$, $p<0.001$. In step two, we entered the proposed interaction effects of preferences for comparison standards and feedback type. Both interactions are in the stated direction, but only the interaction of preferences for social comparison standards and feedback type was significant using one-tailed testing. For participants with higher preferences for social comparison standards, session abortion was significantly more likely after temporal framed feedback and less likely after social framed feedback, $b_{\text{social} \times \text{feedback}}=-0.074$, $p=0.061$. Hence, H1b was supported by these results, while the interaction of feedback-type and preferences for temporal comparisons standards was in the proposed direction but remained insignificant, $b_{\text{temporal} \times \text{feedback}}=0.042$, $p=0.280$. We entered the three-way interaction with fluency as a last step. Only the interaction between preferences for social comparison standards, feedback-type and fluency was significant, $b_{\text{social} \times \text{feedback} \times \text{fluency}}=-0.111$, $p=0.006$, indicating the fit effect as stronger under disfluent conditions, contradicting the stated direction of H3b. All regression coefficients are printed in Table 3 and the interactions of the full model are displayed in Fig. 4.

Table 3 Multilevel regression on learning persistence and performance in experiment 2

	Persistence			Performance		
	Model 1 B(SE)	Model 2 B(SE)	Model 3 B(SE)	Model 1 B(SE)	Model 2 B(SE)	Model 3 B(SE)
Intercept	— 1.236*** (0.229)	— 1.239*** (0.227)	— 1.239*** (0.229)	— 0.227 (0.143)	— 0.228 (0.144)	— 0.233 (0.144)
Feedback level (L1)						
Time to exam	— 0.110** (0.045)	— 0.109** (0.045)	— 0.110** (0.045)	0.071* (0.043)	0.070 (0.043)	0.069 (0.043)
Correct exercises ^a	— 1.241*** (0.454)	— 1.232*** (0.454)	— 1.262*** (0.456)	1.885*** (0.185)	1.887*** (0.186)	1.894*** (0.186)
Session length	0.065*** (0.011)	0.065*** (0.011)	0.066*** (0.011)	-0.051 (0.034)	-0.050 (0.034)	-0.049 (0.034)
Feedback	0.015 (0.037)	0.012 (0.037)	0.008 (0.037)	0.031 (0.036)	0.032 (0.036)	0.032 (0.036)
Fluency	0.037 (0.037)	0.037 (0.037)	0.036 (0.037)	-0.037 (0.036)	-0.037 (0.036)	-0.036 (0.036)
Learner level (L2)						
Temporal RNO	0.003 (0.075)	0.006 (0.075)	-0.001 (0.076)	-0.039 (0.057)	-0.041 (0.057)	-0.044 (0.057)
Social RNO	-0.089 (0.081)	-0.093 (0.081)	-0.090 (0.082)	0.030 (0.061)	0.036 (0.061)	0.036 (0.061)
Cross-level-interactions						
Temporal RNO × Feedback		0.042 (0.038)	0.039 (0.039)		-0.071* (0.038)	-0.074* (0.038)
Social RNO × Feedback		-0.074* (0.040)	-0.070* (0.040)		0.082** (0.037)	0.080** (0.038)
Fluency × Feed-back			-0.013 (0.037)			0.006 (0.036)
Temporal RNO × Fluency			0.057 (0.039)			-0.002 (0.038)
Social RNO × Fluency			-0.025 (0.040)			-0.004 (0.037)
Temporal RNO × Feedback × Flu-ency			0.038 (0.038)			0.045 (0.038)
Social RNO × Feedback × Flu-ency			-0.111*** (0.040)			0.028 (0.037)

N=4690. RNO = reference norm orientation (preferences for comparison standards). Persistence measured as session abort after feedback (L1, binary, 0=no abort, 1=abort), performance as performance on items presented with feedback (L1, binary, 0=wrong, 1=correct). Effect coding of feedback (-1 temporal, 1:social) and fluency (-1: fluent, 1: disfluent).

^aProportion of correct answers p for performance models and proportion of correct answers times proportion of wrong answers $p \times (1 - p)$ for persistence models

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Learning performance. For learning performance on exercises, we also entered the specific exercise as an additional cluster variable. Again, we entered the main effects and control variables in step one. Only the number of previous solved exercises was a significant predictor of performance of the exercises presented directly after feedback, with more correct items predicting a higher probability of success, $b = 1.885$,

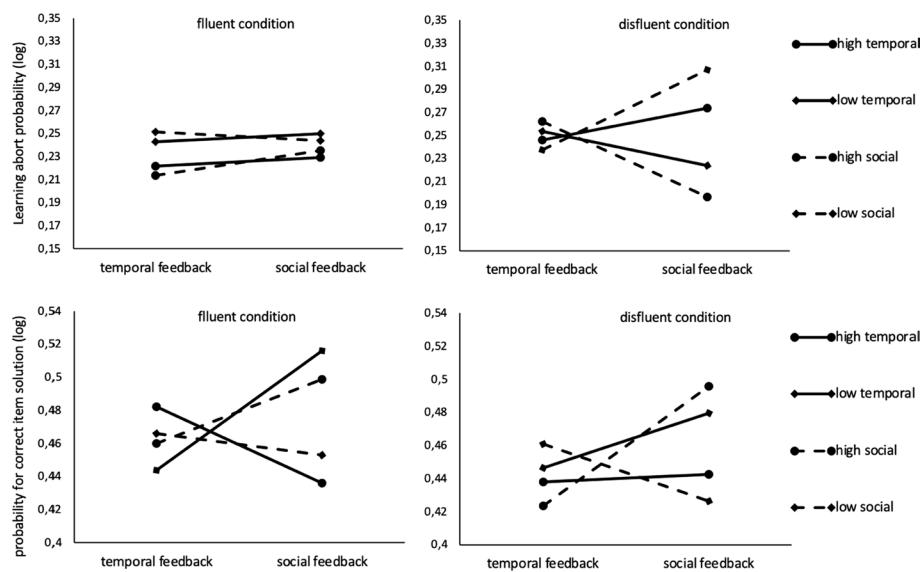


Fig. 4 Effects of preferences for comparison standards and framed feedback under fluent and disfluent conditions. Upper row: Fit effects on abortion probability as an inverted measurement for learning persistence and. Lower row: Fit effects probability for correctly solved exercises as an indicator for learning performance

$p < 0.001$. Both interaction terms in step two were directed as stated and significant using one-tailed testing supporting H2a and H2b. With higher preferences for temporal comparison standards, the probability of solving was higher after temporal feedback and lower after social feedback, $b_{\text{temporal}} \times \text{feedback} = -0.071$, $p = 0.064$. Furthermore, with higher preferences for social comparison standards the probability of solving was lower after temporal feedback and higher after social feedback, $b_{\text{social}} \times \text{feedback} = 0.082$, $p = 0.029$. None of the three-way interactions in step three were significant, but the fit effects remain significant. The regression coefficients of all models are printed in Table 3 and the interaction plots in Fig. 4.

Discussion

Experiment 2 provides further evidence for fit effects of preferences for comparison standards and framed feedback. In this experiment, we observed fit effects on performance as well as on persistence. By changing to a within session design and focusing on item performance (while controlling for item difficulties) we were able to reveal the proposed fit effects on learning performance. However, for the persistence outcome, only the fit effect of preferences for social comparison standards was significant, replicating the findings of our first experiment. However, we want to highlight that the change from a between to a within session design led to less but still highly confounded dependent variables. As in the first experiment, we assume session abortion to be mostly dependent on individuals' time schedules rather than receiving fitting or misfitting feedback. Hence, fit effects were hard to detect. Furthermore, the assessment of preferences for temporal comparison standards seems to be less reliable. This could also be explained by less variance among participants on this dimension as we observed high preferences for temporal comparison standards in general.

Compared to the first experiment we might have faced ceiling effects for temporal comparison standards concealing potential fit effects. Overall, this could explain why we only found a fit of preferences of social comparison standards and framed feedback in a methodical manner. Although, the significant three-way interaction of the observed fit effect with our fluency manipulation was not in the stated direction, we have first empirical evidence for underlying processes of fit effects on learning persistence. However, the fit effects on the item performance were not moderated by the contrast manipulation. It should be noted that, besides the large differences in the reported ratings of our pretesters, the manipulation is a quite subtle one.

General discussion

Feedback can be more or less beneficial for learning outcomes (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). While the large body of literature on feedback interventions focuses on the more effective elaborative feedback guiding learners towards better results, we are interested in improvements of more basal feedback just informing learners about their current performance. For this type of feedback, the present studies provide first evidence for fit effects of preferences for comparison standards. For self-regulated learning activities, we found positive fit effects on learning persistence and performance of preferences for comparison standards and framed feedback. Our general claim that framed feedback towards own preferences has a positive impact on learning outcomes is partly supported by our two studies. Regarding the proposed fit effects on learning persistence, in both studies we only found significant interactions of preferences for social comparison standards and the framed feedback, but no interactions of preferences for temporal comparison standards and the framed feedback. Despite the large standard errors reducing significance, the fit effects, both in the proposed direction, were conceptually replicated with distinguishable manipulations (Schwarz & Strack, 2014; Stroebe & Strack, 2014). For learning performance, we were only able to identify a fit effect of preferences for temporal comparison standards and framed feedback in the second experiment. Furthermore, we revealed a possible underlying process of fit effects of preferences of comparison standards on persistence. For the significant fit effect of preferences for social comparison standards and framed feedback, a three-way interaction with fluency was also significant. Namely, the fit effect was enforced under disfluent conditions. We proposed another direction, but it supports a linkage of fit effects and fluency.

Theoretical implications and future research directions

The literature on preferences for comparison standards traditionally concentrates on the effects of preferences of teachers when evaluating their students. It provides theoretical as well as empirical evidence of the superiority of temporal comparison standards and implies that teachers should be trained towards a more temporal comparison standard (Dickhäuser et al., 2017; Retelsdorf & Günther, 2011; Rheinberg, 2001). However, our approach differs as we are interested in self-regulated learning activities, not with teacher comparing performance but the learner himself. Though, it might be seen as nested within the achievement goals (Dweck & Leggett, 1988; Elliot & McGregor,

2001; Nicholls, 1984), we stick with this construct of preferences for comparison standards as most proximal to our research question of how feedback provided by learning environments is processed. One might argue that social or temporal feedback presented in e-learning software can be compared to teachers and their preferences for comparison standards (Dickhäuser et al., 2017; Lüdtke et al., 2005; Retelsdorf & Günther, 2011; Rheinberg, 2001). Therefore, one could expect positive effects for temporally framed feedback and negative effects for socially framed feedback. This was not the case in both of our studies. Indeed, effects of teachers' comparison standards are linked to specific mindsets: a higher temporal comparison standard of teachers leads to a higher growth mindset of students (Dickhäuser et al., 2017). It seems plausible that this is due to teachers with preferences for temporal comparisons also believing more in individual growth and encouraging students respectively. Learning software does provide feedback without further mindset.

Our replicated finding of fit effects of preferences for social comparison standards and framed feedback on learning persistence could indicate that the preferences of social feedback are more important to be considered as interindividual than a temporal comparison standard. This is in line with existing literature on the effects of teachers' preferences for comparison standards, which explicitly state that temporal comparison standards have positive effects on all students, while social comparison standards are more harmful for low-performing students (Dickhäuser et al., 2017; Lüdtke et al., 2005; Retelsdorf & Günther, 2011; Rheinberg, 2001). It seems plausible to assume that for less beneficial social comparison standards a fit is more important. While temporal feedback might be more or less motivational for all learners (independent of their preferences), social feedback could lead to session abortion with a higher probability for those with low preferences for social comparisons than for those with a high preference. Also, former research revealed stronger effects of social comparisons than temporal ones (e.g., Wolff et al., 2018) and in line with the "overpowering effect of social comparison information" (Van Yperen & Leander, 2014), preferences for social comparisons might also have a stronger impact on learning outcomes than those for temporal comparisons. Further research is needed to replicate our findings and to investigate whether fit effects are limited to preferences for social comparison standards.

In our conceptual framework, we proposed that feedback in line with own preferences is processed with higher ease. To test this mechanism, we additionally manipulated perceptual fluency to alter the strength of fit effects. We found one significant three-way-interaction altering the fit effect of preferences for social comparison standards and framed feedback on learning persistence. The result contradicts our first theoretical assumption as we hypothesized fit effects to be stronger under fluent conditions. Nevertheless, this interaction might not contradict the theoretical linkage of fit effects and ease of processing per se. Regarding these particular results, one could assume that fit can provide easier information processing even when other factors are disfluent. Effects of fluency might be dependent on the default state of fluency (Hansen et al., 2008; Wänke & Hansen, 2015). If learners perceive some high level of flow or fluency experience in general during their studies, a disfluent cue reducing this ease of processing might broaden the space for fit effects to expand their impact. We are cautious about interpreting this result further, as only one of the proposed interactions was significant,

but it is promising first evidence leading to more open questions regarding fluency as an underlying psychological process of fit effects.

For our proposed fit effects on learning performance, the approach of experiment 2 was more promising as we measured direct effects of fitting or misfitting feedback compared to experiment 1. Learning time and especially aggregated performance of items with different difficulties in the first experiment might be too noisy dependent variables in a field experiment to show support for the proposed fit effect. Hence, this might explain why we observed both proposed fit effect on preferences for comparison standards and framed feedback on learning performance only in the second experiment. However, in this second experiment both fit effects were in the proposed direction and significant. The performance of learners on items presented with the feedback was higher for learners with higher preferences for temporal comparison standards if the temporal framed feedback was presented compared to the social framed feedback and also the performance of learners with higher preferences for social comparison standards was enhanced when receiving social framed feedback compared to temporal feedback.

At this point, one might argue whether the occurrence of misfitting feedback or the absence of fitting feedback leads to higher abortion tendencies. Future research revisiting fit effects should address this question with between subject designs comparing fit-, misfit- and control conditions. In general, we cannot say much about the underlying psychological processes of those fit effects on performance, yet. We tested ease of processing as a possible mechanism with an additional fluency manipulation but were not able to see any differences regarding fit effects under fluent or disfluent conditions. This does not rule out ease of processing as an underlying process, but the present data only (partially) support a connection between fluency and fit effects and persistence, which is in line with a recent meta-analysis revealing effects of fluency on judgments of learning but not on objective learning performance (Xie et al., 2018). We would like to highlight another possible mechanism for fit effects on cognition referring to the previously mentioned literature (Macrae et al., 1993; Stangor & McMillan, 1992). If misfitting feedback binds additional cognitive resources in terms of extraneous cognitive load (Sweller, 2010), this effect should be affected by the availability of cognitive resources. Hence, we would suggest investigating fits of preferences for comparison standards and framed feedback on cognitive outcomes using a dual-task paradigm.

In our experiment, we only address the question of whether the feedback-type meets individual preferences motivated by the theoretical considerations of the feedback interventions theory (Kluger & DeNisi, 1996, 1998), where feedback should be in line with goals and standards and support for our fit hypothesis has been provided. However, one could also try to adapt feedback in terms of valence. We already pointed out that especially low performing students are disadvantaged by high social comparison standards of teachers (Dickhäuser et al., 2017; Lüdtke et al., 2005; Retelsdorf & Günther, 2011; Rheinberg, 2001). One might transfer this to our studies on the perception of feedback as well. It would be interesting to investigate whether a temporal comparison standard is more beneficial for all learners, while for high-performers social feedback might not be harmful compared to low performing learners. Furthermore, future research should address whether fitting does also interact with the valence of feedback. Are positive effects of

fitting feedback limited to positive feedback or does fitting feedback enhance the impact of feedback also in terms of aversiveness of negative feedback. Answering this additional open research question would lead to a better understanding of feedback effects on self-regulated learning.

Limitations

In our present research, we solely focus on the effects of temporal and social comparison standards in an actual achievement setting. By doing so, we establish the impact of our finding on actual learning behavior, but were also restricted in opportunities. We only focus on informative feedback as this can be easily framed in digital learning environments. Also, we cannot disentangle if fitting feedback bolsters learning outcomes, misfitting feedback impairs them or how they compete against a control group receiving no informative feedback. Laboratory experiments not limited by the type of chosen design might answer this in future.

We used two different manipulations of feedback providing either feedback on a more aggregated level (experiment 1) or on an exercise-specific level (experiment 2). Participants might have seen feedback in experiment 1 as uninformative as the changes in the learning index were rather small and the social feedback was quite stable. However, this is in fact the nature of social feedback in many classroom situations, a reason for why social comparison standards of teachers have a negative impact on students (Dickhäuser et al., 2017; Rheinberg, 2001). Nevertheless, this cannot be completely disentangled at this point as we switched from a between-session, keeping feedback types constant during learning sessions, to a within-session design, varying feedback types during learning sessions, between experiments 1 and 2.

Furthermore, conducting field experiments with actual learning activities provide certain advantages especially regarding external and ecological validity but can lack internal validity and power. Although we were able to use an extended dataset for our experiments, effects were rather small and statistical significance was only partially achieved, even using one tailed testing due to the specified directions of our hypotheses. However, this should not be interpreted as lack of evidence or relevance. Laboratory experiments focusing on eliminating every single source of error variance might lead to clearer results but findings are not necessarily applicable to human behavior outside the laboratory (Bless & Burger, 2016). Also, with our research conducted *in situ* on actual learning behavior, we demonstrate practical applicability of our supposed fit effects (Berkman & Wilson, 2021).

Practical implications and conclusion

Fit effects are common in psychological research (Cronbach & Snow, 1969; Edwards, 1991; Higgins, 2000; Porter & Umbach, 2006). The traditional research on comparison standards focuses on the main effect of teachers' preferences for comparison standards (Dickhäuser et al., 2017; Lüdtke et al., 2005; Retelsdorf & Günther, 2011; Rheinberg, 2001). Assuming a more differentiated psychological process based on the results provided by Kluger and DeNisi (1998), we investigated preferences for comparison standards of learners to address the unanswered research question of the effects of learners' preferences and framed feedback on actual learning behavior. Our studies supported

those proposed fit effects on relevant learning outcomes. Due to the noisiness of the data, we were only able to find first partial empirical support. With ease of processing, we also presented a possible underlying process of fit effects. First evidence contributes to the assumption of a connection of fluency and fit effects of preferences for comparison standards and framed feedback. Further research is needed to replicate our findings in a more controlled environment and to have a closer look at the underlying processes. Overall, with the present studies, we provided first initial research to build on to fill the research gap on learners' preferences for comparison standards, which calls for a more interindividual perspective of the usage of comparison standards in learning contexts. This is especially relevant for practitioners designing e-learning environments. While feedback is crucial for learning outcomes, we highlight the interindividual perspective of feedback perception. Feedback should be adapted towards the personal preferences of learners to be most beneficial. More concrete, our findings indicate that social feedback, which in general has to be handled with care, can be positive if learners have preferences for such comparisons. Our experiments conducted in an e-learning environment support the tremendous opportunities of adaptive e-learning (Shute & Towle, 2003) and relevance of actual learning data for educational psychological research (Berkman & Wilson, 2021).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41239-022-00358-2>.

Additional file 1. Supplement A. Items used for the assessment of preferences for temporal and social comparison standards. **Supplement B1.** Descriptive statistics for experiment 1 and experiment 2 by gender. **Supplement B2.** Associations of preferences for comparison standards and age.

Acknowledgements

We would like to thank Adriana Fuhs and Emma Martin for assistance with data collection and manuscript preparations.

Author contributions

MPJ came up with the idea, conducted the experiments, analyzed the data and wrote a first draft of the manuscript. JS implemented the feedback manipulations within the learning software and provided assistance for data preparation. OD gave valuable feedback on the experimental design and the existing literature and revised a first draft of the manuscript. All authors read and approved the final manuscript.

Funding

The manuscript is part of a dissertation project supported by grants of the Center of Doctoral Studies in Social Sciences (CDSS) of the Graduate School of Economic and Social Sciences (GESS) of the University of Mannheim and the Konrad-Adenauer-Foundation awarded to MPJ.

Availability of data and materials

The data that support the findings of this study are available from the used learning software (www.cotutor.de) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the learning software (www.cotutor.de).

Declarations

Competing interests

To be transparent on potential conflicts of interest: MPJ and JS are owners of the used e-learning software which is also used for commercial purposes.

Received: 11 March 2022 Accepted: 28 June 2022

Published online: 10 October 2022

References

- Albert, S. (1977). Temporal comparison theory. *Psychological Review*, 84(6), 485–503. <https://doi.org/10.1037/0033-295X.84.6.485>
- Al-Fraiha, D., Joy, M., Masadeh, R., & Sinclair, J. (2020). Evaluating E-learning systems success: An empirical study. *Computers in Human Behavior*, 102, 67–86. <https://doi.org/10.1016/j.chb.2019.08.004>
- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13, 219–235. <https://doi.org/10.1177/1088868309341564>
- Aparicio, M., Bacao, F., & Oliveira, T. (2017). Grit in the path to e-learning success. *Computers in Human Behavior*, 66, 388–399. <https://doi.org/10.1016/j.chb.2016.10.009>
- Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*, 7(3–4), 65–89. https://doi.org/10.1300/J075v07n03_05
- Bangert-Drowns, R. L., Kulik, C.-L.C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213–238. <https://doi.org/10.3102/00346543061002213>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. ArXiv: 1406.5823 [Stat]. <http://arxiv.org/abs/1406.5823>
- Berkman, E. T., & Wilson, S. M. (2021). So useful as a good theory? The practicality crisis in (Social) psychological theory. *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691620969650>
- Bless, H., & Burger, A. M. (2016). A closer look at social psychologists' silver bullet: Inevitable and evitable side effects of the experimental approach. *Perspectives on Psychological Science*, 11(2), 296–308. <https://doi.org/10.1177/1745691615621278>
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79(4), 474–482. <https://doi.org/10.1037/0022-0663.79.4.474>
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563–1569. <https://doi.org/10.1037/a0017021>
- Castillo-Merino, D., & Serradell-López, E. (2014). An analysis of the determinants of students' performance in e-learning. *Computers in Human Behavior*, 30, 476–484. <https://doi.org/10.1016/j.chb.2013.06.020>
- Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In: M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 605–629).
- Covington, M. V., & Omelich, C. L. (1984). Task-oriented versus competitive learning structures: Motivational and performance consequences. *Journal of Educational Psychology*, 76(6), 1038–1050. <https://doi.org/10.1037/0022-0663.76.6.1038>
- Cronbach, L. J., & Snow, R. E. (1969). *Individual Differences in Learning Ability as a Function of Instructional Variables* (No. ED029001). ERIC. <https://files.eric.ed.gov/fulltext/ED029001.pdf>
- Dickhäuser, O., Janke, S., Praetorius, A.-K., & Dresel, M. (2017). The effects of teachers' reference norm orientations on students' implicit theories and academic self-concepts. *Zeitschrift Für Pädagogische Psychologie*, 31(3–4), 205–219. <https://doi.org/10.1024/1010-0652.a000208>
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256–273. <https://doi.org/10.1037/0033-295X.95.2.256>
- Edwards, J. R. (1991). Person-job fit: A conceptual integration, literature review, and methodological critique. In *International review of industrial and organizational psychology, Vol. 6*. (pp. 283–357). John Wiley & Sons.
- Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: Does harder-to-read mean better-to-understand? *Applied Cognitive Psychology*, 28(4), 488–501. <https://doi.org/10.1002/acp.3004>
- Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80(3), 501–519. <https://doi.org/10.1037/0022-3514.80.3.501>
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140. <https://doi.org/10.1177/00187267540070202>
- Flavell, J. C., Over, H., & Tipper, S. P. (2020). Competing for affection: Perceptual fluency and ambiguity solution. *Journal of Experimental Psychology: Human Perception and Performance*, 46(3), 231–240. <https://doi.org/10.1037/xhp0000702>
- Freitas, A. L., & Higgins, E. T. (2002). Enjoying goal-directed action: The role of regulatory fit. *Psychological Science*, 13, 1–6. <https://doi.org/10.1111/1467-9280.00401>
- Godinho, S., & Garrido, M. V. (2021). The in-out effect: Examining the role of perceptual fluency in the preference for words with inward-wandering consonantal articulation. *Psychological Research Psychologische Forschung*, 85(1), 112–120. <https://doi.org/10.1007/s00426-019-01238-7>
- Hansen, J., Dechêne, A., & Wänke, M. (2008). Discrepant fluency increases subjective truth. *Journal of Experimental Social Psychology*, 44(3), 687–691. <https://doi.org/10.1016/j.jesp.2007.04.005>
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement (Reprinted)*. Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 22–34.
- Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 1–46). Academic Press.
- Higgins, E. T. (2000). Making a good decision: Value from fit. *American Psychologist*, 55, 1217–1230. <https://doi.org/10.1037/0003-066X.55.11.1217>
- Hong, J., & Lee, A. Y. (2008). Be fit and be strong: Mastering self-regulation through regulatory fit. *Journal of Consumer Research*, 34, 682–695. <https://doi.org/10.1086/521902>
- Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64(4), 349–371. <https://doi.org/10.1037/0021-9010.64.4.349>

- Keller, J., & Bless, H. (2006). Regulatory fit and cognitive performance: The interactive effect of chronic and situationally induced self-regulatory mechanisms on test performance. *European Journal of Social Psychology*, 36, 393–405. <https://doi.org/10.1002/ejsp.307>
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2950.119.2.254>
- Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, 7(3), 67–72. <https://doi.org/10.1111/1467-8721.ep10772989>
- Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes*, 50(2), 212–247. [https://doi.org/10.1016/0749-5978\(91\)90021-K](https://doi.org/10.1016/0749-5978(91)90021-K)
- Lee, J.-K., & Lee, W.-K. (2008). The relationship of e-Learner's self-regulatory efficacy and perception of e-Learning environmental quality. *Computers in Human Behavior*, 24(1), 32–47. <https://doi.org/10.1016/j.chb.2006.12.001>
- Lüdtke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish–little-pond effect. *Contemporary Educational Psychology*, 30(3), 263–285. <https://doi.org/10.1016/j.cedpsych.2004.10.002>
- Macrae, C. N., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, 23(1), 77–87.
- Mohammadhassan, N., Mitrovic, A., & Neshatian, K. (2022). Investigating the effect of nudges for improving comment quality in active video watching. *Computers & Education*, 176, 104340. <https://doi.org/10.1016/j.compedu.2021.104340>
- Motyka, S., Grewal, D., Puccinelli, N. M., Roggeveen, A. L., Avnet, T., Daryanto, A., de Ruyter, K., & Wetzel, M. (2014). Regulatory fit: A meta-analytic synthesis. *Journal of Consumer Psychology*, 24, 394–410. <https://doi.org/10.1016/j.jcps.2013.11.004>
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91(3), 328–346. <https://doi.org/10.1037/0033-295X.91.3.328>
- Nicholls, J. G. (1989). *The competitive ethos and democratic education*. Harvard University Press.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385–407. <https://doi.org/10.1007/s10648-004-0006-x>
- Porter, S. R., & Umbach, P. D. (2006). COLLEGE MAJOR CHOICE: An analysis of person-environment fit. *Research in Higher Education*, 47(4), 429–449. <https://doi.org/10.1007/s11162-005-9002-3>
- Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, 140(3), 283–302.
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, 8(3), 338–342. <https://doi.org/10.1006/ccog.1999.0386>
- Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, 9(1), 45–48. <https://doi.org/10.1111/1467-9280.00008>
- Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring “fringe” consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, 13(1), 47–60. [https://doi.org/10.1016/S1053-8100\(03\)00049-7](https://doi.org/10.1016/S1053-8100(03)00049-7)
- Retelsdorf, J., & Günther, C. (2011). Achievement goals for teaching and teachers' reference norms: Relations with instructional practices. *Teaching and Teacher Education*, 27(7), 1111–1119. <https://doi.org/10.1016/j.tate.2011.05.007>
- Rheinberg, F. (2001). Teachers Reference-Norm Orientation and Student Motivation for Learning. AERA-conference, Seattle.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Schöne, C., Dickhäuser, O., Spinath, B., & Stiensmeier-Pelster, J. (2004). Zielorientierung und Bezugsnormorientierung: Zum Zusammenhang zweier Konzepte. *Zeitschrift Für Pädagogische Psychologie*, 18(2), 93–99. <https://doi.org/10.1024/1010-0652.18.2.93>
- Schwarz, N., & Strack, F. (2014). Does merely going through the same moves make for a “direct” replication? Concepts, contexts, and operationalizations. *Social Psychology*, 45(4), 305–306.
- Seo, K., Dodson, S., Harandi, N. M., Roberson, N., Fels, S., & Roll, I. (2021). Active learning with online video: The impact of learning context on engagement. *Computers & Education*, 165, 104132. <https://doi.org/10.1016/j.compedu.2021.104132>
- Shih, S.-S., & Alexander, J. M. (2000). Interacting effects of goal setting and self- or other-referenced feedback on children's development of self-efficacy and cognitive skill within the Taiwanese classroom. *Journal of Educational Psychology*, 92(3), 536–543. <https://doi.org/10.1037/0022-0663.92.3.536>
- Shute, V., & Towle, B. (2003). Adaptive E-Learning. *Educational Psychologist*, 38, 105–114. https://doi.org/10.1207/S15326985EP3802_5
- Siebert, J., & Janson, M. P. (2018). CoTutor[Computer software]. <https://www.cotutor.de>
- Son, L. K., & Simon, D. A. (2012). Distributed learning: Data, metacognition, and educational implications. *Educational Psychology Review*, 24(3), 379–399. <https://doi.org/10.1007/s10648-012-9206-y>
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89(6), 845–851. <https://doi.org/10.1037/0022-3514.89.6.845>
- Spiegel, S., Grant-pillow, H., & Higgins, E. T. (2004). Strength during goal pursuit. *European Journal of Social Psychology*, 34, 39–54.
- Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, 111, 42–61. <https://doi.org/10.1037/0033-2909.111.1.42>
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9(1), 59–71. <https://doi.org/10.1177/1745691613514450>
- Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29–47). Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744.004>

- Van Yperen, N. W., & Leander, N. P. (2014). The overpowering effect of social comparison information: On the misalignment between mastery-based goals and self-evaluation criteria. *Personality and Social Psychology Bulletin*, 40(5), 676–688. <https://doi.org/10.1177/0146167214523475>
- Wänke, M., & Hansen, J. (2015). Relative processing fluency. *Current Directions in Psychological Science*, 24(3), 195–199. <https://doi.org/10.1177/0963721414561766>
- Winne, P. H., Jamieson-Noel, D., & Muis, K. (2002). Methodological issues and advances in researching tactics, strategies, and self-regulated learning. *Advances in Motivation and Achievement: New Directions in Measures and Methods*, 12, 121–155.
- Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Journal of Educational Psychology*, 110(7), 1005–1025. <https://doi.org/10.1037/edu0000248>
- Xie, H., Zhou, Z., & Liu, Q. (2018). Null effects of perceptual disfluency on learning outcomes in a text-based educational context: A meta-analysis. *Educational Psychology Review*, 30(3), 745–771. <https://doi.org/10.1007/s10648-018-9442-x>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com