"Your argumentation is good", says the AI vs humans – The role of feedback providers and personalised language for feedback effectiveness

Theresa Ruwe [*], Elisabeth Mayweg-Paus

Humboldt-Universität zu Berlin, Unter den Linden 6, 10099, Berlin, Germany

A R T I C L E   I N F O

A B S T R A C T

Argumentative writing is an important skill for (future) teachers, and this skill can be promoted with feedback. For such feedback to be effective, certain social and non-cognitive aspects must be considered. Building on this, our study investigated the influence of different feedback providers and language styles on learners' motivation, emotions, self-efficacy, and perceptions about the provider and the feedback. A total of $N = 98$ German student teachers participated in the 3 (feedback provider: artificial intelligence (AI) vs educator vs peer) $\times$ 2 (feedback language: personalised vs neutral) experimental between-subjects design. Results showed that student teachers ascribed an AI, compared to educators and peers, with more trustworthiness, particularly regarding its expertise. Against our expectations, we did not find further significant effects, but we found several tendencies that point towards the importance of social and non-cognitive aspects in feedback processes and indicate the need for further research. Conclusively, these aspects should be considered when investigating feedback processes in online learning environments.

1. Introduction

In a dystopian world, artificial intelligence (AI) in various forms takes over and dominates humankind. However, in our present world, AI is being developed to support humans. Discussions about what that support will look like have become heated – particularly after the launch of ChatGPT in 2022. One context of this debate is in educational settings, where the application of AI has great potential (Chen, Xie, Zou, & Hwang, 2020; Chiu, Xia, Zhou, Chai, & Cheng, 2023; Ouyang, Zheng, & Jiao, 2022). AI can, for example, save educators resources (Wilson et al., 2021; Zhu, Liu, & Lee, 2020) by automatically providing feedback on learners' argumentative writing skills (Wambsganns et al., 2021). Fostering argumentation skills (*learning to argue*) is considered crucial in the 21st century (Redecker, 2017), as these skills facilitate deep and elaborated engagement with information when interacting with others (Jonassen & Kim, 2010) and during knowledge construction (Osborne, 2010) (*arguing to learn*). Thus, in the context of teacher education, argumentation skills are crucial from student teachers' perspectives of being both the learner acquiring these skills and the (future) teacher promoting these skills. Further, in dealing with AI, argumentation skills are again a crucial competence: People must learn how to interact with (information from) AI systems (Ng, Leung, Chu, & Qiao, 2021). Finally,

feedback seems to be promising for fostering argumentation skills (e.g., Latifi, Noroozi, Hatami, & Biemans, 2019). Yet, providing such feedback is often impossible for educators, as resources are scarce. AI systems that can support educators in feedback processes can thus potentially facilitate the promotion of argumentation skills.

In recent years, research interest in the social and non-cognitive aspects of online learning processes like feedback has grown (Schneider, Beege, Nebel, Schnaubert, & Rey, 2021). In this study, we consider that key social cues important in the feedback process include the feedback provider and their language style, and how they affect key non-cognitive outcomes of feedback, namely students' motivation, emotions, self-efficacy, and perceptions of the provider and the feedback. As such, we address several gaps in the literature. First, we investigate the role of feedback provided by AI systems compared to humans, which is crucial considering current developments. Particularly, AI systems are emerging in the field of feedback; they provide written, elaborated feedback on texts rather than just corrective feedback on multiple-choice answers (Cavalcanti, Ferreira, & Gomes, 2021). Their ability to provide written texts, i.e., communicate with natural language, makes AI systems more human (Kaplan, Kessler, Brill, & Hancock, 2023) and subject to evaluations by their users. This indicates the importance of researching their role in learning through feedback

---

* Corresponding author.
E-mail address: theresa.ruwe@hu-berlin.de (T. Ruwe).

interactions. Second, we focus on a social aspect manifested in language, namely personalised language (Moreno & Mayer, 2004), which impacts learning and is easy to implement in practice; yet personalised language has not received a lot of attention in feedback contexts. Personalisation, as defined by Moreno and Mayer (2004), has been researched in multimedia learning environments and as a way of instructing effectively and is considered beneficial in these contexts. It is seen as a social aspect that can enhance the learning experience (Schneider et al., 2021) and could thus be valuable when used with feedback in general, and it may also constitute an appropriate use of language for an AI system when providing written feedback. Thus, the following study was designed to investigate the influence of the feedback provider, i.e., educators vs peers vs AI, and the feedback language, i.e., personalised vs neutral language, on the effectiveness of feedback as determined by learners' motivation, self-efficacy, and emotions as well as their perceptions of the provider and the feedback.

2. Theoretical background

2.1. What is Feedback and why is it important in education?

Lipnevich and Panadero (2021) defined feedback as information about several components from several sources that works best if learners actively engage with it. Conclusively, feedback is not only concerned with the information itself, but rather encompasses the providers' and recipients' (re)actions to it. Complementing this definition, Panadero and Lipnevich (2022) suggested a new feedback model, the *MISCA* (*message, implementation, student, context, and agents*) model, capturing the breadth of relevant factors. Their work has highlighted the need to continue research while considering changes and developments to these factors. Accordingly, our study investigates changes in digital, written feedback processes (*context* and *implementation*) regarding language style (*message*) and feedback provider (*agents*) and their effect on *students'* non-cognitive reactions (see Fig. 1).

Particularly in digital environments, not only *cognitive* but also *affective* and *social* aspects should be investigated, as they are still under-researched despite their importance for feedback effectiveness (Evans, 2013; Lawson et al., 2021; Schneider et al., 2021). Overall, research on feedback effectiveness is complex and highly variable, as many aspects can influence it – positively or negatively (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008; Wisniewski, Zierer, & Hattie, 2020). Our study particularly focusses on social aspects and their relation to feedback effectiveness, which is determined by cognitive (e.g., improved quality of argumentative writing (Latifi et al., 2019)), affective (e.g., motivation (Wisniewski et al., 2020) or emotion (Molloy, Borrell-Carrió, & Epstein, 2013; Rowe, 2017)), or relational aspects (e.g., affecting trust (Davis & Dargusch, 2015)) as summarised by Henderson, Ryan, and Phillips (2019). The impact of feedback might even be multifaceted and simultaneously reveal itself on different levels (e.g., peer feedback can improve the quality of argumentative texts (Latifi

et al., 2019), but might also elicit resistance (Hovardas, Tsivitanidou, & Zacharia, 2014)). Therefore, we summarise that feedback effectiveness can be reflected by students' perceptions, emotions, motivation, and self-efficacy as well as performance; we focus on the non-cognitive aspects (see Fig. 1) which have been underrepresented in research (e.g., Evans, 2013; Fong & Schallert, 2023).

2.2. The influence of social cues on feedback effectiveness

What role do social cues play in the context of learning? Social cues elicit social presence (Kreijns, Xu, & Weidlich, 2022; Short, Williams, & Christie, 1976), which then positively impacts learning by increasing interest and engagement (Seitchik, Brown, & Harkins, 2017). In this vein, social presence has been found beneficial for non-cognitive learning outcomes (Richardson, Maeda, Lv, & Caskurlu, 2017; Russo & Benson, 2005), meaning that motivation, emotions, self-efficacy, and perceptions can be influenced by social cues. This process is called *social facilitation*, but researchers refer to *social inhibition* as well: The presence of others can inhibit learning (Belletier, Normand, & Huguet, 2019; Zajonc, 1965). Thus, it seems important to consider whether and how social cues, like the feedback provider and language style, impact the non-cognitive aspects of feedback effectiveness.

2.2.1. The feedback provider as a social cue

Feedback can be provided by different sources with different characteristics. Traditionally, feedback is provided by the educator, but it can likewise be provided by peers or AI (Hattie & Timperley, 2007).

The educator is characterised as a human expert with authority. Overall, feedback from such a source is likely to be accepted and related to motivation and feedback perceptions, among other outcomes (Lechermeier & Fassnacht, 2018), potentially because of the trustworthiness and credibility of experts (Metzger, Flanagin, Eyal, Lemus, & Mccann, 2016). Nevertheless, the educators' status comes with the *curse of expertise* (Hinds, 1999). This means that recipients' understanding of the information might be hindered by the language the educator uses. Furthermore, the authority might have hindering effects on the learning process (Carless, 2006; Jonsson, 2012). Overall, though, it is generally assumed that educator feedback is preferred over peer feedback (e.g., Kwok, 2008; Zhang, 1995) because peers lack expertise and authority, even though peer feedback is welcome (Kwok, 2008; Tsui & Ng, 2000). In this vein, peers have a similar status and use similar language, which eases understanding (e.g., Cho, Chung, King, & Schunn, 2008), potentially making peer feedback more effective. Yet, students often resist accepting peers' advice (Hovardas et al., 2014; Strijbos, Narciss, & Dünnebier, 2010), likely because a student's use of peer feedback highly depends on the expertise of the peer giving the feedback (e.g., Hovardas et al., 2014; Van der Kleij & Lipnevich, 2021; Van Gennip, Segers, & Tillema, 2010).

AI can appear in different modes, such as robotic or embedded AI, but in this study we focus on AI as a learning system (Glikson & Woolley,
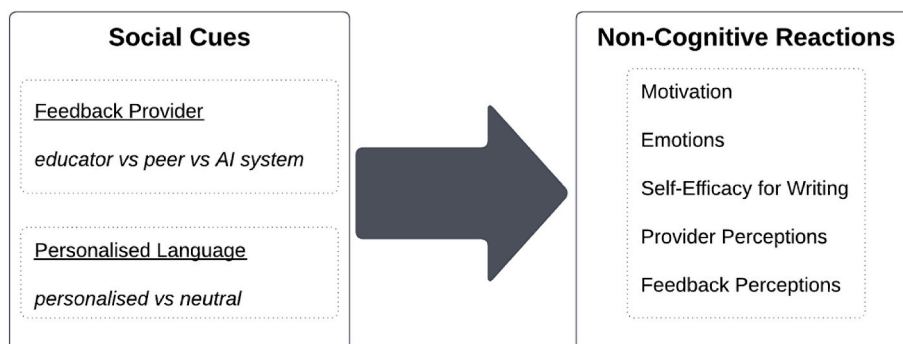

Social Cues	Non-Cognitive Reactions
Feedback Provider *educator vs peer vs AI system* **Personalised Language** *personalised vs neutral*	Motivation Emotions Self-Efficacy for Writing Provider Perceptions Feedback Perceptions

Fig. 1. Overview of the variables under investigation and their assumed relationship.

2020). In the educational context, automated feedback from AI is implemented for different reasons: It can facilitate self-regulated learning or motivation (Cavalcanti et al., 2021; Chiu et al., 2023; Wilson et al., 2021) and can be particularly useful for improving argumentative writing (Zhu et al., 2020). Further, AI systems or automated feedback can save educators resources (Cavalcanti et al., 2021; Wilson et al., 2021).

However, when considering social presence, specifically how people respond to interactions with other humans vs non-human systems (like computers), Okita, Bailenson, and Schwartz (2008) found that interacting with humans showed more facilitating effects than computers, suggesting that learners put more effort into learning when interacting with a human. But, according to Reeves and Nass (1996) and research building on their studies, people are predisposed to apply the same dynamics in human-human and human-computer interactions. Thus, it seems important to consider the dynamics of human-computer interactions in social feedback interactions (Ajjawi & Boud, 2017); particularly in the context of this study, since an AI system's use of language can be an anthropomorphic characteristic (Kaplan et al., 2023).

In this vein, an AI agent's communication style influences its trustworthiness (Kaplan et al., 2023; Rheu, Shin, Peng, & Huh-Yoo, 2021). Regarding an AI's expertise in education, a twofold image is revealed: On the one hand, educators experience AI positively, for example, in its ability to support scientific writing (Kim & Kim, 2022). On the other hand, they mention several issues concerning the AI's transparency about its decision-making, e.g., rating an essay (Shin, Zhong, & Biocca, 2020), or the ease of implementing digital technologies for providing feedback (Clark-Gordon, Bowman, Hadden, & Frisby, 2019). According to students, they perceive the social aspects of AI for providing feedback on argumentative writing as motivating and beneficial for learning (Wambsganns et al., 2022). Thus, while AI is evaluated positively in many respects, users may lack trust in its abilities (Shin et al., 2020), similar to how peer feedback is evaluated.

Research on the antecedents of trust in AI is still in its infancy (Kaplan et al., 2023), but two possible trust trajectories have been discussed (Glikson & Woolley, 2020): On the one hand, initial trust in an AI is high due to a positivity bias towards new technologies (Parasuraman & Manzey, 2010) but decreases when expectations about an AI system's abilities do not hold true (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). People could see an AI system as objective and unbiased and referring to huge amounts of data (see Swiecki et al., 2022). On the other hand, the scepticism towards AI in education might be responsible for low initial trust, which might increase with experience (Kim & Kim, 2022). In this vein, explainable AI, or increasing transparency, would help users trust an AI system's decisions (Kaplan et al., 2023; Khosravi et al., 2022; Shin, 2021). Whereas trust in AI for performing rather technical tasks is higher than in humans, the opposite is true for emotional tasks (Glikson & Woolley, 2020). For example, (teacher) students could be angry when receiving technical feedback from an AI due to their scepticism or they could even enjoy it as they are surprised by the AI's skills.

Overall, whereas educators and peers have their humanness in common, educators and AI share authority, and expertise differentiates all three possible feedback providers.

2.2.2. (Personalised) language as a social cue

Language is a social cue (Holtgraves, 2001; Holtgraves & Kashima, 2008). The use of language is crucial for how people perceive feedback and the feedback provider as well as for people's affective reactions to and their actual use of feedback (Bromme & Jucks, 2018; (Lipnevich, Berg, & Smith, 2016); Winstone, Nash, Parker, & Rowntree, 2017). Even slight nuances in language may dictate whether and how feedback is used (Winstone et al., 2017). Even though not yet investigated in the context of feedback, personalised language is an example of such a nuance: The personalisation effect describes how the use of *you/your*

instead of *the* increases learning (Moreno & Mayer, 2004). It is assumed that personalised language acts as a social cue and thus facilitates connecting new information to the self via self-referencing (Rogers, Kuiper, & Kirker, 1977), whereas from a motivational-emotional view, personalised language elicits social presence (Moreno & Mayer, 2004). Social presence – as outlined in 2.2 – can positively affect non-cognitive elements of feedback effectiveness.

In this vein, by increasing social presence, feedback written in personalised language may lead to increasing feedback effectiveness, for example resulting in higher motivation. The effects of personalisation on motivation have been tested, but results are inconclusive (Ginns, Martin, & Marsh, 2013): For example, Kartal (2010) found medium effects of personalised compared to neutral online learning materials; Zander, Reichelt, Wetzel, Kämmerer, and Bertel (2015) found descriptive but no significant differences between personalised and neutral language. The effects of personalisation hold true for computer agents (Moreno & Mayer, 2004): When instructed by an agent using personalised language, learning was rated significantly less difficult, and performance improved. Furthermore, students who received personalised instructions evaluated the agent as more friendly and helpful. It can thus be assumed that personalised language affects not only perceptions of the information itself but of the provider of the information. Research on personalised language has primarily been conducted in multimedia online learning environments, for example focussing on formulating instructions. This study transfers these approaches to online feedback processes in which personalisation could be a valuable addition – particularly when considering different feedback providers.

There is evidence for the interplay between the language style used to convey information and the provider. For example, people receiving information have certain expectations towards a source's use of language that affect their reactions to the respective information (*language expectancy theory*; LET; Burgoon, Denning, & Roberts, 2002). Experts, for example, can draw from a broader range of language styles, e.g., formal and informal, and still provide credible information, while low-trustworthy sources can only use limited language styles to make their message credible and effective. As outlined above (see 2.2.1), educators are perceived as experts, but peers are not often seen as experts for providing feedback (Hovardas et al., 2014; Strijbos, Narciss, & Dünnebier, 2010), and, similarly, AI systems are rarely considered experts. This would mean that the trust students have in an educator's feedback should be independent of the educator's use of language, i.e., they can use formal (=neutral) or informal (=personalised) language to appear trustworthy, while students would have stricter expectations for the language used by peers and AI systems. As long as the feedback falls within this certain expected range, the feedback information is more likely to be effective and the feedback providers are more likely to be perceived positively (Burgoon et al., 2002, 2016). In addition, these expectations can even be exceeded, such as when the communication is better than expected (*expectation violation theory*; EVT; Burgoon et al., 2016).

Conclusively, attending to personalisation of language in the context of AI systems and feedback providers is crucial when it comes to understanding how feedback can become more effective in higher education.

2.3. The impact of feedback on non-cognitive aspects

Social aspects can influence the effectiveness of feedback, but which aspects are affected by it? As outlined above, feedback effectiveness is partly cognitive (e.g., performance), which has been the focus of considerable literature investigating the effects of feedback on learners' performance (e.g., Wisniewski et al., 2020). Such benefits have even been shown for argumentative writing skills (Latifi et al., 2019) and with feedback from AI (Cavalcanti et al., 2021). Thus, we shift our focus to non-cognitive aspects of feedback effectiveness (e.g., emotions, motivation, self-efficacy, perceptions) which affect the use of feedback

(Henderson et al., 2019).

- **Achievement Emotions.** In general, feedback and the provider of feedback can be sources of a broad range of achievement emotions (e.g., Fong & Schallert, 2023; Molloy et al., 2013; Rowe, 2017). Potentially, social presence plays a role in this context, as it affects students' affective reactions (Richardson et al., 2017; Russo & Benson, 2005). Achievement emotions can be directed at an outcome (e. g., performance) or ongoing activity (e.g., learning with feedback) (Pekrun, 2006). According to Pekrun's (2006) control-value theory (CVT), emotions are a function of perceived control and task value. The task value can, for example, depend on the learning material and its usefulness, e.g., feedback and its provider. Enjoyment and anger are activity-related emotions that likewise depend on the task value (Pekrun, 2006) and will thus be included in this study.
- **Motivation.** According to Ryan and Deci (2000), learners' needs for competence (e.g., being skilled at the task), relatedness (e.g., being accepted), and autonomy (e.g., choosing one's actions) need to be fulfilled to elicit motivation. These needs can be threatened or fulfilled (Leenknecht et al., 2021) by feedback processes. The feedback provider, e.g., the relationship between the provider and the recipient, or personalised language, i.e., the words the provider uses, can, for example, fulfil the need for relatedness and thus motivate (Ten Cate, 2013). Educator feedback compared to that of peers can potentially threaten learners' autonomy (Miao, Badger, & Zhen, 2006).
- **Self-efficacy.** Self-efficacy is the belief about one's capabilities to successfully perform a particular behaviour (Bandura, 1986), e.g., argumentative writing (Putra, Saukah, Basthomi, & Irawati, 2020). Feedback can be a source of self-efficacy (Van Dinther, Dochy, & Segers, 2011; Wang & Wu, 2008), but observing changes in it after feedback is difficult (see Rowe, 2017). Bong and Skaalvik (2003) argued that the provider of feedback must be trustworthy in order to persuade learners of their capabilities, i.e., increase their self-efficacy beliefs in the respective domain. Ruegg (2018) compared students' writing self-efficacy beliefs after peer vs educator feedback and found that educator feedback had a significantly higher effect on writing self-efficacy than peer feedback. Wambsganss, Janson, and Leimeister (2022) pointed at the lack of research on AI systems and self-efficacy beliefs when writing argumentative texts, even though AI agents can be a source of such beliefs (Sikström, Valentini, Sivunen, & Kärkkäinen, 2022). Furthermore, Van Dinther et al. (2011) pointed at the importance of investigating different feedback contexts and types (e.g., language styles) and their effects on self-efficacy beliefs in the respective contexts.
- **Feedback and Provider Perceptions.** In feedback processes, learners have to rely on information from others (Harris, 2012). For feedback to be effective, learners must perceive the feedback information as well as the provider in a certain way (e.g., Eva et al., 2012; Ilgen, Fisher, & Taylor, 1979; Winstone et al., 2017). A learner's epistemic trust in the feedback provider influences whether they will use or ignore the feedback (Carless, 2012; Davis & Dargusch, 2015; Hendriks, Kienhues, & Bromme, 2015). Trust is facilitated by characteristics of the provider, e.g., their expertise and experience (Lucassen & Schraagen, 2011), characteristics in which educators, peers, and AI systems differ (see 2.2.1). Feedback perceptions "capture[s] how students comprehend, perceive, and value a feedback message and how they experience and receive feedback" (Van der Kleij & Lipnevich, 2021, p 349). Variability in students' perceptions has been shown to be high (Van der Kleij & Lipnevich, 2021). The feedback provider might influence feedback perceptions and vice versa (Douglas, Salter, Iglesias, Dowlman, & Eri, 2016; Telio, Regehr, & Ajjawi, 2016).

2.4. Hypotheses

Building on the theoretical and empirical background outlined above, we assumed and tested three hypotheses. We first consider that feedback effectiveness partly depends on the provider of the feedback (Evans, 2013). Characteristics of the provider, i.e., expertise, authority, and humanness, inhibit or facilitate learners' (re)actions to feedback (see 2.2.1), leading to the first hypothesis:

- **Hypothesis 1: Feedback Provider and Non-Cognitive Aspects.** Feedback from educators will have the most beneficial effects while feedback from AI systems will have the least beneficial effects on students' motivation, self-efficacy, emotions, and perceptions of the feedback and the provider. Furthermore, we assume that feedback from humans and authorities will have more beneficial effects on the abovementioned aspects.

Furthermore, language has an influence on non-cognitive aspects in learning contexts (see 2.2.2; Bromme & Jucks, 2018; Lipnevich et al., 2016). Personalised, informal language can have facilitating effects on learners (Moreno & Mayer, 2004), which forms the basis of the second hypothesis:

- **Hypothesis 2: Feedback Language and Non-Cognitive Aspects.** Compared to personalised feedback, neutral feedback will lead to the least beneficial effects on students' motivation, self-efficacy, emotions, and perceptions of the feedback and the provider.

The MISCA model states that the various aspects of feedback processes cannot be considered in isolation (Panadero & Lipnevich, 2022). How the feedback process is perceived thus depends on characteristics of the feedback and its provider. In accordance with the LET and EVT (see 2.2.2), the interaction of language style and the source influences recipients' reactions to the information. Accordingly, trust in feedback from experts, i.e., educators, is rather independent of the language style, whereas trust in feedback from peers and AI systems (as less credible sources) depends on the language style (Burgoon et al., 2002). In this vein, the language style used by the feedback providers can be perceived as unexpected in a positive way, leading information recipients to evaluate the feedback and its provider more positively (Burgoon et al., 2016). For example, the language style can be more informal than expected, but in a way that positively affects the information recipients. Thus, sources are expected to use a certain style of language (e.g., experts can use various styles, whereas non-experts are restricted to a narrower range of styles), and the recipient's perception of the information and the provider is affected by whether these expectations are met. Overall, inferring a direction of this interaction effect for feedback provider and language style is difficult, and we state the third hypothesis:

- **Hypothesis 3: Interaction of Feedback Language and Provider and Non-Cognitive Aspects.** The interaction of language style and the provider influences students' motivation, self-efficacy, emotions, and perceptions of the feedback and the provider.

3. Methods

3.1. Participants and design

According to an a priori power analysis with small to medium effects and 80% power in G*Power (Faul, Erdfelder, Buchner, & Lang, 2009), 96 student teachers were targeted. Via different channels (e.g., mailing lists and social media), we recruited 129 student teachers from teacher education programmes at German universities covering a wide range of subjects within language, natural sciences, humanities, and arts. Of these, $n = 25$ participants dropped out, not affecting the distribution of

the experimental conditions, i.e., dropouts were not due to the respective experimental condition ($\chi^2(2) = 3.99$, $p = .136$). Additionally, we excluded incomplete data from six participants, leading to a final sample of $N = 98$ (71.4% female, 0.04% diverse/not specified; $M_{age} = 23.38$, $SD_{age} = 5.47$; 72.4% at bachelor's degree level; $M_{Semester} = 4.76$, $SD_{Semester} = 2.45$; 99% German native speakers) representative of all German student teachers according to the Federal Statistical Office (Destatis, 2023).

The participants completed the 3×2 between-subjects study with the factors *feedback provider* (peer vs educator vs AI) and *feedback language* (personalised vs neutral) online (unipark.com by Questback EFS Survey). They were randomly assigned to one of the six experimental conditions.

3.2. Measures

3.2.1. Motivation

To assess participants' motivation, we translated Guay, Vallerand, and Blanchard's (2000) Situational Motivation Scale (SIMS). The SIMS was developed to assess situational motivation in field and laboratory settings. It builds on self-determination theory (SDT; Ryan & Deci, 2000), which is important in the context of feedback (Ajjawi, Boud, Henderson, & Molloy, 2019; Evans, 2013; Fong & Schallert, 2023; Ten Cate, 2013) and thus fits our learning situation. The four subscales (i.e., intrinsic motivation, identified regulation, external regulation, and amotivation) include four items each (e.g., "Because I am doing it for my own good" (identified regulation)). They were assessed on a 7-point Likert-type scale from 1 (= *corresponds not all*) to 7 (= *corresponds exactly*). Cronbach's α ranged between 0.62 and 0.90 for the subscales intrinsic motivation (first measurement: 0.85; second measurement: 0.87), identified regulation (first measurement: 0.62; second measurement: 0.73), external regulation (first measurement: 0.93; second measurement: 0.90), and amotivation (first measurement: 0.73; second measurement: 0.83).

3.2.2. Self-efficacy for writing

To assess self-efficacy for writing, we adapted Bruning, Dempsey, Kauffman, McKim, and Zumbrunn (2013) Self-Efficacy for Writing Scale (SEWS) according to Putra et al.'s (2020) adaptation for argumentative writing to match the specific context. In our study, we included only the two subscales of ideation (eight items, e.g., "I can think of many ways to support my arguments in writing") and self-regulation (five items, e.g., "I can control my frustration when I write"). Participants indicated their self-efficacy on a continuous scale ranging from 0 (= *no confidence*) to 100 (= *complete confidence*). Cronbach's α for ideation (first measurement: 0.90; second measurement: 0.92) and regulation (first measurement: 0.80; second measurement: 0.85) was good to excellent.

3.2.3. Achievement emotions

For assessing participants' enjoyment and anger, we employed two respective subscales of the learning-related Achievement Emotions Questionnaire – Short (AEQ-S; Bieleke, Gogol, Goetz, Daniels, & Pekrun, 2021). Compared to the original questionnaire (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011), the short version benefits from shorter administration times while being equally reliable and valid (Bieleke et al., 2021) and was employed for this reason. Furthermore, the learning-related emotions as compared to test- and class-related emotions are relevant for the investigated context (Bieleke et al., 2021). The four items each (e.g., "I enjoy the challenge of working with the feedback" (enjoyment)) were assessed on 5-point Likert scales ranging from 1 (= *strongly disagree*) to 5 (= *strongly agree*). We adapted the items to fit the context (i.e., referred to feedback instead of material). Cronbach's α for the subscales were 0.68 (enjoyment) and 0.71 (anger).

3.2.4. Provider perceptions

Student teachers' perceptions of the feedback providers, i.e., their

trustworthiness, were assessed with the Muenster Epistemic Trustworthiness Inventory (METI; Hendriks et al., 2015). This scale was applied as it focusses on epistemic trustworthiness, which plays a role in feedback environments where key aspects include not only trustworthiness but the acceptance of information (see 2.3). The METI includes three subscales and consists of 16 items (i.e., expertise (seven items), integrity (five items), benevolence (four items)). The items, i.e., antonyms, were assessed on a 7-point semantic differential scale (e.g., 1 = *professional* vs 7 = *unprofessional* (expertise)). Participants in the AI conditions had the option to rate the items as *not applicable* since not all items were applicable. Cronbach's α for the subscales was good to excellent with 0.93 (expertise), 0.80 (integrity), and 0.81 (benevolence).

3.2.5. Feedback perceptions

For assessing participants' perceptions of the feedback, we applied the Feedback Perceptions Questionnaire (FPQ; Strijbos, Pat-El, & Narciss, 2010). Even though the FPQ was validated in a peer feedback context, it is intended to be used for further contexts such as online feedback processes (Strijbos, Pat-El, & Narciss, 2010) and is thus ideal for this study. We included nine items from three subscales (fairness, usefulness, and acceptance, e.g., "I am satisfied with this feedback" (fairness)). These were assessed on a 10-point bipolar scale ranging from 1 (= *fully disagree*) to 10 (= *fully agree*). Cronbach's α was good to excellent (fairness: 0.87; usefulness: 0.93; acceptance: 0.87).

3.2.6. Further variables

Additionally, we included control variables: To gain insights into the participants' attitudes towards online learning, we employed the Perceived Usefulness (PU) subscale of the Technology Acceptance Model 3 (TAM3; Venkatesh & Bala, 2008) as well as the internet-specific epistemological belief scale (ISEB; Bråten, Strømsø, & Samuelstuen, 2005). PU was assessed with four items (e.g., "Using online learning enhances my effectiveness at university") on a 7-point Likert-type scale (1 = *completely disagree* to 7 = *completely agree*) and yielded an internal consistency of Cronbach's α = 0.91. We employed the two subscales of the ISEB dealing with internet communication (six and four items, e.g., "I would rather get feedback on my work face-to-face than on the Internet"). We assessed these on a 10-point scale from 1 (= *completely disagree*) to 10 (= *completely agree*) and reverse coded negatively phrased items. Cronbach's α for both subscales was 0.73. To control for participants' prior knowledge about the topic, we adapted the content knowledge subscale of the TPACK.xs (short assessment instrument for Technological Pedagogical Content Knowledge) by Schmid, Brinaza, and Petko (2020). Its four items (e.g., "I have sufficient knowledge about Kounin's classroom management techniques") were assessed on a 5-point Likert scale ranging from 1 (= *completely disagree*) to 5 (= *completely agree*). Cronbach's α was 0.94.

3.3. Procedure

The study was divided into two parts (see Fig. 2). First, participants were asked to respond to the quantitative measures. Then, they received information about the topic (Kounin's classroom management techniques) and were asked to write an argumentative text (Appendix A). This topic was chosen because it illustrates a straightforward concept that is widely established in teacher education. It provides sufficient scientific evidence to incorporate in argumentative writing and is relevant to student teachers (see Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2016). Discussing the appropriateness for online environments is, particularly after the pandemic, a current debate in public, so participants were likely to have an opinion.

Each participant's text was given feedback on its argumentative quality; the feedback was provided by the first author, who was blind to the participants' experimental conditions. The first author assessed the argumentative quality of the texts by indicating positive or negative valence of seven assessment categories: a complete argument, consisting
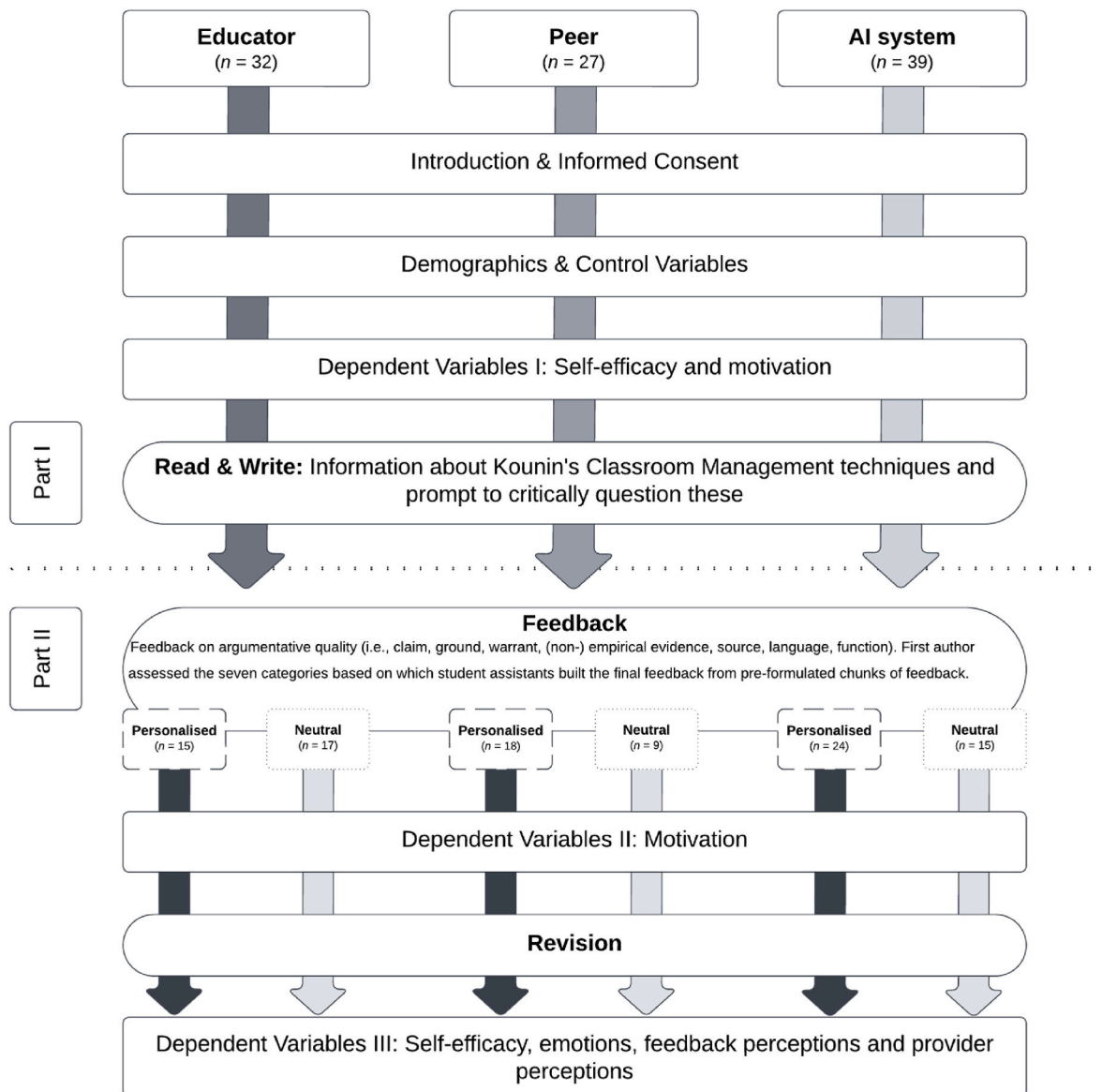
Fig. 2. Flow Diagramme illustrating the Procedure of the study.

of a (1) claim, (2) ground, and (3) warrant, as well as (4) (non-)empirical evidence, (5) source, (6) language, and (7) function. The categories were derived from research on student teachers' argumentative skills (Lytzerinou & Iordanou, 2020), from the central literature on argumentation (Toulmin, 1958), and from feedback research (Narciss, 2012; Shute, 2008). Building on this, two student assistants composed the final feedback according to the experimental condition using generic chunks of feedback that could be easily processed by the recipients (Shute, 2008). The chunks were explicit (e.g., providing examples) and spelled out. The chunks for the seven assessment categories were either personalised or neutral and were either positive or negative (see Table 1 for an example).

The feedback was sent to the participants via the lab's email address. It was provided in a pdf document and designed according to a popular online learning platform. Thus, participants did not know the alleged feedback provider's name, gender, or any other characteristics; The alleged provider remained completely anonymous apart from their status group according to the experimental condition (i.e., educator, peer, AI). One day after they finished the first survey, participants received their feedback and instructions on how to continue with the second part of the study. In the final part of the investigation,

Table 1
Translation of example for feedback chunks (positive chunk for Category "function").

Personalised
Your essay should take your stand on a statement and question it argumentatively. This means that you should use your arguments to critically examine the statement. In my opinion you have implemented this well and worked out your point of view with your arguments.

Neutral
The essay should take a position on a statement and question it argumentatively. This means that arguments should be used to critically examine the statement. This was implemented well, and the point of view was worked out with arguments.

participants were asked to revise their text and complete the questionnaires to assess the dependent variables.

Prior to our experiment, we pilot tested the study with five student teachers (80% female, $M_{age} = 23.2$, $SD_{age} = 3.83$). The pilot tests verified the comprehensibility of the study including its materials and functioned as training for the first author regarding using the assessment categories.

3.4. Ethics

The study (including planning, conducting, analysing, storing etc.) complied with APA (American Psychological Association) ethical standards for research with human subjects as well as with the European Commission's General Data Protection Regulation. All student teachers provided their written informed consent. The study's purpose was clarified at the end of the survey. For their time, participants were reimbursed with 15€.

3.5. Statistical analyses

Due to the complexity of the data combined with the sample size, linear mixed effects models did not converge, and we decided to run basic linear regression models instead (Brauer & Curtin, 2018). In our models, peers/humans and personalised language served as intercepts, i. e., reference categories. We built one linear model per hypothesis and ran the usual model checks in which no abnormalities occurred. We included both the overall mean values and those of the respective subscales for each dependent variable. The models were built in R (R Core Team, 2022). We set an α error of 0.05, but to avoid alpha error inflation due to multiple testing, we adjusted it to 0.01 according to the Bonferroni correction (0.05/5 dependent variables). In conclusion, since this method is conservative, findings approaching significance ($α < .05$) are interpreted as tendencies that could be of practical importance.

3.5.1. Preparatory analyses

We checked the equality of our measures across the experimental groups. Age ($F(1, 138.6) = 4.77, p < .05, η_p^2 = .05$) and semester ($F(1, 37.8) = 6.7, p < .05, η_p^2 = .07$) showed mean differences across the feedback language conditions. To account for the unequal distribution, we included both variables as control variables in the models with feedback language as a predictor.

When comparing the humanness means, we found significant differences between the feedback language conditions for age ($F(1, 147.9) = 5.09, p < .05, η_p^2 = .05$), semester ($F(1, 37.54) = 6.68, p < .05, η_p^2 = .07$), and previous content knowledge ($F(1, 4.18) = 4.54, p < .05, η_p^2 = .05$). For the authority means, we found significant differences between the feedback language conditions for age ($F(1, 153.4) = 5.33, p < .05, η_p^2 = .05$), semester ($F(1, 37.9) = 6.8, p < .05, η_p^2 = .07$), and previous content knowledge ($F(1, 4.78) = 5.23, p < .05, η_p^2 = .05$). Motivation means differed in the interaction ($F(1, 4.25) = 4.26, p < .05, η_p^2 = .04$). Thus, we controlled for all these variables in the models with feedback language or the interaction of feedback language and provider as predictors in the respective analyses.

3.6. Manipulation checks

We asked participants for their opinions about the setting as well as the feedback. Perceptions regarding the setting were assessed with three items ranging from 1 (positive end) to 8 (negative end). Participants rated the setting as being quite realistic, enjoyable, and useful ($M = 3.2, SD = 1.57$). According to $χ^2$ analyses, there were no differences between the experimental conditions for either item. The feedback manipulation was assessed with eight items on the same scale. There were significant differences between the experimental conditions of feedback language for the items that assessed the degree of formality and friendliness of the feedback. Neutral feedback was perceived as being more formal ($t(94.09) = −2.02, p < .05; M_{neutral} = 2.49, M_{personalised} = 3.09$) and less friendly ($t(80.19) = 1.9, p < .10; M_{neutral} = 2.76, M_{personalised} = 2.16$) than personalised feedback,[1] confirming the personalisation effect (Moreno & Mayer, 2004).

Furthermore, we checked whether participants actually used their feedback by comparing their initial texts with their revised texts. On average, participants wrote 293.43 words ($SD = 118.10, Min = 204, Max = 802$) in their initial and 360.52 ($SD = 137.29, Min = 208, Max = 1002$) words in their revised texts. They received $M = 3.37$ ($SD = 0.98$) negative chunks before the revision and $M = 1.58$ ($SD = 0.9$) after the revision and were, thus, significantly better after the revision ($t(192.53) = 13.33, p = < .01$).

Overall, additional sources (e.g., Google) were used by 19.4% of participants for writing text and 28.6% for revising text, whereas they were used by 24.5% for contextual help and 4.1% for structural help. Further, 90.70% of participants were not reminded of a person they knew while participating which indicates the anonymity of the feedback process. Participants' perceived usefulness of online learning environments was average ($M = 4.8, SD = 1.44$). Before completing the study, participants' content knowledge about Kounin's classroom management techniques was rather low ($M = 1.74, SD = 0.97$).

4. Results

4.1. Main analyses

Fig. 3 shows an excerpt of the descriptive values of participants' perceptions of the feedback providers. A full table including all experimental conditions as well as the descriptive values for the other dependent variables across the experimental conditions can be found in Supplementary Material A.

4.1.1. Hypothesis 1 - Feedback Provider and Non-cognitive aspects

The first hypothesis was confirmed regarding provider perceptions: There was a significant difference in participants' perceptions of the different providers' trustworthiness (see also Fig. 3): We found that the AI system was perceived as significantly more trustworthy regarding its expertise compared to peers and educators ($F(2, 95) = 6.17, R^2 = 0.12, p < .01, η_p^2 = .12$; Table 2)[2] as well as compared to humans ($F(1, 96) = 11.1, R^2 = 0.10, p < .01, η_p^2 = .10$; Table 3). The feedback provider's authority tended to have an influence on perceived expertise ($F(1, 96) = 6.60, p = .012, η_p^2 = .06$; Table 4). All three providers ($F(2, 95) = 3.43, p = .036, η_p^2 = .07$; Table 2), humanness ($F(1, 96) = 6.33, p = .014, η_p^2 = .06$; Table 3), and authority ($F(1, 96) = 3.56, p = .062, η_p^2 = .04$; Table 4) were partly close to but did not significantly influence overall trustworthiness. The provider in any constellation did not affect the subscales
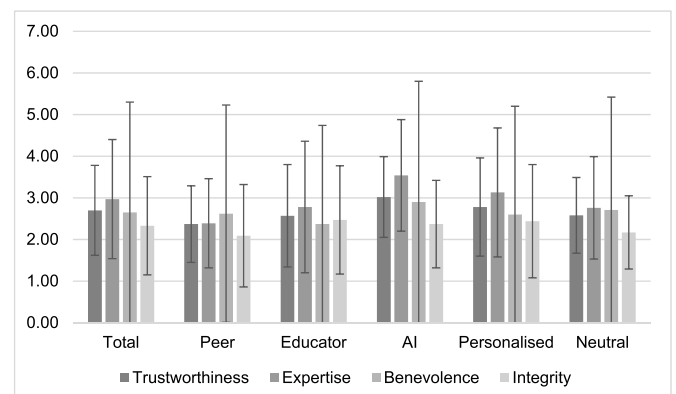

Fig. 3. Descriptive means and standard Deviations of the provider perceptions (METI) Distinguished between Selected experimental factors.

---

[1] The antonyms were defined as 1 = *formal* and 8 = *informal* as well as 1 = *friendly* and 8 = *unfriendly*, respectively.

[2] According to Cohen (1988), $η_p^2 = .010$ indicates a small, $η_p^2 = .059$ a moderate, and $η_p^2 = .138$ a large effect.

Table 2

Basic linear models for comparison of feedback provider on provider perception (METI) (H1).

	Est.	SE	t	p
Provider Perception ~ Feedback Provider ($F(2, 95) = 3.43, p < .05, \eta_p^2 = .07$)				
Intercept	2.37	0.20	11.71	<.001***
Educator	0.21	0.27	0.75	.455
AI	0.66	0.26	2.49	.01*
Expertise ~ Feedback Provider ($F(2, 95) = 6.17, R^2 = 0.12, p < .01, \eta_p^2 = .12$)				
Intercept	2.39	0.26	9.12	<.001***
Educator	0.39	0.36	3.38	.273
AI	1.15	0.34	3.38	.001**
Benevolence ~ Feedback Provider ($F(2, 95) = 1.41, p = .248, \eta_p^2 = .03$)				
Intercept	2.61	0.26	10.09	<.001***
Educator	−0.24	0.35	−0.69	.489
AI	0.29	0.34	0.87	.387
Integrity ~ Feedback Provider ($F(2, 95) = 0.78, p = .460, \eta_p^2 = .02$)				
Intercept	2.09	0.23	9.17	<.001***
Educator	0.38	0.31	1.21	.228
AI	0.28	0.30	0.94	.350

Note. *p ≤ .05, **p ≤ .01, ***p ≤ .001.

Table 3

Basic linear models for comparison of humanness on provider perception (METI) (H1).

	Est.	SE	t	p
Provider Perception ~ Humanness ($F(1, 96) = 6.33, p < .05, \eta_p^2 = .06$)				
Intercept	2.48	0.14	18.16	<.001***
AI	0.54	0.22	2.52	.014*
Expertise ~ Humanness ($F(1, 96) = 11.1, R^2 = 0.10, p < .01, \eta_p^2 = .10$)				
Intercept	2.60	0.18	14.67	<.001***
AI	0.94	0.28	3.33	.001**
Benevolence ~ Humanness ($F(1, 96) = 2.36, p = .128, \eta_p^2 = .02$)				
Intercept	2.48	0.17	14.20	<.001***
AI	0.43	0.28	1.54	.128
Integrity ~ Humanness ($F(1, 96) = 0.09, p = .760, \eta_p^2 < .01$)				
Intercept	2.30	0.15	14.83	<.001***
AI	0.08	0.25	0.31	.760

Note. *p ≤ .05, **p ≤ .01, ***p ≤ .001.

Table 4

Basic linear models for comparison of authority on provider perception (METI) (H1).

	Est.	SE	t	p
Provider Perception ~ Authority ($F(1, 96) = 3.56, p = .062, \eta_p^2 = .04$)				
Intercept	2.37	0.20	11.57	<.001***
Authority	0.45	0.24	1.89	.062
Expertise ~ Authority ($F(1, 96) = 6.60, p < .05, \eta_p^2 = .06$)				
Intercept	2.39	0.27	8.92	<.001***
Authority	0.81	0.31	2.57	.012*
Benevolence ~ Authority ($F(1, 96) = 0.03, p = .869, \eta_p^2 < .01$)				
Intercept	2.61	0.26	10.00	<.001***
Authority	0.05	0.31	0.17	.869
Integrity ~ Authority ($F(1, 96) = 1.46, p = .230, \eta_p^2 = .02$)				
Intercept	2.09	0.23	9.21	<.001***
Authority	0.32	0.27	1.21	.230

Note. *p ≤ .05, **p ≤ .01, ***p ≤ .001.

of benevolence and integrity (Tables 2–4). Notably, in the AI condition, on average 4.08 participants chose the option *not applicable* for each item (expertise: 0.83/item, benevolence: 6.00/item, integrity: 5.40/item).

Furthermore, self-efficacy, motivation, emotions, and feedback perceptions were not predicted by the feedback provider. Yet, the effect of humanness on self-efficacy ideation was close to significance ($F(1, 96) = 6.36, p = .013, \eta_p^2 = .06$), with the AI system promoting self-efficacy beliefs. Likewise, the influence of authority on enjoyment was almost significant ($F(1, 96) = 5.85, p = .017, \eta_p^2 = .06$); compared to peers, participants experienced more enjoyment with feedback from authorities. See Supplementary Material B for summaries of the statistical analyses.

4.1.2. Hypothesis 2 - feedback type and non-cognitive aspects

To begin with, while there were no changes in participants' self-efficacy beliefs before and after the feedback ($t = -1.63, p = .105, M_{SEWS\_I} = 6.48, M_{SEWS\_II} = 6.85$, Cohen's $d^3 = -0.23$), their motivation decreased ($t = 2.74, p < .01, M_{SIMS\_I} = 4.56, M_{SIMS\_II} = 4.15$, Cohen's $d = 0.39$). Nevertheless, contrary to our second hypothesis, language had no significant influence on overall motivation ($F(3, 94) = 2.80, R^2 = 0.08, p = .044, \eta_p^2 = .06$). Similarly, the feedback language also did not influence the respective subscales (Table 5).

Furthermore, contrary to our assumption, the feedback language did not affect participants' writing self-efficacy ($F(3, 94) = 1.03, p = .384, \eta_p^2 = .02$), their emotions (enjoyment: $F(3, 94) = 0.16, p = .925, \eta_p^2 < .01$; anger: $F(3, 94) = 0.53, p = .664, \eta_p^2 < .01$), or their perceptions of the feedback ($F(3, 94) = 0.54, p = .653, \eta_p^2 < .01$) or the feedback provider (F

Table 5

Basic linear models for comparison of feedback language on motivation (SIMS) (H2).

	Est.	SE	t	p
Situational Motivation ~ Feedback Language + Age + Semester (F(3, 94) = 2.80, p = .044, η2p = .06)				
Intercept	0.31	0.48	0.66	.510
Neutral	−0.55	0.22	−2.51	<0.05*
Age	−0.01	0.02	−0.39	.697
Semester	−0.07	0.04	−1.53	.129
Amotivation ~ Feedback Language + Age + Semester (F(3, 94) = 1.27, p = .288, η2p = .02)				
Intercept	−0.25	0.58	−0.43	.667
Neutrals	−0.46	0.27	−1.70	.093
Age	0.03	0.02	1.11	.270
Semester	−0.06	0.05	−1.19	.236
Intrinsic Motivation ~ Feedback Language + Age + Semester (F(3, 94) = 1.97, p = .124, η2p = .04)				
Intercept	0.07	0.55	0.13	.894
Neutral	−0.34	0.25	−1.33	.188
Age	−0.04	0.02	−1.57	.120
Semester	0.01	0.05	0.21	.836
Identifying Regulation ~ Feedback Language + Age + Semester (F(3, 94) = 2.49, p = .065, η2p = .03)				
Intercept	0.24	0.54	0.44	.660
Neutral	−0.56	0.25	−2.22	.029
Age	0.00	0.02	0.21	.835
Semester	−0.10	0.05	−2.08	.041
External Regulation ~ Feedback Language + Age + Semester (F(3, 94) = 2.48, p = .066, η2p = .05)				
Intercept	1.20	0.86	1.39	.169
Neutral	−0.87	0.40	−2.17	<0.05*
Age	−0.03	0.04	−0.74	.458
Semester	−0.11	0.08	−1.40	.164

Note. *p ≤ .05, **p ≤ .01, ***p ≤ .001.

---

[3] According to Cohen (1988), the conventional effect sizes for *t*-tests are: 0.2 (small effect), 0.5 (moderate effect), and 0.8 (large effect).

$(3, 94) = 0.33$, $p = .800$, $\eta_p^2 < .01$); feedback language also did not affect the respective subscales (see Supplementary Material C). Hypothesis 2 was not supported.

4.1.3. Hypothesis 3 - Interaction of Feedback Language and Provider and Non-cognitive aspects

The interaction of feedback language and provider was not significant for non-cognitive aspects of feedback effectiveness (Supplementary Material D). Hypothesis 3 was thus rejected.

5. Discussion

The findings provide insights into the impact of feedback provider and language on non-cognitive aspects of feedback effectiveness. Our first hypothesis was accepted regarding participants' perceptions of the feedback provider: We found that trustworthiness, particularly regarding the expertise of the provider, was influenced by the feedback provider. The AI system by itself or as a counterpart to humans was ascribed more expertise for providing feedback than educators and peers. Hypothesis 2 investigated the effects of (personalised) language on non-cognitive aspects and found no significant effects. Nor were there any effects of the interaction of feedback language and provider on non-cognitive feedback effectiveness (Hypothesis 3).

Overall, participants were able to work with the feedback when revising their argumentative text, thus indicating a high validity of the setting. Regarding Hypothesis 1, the results point at the importance of trust in feedback processes. Surprisingly, we found that the AI system was ascribed more trustworthiness across different analyses than were educators and peers. Ascribing humans trust requires information about them (e.g., information about their expertise or experience) (Lucassen & Schraagen, 2011), and, similarly, researchers have argued that for users to trust an AI system, they must have information about the AI, e.g., transparency about its functioning (Hoff & Bashir, 2015; Kaplan et al., 2023; Lee & See, 2004). But we offered no such information in this study.

As mentioned in the theoretical background, different trust trajectories are suggested when interacting with AI: We found high initial trust compared to that in humans, indicating that participants had a potential positivity bias towards the AI (Parasuraman & Manzey, 2010). This is in line with colloquial prejudice about AI systems in assessment (Chiu et al., 2023; Swiecki et al., 2022). First, the benefits of AI in decision-making (e.g., bias-free decisions) prevail and give the AI the benefit of the doubt. Compared to human feedback providers, an AI would be assumed to not have its personal taste affect its assessment, which would make it highly competent, e.g., having high expertise to provide feedback. Nonetheless, in this vein, we must consider that participants' trustworthiness perceptions might have decreased if allowed to continue interacting with the AI system (see 2.2.1). In this study, there was only one interaction point, leaving it unknown whether the more participants interact with the system, the more their perceptions of its trustworthiness are affected.

Another colloquial benefit is that an AI system is able to integrate more data into its decision-making progress than humans ever could, thereby leading to higher ratings of expertise. In addition, the AI system's use of language is an antecedent of trust. User-related aspects like participants' competences and experiences, their attitudes, and their personality might justify the high trust (Kaplan et al., 2023). Their experiences and familiarity with AI as well as their culture may be beneficial for the development of trust in AI systems (Kaplan et al., 2023). At the same time, AI systems that generate and assess text are currently not the most common AI systems to be applied in higher education contexts (Cavalcanti et al., 2021), meaning that participants' experience and familiarity with such systems might still be rather low.

Separate from but related to trust, in this study authority was not found to significantly predict trustworthiness (regarding expertise), but we found a tendency in line with the literature. This reasoning complements the findings above as well as the descriptive tendencies. Overall, these findings point at the importance of considering trust in feedback processes.

Regarding emotions, there was a tendency towards authority to predict enjoyment. According to the CVT (Pekrun, 2006), enjoyment is a function of perceived control and task value. The results indicate that the task value of feedback from an authority might be higher compared to feedback from peers. Increases in task value could be ascribed to social presence, which was elicited by all providers (and not significantly more by authorities). Nevertheless, since the task apparently was valued as being rather positive, anger did not arise (Pekrun, 2006). Future research should thus investigate the meaning of the task value. In addition, some methodological issues regarding emotions, such as their retrospective assessment (Robinson & Clore, 2002) and the rather low reliability of the enjoyment subscale (see 3.2.3) may also help explain the absence of significance.

The lack of significance regarding self-efficacy beliefs might be explained by social presence, i.e., no differences in social presence were elicited by the three feedback providers. Additionally, as mentioned by Rowe (2017), seeing changes in self-efficacy immediately after feedback might be difficult, as self-efficacy beliefs are rather stable. For future research, we would thus recommend longitudinal designs. Nevertheless, we found a tendency of feedback from AI compared to that from humans to increase participants' self-efficacy ideation. As outlined above, AI system seems to be a trusted source of feedback and therefore a trusted source of self-efficacy (Bong & Skaalvik, 2003). According to Bandura (1997), self-efficacy beliefs depend, among other factors, on social persuasions, and these are more likely to come from trusted sources. This is another indicator that social considerations are important in feedback processes.

Neither motivation nor feedback perceptions were affected by the feedback providers. First, effects of the feedback process (including the provider) on motivation tend to be rather low (Wisniewski et al., 2020) because of the possibility that motivation will be negatively influenced by the feedback's/provider's threat to learners' needs (Ryan & Deci, 2000; Wisniewski et al., 2020). In this study, the providers could have threatened the need for relatedness or autonomy (Miao et al., 2006; Ten Cate, 2013) and, thereby, diminished the positive effects on motivation. The absence of significance might also be explained by methodological reasons, like the assessment of motivation in a fictious setting (Touré-Tillery & Fishbach, 2014). Second, feedback perceptions are assumed to be influenced by feedback providers due to the importance of the relationship between provider and recipient (e.g., Douglas et al., 2016), i.e., the social presence. This relationship may have been difficult to establish, as the study's setting was anonymous, and participants did not have the opportunity to contact the feedback providers.

Hypothesis 2 tackled the question of how the language style affects non-cognitive feedback effectiveness. First and against our expectations, personalised feedback language did not significantly affect motivation. According to the SDT (Ryan & Deci, 2000), learners' needs for competence, autonomy, and relatedness were not fulfilled – potentially due to the fictional setting – but the descriptive values indicate potential differences. This finding contrasts with the personalisation effect (Moreno & Mayer, 2004) that was confirmed in the material. Nevertheless, we mentioned above that results in terms of motivation are inconclusive (Ginns et al., 2013). Similar to Zander et al. (2015), our results hint at the effect of personalised language on motivation, indicating the importance of considering it in future research.

Against our expectations, we did not find effects of language on self-efficacy, emotions, and feedback perceptions due to social presence and reduced cognitive load (Moreno & Mayer, 2004). For one, social presence (from the participants' perspective) might not have differed between the two experimental conditions. Second, social presence depends on interactivity (Kim, Kwon, & Cho, 2011), which could have been insufficient in this study to elicit effects. Methodologically, the absence of significance for self-efficacy could have been due to the short duration

of the intervention (Pajares, 2003), and that for emotions could have been related to the retrospective assessment (Robinson & Clore, 2002). Furthermore, for self-efficacy, emotions, and feedback perceptions, effects could have arisen in comparison to no feedback (Hattie & Timperley, 2007) or when considering the valence of the feedback (Hattie & Timperley, 2007; Kluger & DeNisi, 1996), but these situations were not assessed in the current study.

Hypothesis 3 addressed the question of how the interaction of feedback language and provider affects non-cognitive feedback effectiveness; the results did not indicate any interaction effect. In accordance with the LET, recipients expect the information provider to use appropriate language. Transferred to our findings, the feedback providers' language styles could have been viewed as appropriate (Burgoon et al., 2002). This could have been due to participants characterising the providers as competent to provide feedback and, thus, accepting a broader range of language styles as appropriate. This is, for example, in line with findings from the first hypothesis, in which the AI system was characterised as trustworthy regarding its expertise.

5.1. Practical implications

From a practical point of view, our findings highlight the importance of establishing trust in feedback processes in higher education. The relationship of feedback recipient and feedback provider as well as their individual characteristics should be considered when designing feedback processes. The findings indicate that an AI system can be implemented as a feedback provider; however, when doing this, attention must be paid to the respective people interacting with it, as they should consider the level of trustworthiness they have in the AI as well as potential information that could cause harm. Furthermore, aspects like AI literacy, experience, and culture play important roles and influence trust in AI systems (Kaplan et al., 2023). For example, as discussed, the frequency of interactions with an AI system can influence users' trustworthiness perceptions. Our study showed that the sole presence of an alleged AI system without any reference to its technological functioning did influence learners' perceptions, which highlights the need for fostering AI and feedback literacy to protect learners from manipulation.

For feedback practitioners, we recommend using personalised language, as more formal language styles tend to be detrimental for motivation. Furthermore, this is easily implemented and, considering the methodology, e.g., the assessment of motivation in fictional settings, the non-significant finding is still of practical relevance.

In general, it seems important to consider the social and non-cognitive aspects of learning with feedback, as effects were found for different aspects, even if only as tendencies. When implementing such processes, these social and non-cognitive aspects should be addressed to create awareness and hinder the potential negative effects of feedback. The study showed that social cues are influential for feedback effectiveness in terms of trustworthiness and, thus, should not be neglected.

5.2. Theoretical implications

Theoretically, this study adds to a growing field of research. We can tentatively confirm that, in accordance with the MISCA model (Panadero & Lipnevich, 2022), characteristics of the feedback process, namely the *message*, i.e., language style, the *agents*, i.e., the feedback provider, as well as the *student*, i.e., their non-cognitive reactions, *implemented* in digital learning environments *(context)* play a role for feedback effectiveness. Thus, in future research the complexity of the feedback process should be considered. Further research should address the question of whether and how provider-related effects can be acted upon to make feedback even more successful (e.g., those aspects influencing the trustworthiness). For example, future research might focus on promoting trust and deepening the understanding of the encounter between feedback providers and recipients. Additionally, future studies should make sure to consider AI systems in learning environments.

Furthermore, it would be interesting to investigate the effect of feedback provider and language style on skill acquisition. In our study, we focussed on participants' non-cognitive reactions and did not particularly investigate the extent to which their argumentative writing skills improved. Taking a closer look at the impact of the feedback, its provider, and language across multiple time points on learners' skill development would inform the didactical design of feedback processes.

Also, we can infer that it might be promising to integrate AI systems in educational research as their application is becoming more and more common, but research on their influence in this context is not as thorough. Similarly, our focus on social and non-cognitive aspects in learning environments is justified, and new questions arise as these aspects continue to change. We showed that participants' perceptions of the feedback provider were influenced by social cues and that social cues showed tendencies for influencing emotions, motivation, and self-efficacy. Further insights into how non-cognitive aspects are influenced are crucial for theory and practice and should, thus, be subject to further research. For example, transparency, e.g., meta-information about an AI, as mentioned above, could influence learners' reactions (Shin et al., 2020) and should be investigated further.

5. 3. Limitations

The launch of ChatGPT has stirred a huge discussion about the use of AI (in education). Peoples' attitudes towards or their experiences with an AI system might quickly change. The results might look different now, which is why they should be considered with caution.

To begin with, we investigated written feedback provided anonymously in a digital learning environment on argumentative texts. This specific context (e.g., a sample of German student teachers) limits the generalisability of the results and thus points at the importance of investigating the role of social cues in other contexts. Furthermore, it is important to note that effects depend on the individual interaction of a specific system or a specific human with the learner, and effects might also depend on the culture of the participants (Kaplan et al., 2023) as well as the abovementioned development regarding AI in education. Second, the methodological limitations of our study (e.g., assessment of motivation in fictional settings (Touré-Tillery & Fishbach, 2014), short duration of intervention (Pajares, 2003)) indicate that future studies in the field could provide deeper insights into the assumed relationships. Along this line, the reliability of some subscales was below acceptable, which could potentially threaten the validity of the results; thus, for now we suggest caution in interpreting the results. In addition, the sample size and small number of participants per experimental condition may have been insufficient for the study. Finally, feedback can only be considered effective when performance improves. Our manipulation check showed that participants' texts significantly improved after the feedback, but in our analysis we focussed on the non-cognitive aspects; investigating learners' actual performance and the cognitive learning outcomes is important for future research.

6. Conclusion

So, how do the feedback provider and the personalisation of their language style influence the effectiveness of feedback in terms of non-cognitive aspects? According to our results, the feedback provider influences feedback effectiveness. In particular, the study showed that the trustworthiness of the feedback depends on the provider, as the AI system was perceived as more trustworthy regarding its expertise compared to humans (i.e., peer and educator). To implement feedback in online learning environments, this means that feedback providers and their characteristics (e.g., expertise) should be considered. Furthermore, this result indicates that an AI system can be applied in feedback processes without expecting drawbacks regarding its influence on learners' evaluations of its trustworthiness.

Nevertheless, these findings must be used with caution. A range of other variables, like characteristics of the recipient, might have

influenced the results. Furthermore, we found tendencies that indicate further differences in feedback effectiveness due to the feedback provider and language. For online feedback processes, this means that the feedback provider and the language they use do potentially affect learners' non-cognitive reactions (i.e., trustworthiness). Thus, this again confirms that such social cues should be considered when designing feedback processes. The study may become particularly relevant as AI applications are increasingly implemented in educational contexts.

Overall, when designing feedback environments, the influential nature of social cues like the feedback provider on learners' reactions should be considered.

Statement on open data and ethics

Participants provided their written informed consent before the study. Due to ethical reasons, raw data cannot be made available.

Author note

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. This work was supported by the Einstein Centre Digital Future (ECDF).

CRediT authorship contribution statement

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary Material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.caeai.2023.100189.

Acronyms

AI	Artificial Intelligence
CVT	Control Value Theory
EVT	Expectation Violation Theory
LET	Language Expectancy Theory
MISCA	Message, Implementation, Student, Context, Agent

Appendix A. Instructions for Writing the Argumentative Text

German	„Kounins Techniken zur Klassenführung sind hilfreich in online Klassenräumen."
Original	Setzen Sie sich kritisch mit diesem Statement auseinander. Wenden Sie dafür Kounins Techniken zur Klassenführung auf eine beispielhafte Situation in einem online Klassenraum an und argumentieren Sie deren Relevanz. Nehmen Sie Stellung, inwieweit die Techniken dabei unterstützen, eine derartige Unterrichtssituation in den Griff zu bekommen.
	Denken Sie daran, starke und vollständige Argumente zu formulieren. Beachten Sie weiterhin, dass Sie beispielsweise verschiedene Perspektiven berücksichtigen und Ihre Aussagen belegen können. Sie können auch weitere Quellen einbeziehen. Behalten Sie die Aufgabenstellung während der Bearbeitung im Hinterkopf. Sie können mit der Umfrage fortfahren, wenn Sie min. 200 Wörter geschrieben haben.
Translation	"Kounin's classroom management techniques are helpful in online classrooms."
	Critically reflect on this statement. Apply Kounin's classroom management techniques to an example situation in an online classroom and argue their relevance. Comment on the extent to which the techniques help to manage such a classroom situation.
	Remember to formulate strong and complete arguments. Further note that you can, for example, consider different perspectives and provide evidence for your statements. You can also include additional sources. Keep the assignment in mind while you are working on it. You can continue with the survey when you have written at least 200 words.

References

Ajjawi, R., & Boud, D. (2017). Researching feedback dialogue: An interactional analysis approach. *Assessment & Evaluation in Higher Education, 42*(2), 252–265. https://doi.org/10.1080/02602938.2015.1102863

Ajjawi, R., Boud, D., Henderson, M., & Molloy, E. (2019). Improving feedback research in naturalistic settings. In M. Henderson, R. Ajjawi, D. Boud, & E. Molloy (Eds.), *The impact of feedback in higher education: Improving assessment outcomes for learners* (pp. 245–265). Palgrave Macmillan Cham. https://link.springer.com/chapter/10.1007/978-3-030-25112-3_14

Bandura, A. (1986). *Social Foundations of Thought and action: A social cognitive theory.* Prentice-Hall, Inc.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* Freeman.

Belletier, C., Normand, A., & Huguet, P. (2019). Social-facilitation-and-impairment effects: From motivation to cognition and the social brain. *Current Directions in Psychological Science, 28*, 1–6. https://doi.org/10.1177/0963721419829699

Bieleke, M., Gogol, K., Goetz, T., Daniels, L., & Pekrun, R. (2021). The AEQ-S: A short version of the achievement emotions questionnaire. *Contemporary Educational Psychology, 65*, Article 101940. https://doi.org/10.1016/j.cedpsych.2020.101940

Bong, M., & Skaalvik, E. M. (2003).). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Bråten, I., Strømsø, H. I., & Samuelstuen, M. S. (2005). The relationship between internet-specific epistemological beliefs and learning within internet technologies. *Journal of Educational Computing Research, 33*(2), 141–171. https://doi.org/10.2190/E763-X0LN-6NMF-CB86

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of non independent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods, 23*(3), 389–411. https://doi.org/10.1037/met0000159

Bromme, R., & Jucks, R. (2018). Discourse and expertise: The challenge of mutual understanding between experts and laypeople. In M. F. Schober, D. N. Rapp, &

M. A. Britt (Eds.), *The Routledge handbook of discourse processes* (pp. 222–246). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315687384-13.

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology, 105*(1), 25–38. https://doi.org/10.1037/a0029692

Burgoon, J. K., Bonito, J. A., Lowry, P. B., Humpherys, S. L., Moody, G. D., Gaskin, J. E., et al. (2016). Application of expectancy violations theory to communication with and Judgments about Embodied agents during a decision-making task. *Journal of Human Computer Studies, 91*, 24–36. https://doi.org/10.1016/j.ijhcs.2016.02.002

Burgoon, M., Denning, V. P., & Roberts, L. (2002). Language expectancy theory. *The Persuasion Handbook: Developments in Theory and Practice*, 117–136.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*, 219–233. https://doi.org/10.1080/03075070600572132

Carless, D. (2012). Trust and its role in facilitating dialogic feedback. In D. Boud, & E. Molloy (Eds.), *Feedback in higher and professional education* (pp. 90–103). Routledge.

Cavalcanti, M. C., Ferreira, J. M., & Gomes, I. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers & Education: Artificial Intelligence, 2*, Article 100027. https://doi.org/10.1016/j.caeai.2021.100027

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers & Education: Artificial Intelligence, 1*, Article 100002. https://doi.org/10.1016/j.caeai.2020.100002

Chiu, T., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers & Education: Artificial Intelligence, 4*, Article 100118. https://doi.org/10.1016/j.caeai.2022.100118

Cho, K., Chung, T. R., King, W. R., & Schunn, C. D. (2008). Peer-based computer-supported knowledge refinement: An empirical investigation. *Communications of the ACM, 51*(3), 83e88.

Clark-Gordon, C. V., Bowman, N. D., Hadden, A. A., & Frisby, B. N. (2019). College instructors and the digital red pen: An exploratory study of factors influencing the adoption and non-adoption of digital written feedback technologies. *Computers & Education, 128*, 414–426. https://doi.org/10.1016/j.compedu.2018.10.002

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic Press. https://doi.org/10.4324/9780203771587

Davis, S. E., & Dargusch, J. M. (2015). Feedback, Iterative processing and academic trust – teacher education students' perceptions of assessment feedback. *Australian Journal of Teacher Education, 40*(1). https://doi.org/10.14221/ajte.2015v40n1.10

Destatis. (2023). *Statistik der studenten. Statistisches Bundesamt Deutschland*. https://www-genesis.destatis.de/genesis/online?operation=statistic&levelindex=0&levelid=1695804866463&code=21311#abreadcrumb.

Douglas, T., Salter, S., Iglesias, M., Dowlman, M., & Eri, R. (2016). The feedback process: Perspectives of first and second year undergraduate students in the disciplines of education, health science and nursing. *Journal of University Teaching and Learning Practice, 13*(1). https://doi.org/10.53761/1.13.1.3

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*(6), 697–718.

Eva, K. W., Armson, H., Holmboe, E., Lockyer, J., Loney, E., Mann, K., et al. (2012). Factors influencing responsiveness to feedback: On the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education, 17*, 15–26. https://doi.org/10.1007/s10459-011-9290-7

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research, 83*(1), 70–120. https://doi.org/10.3102/0034654312474350

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

Fong, C. J., & Schallert, D. L. (2023). "Feedback to the future": Advancing motivational and emotional perspectives in feedback research. *Educational Psychologist, 58*(3), 146–161. https://doi.org/10.1080/00461520.2022.2134135

Ginns, P., Martin, A. J., & Marsh, H. W. (2013). Designing instructional text in a conversational style: A meta-analysis. *Educational Psychology Review, 25*, 445–472. https://doi.org/10.1007/s10648-013-9228-0

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *The Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and Extrinsic motivation: The situational motivation scale (SIMS). *Motivation and Emotion, 24*(3), 175–213. https://doi.org/10.1023/A:1005614228250

Harris, P. L. (2012). *Trusting what you're told*. Belknap of Harvard University Press.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Henderson, M., Ryan, T., & Phillips, M. (2019). The challenges of feedback in higher education. *Assessment & Evaluation in Higher Education, 44*(8), 1237–1252. https://doi.org/10.1080/02602938.2019.1599815

Hendriks, F., Kienhues, D., & Bromme, R. (2015). Measuring Laypeople's trust in experts in a digital age: The Muenster epistemic trustworthiness Inventory (METI). *PLoS One, 10*(10). https://doi.org/10.1371/journal.pone.0139309

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on predictions of novice-performance. *Journal of Experimental Psychology: Applied, 5*(2), 205–221.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Holtgraves, T. M. (2001). *Language as social action: Social psychology and language use*. Psychology Press.

Holtgraves, T. M., & Kashima, Y. (2008). Language, meaning, and social cognition. *Personality and Social Psychology Review, 12*(1), 73–94. https://doi.org/10.1177/1088868307309605

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education, 71*, 133–152. https://doi.org/10.1016/j.compedu.2013.09.019

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

Jonassen, D. H., & Kim, B. (2010). Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology Research & Development, 58*(4), 439–457. https://doi.org/10.1007/s11423-009-9143-8

Jonsson, A. (2012). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education, 14*(1), 63–76. https://doi.org/10.1177/1469787412467125

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in artificial intelligence: Meta-Analytic findings. *Human Factors, 65*(2), 337–359. https://doi.org/10.1177/00187208211013988

Kartal, G. (2010). Does language matter in multimedia learning? Personalization principle revisited. *Journal of Educational Psychology, 102*, 615–624. https://doi.org/10.1037/a0019345

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., et al. (2022). Explainable artificial intelligence in education. *Computers & Education: Artificial Intelligence, 3*, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

Kim, N. J., & Kim, M. K. (2022). Teacher's perceptions of using an artificial intelligence-based educational tool for scientific writing. *Frontiers in Education*. https://doi.org/10.3389/feduc.2022.755914

Kim, J., Kwon, Y., & Cho, D. (2011). Investigating factors that influence social presence and learning outcomes in distance higher education. *Computers & Education, 57*(2), 1512–1520. https://doi.org/10.1016/j.compedu.2011.02.005

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, II9*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86*(3), 643–680. https://doi.org/10.3102/0034654315626799

Kreijns, K., Xu, K., & Weidlich, J. (2022). Social presence: Conceptualization and measurement. *Educational Psychology Review, 34*(1), 139–170. https://doi.org/10.1007/s10648-021-09623-8

Kwok, L. (2008). Students' perceptions of peer evaluation and teachers' role in seminar discussions. *Electronic Journal of Foreign Language Teaching, 5*(1), 84–97.

Latifi, S., Noroozi, O., Hatami, J., & Biemans, H. J. A. (2019). How does online peer feedback improve argumentative essay writing and learning? *Innovations in Education & Teaching International, 58*(2), 195–206. https://doi.org/10.1080/14703297.2019.1687005

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021). Recognizing the emotional state of human and virtual instructors. *Computers in Human Behavior, 114*. https://doi.org/10.1016/j.chb.2020.106554

Lechermeier, J., & Fassnacht, M. (2018). How do performance feedback characteristics influence recipients' reactions? A state-of-the-art review on feedback source, timing, and valence effects. *Management Review Quarterly, 68*(2), 145–193. https://doi.org/10.1007/s11301-018-0136-8

Leenknecht, M., Wijnia, L., Köhlen, M., Fryer, L., Rikers, R., & Loyens, S. (2021). Formative assessment as practice: The role of students' motivation. *Assessment & Evaluation in Higher Education, 46*(2), 236–255. https://doi.org/10.1080/02602938.2020.1765228

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *HumanFactors: The Journal of the Human Factors and Ergonomics Society, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lipnevich, A. A., Berg, D. A., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. T. L. Brown, & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 169–185). Routledge.

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and Theories: Descriptions, definitions, and conclusions. *Frontiers in Education*. https://doi.org/10.3389/feduc.2021.720195

Lucassen, T., & Schraagen, J. M. (2011). Factual accuracy and trust in information: The role of expertise. *Journal of the American Society for Information Science and Technology, 62*(7), 1232–1242. https://doi.org/10.1002/asi.21545

Lytzerinou, E., & Iordanou, K. (2020). Teachers' ability to construct arguments, but not their perceived self-efficacy of teaching, predicts their ability to evaluate arguments. *International Journal of Science Education, 42*(4), 617–634. https://doi.org/10.1080/09500693.2020.1722864

Metzger, M., Flanagin, A., Eyal, K., Lemus, D., & Mccann, R. (2016). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the Contemporary media environment. *Annals of the International Communication Association, 27*(1), 293–335. https://doi.org/10.1080/23808985.2003.11679029

Miao, Y., Badger, R., & Zhen, Y. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*(3), 179–200. https://doi.org/10.1016/j.jslw.2006.09.004

Molloy, E., Borrell-Carrió, F., & Epstein, R. (2013). The impact of emotions in feedback. In D. Boud, & E. Molloy (Eds.), *Feedback in higher and professional education* (pp. 60–81). Routledge.

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

Narciss, S. (2012). Feedback strategies. In N. Seel (Ed.), *Encyclopedia of the learning sciences* (pp. 1289–1293). Springer Science & Business Media, LLC.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers & Education: Artificial Intelligence, 2*. https://doi.org/10.1016/j.caeai.2021.100041

Okita, S., Bailenson, J., & Schwartz, D. L. (2008). Mere belief in social action improves complex learning. In G. Kanselaar, V. Jonker, P. A. Kirschner, & F. J. Prins (Eds.), *International perspectives in the learning sciences: Cre8ing a learning world. Proceedings of the Eighth International conference for the learning sciences – ICLS 2008* (pp. 132–139). International Society of the Learning Sciences. https://doi.org/10.22318/icls2008.2.132.

Osborne, J. F. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*, 463–468. https://doi.org/10.1126/science.1183944

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies, 27*, 7893–7925. https://doi.org/10.1007/s10639-022-10925-9

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly, 19*, 139–158. https://doi.org/10.1080/10573560308222

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and Typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, Article 100416. https://doi.org/10.1016/j.edurev.2021.100416

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 52*(3), 381–410.

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., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*, 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Putra, I., Saukah, A., Basthomi, Y., & Irawati, E. (2020). The predicting power of self-efficacy on students' argumentative writing quality. *The Journal of Asia TEFL, 17*, 379–394. https://doi.org/10.18823/asiatefl.2020.17.2.5.379

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

Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu (No. JRC107466)*. Joint Research Centre (Seville site).

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

Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human-Computer Interaction, 37*(1), 81–96. https://doi.org/10.1080/10447318.2020.1807710

Richardson, J. C., Maeda, Y., Lv, J., & Caskurlu, S. (2017). Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis. *Computers in Human Behaviour, 71*, 402–417. https://doi.org/10.1016/j.chb.2017.02.001

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128*(6), 934–960. https://doi.org/10.1037/0033-2909.128.6.934

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*(9), 677–688. https://doi.org/10.1037/0022-3514.35.9.677

Rowe, A. D. (2017). Feelings about feedback: The role of emotions in assessment for learning. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education* (pp. 159–172). Springer. https://doi.org/10.1007/978-981-10-3045-1_11.

Ruegg, R. (2018). The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy. *Language Learning Journal, 46*(2), 87–102. https://doi.org/10.1080/09571736.2014.958190

Russo, T., & Benson, S. (2005). Learning with invisible others: Perceptions of online presence and their relationship to cognitive and affective learning. *International Forum of Educational Technology & Society, 8*(1), 54–62.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Schmid, M., Brinaza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education, 157*. https://doi.org/10.1016/j.compedu.2020.103967

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2021). The cognitive-affective-social theory of learning in digital environments (CASTLE). *Educational Psychology Review, 34*(1), 1–38. https://doi.org/10.1007/s10648-021-09626-5

Seitchik, A. E., Brown, A. J., & Harkins, S. G. (2017). Social facilitation: Using the molecular to inform the molar. In S. G. Harkins, K. D. Williams, & J. Burger (Eds.), *The Oxford handbook of social influence* (pp. 183–220). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199859870.013.8.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies, 146*. https://doi.org/10.1016/j.ijhcs.2020.102551

Shin, D., Zhong, B., & Biocca, F. (2020). Beyond user experience: What constitutes algorithmic experiences. *International Journal of Information Management, 52*, 1–11. https://doi.org/10.1016/j.ijinfomgt.2019.102061

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189. https://doi.org/10.3102/0034654307313795

Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers and Education, 188*. https://doi.org/10.1016/j.compedu.2022.104564

Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Unravelling Peer Assessment, 20*(4), 291–303. https://doi.org/10.1016/j.learninstruc.2009.08.008

Strijbos, J.-W., Pat-El, R., & Narciss, S. (2010). Structural validity and Invariance of the feedback perceptions questionnaire. *Studies In Educational Evaluation, 68*, Article 100980. https://doi.org/10.1016/j.stueduc.2021.100980

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., et al. (2022). Assessment in the age of artificial intelligence. *Computers & Education: Artificial Intelligence, 3*, Article 100075. https://doi.org/10.1016/j.caeai.2022.100075

Telio, S., Regehr, G., & Ajjawi, R. (2016). Feedback and the educational alliance: Examining credibility judgements and their consequences. *Medical Education, 50*(9), 933–942.

Ten Cate, O. T. J. (2013). Why receiving feedback collides with self determination. *Advances in Health Sciences Education, 18*, 845–849. https://doi.org/10.1007/s10459-012-9401-0

Toulmin, S. (1958). *The uses of argument*. Cambridge University Press. https://doi.org/10.1017/CBO9780511840005

Touré-Tillery, M., & Fishbach, A. (2014). How to measure motivation: A Guide for the experimental social Psychologist. *Social and Personality Psychology Compass, 8*(7), 328–341. https://doi.org/10.1111/spc3.12110

Tsui, A. B., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing, 9*(2), 147–170. https://doi.org/10.1016/S1060-3743(00)00022-9

Van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review, 6*(2), 95–108. https://doi.org/10.1016/j.edurev.2010.10.003

Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20*(4), 280–290. https://doi.org/10.1016/j.learninstruc.2009.08.010

Van der Kleij, F. M., & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability, 33*, 345–373. https://doi.org/10.1007/s11092-020-09331-x

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research Agenda on interventions. *Decision Sciences, 39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Wambsganss, T., Janson, A., & Leimeister, J. M. (2022). Enhancing argumentative writing with automated feedback and social comparison nudging. *Computers & Education, 191*, Article 104644. https://doi.org/10.1016/j.compedu.2022.104644

Wambsganss, T., Kueng, T., Soellner, M., & Leimeister, J. M. (2021). ArgueTutor: An adaptive dialog-based learning system for argumentation skills. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–13). https://doi.org/10.1145/3411764.3445781

Wang, S. L., & Wu, P. Y. (2008). The role of feedback and self-efficacy on web-based learning: The social cognitive perspective. *Computers & Education, 51*(4), 1589–1598. https://doi.org/10.1016/j.compedu.2008.03.004

Wilson, J., Ahrendt, C., Fudge, E. A., Raiche, A., Beard, G., & MacArthur, C. (2021). Elementary teachers' perceptions of automated feedback and automated scoring: Transforming the teaching and learning of writing using automated writing evaluation. *Computers & Education, 168*, Article 104208. https://doi.org/10.1016/j.compedu.2021.104208

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.03087

Zajonc, R. B. (1965). Social facilitation. *Science, 149*(3681), 269–274. https://doi.org/10.1126/science.149.3681.269

Zander, S., Reichelt, M., Wetzel, S., Kämmerer, F., & Bertel, S. (2015). Does personalisation promote learners' attention? An Eye-Tracking study. *Frontline Learning Research, 3*(4), 1–13. https://doi.org/10.14786/flr.v3i4.161

Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing, 48*(3), 209–222. https://doi.org/10.1016/1060-3743%2895%2990010-1

Zhu, M., Liu, O. L., & Lee, H. S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education, 143*, Article 103668. https://doi.org/10.1016/j.compedu.2019.103668