

Please quote as: Wambsganss, T.; Janson, A; Käser, T.; Leimeister, J. M. (2022).
Improving Students Argumentation Learning with Adaptive Self-Evaluation Nudging.
Proceedings of the ACM on Human-Computer Interaction (PACM), 6 (CSCW2), 1-31.

Improving Students Argumentation Learning with Adaptive Self-Evaluation Nudging

THIEMO WAMBSGANSS, EPFL, Switzerland

ANDREAS JANSON, University of St.Gallen, Switzerland

TANJA KÄSER, EPFL, Switzerland

JAN MARCO LEIMEISTER, University of St.Gallen, Switzerland and University of Kassel, Germany

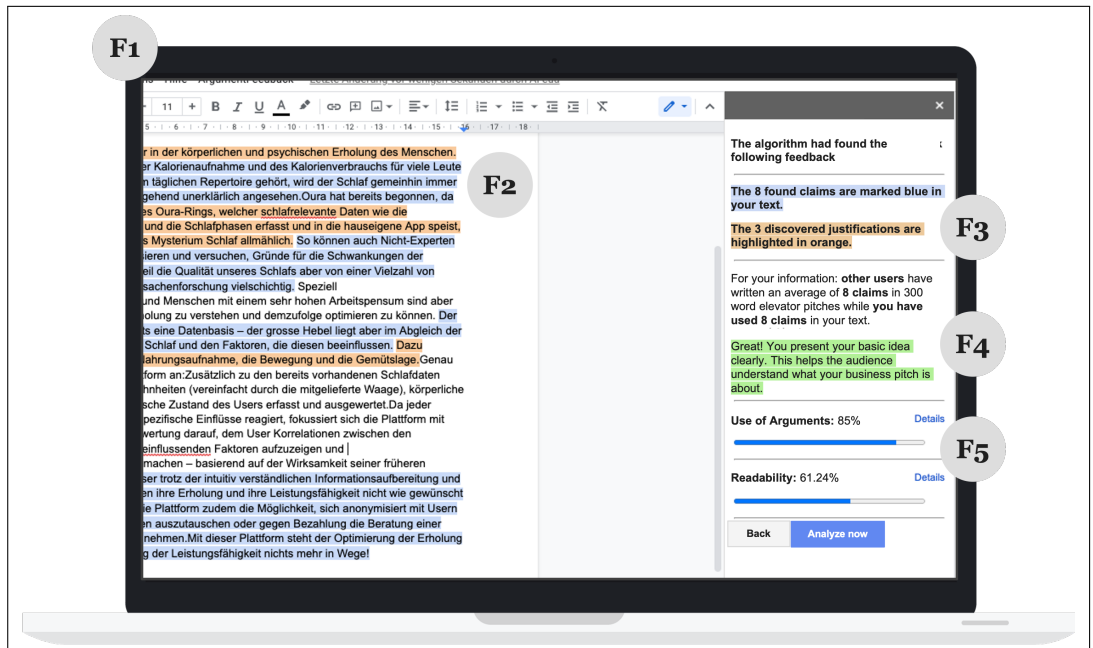


Fig. 1. Screenshot of our intelligent argumentation writing support system: a student conducts a persuasive writing exercise and is nudged with intelligent argumentation self-evaluation based on the argumentation quality of her text to increase augmentation skill learning. For the sake of this paper we translated the user interface into English.

Authors' addresses: Thiemo Wambsganss, thiemo.wambsganss@epfl.ch, EPFL, Lausanne, Switzerland; Andreas Janson, andreas.janson@unisg.ch, University of St.Gallen, St.Gallen, Switzerland; Tanja Käser, tanja.kaeser@epfl.ch, EPFL, Lausanne, Switzerland; Jan Marco Leimeister, janmarco.leimeister@unisg.ch, University of St.Gallen, St.Gallen, Switzerland, University of Kassel, Kassel, Germany.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2022/11-ART520 \$15.00

<https://doi.org/10.1145/3555633>

Recent advantages from computational linguistics can be leveraged to nudge students with adaptive self-evaluation based on their argumentation skill level. To investigate how individual argumentation self-evaluation will help students write more convincing texts, we designed an intelligent argumentation writing support system called ArgumentFeedback based on nudging theory and evaluated it in a series of three qualitative and quantitative studies with a total of 83 students. We found that students who received a self-evaluation nudge wrote more convincing texts with a better quality of formal and perceived argumentation compared to the control group. The measured self-efficacy and the technology acceptance provide promising results for embedding adaptive argumentation writing support tools in combination with digital nudging in traditional learning settings to foster self-regulated learning. Our results indicate that the design of nudging-based learning applications for self-regulated learning combined with computational methods for argumentation self-evaluation has a beneficial use to foster better writing skills of students.

CCS Concepts: • **Computing methodologies** → **Natural language processing**; • **Human-centered computing** → **Natural language interfaces**; **Field studies**; • **Applied computing** → **Interactive learning environments**.

Additional Key Words and Phrases: educational applications, argumentation learning, adaptive learning, digital nudging

ACM Reference Format:

Thiemo Wambsganss, Andreas Janson, Tanja Käser, and Jan Marco Leimeister. 2022. Improving Students Argumentation Learning with Adaptive Self-Evaluation Nudging. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 520 (November 2022), 31 pages. <https://doi.org/10.1145/3555633>

1 INTRODUCTION

The ability to memorize existing information is becoming less important in today's environment as most information is easily available. Instead, the ability to arrange and absorb information is becoming increasingly important. As a result, work profiles are shifting toward transdisciplinary, ambiguous, and creative jobs [102]. When it comes to the composition of skills and knowledge imparted, educational institutions must change their curriculum [94]. Metacognition abilities, such as critical thinking, teamwork, and problem-solving, have become more crucial to teach to students [29]. International institutions such as the Organization for Economic Cooperation and Development (OECD) and the World Economic Forum (WEF) have advocated for a shift in student education to prepare future workers with the necessary skills for digitization [114]. In fact, the OECD included metacognition skills as a major element in their Learning Framework 2030 [64]. One example of these competencies is the skill of arguing in a structured, reflective, and well-formed way [96]. Argumentation is not only an important aspect of our daily speech and thinking, but it also supports a variety of other metacognitive skills like communication, teamwork, and problem-solving. [48]. The ability to construct convincing arguments has been recognized as the foundation for persuading an audience of novel ideas since Aristotle's studies, and it plays a major role in analyzing different viewpoints and productive, democratic civil discourse, such as citizens determining whether certain news is fake or not [22].

However, due to a lack of individual and personal learning experiences (i.e., [38, 105]), teaching and training argumentation abilities in educational institutions are inherently hampered. Not just in the distance-learning scenarios of Massive Open Online Courses (MOOCs, [78]), but also in typical large-scale lectures at universities, students are challenged with unequal educator-student ratios. Large class numbers in high schools, mass lectures at universities with more than 100 students per speaker, and MOOCs with more than 1000 participants all obstruct individual interaction between students and educators [78, 118]. This trend has been aided by the rise of MOOCs in recent years. In 2017, 33.1 percent of learners around the world took at least one online course, up from 24.8 percent in 2012 [51]. With Coursera, the largest MOOC provider, launching a successful Initial

Public Offering (IPO) in March 2021, more distance learning settings are expected to grow [97]. Several studies have found that a lack of individualized support leads to poor learning outcomes, high dropout rates, and unhappiness with the entire learning experience, particularly when it comes to gaining argumentation skills [9, 27, 40].

Utilizing recent advances in computational linguistics is one possibility for providing students with individual self-evaluation and adaptive feedback [71]. Methods from natural language processing (NLP) and machine learning (ML) have been successfully used to predict argumentation quality in given texts [50, 54]. In this context, argumentation mining (AM) research is a crucial field for the development of support systems that identify arguments in unstructured texts [54]. AM can be used to score the quality of a text and provide tailored feedback about textual documents [16, 55, 81, 83]. For at least 35 years, scientists (i.e., [79]), especially those from the fields of educational technology, computer-supported collaborative work (CSCW) and human-computer interaction (HCI), have designed systems to support the active teaching of argumentation for students with input masks, representational guidance, or adaptive feedback to enhance students' learning of argumentation (e.g., [2, 20, 57, 66, 69, 110]). Nevertheless, the design and adoption of current argumentation learning systems are lagging behind the recent technological developments in HCI and computational linguistics research. This eventually leads to a lack of literature and design knowledge from an interdisciplinary perspective for building an accessible, intelligent, and learner-centered argumentation tool that enables self-regulated learning [119], e.g., to support collaborative scenarios [57]. Thus, learners still struggle to train argumentation skills due to a lack of adaptive and instant evaluations in their individual learning process since current approaches either fail to provide scalable and formative feedback on the argumentation skills of students or are not student-centered [50, 73]. AM algorithms have been embedded into learning exercises to support students when learning argumentation skills in peer review scenarios [108, 110] or in online discussions [57]. However, even though much literature on argumentation systems exists, insights on specific learner-centered design features with a more nuanced view on certain design characteristics in combination with self-evaluation based on recent computational methods, such as AM, are rare. Moreover, studies that confirm the recently investigated positive effects on adaptive argumentation feedback based on AM are lacking [2, 110].

To consider learner-centricity, we drew on the concept of nudging, originally a concept from behavioral economics [92], to nudge users to utilize computational methods in their learning process. In fact, the soft-paternalistic paradigm of nudging users towards an intended behavior has shown positive effects on decision making in digital environments [13], e.g., for nudging towards secure [37, 70, 121] and privacy-friendly internet behavior [77]. Nudging is based on psychological effects that influence or counteract behavior to support users in their decision-making, especially by exploiting biases and heuristics of individuals [59]. Embedding both nudges in traditional learning environments and digital nudges in learning tools have shown positive effects on learning outcomes, especially for self-regulated learning such as self-evaluation [13, 37, 61, 115]. For example, [61] have found that push notifications about missing submissions nudge students to improve assignment adherence and course grades, and [28] have found that nudging students to give motivational advice to their peers (e.g., how to stop procrastinating) leads to higher academic grades. However, as much literature on argumentation learning exists, extant research falls short of providing a computational approach combined with a student-centered and theory-based design along with rigorous empirical evaluations to compare distinct features such as a self-evaluation nudge to help students learn how to argue (i.e. [2, 108, 110, 113]). An interdisciplinary HCI perspective combined with recent methods from AM to form a learner-centered argumentation writing support system that provides students with the opportunity to monitor and evaluate themselves based on nudging theory is missing [2, 41, 110, 113].

Drawing on nudging theory as a guiding design principle [92], we aimed to contribute by investigating, designing, and evaluating a novel student-centered learning system called ArgumentFeedback to enable students to self-evaluate their argumentation writing skills independently [6, 8, 120]. We followed a rigorous theory-motivated design approach, where we systematically searched literature in the field of educational technology and HCI following [18, 103] to carefully derive requirements for the design of ArgumentFeedback. Based on these design insights, we followed the methodology of [59] to design different nudges for argumentation skill learning based on self-evaluation [17]. We iteratively evaluated our design with 83 students in three empirical studies to investigate if and how individual argumentation self-evaluation based on ML in combination with digital nudging helps students write more convincing texts.

To develop an intelligent writing support tool based on AM, we (1) constructed a novel, theory-driven argumentation annotation scheme for student-written business model pitches, (2) compiled a corpus of 200 student-written texts, (3) trained a long short-term memory (LSTM) model predict the individual argumentation skill level, and (4) embedded the model into our student-centered and theory-informed design based on nudging theory [59, 117]. ArgumentFeedback provides students with individual argumentation self-evaluation by highlighting the argumentative components, by providing an adaptive learning dashboard when writing persuasive business pitches, and by adaptively nudging students to write more argumentative texts based on social comparison [17, 58, 58].

To determine the impact of adaptive self-evaluation with digital nudging for argumentative writing on students' argumentation skills, we evaluated our learning system in one qualitative study (Study 1), one laboratory experiment (Study 2), and one field experiment (Study 3). We observed that participants who received an intelligent self-evaluation nudge including a social comparison and an adaptive feedback message when writing a persuasive business pitch wrote more formally argumentative texts compared to students who received only general self-monitoring. Furthermore, the perceived persuasiveness of the texts was significantly higher than of the texts from students using the other tool. Moreover, the measured intention to use, the perceived ease of use, the perceived usefulness [100], and the self-efficacy [6] provide promising results for the usage of social nudging in combination with adaptive feedback for effective self-regulated argumentative skill learning. The results provide evidence that a student-centered learning tool based on digital nudging combined with ML and NLP feedback helps students write more persuasive texts.

Our work makes three major contributions. First, based on recent advances in computational linguistics and digital nudging theory for self-regulated argumentation skill learning and metacognition skills in general, we offer insight into a comprehensive design of a student-centered writing support tool. Second, we thoroughly evaluated our tool ArgumentFeedback in three empirical trials with a total of 83 students, demonstrating the effectiveness of social nudging for argumentation learning. The findings shed light on the possibilities of using NLP and ML in combination with digital nudging to build student-centered learning support that promotes self-regulated argumentative writing throughout a student's learning journey. Our study, thus, confirms recent findings of adaptive argumentation learning support (e.g., [2, 57, 110]) and transfers them to 1) a new pedagogical domain and 2) combines them with another design paradigm, namely digital nudging. Finally, our findings provide a case study on how to enhance metacognition skills in a scalable and individual format in large-scale scenarios.

2 RELATED WORK AND CONCEPTUAL BACKGROUND

Our research is related to studies on argumentation models, computer-supported argumentation learning systems, and AM algorithms, as well as literature about digital nudging and self-regulated learning.

2.1 Argumentation Skills and Underlying Theoretical Models

Argumentation is an omnipresent foundation of our daily communication and thinking [47]. In general, argumentation aims at increasing or decreasing the acceptability of a controversial standpoint [26]. Logically, structured arguments are a required precondition for persuasive conversations, general decision-making, and drawing acknowledged conclusions [25]. Across multiple fields, such as law, science, politics, or management, individuals must support their claims with essential facts, argue in support of conclusions that are derived from those facts, and counter the claims of the opponent in a principled way to convince others of their position or justify a conclusion [104]. Decades ago, research found that humans are generally deficient in argumentation (e.g., [10, 56, 98]). They often do not recognize the difference between merely expressing an opinion versus making a claim based on facts. Moreover, they do not rebut the arguments of others but ignore points of conflict and continue to establish their own argument [10]. Thus, research in fields such as HCI or educational technology has shown a growing interest in developing argumentation systems to support individuals [2, 26, 110, 113].

Theories of argumentation have a long history in philosophy, linguistic research, and mathematics. One of the first foundations was provided by Aristotle with his theory of persuasion. Afterwards, many frameworks and “rules” of argumentation have been proposed and identified (e.g., [47, 93, 96, 106]).

In his fundamental theory, Aristotle distinguished between three interconnected principles of persuasion: *logos* (logic and proof of argumentation), *ethos* (authority and credibility of the speaker), and *pathos* (empathy and vivid language). Also authors in CSCW have built up on these three elements [57, 93]. Accordingly, an individual can build on these three strategies to persuade an opponent. *Logos* focuses on the general formality and structure of argumentation. *Ethos* is largely dependent on the individual and their relation to the opponent. *Pathos* is about context-related emotions and the strength of languages.

Most theoretical and practical approaches to argumentation vary in their level of detail, perspective, and specific context of applicability. Nevertheless, several scholars (e.g., [73–75]) identified that the logic of argumentation is probably the foundation across the different theoretical argumentation approaches. In general, argumentation theory agrees on the importance of *logos* as a basis for proper argumentation. An argumentation discourse should fundamentally consider all relevant facts, claims should be well-grounded and supported by premises, and both supporting and conflicting claims should be taken into account [47, 96]. Most argumentation theories formulate formal argumentation models, which address the logical part of Aristotle’s theory of persuasion. One of the most prominent argumentation models is the Toulmin model [93, 95, 96]. Accordingly, an argument consists of several components, such as a claim and at least one premise. The claim is the central component and the statement of an argument, which is justified by premises. According to Toulmin’s argumentation theory, a “good” argument involves a logical structure built on ground, claim, and warrant, whereas the grounds are the evidence used to prove a claim [95]. For example, according to the Toulmin model, each argument can be broken down into six parts (claim, premise/ground, warrant, backing, rebuttal, qualifier). Each of these argument components fulfills a specific argumentative role and complements each other to create the argument. Nevertheless, it is commonly considered that “claim” and “premise” (or also called “evidence”) are the main components of every argument, and the rest are supporting sub-argument parts that may or may not exist in an argument [73, 86]:

- **Claim:** the central point or conclusion of the argument. The ultimate goal is to convince the opponent of the truthfulness of the claim.

- **Premise (fact, evidence, data):** the support or rationale for the claim. This is the data or evidence that the arguer uses to explain and support their claim.

2.2 Computer-Supported Argumentation Learning Systems

Approaches for teaching argumentation are limited. [42] identified three major challenges for teaching it: *"Teachers lack the pedagogical skills to foster argumentation in the classroom, so there exists a lack of opportunities to practice argumentation; external pressures to cover material leaving no time for skill development; and deficient prior knowledge on the part of learners"*. Therefore, many authors have claimed that fostering argumentation skills should be assigned a more central role in our formal educational system [23, 49]. Most students learn to argue in the course of their studies simply through interactions with their classmates or teachers. In fact, individual support of argumentation learning is missing in most learning scenarios. To train argumentation, it is of great importance for the individual student to receive continuous feedback and tutoring throughout their learning journey [38, 105].

Hence, researchers, especially from the fields of educational technology, CSCW and HCI, have analyzed how technology-based learning systems can address this gap and enhance students' learning of argumentation. The application of computer-supported tools in education bears several advantages, such as consistency, scalability, perceived fairness, widespread use, and better availability compared to human teachers. Computer-based argumentation systems can help relieve some of the burdens on teachers to teach argumentation by supporting learners in creating, editing, interpreting, or reviewing their own arguments [74]. This has been investigated across a variety of fields, including law [69], science [66, 90], conversational argumentation [20, 112], business reviews [110], and online debates [57]. Different computer-supported approaches have been used in education. Intelligent tutoring systems (ITS) and computer-supported collaborative learning (CSCL) [44] are of special relevance for argumentation learning since argumentative discussions and debates have been identified as a key for collaborative learning settings. Following [74, 110], three different technology-based argumentation learning systems in the field of CSCL and ITS can be distinguished:

- **Discussion scripting approaches:** Students are provided with structured elements for argumentation learning to stimulate interactions based on script theory of guidance [31, 74].
- **Representational guidance approaches:** Students are supported by being provided with representations of their argumentation structures with the objective to foster individual reasoning, collaboration, and learning. A typical example is to help students represent their argument structure in the form of node-and-link graphs (e.g., [62, 69, 89]).
- **Adaptive support approaches:** Students are provided with pedagogical feedback on their actions with hints and recommendations to encourage and guide future activities in the writing processes. Typical approaches use an automated evaluation to indicate whether an argument is syntactically and semantically correct (e.g., [2, 57, 69, 85, 87, 107, 110, 113]).

We propose to combine recent advances in NLP, ML, and AM to evaluate new forms of HCI and computer-supported learning systems based on digital nudging theory for self-regulated argumentation skill learning [6, 92, 120]. Our aim is to not only address the educational task by providing feedback with recent computational methods but also focus on the learner through a user-centered and theory-motivated design to enable an adaptive and individual learning experience independent of an instructor, time, and place [119].

AM is a relatively new research field in computational linguistics that focuses on the extraction and analysis of arguments from textual corpora and following and analyzing the lines of argumentation, i.e., the interplay between arguments. Mainly since 2007, scientists have started

to publish studies on AM in legal texts, online reviews, or debates [11, 35, 50, 60]. AM tries to analyze the arguments of a given text based on a defined argumentation structure (often based on [96]). The identification of argumentation structures can be done on three different levels: the first level is to detect a sentence containing an argument to differentiate argumentative from non-argumentative text units [34]. The second level deals with classifying argument components into claims and premises [60, 84]. The third level is the identification of argumentative relations [68, 84]. Researchers have developed an increasing interest in intelligent writing assistance [81, 86] since it enables adaptive argumentative writing support with tailored feedback about arguments in texts [50, 55]. However, the complexity of using this technology in combination with digital nudging to provide students with intelligent self-evaluation has, so far, been poorly assessed [50]. In our approach, we focus on the first two subtasks to assess the argumentation level of a student to provide individual self-evaluation in combination with social nudging [59].

2.3 Digital Nudging to Foster Students' Self-Regulated Learning

In 2009, [91] introduced the concept of *nudging* to suggest guiding user behavior through the targeted design of user interfaces. They define a nudge as *"any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any option or significantly changing their economic incentives"*. The aim is to build on the known systematic biases in human decision-making to support people in making better decisions [14]. Based on insights from behavioral economics, psychological effects that influence behavior are applied or counteracted to support users in their decision-making process [92]. The concept of influencing people with small design interventions to nudge them towards a certain behavior has been applied in several domains, such as information systems (e.g., [59, 117]) or HCI (e.g., [1, 12, 52]). More recently, the positive effects of nudging have been also used for the design of digital learning systems. [61] have found that push notifications about missing submissions nudged students to improved assignment adherence and course grades compared to students who did not receive the nudge. Moreover, [28] have found that nudging students to give motivational advice to peers (e.g., how to stop procrastinating) leads to higher academic grades. However, to the best of our knowledge, the concept of nudging has not been investigated for argumentation skill learning or argumentative writing support, e.g., to nudge users to write more persuasive texts.

We believe that embedding the concept of nudging for individual writing feedback could help students learn how to argue in a more effective way. We back up our hypothesis on self-regulated learning theory [8]. This theory supports our underlying hypothesis that individual and personal feedback including digital nudging on a student's argumentation motivates or engages the student to improve their skill level. Self-regulated learning theory reflects that students learn better with formative feedback and goal setting [6, 120]. For students in a learning process, especially, critical reflection through self-monitoring and self-evaluation is an important component for effective learning [120]. It can be an initial trigger for a student's learning process and thus the creation of new knowledge structures. However, the right portion of self-monitoring and self-reflection in combination with a learning goal is important for students to learn effectively [6, 120]. The paradigm of nudging could provide guidance in better designing a self-evaluation mechanism [12], e.g., in how to *frame*, *anchor*, or *socially compare* the user's behavior to nudge them towards more persuasive writing [59, 92]. For example, the concept of social comparison [17], also-called social nudge [7], might lead to a combination of learning feedback and goal setting (e.g., by stating a feedback message such as *"other users have used more arguments in the same task"*). For instance, nudges that include social comparisons with other learners have led to better exam performance and an increase in study time [63]. Theoretically, the underlying mechanism is also rooted in social cognitive theory [5] and is thus, a prime candidate for positive behavior change in learning

processes [6]. Therefore, we aim to investigate if and how individual argumentation self-evaluation based on ML combined with digital nudging will help students write more convincing texts. To do so, we designed a novel pedagogical scenario based on a new annotated argumentation corpora of student-written texts to provide students with adaptive self-evaluation in a real-world writing exercise.

3 DESIGN AND DEVELOPMENT OF ARGUMENTFEEDBACK

In order to investigate how to design an intelligent argumentation writing support tool with self-evaluation based on ML to influence students' argumentation writing skills, we designed ArgumentFeedback. ArgumentFeedback consists of two main components: 1) the theory-based self-evaluation interface based on digital nudging theory and 2) the adaptive feedback algorithm to access the individual argumentation level of students. An overview of the basic concept of ArgumentFeedback is depicted in Figure 2. The evaluation of ArgumentFeedback based on three user studies is explained in Section 4.

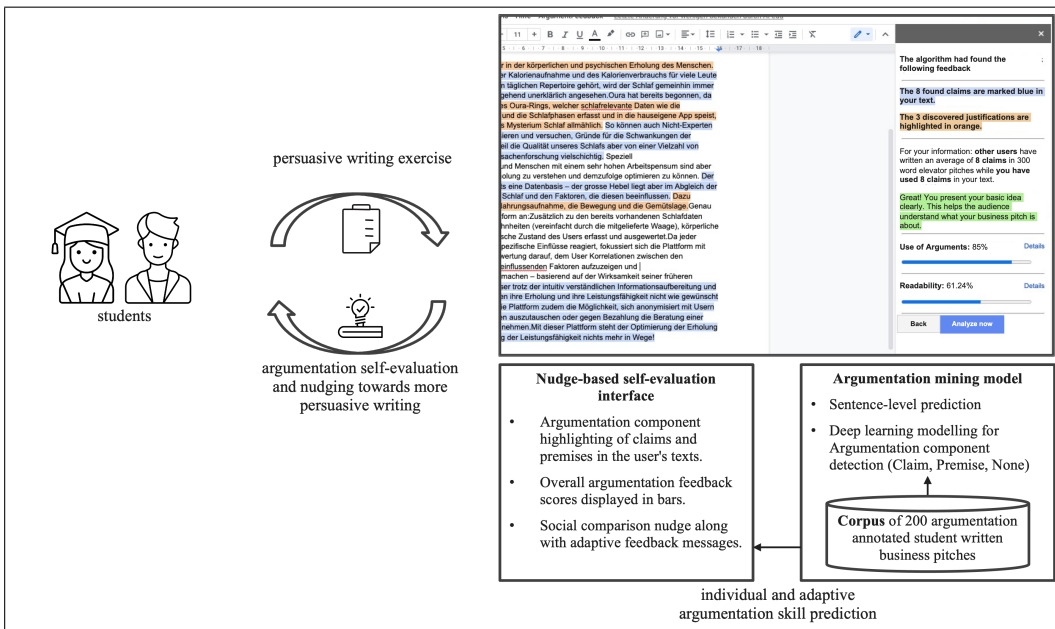


Fig. 2. Basic concept of ArgumentFeedback: students are conducting a persuasive writing exercise in a common writing editor and are nudged with an intelligent self-evaluation towards more argumentative writing.

3.1 Self-Evaluation Interface for Individual Argumentation Writing Support

3.1.1 Deriving Requirements for Nudging Students towards more Persuasive Writing. To build a theory-motivated learning tool, our objective was to apply a rigorous theory-driven approach. We followed [18] and [103] to conduct a systematic literature review with the aim of deriving a set of theory requirements.

As a result, we (A) specified the scope of the review, (B) conceived the topic, (C) conducted a literature search, and (D) examined the findings in terms of requirements. We concentrated our

research on studies that demonstrated the successful adoption of learning tools for argumentation skills in step A, defining the review scope. In addition, our purpose is to establish requirements on a conceptual level, with a focus on an espousal of position and a representative coverage [18]. We defined two main areas for deriving requirements in step B (conceptualizing the topic): educational technology and HCI. We first focused on these literature streams since developing a learning tool for argumentation skills is a complex endeavor that is studied by psychologists, pedagogues, and computer scientists using various methodologies. To find relevant publications (step C), we used Google Scholar to run a keyword search. We chose Google Scholar because it offers advanced full-text searching and a variety of filtering options for academic publications. We used the following search terms: „Argumentation Learning”, „Argumentation Model”, „Reasoning Skills”, „Debating Skills”, „Collaborative Learning”, „Skill Learning”, „Learning Theory“, „Learning Tool” and „Learning Systems”. Initially, we received over 1500 hits. Therefore, we defined criteria for inclusion and exclusion and reviewed the titles and abstracts of our search results in a first step. We only included studies that dealt with or contributed to a kind of learning tool in the field of argumentation learning, such as an established pedagogical theory. Several works dealing with reasoning in other fields of study were omitted. On this basis, we selected 75 papers for a more intensive analysis.

We summarized similar topics of these contributions as literature issues and formed three clusters from them, which served as design requirements for our argumentation writing support tool. The theory requirements were then instantiated and iteratively refined in the course of our evaluation studies (Study 1 and Study 2) with 37 users. In the course of the evaluation, a fourth design requirement was derived. Finally, four principles (displayed in Table 1) informed the design of the final version of our tool, ArgumentFeedback. Based on the guidance of [36] we derived requirements in the form of design principles that guided the design of our learning tool. Our design requirements are focused on providing design principles for educational designers to build intelligent writing support systems. According to [36], these requirements are formulated in a more abstract way for the general class of writing support systems for educational purposes. Hence, we can ensure that designers with other pedagogical backgrounds can also apply the principles or compare their system designs with our requirements. The design requirements were implemented as functionalities (F1- F8) in our nudging-based argumentation learning system.

Next, our goal was to derive design knowledge for designing a novel nudge for argumentation skill learning. Based on self-regulated learning theory [120], our aim was to provide students with self-evaluation based on their current argumentation skill level. In order to achieve our goal, we followed the proposed methods for designing digital nudges by [76]. The methods provide four practical steps on how to design a digital nudge for different scenarios, starting with 1) define the goal, 2) understand the users, 3) design the nudge, and 4) test the nudge.

1) Define nudging goal: The first step was to define the goal that was aimed to be achieved with digital nudging. Our objective was to provide students with a tool that helped them improve their argumentation skills based on self-evaluation independent of instruction, time, and location. To quantify the desired user behavior, we aimed to design a feature that enabled users to write texts with a higher level of argumentation.

2) Understand the users: Next, the decision behavior of the users was examined, and the psychological effects were investigated in the context of argumentation skill and digital nudge. Building on the derived theory-based design requirements on how to build an adaptive argumentation writing support system (see Table 1), our aim was to better understand the human heuristics and biases in our particular scenario. Hence, we performed a general literature review on digital nudging and digital nudges in educational settings specifically. We investigated the found literature under the criteria of a) the underlying psychological effects (heuristics) of the intended

	Design Requirement from Literature	Exemplary Guiding Literature
1)	Provide colored in-text argumentation highlighting to enable students to self-monitor their argumentation structures in their written text.	[2, 8, 16, 80, 110, 120]
2)	Provide argumentation theory explanation and an argumentation learning goal (e.g., "write your texts as persuasive as possible") before and during the writing exercise.	[6, 80, 96]
3)	Provide summarizing argumentation scores in a learning dashboard with transparent explanations to enable students to receive a general overview of their argumentation skill level.	[80, 110]
4)	Provide students with adaptive feedback messages to guide them with feedback and self-evaluation based on their individual argumentation skill level.	[38, 105, 108, 113]

Table 1. Theory-based design requirements on how to build an intelligent argumentation writing support system

user behavior, b) the empirical proof of the effect, and c) the insights on digital nudges to influence the user behavior. Following [13, 58], we found seven heuristics to be suitable for the context of argumentation skill learning, namely *framing*, *status quo bias*, *social norms*, *loss aversion*, *anchoring and adjustment*, *hyperbolic discounting*, *decoupling*, and *priming*. After a more detailed evaluation and a workshop with two senior researchers who had much experience in user-interface design, digital nudging, and educational research, the concepts of *framing*, *social norms* and *anchoring and adjustment* were found to be most suitable for nudging students towards adaptive self-monitoring and self-evaluation based on recent methods from AM.

Along these lines, framing can be used to show specific consequences, implications, or eventualities of choices. For example, the technique of reinforcement can be used to strengthen the user's awareness of the desired behavior [13]. The psychological effect of social norms can be applied to make comparisons with other people. In this regard, the nudging technique of social influence is used in literature [13]. Social nudging is based on the psychological effect of an individual's tendency to conform to the expectations of others [17]. Hence, user behavior can be influenced by comparing one's performance to others. Finally, by nudging users through *anchoring and adjustment*, additional information is provided to students, which is perceived as a reference. Thus, decision-making can be facilitated, and users can be encouraged to perform a certain behavior [13].

3) Design the nudge: Based on these three derived concepts from literature, we built nine low-fidelity prototypes with different design instantiations of the three nudge groups (three design instantiations for each psychological effect). For example, for social nudging, we designed three different nudge instantiations for argumentation skill monitoring and evaluation. One nudge instantiation provided the users with a relative social comparison of their numbers of arguments (e.g., "80% of other users have used at least 6 arguments"). Moreover, users were provided with an absolute comparison by indicating how many more arguments others have written (e.g., "Other users have used 2 arguments more on average") and, thirdly, by a nudge indicating how many arguments others have written (e.g., "Others have used 8 arguments on average").

4) Test the nudge: The different designs were then tested in qualitative evaluations with thirteen students from our university to receive feedback about the different designs (see Study 1 in 4). Based on these evaluations, we found that targeted information on the argumentative quality of their text facilitated the decisions of students to improve their written text argumentatively. Therefore, we implemented an informative framing nudge to provide students with self-monitoring (control group in Study 3). Second, the students stated that the social nudge was best suited to provide them with self-evaluation for argumentation skill learning. Most students mentioned that a social comparative feature based on the historical number of arguments others users had written for the same witting task in combination with an individual feedback message encouraged and motivated them the most to write more arguments. These two nudge designs were used in the evaluations later on (i.e., Figure 4).

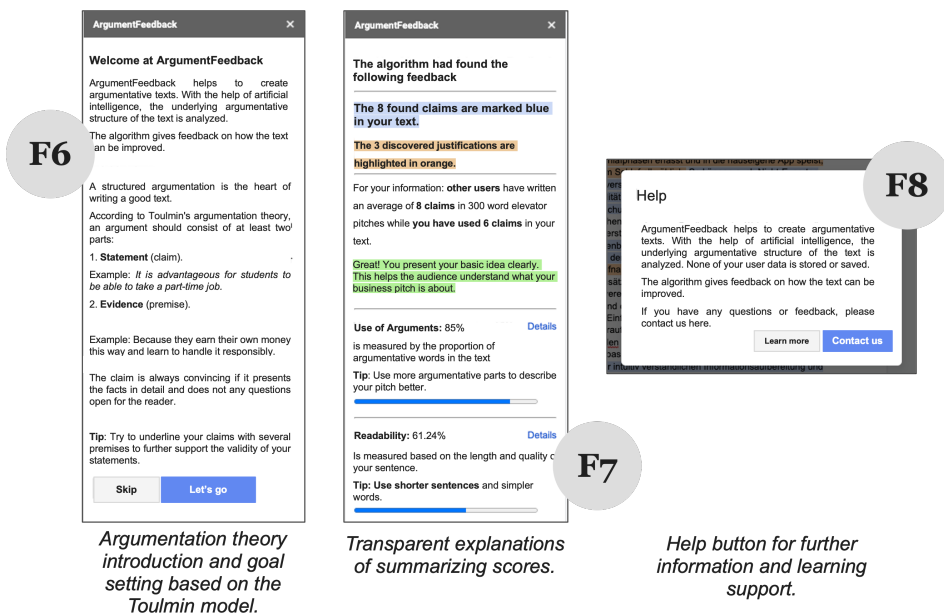


Fig. 3. Additional functionalities of ArgumentFeedback. Left: argumentation theory introduction and goals setting based on the Toulmin model. Middle: transparent and adaptive explanations of summarizing scores. Right: help button for further information and learning support.

3.1.2 User Interaction of ArgumentFeedback. Following the above-mentioned design requirements and the insights from nudging literature, ArgumentFeedback is built as a Google Doc's add-on that enables students to receive adaptive argumentation evaluation whenever and wherever they want. A screenshot of ArgumentFeedback and the main functionalities (e.g., F1 - F5) can be regarded in Figure 1. The interface of ArgumentFeedback is kept rather simple to not distract users from the educational writing task. Hence, the learning tool is displayed on the right side of the text input field by providing adaptive argumentation self-evaluation. When starting ArgumentFeedback, students first receive an introduction to the functionality of the tool. This ensures that first-time users are aware of the different functionalities of the writing support system. By referring to the Toulmin argumentation model, students are informed about the basics of argumentation theory (see F6 in

Figure 3) [96]. We integrated ArgumentFeedback in a popular writing editor. This comes with the advantage that users can not only write their texts in a familiar environment with several useful writing functions (such as spell check) but also then with ArgumentFeedback being thus more easy-to-access and easy-to-use on all kinds of advances through a responsive web-based design. In fact, to the best of our knowledge, with ArgumentFeedback, we present the first argumentation learning tool that is integrated into a common writing environment (such as Google Docs or Microsoft Office). Thus, we believe the availability as an add-on to a known cloud platform will foster its long-term usage and adaption compared to a stand-alone research tool. After introducing the user to the function of the tool and providing a short overview of argumentation theory, ArgumentFeedback displays a simple self-evaluation dashboard to the user (Figure 1: F3, F4, and F5). After the user writes their text, they can select a certain text part and click on "now analyze". Afterwards, they will receive an adaptive argumentation evaluation based on their argumentation skills. To prevent information overload, we incorporated the function of only analyzing selected text paragraphs. The argumentation monitoring and evaluation consist of three different parts: 1) the argumentation component highlighting of the claims and premises in the user's texts (F2), 2) the overall argumentation feedback scores displayed in bars (F5), and 3) the social comparison (nudge) along (F3) with the adaptive feedback message (F4). Through the argumentation highlighting, students can monitor their argumentation skill level. Claims are highlighted in blue and supporting premises in orange (see F2). Through the color-coding, they can easily assess if a claim is supported or not and sufficiently backed up with a premise. This function provides the user with clear action steps on how to improve the persuasiveness and formal quality of their texts. Moreover, best practices and explanations about argumentation and argumentation theory are provided by clicking on the "Help" button (see F8 in Figure 3).

The two summarizing scores (readability and persuasiveness) (F5) provide the student with a ranking of their text to provide general feedback (calculation explained in 3.2.3). By clicking on the scores or "details", the methodology for calculating the scores, concrete hints, action steps, and explanations on how the student can increase their score level will be shown (see F7 in Figure 3). These action steps provide the user with orientation and context to improve their writing quality [38, 80].

The social nudge (F3 and F4) counts the number of claims in the user's document and provides a comparison based on historical data of the underlying knowledge base or task. In our case, we retrieved a corpus of 200 student-written business pitches. On average, students wrote eight arguments (claims) in a 300-word pitch. If students had more claims than the average, they received a positive feedback message (highlighted in green), such as: "Very nice. You displayed your idea with a sufficient amount of arguments." (see F4 in Figure 1). However, if a student had less than eight claims, he or she would receive a feedback message highlighted in yellow saying: "Maybe you could write some more arguments in your texts to improve the persuasiveness of your idea". The intelligence of ArgumentFeedback is based on a trained AM model.

3.2 Argumentation Mining Algorithm for Adaptive Argumentation Self-Evaluation

To build an adaptive argumentation self-evaluation system based on recent computational methods, we collected and annotated our own corpus to train and tune a model that provides students with adaptive argumentation self-evaluation.

3.2.1 Building a Corpus of Persuasive Business Pitches. A major prerequisite for developing NLP methods that are able to identify argument components and argumentative relations in written texts is the availability of annotated corpora. Our aim was to provide students with argumentation self-evaluation when writing a persuasive business pitch. However, no suitable corpus exists in

literature that is based on a rigorous annotation scheme, consists of sufficient size to train a model with acceptable predictive accuracy, and fits our pedagogical scenario in the German language [50]. Business model pitches - also called entrepreneurial or business pitches [72] - are described as “a brief description of the value proposition of an idea or company” [19] with the objective to convince a group of stakeholders of the novelty of an idea. A business model pitch usually lasts between 30 seconds and two minutes [19]. The formulation of persuasive business model pitches is increasingly used in modern pedagogical scenarios, e.g., to train the entrepreneurship mindset or agile work (i.e., [65]). Students are asked to write a concise but persuasive summary of the “*what, why, and how*” of their (business) idea in order to convince a peer. This pedagogical scenario is domain independent, easy to implement in different settings (e.g., in MOOCs), and can be utilized to train skills such as logical argumentation. In fact, in their study about entrepreneurial business pitches, [30] found that “*the lack of rational arguments determines the failure of the entrepreneur’s efforts to be persuasive, regardless of the emotional appeals that are introduced into the pitch*”. Therefore, [30] calls for more emphasis on logical argumentation chains in business pitches.

However, linguistic research on business model pitches is a growing but still small field [24]. Therefore, it is not surprising that no pitch corpus exists that is annotated for argumentation discourse structures based on an appropriate argumentation scheme [50]. We propose a new annotation scheme to model argument components in persuasive business model pitches. We based our annotation scheme on the model of [95] and the studies of [15, 43, 83, 86, 111].

Hence, we gathered a corpus of 200 student-written business model pitches in German. The data was collected in a mandatory business model innovation lecture at a Western European university. In this lecture, around 200 students developed and presented a new business model. Students were asked to write a concise but persuasive pitch about the “*what, why, and how*” of their novel business idea in order to convince peer students. Afterwards, the students received peer feedback from three fellow students on the persuasiveness of their business model pitch. The business pitches were collected from 2019 to 2020 according to the ethical guidelines of our university and with approval from the students to utilize the writings for scientific purposes.

Our objective was to model the argumentation level of student-written business model pitches by capturing argument components. The majority of the pitches in our corpus followed the same structure. They described a novel business model and then provided convincing statements backed by examples, statistics, user-centered descriptions, quotes, or intuitions. For capturing the argument components, we followed established models in argumentation theory that provide detailed definitions of argument components (e.g., [86, 95]). These theories generally agree that a basic argument consists of multiple components and that it includes a *claim* that is supported or attacked by at least one *premise*. Also, in the student-written business model pitches, we found that a *claim* was the central component of an argument. A claim is a controversial statement (e.g., claiming a strength or novelty of a business model) that is either true or false and should not be accepted by the stakeholder without additional support or backing. In business model pitches, authors usually start or conclude with an overall idea and topic of the business model. Similar to the persuasive student essays corpus by [86], we modeled this statement as a *major claim*. Usually, the major claim is present in the introduction or conclusion of the pitch - or in both. In the introduction, it often represents a general claim of the novelty of the business idea, whereas, in the conclusion, the major claim often summarizes or repeats the argumentation according to the author’s business model idea. The major claim is then backed up by several other claims to manifest its validity. The *premise* supports the validity of the *claim* (e.g., by providing a statistic, analogy, user-centered example, or a value-based intuition). It is a reason given by the author to persuade the reader of their *claim*.

Two native German speakers annotated the business pitches independently from each other for the *major claim*, *claims*, and *premises* according to the annotation guidelines we specified.

Inspired by [86, 111], our guidelines consisted of 16 pages, including definitions and rules for what an argument is composed of, which annotation scheme was to be used, and how argument components were to be judged. After constructing the annotation guidelines, the results were discussed and validated by two independent senior researchers concerning the criteria of robustness, conciseness, extensibility, and comprehensibility. Several private training sessions and three team workshops were performed to resolve disagreements among the annotators and to reach a common understanding of the annotation guidelines. We used the *tagtog* annotation tool. After the first 50 pitches had been annotated by both annotators, we calculated the inter-annotator agreement (IAA) scores. As we obtained satisfying results, we proceeded with a single annotator who marked the remaining 150 documents. The exact annotation guidelines as well as the entire corpus is published and can be accessed in [109].

3.2.2 Inter-Annotator Agreement. To evaluate the reliability of the argument component annotations, we followed the approach of [83]. Since there were no predefined markables, the annotators not only had to identify the *type of argument component* but also its *boundaries*. In order to assess the latter, we use Krippendorff's α_U [46], which allows for assessing the reliability of an annotated corpus considering the differences in the markable boundaries. To evaluate the annotators' agreement in terms of the selected category of an argument component for a given sentence, we calculated the percentage agreement and two chance-corrected measures, multi π [32] and Krippendorff's α [45].

	%	Multi- π	Krip. α	Krip. α_U
Major claim	0.9948	0.9673	0.9673	0.5186
Claim	0.8729	0.7087	0.7088	0.5002
Premise	0.8768	0.7454	0.7455	0.5356

Table 2. Inter-annotator agreement of argument component annotations

Table 2 displays the resulting IAA scores. We obtained an IAA of 87.3% for the claims and 87.7% for the premises. The corresponding multi- π scores were 0.71 and 0.75. Regarding Krippendorff's α , a score of 0.71 and 0.75 was obtained, indicating a substantial agreement for both categories. With a score of 0.50 and 0.54, the unitized α of both the claim and premise annotations was somewhat smaller compared to the sentence-level agreement. Thus, the boundaries of argument components were less precisely identified in comparison to the classification into argument types. Yet the scores still suggested that there was a moderate level of agreement between the annotators. Finally, with an IAA of 99.5% and a score of 0.97 for both multi- π and Krippendorff's α , we obtained an almost perfect agreement for the major claims. Hence, we concluded that the annotation of the argument components in the student-written business model pitches was reliably possible.

The final corpus consisted of 200 student-written business pitches in German that were composed of 3,207 sentences with 61,964 tokens in total. Hence, on average, each document had 16 sentences and 305 tokens. A total of 262 major claims, 1,270 claims, and 1,481 premises were annotated. 1,069 textual spans were identified as not being an argument component ("None").

3.2.3 Modeling Argumentation Structures in Persuasive Business Pitches. After building and examining our corpus, we used the raw data to train an ML model. In order to give students individual arguments self-evaluation during the writing process, our goal was to integrate a classification algorithm into the back end of an argumentative writing support system. The task is defined as sentence-based, with each sentence being either a *claim*, a *premise*, or *non-argumentative*. In order to classify the argumentative elements of a given text, we trained and adjusted the hyperparameters of an LSTM model[39], as several other researcher (e.g., [16]) had achieved satisfactory results

with this approach. We tokenized the texts and created word embeddings from them (GloVe). The texts were tokenized and converted into word embeddings (GloVe) (one-hot encoded labels). Using an 80:20 stratified split, the data set was divided into training and test sets. Eight layers and 60 units per layer made up the LSTM design, which had a dropout rate of zero. According to the test results, the component classification had an accuracy of 54.12 percent, a precision of 55.90 percent, and a recall of 54.12 percent. We compared our approach to a model called a Bidirectional Encoder Representations from Transformers (BERT) model [21]. The best results were obtained with a learning rate of $1e-5$, a batch size of 16, and a training period of more than 25 epochs for the BERT model. The LSTM's accuracy of 54.12% compares poorly to other studies on AM component classification on student-written text from a technical standpoint. For instance, [110] recently achieved an accuracy of 65.4% for the classification of argumentation components in German student-written peer reviews. However, as described in Section 4, our qualitative research supported the notion that our model appeared to be sufficiently solid for students to receive a satisfying feedback to their argumentative pitches.

Furthermore, we calculated two summary scores to provide students with an overview of the quality of their argumentation based on previously extracted argumentative structures, including **Readability**: How readable is the text based on the Flesch Reading Ease score [33]?

Argumentativeness: How is the proportion of argumentative parts in the text compared to non-argumentative parts?

4 EXPERIMENTAL EVALUATION

In this section, we describe the experimental evaluation of our study. Our goal was to answer how can an individual argumentation self-evaluation tool based on ML to influence students' argumentation writing skills be designed. Hence, we conducted three evaluation studies according to [82, 99]. We 1) qualitatively evaluated (Study 1) the nudge design of our tool, 2) performed a proof-of-concept evaluation (Study 2) concerning the functionality and design of ArgumentFeedback, and 3) performed a proof-of-value field-experimental evaluation (Study 3) to investigate if and how individual argumentation self-evaluation in combination with social nudging and adaptive feedback will help students write more convincing texts. An overview of the experimental evaluation and the three studies can be found in Table 3. In total, we evaluated our tool with 83 students. All participants were bachelor's or master's students of Business Innovation and were recruited at a Western European German-speaking university. We ensured that no student participated in more than one study to control for intervention effects.

4.1 Study 1: Qualitative Design Evaluation

To ensure the validity of our design and incorporate the change requests from users about the functionality and the instantiations of the different nudges and pre-test effects for argumentation skill learning, we performed two studies (Study 1 and Study 2). The objectives of these studies were to verify if the theory-derived design requirements were of value to the students, to measure their completeness, and to identify student-based design requirements and requests.

4.1.1 Study Design. The aim of Study 1 was to evaluate the alpha version of ArgumentFeedback as a first prototype in Google Docs. Our objective was to ensure that the design of the argumentation self-monitoring nudge leads to the intended effects and positive perceptions of the users. Therefore, we performed a qualitative evaluation, where we asked students to use our application to write a short persuasive paragraph of 50 to 100 words about a business idea. Afterwards, the students were interviewed about their experience with the tool and their perception of particular design features, such as the argumentation highlighting, the theory explanation, and the summarizing scores (to

Study	Objective	Design	n	Results
Study 1	<ul style="list-style-type: none"> - Ensure all theory-derived requirements are correctly addressed and instantiated in our design - Ensure that the design of the two nudges (argumentation self-monitoring and argumentation self-evaluation) leads to the intended effects and positive perceptions of the users 	Qualitative evaluation based on user observation and post-questionnaire	13	<ul style="list-style-type: none"> - Fourteen change requests regarding the design and functionalities of ArgumentFeedback (e.g., concerning a "help" function, the size and color of the analyze button, and requests about the explanation and feedback texts of the two summarizing scores) - Social comparative features based on the historical number of arguments in combination with an individual feedback message lead to high user satisfaction
Study 2	<ul style="list-style-type: none"> - Proof-of-concept evaluation of final version of ArgumentFeedback in a real-world educational writing scenario in an artificial and controlled set-up - Ensure that technical artifacts provide high-quality argumentation self-evaluation 	Laboratory experiment	24	<ul style="list-style-type: none"> - Perceived usefulness and perceived ease of use show promising results for using an intelligent writing support system based on our design for supporting students in writing argumentative texts [100] - The positive measured self-efficacy of the students might indicate that ArgumentFeedback increases self-regulated learning (i.e., [6])
Study 3	<ul style="list-style-type: none"> - Empirically investigate the effects of intelligent argumentation self-evaluation in the field - Evaluate to what extent an intelligent writing support system in combination with self-evaluation nudging influences the argumentation writing skills of students in a large-scale learning setting 	Field experiment	46	<ul style="list-style-type: none"> - Intelligent argumentation self-evaluation including social comparison for students' argumentation skills on business pitches feedback helps them to write more convincing texts - Students using ArgumentFeedback including an intelligent self-evaluation nudge wrote texts with a better formal and perceived quality of argumentation to the ones receiving only an (intelligent) self-monitoring nudge.

Table 3. Overview of the experimental evaluation of ArgumentFeedback in three subsequent studies

control the theory-based design requirements in Table 1). Moreover, participants were explicitly asked about the perception and the use of different digital nudges, which were prototypically instantiated in different versions of ArgumentFeedback. For example, the participants were asked to indicate how they felt when they read the written social comparison information.

In total, thirteen students participated in Study 1 (eight were graduate students, seven were undergraduates). Seven were male, and six were female with an average age of 24.0 (SD= 2.16).

4.1.2 Results. Based on the qualitative answers, the observed usage, and the reported experience, several change requests for design instantiations were collected. For example, based on several user comments, the theory-explanations in ArgumentFeedback were shortened (e.g., "I would cut the 'for your information', I think that is redundant."), and an introduction to the usage of the tool was added to provide the users with a quick overview of the core functionalities (e.g., "An introduction to the learning tool would have helped me to get better familiar with the usage of the tool"). Concerning the nudges, eleven of the thirteen students (84 percent) mentioned that the social comparative feature based on the historical number of arguments that the other users had written for the same writing task combined with an individual feedback message would encourage and motivate them the most to write more arguments. One student mentioned, for example, "A social comparison would probably grab my ambition and I would try to include more arguments. Nevertheless, there is the 'danger' that you get too hung up on the number of arguments and the text lose importance", another mentioned "I think a social comparison is very good, because on the one hand a comparison is made with others and on the basis of that, a tip is then given to me. The wording needs to be very motivating in my opinion.", or "A social comparison would help me seeing my text from another perspective. This offers something only another person could do: giving inputs that you could never come up with on your own since writing a text is often an obstacle in the correction process."). Hence, we incorporated that into our design and scripted different motivational messages based on the argumentation evaluation. Moreover, several smaller design requests were captured, such as including a "help" function, the size and color of the *analyze* button, and requests about the explanation and feedback texts of the two summarizing scores. In total, we captured fourteen change requests based on Study 1 (five general requests, two about the introduction, two about the text explanations, three about the learning dashboard, and two about the digital nudges).

4.2 Study 2: Proof-of-Concept Laboratory Experiment

The aim of Study 2 was to empirically evaluate the application of ArgumentFeedback in a real-world educational writing scenario. Hence, after we incorporated all of the change requests from Study 1, we designed a laboratory experiment where students were asked to conduct a persuasive writing exercise and provided adaptive argumentation self-evaluation with social comparison and an adaptive feedback message (similar to TG in Study 3 - see Figure 4).

In total, 24 students participated in this study. The participants were, on average, 24.31 years old (SD= 1.99); 11 were male, and 10 were female. 20 were studying at the master's level, and four were at the bachelor's level. All students were business students. We controlled the experiment so that no student from Study 1 also took part in Study 2.

4.2.1 Study Design. The experiment consisted of two parts: 1) an argumentative writing task and 2) a post-survey.

1) Argumentative writing task: The experiment started with the students receiving a link to Google Docs, where they were asked to open our ArgumentFeedback add-on and become familiar with the tool. Next, they retrieved a business pitch of around 300 words. The students were asked to rewrite the pitch with the objective to increase its persuasiveness in order to convince a potential audience of investors. We asked all participants to use ArgumentFeedback to monitor and evaluate the argumentation level and the persuasiveness of their text. Providing the students with a pre-written text with incorporated argumentative flaws ensured that the students could focus on the argumentative part of the task and not on brainstorming their own novel business model. During

the task, students received intelligent argumentation evaluation on the text, including a social comparison and an adaptive feedback message.

2) Post-survey: In the post-survey, we measured the perceived ease of use, the intention to use, and the perceived usefulness of the participants following the technology acceptance model of [100, 101]. Example items for the three constructs were: "*The use of the argumentation tool enables me to write better argumentative texts*" or "*I would find the tool to be flexible to interact with.*" Moreover, we measured the self-efficacy of students for the task of argumentation skill learning based on three items following [6] to control for self-regulated learning. The items included, "*In comparison to other users, I will write a good argumentative text*", "*I am sure that I could write a very good argumentative text*", and "*I think I now know quite a bit about argumentative writing.*" Both constructs were measured with a 1- to 7-point Likert scale (1: totally disagree to 7: totally agree, with 4 being a neutral statement). Also, we asked two questions to control for the digital nudge, such as "*I noticed on the analysis page the information on how many arguments I used compared to other users.*" and "*The information about how many arguments I used compared to other users motivated me to write more arguments.*". Finally, we captured the demographics and asked three qualitative questions: "*What did you particularly like about the use of the argumentation tool?*", "*What else could be improved?*" and "*Do you have any other ideas?*" In total, we asked 15 questions in the post-survey.

4.2.2 Results. The *perceived usefulness* of students using ArgumentFeedback had a mean value of 4.65 (SD= 1.42). The *perceived ease of use* of our tool was rated with a mean value of 5.31 (SD= 1.81). All of the results were positive when compared to the midpoint scale of 4, indicating a positive technology acceptance of our writing support tool in the particular pedagogical scenario [100, 101]. Hence, we concluded that our design instantiations based on the derived theory requirements were satisfying for the user. Moreover, participants of Study 2 judged the self-efficacy with a mean value of 4.62 (SD= 1.23). Moreover, the two control questions for the evaluation of the social comparison nudge were judged with a mean of 5.94 (SD= 1.48), representing a positive result. Finally, we analyzed the qualitative answers and clustered similar responses into categories. In summary, the adaptive self-evaluation based on in-text highlighting and the summarized scores were positively mentioned several times. At the same time, students also mentioned that, sometimes, wrong argumentative components seemed to be highlighted by ArgumentFeedback. Some also asked for more details on the functionality of the tool, e.g., information about the accuracy of the feedback algorithm. We incorporated this qualitative feedback, e.g., by improving some transparent explanations about the functionality of our tool.

4.3 Study 3: Proof-of Value Field Experiment

After evaluating the qualitative design of our approach (Study 1) and conducting a proof-of-concept study with ArgumentFeedback (Study 2), our objective was to empirically evaluate and investigate the effects of intelligent self-evaluation in the field. Hence, we designed a pedagogical scenario in which participants were asked to write a persuasive business pitch of 300 words. We manipulated the way students received self-monitoring and adaptive self-evaluation based on the persuasiveness of their pitches. Study 3 was conducted as a field experiment in a large-scale lecture about information management at a Western European University. Participants were randomly assigned to one of two groups, a treatment group, and a control group. Students in the control group used a version of ArgumentFeedback which provided them with general argumentation self-monitoring based on syntactic rules (see Figure 4). The syntactical rules of the CG were the same two summarizing scores (readability and persuasiveness) from the treatment group.

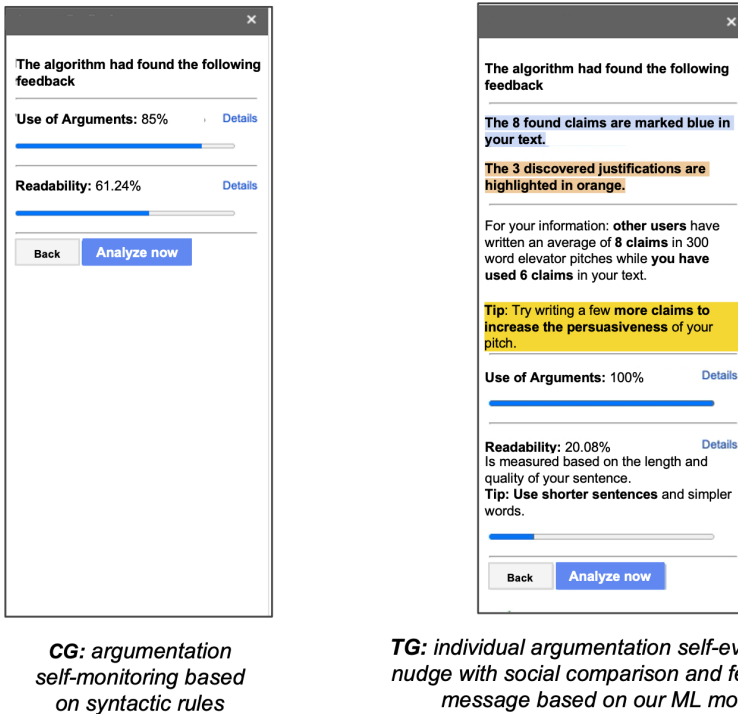


Fig. 4. Overview of both experimental groups: participants of the control group (CG) received argumentation self-monitoring based on syntactical rules while conducting an argumentative writing exercise. Participants in the treatment group (TG) were nudged with an intelligent self-evaluation interface based on social comparison and an adaptive feedback message based on our trained ML model.

4.3.1 Study design. The field experiment was conducted in a lecture where students created and presented digital business models. In one project, the students were required to provide a new business concept based on a value proposition canvas [67]. A 300-word persuasive business pitch on the value proposition of their business idea was also required as part of the assignment. The assignment was to create a pitch that would persuade potential investors. Although participation in the assignment was not required in order to complete the class, doing well on it would increase students' final grades (by around two percent). The persuasiveness of the business pitch, for example, had no bearing on the assignment's grading and, as a result, no bearing on the final grade. We asked the students to write their business pitches on the Google Docs platform. Here, we provided them with different levels of feedback on the argumentation level of their business pitch. Nevertheless, we ensured that the students also had the opportunity to conduct the exercise in any other writing tool of their choice due to data privacy reasons and the ethical standards of our university. Students who chose to not use Google Docs in combination with ArgumentFeedback were excluded from the sample.

In total, 46 students from the course successfully completed our experiment (including pre-test, argumentative writing task and post-test). After randomization, we counted 21 results in the control group and 25 in the treatment group. Participants of the control group had an average age of 24.17 (SD= 3.56, 9 males, 12 females) and participants in the treatment group had an average age of 23.04

(SD= 2.32, 13 males, 12 females). The persuasive writing task took an average of 30 to 45 minutes. All participants of Study 3 had neither participated in Study 1 nor in Study 2. The experiment consisted of three main parts: 1) pre-test, 2) persuasive writing exercise and 3) post-test. The pre-and post-phases were consistent for all participants. In the writing phase, the treatment group used ArgumentFeedback with adaptive self-evaluation based on social comparison and an individual feedback message to conduct the persuasive writing exercise, whereas participants of the control group conducted the same exercise using ArgumentFeedback with argumentation self-evaluation based on syntactic rules.

1) Pre-test: The experiment started with a pre-survey of eight questions. We tested two different constructs to assess whether the randomization was successful. First, we asked four questions to test the personal innovativeness in the domain of information technology of the participants following [3]. The exemplary items were, *"I like to experiment with new information technologies,"* or *"If I heard about a new information technology, I would look for ways to experiment with it,"* *"In general, I am hesitant to try out new information technologies,"* and *"Among my peers, I am usually the first to try out new information technologies."* Second, we tested the construct of feedback-seeking of the individuals following [4]. Example items were: *"It is important for me to receive feedback on my performance"* or *"I find feedback on my performance useful."* Both constructs were measured with a 1- to 7-point Likert scale (1: totally disagree to 7: totally agree, with 4 being a neutral statement).

2) Persuasive writing exercise: In the persuasive writing exercise of the experiments, we asked the participants to write a persuasive business pitch of 300 words. More information about the structure and the definition of a persuasive business pitch can be found in Section 3.2.1. Students in the treatment group used ArgumentFeedback with adaptive self-evaluation based on a social comparison to conduct the same exercise. Participants of the control group conducted the same exercise using ArgumentFeedback with self-evaluation based on syntactic rules.

3) Post-test: Similar to Study 2, in the post-survey, we measured the perceived ease of use, the intention to use, and the perceived usefulness of the participants following the technology acceptance model of [100, 101] to ensure the validity of our design instantiation (exemplary item for intention to use: *"Imagine the tool would be available in your next course, would you use it?"*). All constructs were measured with a 1- to 7-point Likert scale (1: totally disagree to 7: totally agree, with 4 being a neutral statement). Finally, we captured the demographics and asked, again, three qualitative questions: *"What did you particularly like about the use of the argumentation tool?"*, *"What else could be improved?"*, and *"Do you have any other ideas?"* In total, we asked 13 questions in the post-survey of Study 3.

4.3.2 Measurement of Argumentation Quality. Following recent studies on adaptive argumentation learning [108, 110], our major goal was to assess the quality of both groups' written texts. As a result, we looked at two primary variables: 1) the formal quality of the argumentation, and 2) the perceived quality of the argumentation [108, 110].

1) Formal quality of the argumentation: The written business pitches were analyzed for their formal quality of argumentation. As [110], we applied the annotation scheme for argumentative knowledge construction described by [116]. This annotation scheme has been applied in various studies and has proven high objectivity, reliability, and validity (e.g., [88, 108, 110]). To measure the formal quality of the argumentation, the annotator had to distinguish between a) *unsupported claims*, b) *supported claims*, c) *limited claims*, and d) *supported and limited claims*. A more precise description of the scheme can be found in [116]. Therefore, we trained two annotators based on the 16-page annotation guidelines from our corpus to assess the argumentation components of the persuasive pitches. More information about the guidelines, the annotation study, and the inter-annotator agreements can be found in Section 3.2.2. The formal quality of the argumentation

of the individual user was then defined by the number of arguments written by a user during the writing phase. Following [88], only *supported*, *limited*, and *supported and limited claims* were counted as argumentation.

2) Perceived quality of the argumentation: As done in prior studies [108, 110], we also captured the perceived quality of the argumentation. The two annotators subjectively assessed how persuasive the given argumentation was on a Likert scale from 1 to 7 points (1: not very persuasive, 7: very persuasive). We took the mean of both annotators as a final variable for the formal and the perceived quality of the argumentation of the texts.

4.3.3 Results. Our objective was to investigate the difference between the argumentation self-monitoring nudge based on syntactical rules (CG) and the ML-based argumentation self-evaluation nudge (TG) on the students' argumentative writing skills and answer behavior. Moreover, we aimed to control for the design of our intelligent argumentation writing support tool and its impact on the students' perceived user-experience. Hence, we compared both the formal and perceived quality of the argumentation between the written texts of the control groups and the treatment group. In particular, we used a double-sided t-test to evaluate whether the means of the constructs are significantly different. We compared the results of the perception measures, such as intention to use, perceived usefulness, and ease of use to the midpoints scale to validate our design intervention of ArgumentFeedback as done in [53]. Finally, we evaluated the differences in the means of the two constructs included in the pre-test and the manipulation checks to ensure that the randomization resulted in randomized groups and to account for the effects of interfering variables with our limited sample size. We received p values larger than 0.05 between the treatment and the control group for both dimensions, including personal innovativeness and feedback-seeking of individuals (personal innovativeness $p = 0.4697$, feedback-seeking of individuals $p = 0.5047$). This demonstrated that no significant difference in the mean values for these two constructs existed between the groups.

Argumentation writing quality

Group	formal argumentation quality (number of formally correct arguments)	perceived quality of argumentation (1-7 Likert scale, 1:low, 7: high)
Mean TG	4.875	4.5
Mean CG	3.6428	3.738
SD TG	1.2445	0.8597
SD CG	1.2056	0.9568
p-value	0.0016**	0.0079**

Table 4. Results on the argumentation writing quality between both groups (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$)

The average number of arguments in the texts from the participants using ArgumentFeedback with intelligent self-evaluation (TG) for the persuasive writing exercise was 4.875 (SD= 1.4). The participants in the control group wrote their texts with a mean of 3.64 arguments (SD= 1.20) (see Table 4 and Figure 5). A double-sided t-test confirmed that the treatment group wrote texts with a statistically significantly higher quality of formal argumentation: t-value= -3.369 and $p = 0.0016$. These results indicate that intelligent self-evaluation based on social comparison and an adaptive feedback message on the students' argumentation in their business pitches helped them write more formally convincing texts.

For the perceived quality of the argumentation (see Table 4), we found that, on a Likert scale from 1 to 7 points (1: not very persuasive, 7: very persuasive, 4: neutral), texts from the TG achieved

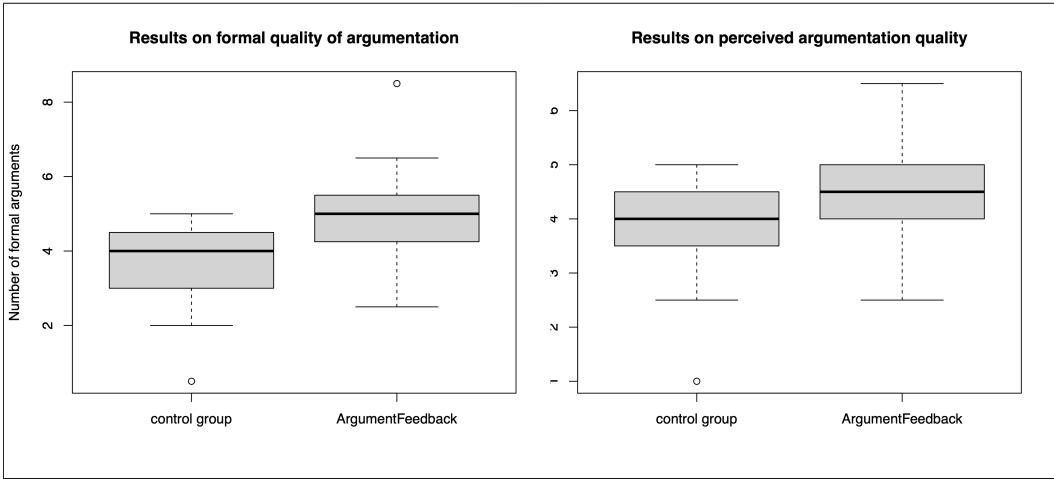


Fig. 5. Results on formal (left) and perceived (right) quality of argumentation between both tools.

an average value of 4.5 (SD= 0.85). For the CG, we measured a perceived quality of 3.73 (SD= 0.95). The double-sided t-test confirmed that the text of both groups significantly differ in the perceived argumentation level: $t\text{-value} = 2.7934$, $p = 0.0079$.

User perceptions and qualitative user feedback

Finally, we evaluated the students' perception in order to check for the validity of our design instantiation of ArgumentFeedback. Thus, we calculated the mean across all of the groups for the perceived ease of use, perceived usefulness, and intention to use and compared them to the midpoints. All results were greater than the neutral value of 4, indicating a positive value for the design and the pedagogical scenario. A high perceived usefulness (mean= 4.2647, SD= 1.3227), intention to use (mean= 4.4975, SD= 1.2966), and ease of use (mean= 5.0096, SD= 1.1546) are especially important for learning tools to ensure students are experiencing the usage of the tool as a benefit and that they find it easy to interact with. This will foster the motivation, engagement, and adoption of the learning application. We also asked open questions in our survey to receive the participants' opinions about the tool they used. The general attitude for ArgumentFeedback was quite positive. Participants positively mentioned the intelligent self-evaluation, the social comparison, and the integration of Google Docs several times (e.g., "The argumentation tool was very clear and accurate. The green highlighted tip once again motivated you to revise and improve your pitch. It was pleasant to work with because it also didn't take up too much space on the screen. The hints were helpful."). However, participants also criticized the feedback accuracy, the granularity of the feedback, and suggested to provide concrete argument suggestions to provide more feedback on how to improve the argumentativeness (e.g., "Recognizing justifications could be improved. Better tips to improve readability."). We translated the responses from German and categorized the most representative responses in Table 5.

5 DISCUSSION OF FINDINGS AND THEORY IMPLICATIONS

In this study, we aimed to investigate, design and evaluated a novel student-centered learning system called ArgumentFeedback to enable students to self-evaluate their argumentation writing skills independently. We evaluated our learning system in one qualitative study (Study 1), one laboratory experiment (Study 2), and one field experiment (Study 3). We observed that participants

Cluster	Exemplary User Response
Social comparison	<i>"The argumentative comparison with peers serves as an incentive to write at least as many or more arguments."</i>
Summarizing scores	<i>"Readability as a criterion is often neglected for argumentative texts. The feedback on this was very helpful for me to improve my pitch."</i>
Argumentative feedback	<i>"I was encouraged to write more arguments without stretching the text too much, as the ratio of the length of the text to the arguments is taken into account. Likewise, one is motivated to shorten the sentences. It is also good that tips are given for the categories in order to improve the evaluation." "It was not clear to me what was required in the readability. On the one hand, you should write short sentences, but on the other hand, you should include many arguments in them, this seemed difficult to me in the implementation." "The evaluation of the tool is more focused on the quantified design of the text, e.g. how many claims do you have in your text, but not the qualitative design of the argument, this should be improved again, as quality should be put before quantity."</i>
Dynamics of the interaction	<i>"In general, the tool should be more dynamic if possible. Going through an analysis every time is a bit cumbersome (tagging, etc.). Automatic updates of the feedback and metrics would improve the user experience a lot."</i>
Feedback accuracy	<i>"I believe I have always substantiated my claims, however the tool did not identify them. This has surprised me, since I have used extra words like in the given example."</i>

Table 5. Representative examples of qualitative user responses

who received an intelligent self-evaluation nudge including a social comparison and an adaptive feedback message when writing a persuasive business pitch wrote more formally argumentative texts compared to students who received only general self-monitoring. Furthermore, the perceived persuasiveness of the texts was significantly higher than of the texts from students using the alternative tool. The student perceptions (usefulness, intention to use and ease of use) and the qualitative data acquired in our experiments indicate a positive effect of an self-evaluation nudge including a social comparison and an adaptive feedback message based on argumentation mining. Therefore, this study expands prior research around formative feedback and digital nudging in computer-supported learning and HCI that used mainly exploratory or non-automated approaches (e.g., [2, 110, 113]).

Our work comes with important theoretical implications. The results provides further evidence for the findings of recent studies on argumentation skill learning in HCI (e.g., [2, 57, 108, 110]) in a) a new pedagogical domain and scenario (business model pitches) and b) based on digital nudging theory. Hence, for practitioners and designers we provide insights (in the form of design requirements and empirical results) on how to design intelligent writing support systems based on recent advantages in NLP and ML. The self-evaluation nudge, along with social comparison and feedback messages on students' argumentation skills, seemed to help students to write more persuasive texts compared to only monitoring their argumentative writing based on syntactical rules. Educational designers can build on our findings to embed our design features and requirements in writing support tools in other domains or languages to support skill learning at scale. Especially for the growing field of MOOCs, our study might shed additional light on the mechanisms, the

design and effects of nudging mechanisms which can be applied in online writing tasks to improve learning outcomes.

Moreover, our results contribute to nudging theory [92]. The findings signify that the social comparison contributes to, on the one hand, providing learning feedback and, on the other hand, necessary goal setting [17]. This is also reflected in the numerous qualitative comments we received about the peer comparison by user. Students seem to be nudged in their argumentative writing behavior, as social comparisons with other learners lead to better learning outcomes, which is even more relevant when being confronted with unusual tasks such as argumentation tasks. Moreover, our study contributes to literature about social conformity [17]. The results are in line with the results on social nudging for learning, e.g., with [63], who found that social comparison leads to higher exam performance and an increase in study time [63]. The correct level of self-evaluation on a student's skills, such as argumentation skills, seems to lead to high self-efficacy and thus higher learning outcomes. Additionally, our findings are in line with self-regulated learning theory [120]. The results seem to confirm the underlying mechanism rooted in social cognitive theory [5] that intelligent self-evaluation with social comparison and adaptive monitoring leads to positive behavior changes in learning processes [6]. For argumentation skill learning, this result is especially novel. Past research on adaptive argumentation learning tools based on computational methods, such as [2, 110, 113], have mostly focused on adaptive argumentation monitoring, e.g. based on in-text highlighting [2, 16, 55] or an argumentation discourse monitoring [110]. Social comparisons for argumentative writing were mostly neglected or even showed that users did not prefer to compare themselves with peers, e.g., [112]. Moreover, traditional argumentation learning theory, such as the representational guidance theory by [89], solely focuses on argumentation monitoring, e.g., supporting argumentation learning by providing representations of argumentation structures with the objective to stimulate and improve individual reasoning, collaboration, and learning [69, 90]. Our research provides novel insights on how to support students' argumentative writing based on self-evaluation and digital nudging more efficiently.

Furthermore, our work comes with several practical implications, especially in HCI research and computer-supported collaborative learning. We provide insights on a holistic design of a student-centered writing support tool based on recent advantages in computational linguistics and digital nudging theory for self-regulated argumentation skill learning and metacognition skills. Moreover, we demonstrate the effectiveness of social nudging for argumentation learning through evaluating our tool ArgumentFeedback in three empirical studies with a total of 83 students. The results provide insight into the potential of designing student-centered learning tools based on NLP and ML combined with digital nudging to foster self-regulated argumentative writing in a student's learning journey. Our research shows a case example of supporting metacognition skills in a scalable and individual way in possible large-scale scenarios. We contribute to computer-supported learning by showing that interactions with a writing feedback tool in a common writing environment have the potential to impact students' learning processes, resulting in increased levels of skill development. The qualitative results indicate that this new technology might be able to offer intelligent feedback for self-regulated learning in a more natural and adaptive way. Especially in distance-learning scenarios such as MOOCs and in common mass lectures at public universities, automated writing feedback in common writing environments such as Google Docs could be a beneficial addition to current learning environments at scale. Thus, we provide design knowledge for other researchers and educators to design and compare similar tools for the support of metacognition skills of students based on computational methods and nudging theory.

6 LIMITATIONS AND FUTURE WORK

Nevertheless, our work faces several limitations. First of all, the accuracy of our AM algorithm leaves space for improvement. Our model only predicted argumentation components with an accuracy of 54.12%. Nevertheless, we already found significant effects of argumentation in-text highlighting in combination with a social nudge on students' argumentation writing qualities. This could indicate that even "low performing" prediction models could already contribute to students' learning outcomes. We hypothesize that receiving adaptive feedback, even if erroneous, triggers students to more reflect on their argumentative writing compared to non-adaptive argumentation feedback. Future literature could dig deeper into this phenomenon by investigating this hypothesis. Nevertheless, we believe with higher accuracy the impact of our tool ArgumentFeedback on students' argumentative writing abilities should be even greater. Moreover, in our study, we only provided insights on the short-term effects of our design interventions. In Study 3, we showed that self-evaluation has a short-term influence on a student's argumentation skills. For future work, we suggest measuring the long-term learning effects on students' skills based on different nudging interventions, e.g., adaptive nudging to prevent habituation or wear-out effects.

Next, we compared the design of our nudge against a non-nudging based version in combination with syntactical feedback. For future research, it would be interesting to compare our nudging-based argumentation learning tool against established writing support systems such as Grammarly and investigate the similarities and differences in the usage and interaction with these tools in students writing processes. Additionally, we want to highlight the ethical limitations of our study. Regarding the implementation of our intelligent writing support system, we do not want to replace human educators by any means, as we believe that skilled teachers will always be able to provide better individual writing support than ArgumentFeedback. However, we hope that through our system, human educators can focus more on detailed questions and devote more time to difficult cases. Also, we see several data privacy concerns about integrating our tool in a common writing editor, since personal data, e.g., about the argumentation skill level of students, might be exposed to large stock-based tech companies. Hence, we call for future discussions on how to overcome the trade-off of making novel user-centered learning applications widely accessible and easy to use, e.g., by integrating them in known cloud environments without exposing learner data to third parties.

Furthermore, writing support tools in general can be misused by students, e.g., by cheating the algorithm or skating the exercises. Although, we did not observe such behaviour in our quantitative and qualitative data, a rigorous observation study in an educational scenario (e.g., with eye-tracking) of potential misuse and harm of such systems is needed in future research.

All in all, we aim to offer insights on the learner-centered design to further improve educational applications based on advances in computational methods. With further progress in NLP and ML, we hope our work will attract researchers to design more intelligent forms of HCI systems for other learning scenarios or metacognition skills and thus contribute to the OECD Learning Framework 2030 towards a metacognition-skill-based education.

7 CONCLUSION

In our research, we (1) constructed a novel, theory-driven argumentation annotation scheme for student-written business model pitches, (2) compiled a corpus of 200 student-written texts, (3) trained a LSTM to predict the individual argumentation skill level, and (4) embedded the model in our student-centered and theory-informed design based on nudging theory. We evaluated our final tool, ArgumentFeedback, in three studies and observed that participants who received an intelligent self-evaluation nudge including social comparison and an adaptive feedback message when writing a persuasive business pitch wrote more formally argumentative texts compared to

students who received only general self-monitoring. The measured intention to use, the perceived ease of use, and the perceived usefulness provide promising results for the usage of social nudging in combination with adaptive feedback for effective self-regulated argumentative skill learning. The results suggest that a student-centered learning tool based on digital nudging combined with ML and NLP feedback helps students write more persuasive texts even with a low-performing model.

REFERENCES

- [1] Alessandro Acquisti, Saranga Komanduri, Civis Analytics, Pedro Giovanni Leon, Banco De Mexico, Florian Schaub, Yang Wang, Shomir Wilson, ; L F Cranor, N Sadeh, ; Y Wang, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Norman Sadeh, and Manya Sleeper. 2017. 44 Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online ACM Reference Format. *Comput. Surveys* 50, 3 (2017). <https://doi.org/10.1145/3054926>
- [2] Tazin Afrin, Omid Kashefi, Christopher Olshefski, Diane Litman, Rebecca Hwa, and Amanda Godley. 2021. Effective Interfaces for Student-Driven Revision Sessions for Argumentative Writing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. <https://doi.org/10.1145/3411764.3445683>
- [3] Ritu Agarwal and Elena Karahanna. 2000. Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly* 24, 4 (12 2000), 665. <https://doi.org/10.2307/3250951>
- [4] S. J. Ashford. 1986. Feedback-Seeking in Individual Adaptation : A Resource Perspective. *Academy of Management Journal* 29, 3 (9 1986), 465–487. <https://doi.org/10.2307/256219>
- [5] Albert Bandura. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84, 2 (3 1977), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- [6] Albert Bandura. 1991. Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes* 50, 2 (12 1991), 248–287. [https://doi.org/10.1016/0749-5978\(91\)90022-L](https://doi.org/10.1016/0749-5978(91)90022-L)
- [7] Torben Barev, Melanie Schwede, and Andreas Janson. 2021. The Dark Side of Privacy Nudging – An Experimental Study in the Context of a Digital Work Environment. *HICSS 2021 Proceedings* (2021). <https://doi.org/10.24251/HICSS.2021.500>
- [8] Robert A. Bjork, John Dunlosky, and Nate Kornell. 2013. Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology* 64, November 2012 (2013), 417–444. <https://doi.org/10.1146/annurev-psych-113011-143823>
- [9] Christopher G. Brinton, Ruediger Rill, Sangtae Ha, Mung Chiang, Robert Smith, and William Ju. 2015. Individualization for education at Scale: MIIC design and preliminary evaluation. *IEEE Transactions on Learning Technologies* 8, 1 (3 2015), 136–148. <https://doi.org/10.1109/TLT.2014.2370635>
- [10] Ruth M. J. Byrne. 1989. Everyday Reasoning with Conditional Sequences. *The Quarterly Journal of Experimental Psychology Section A* 41, 1 (2 1989), 141–166. <https://doi.org/10.1080/14640748908402357>
- [11] Elena Cabrio and Serena Villata. 2012. Natural Language Arguments: A Combined Approach.. In *20th European Conference on Artificial Intelligence (ECAI 2012)*. Montpellier, France.
- [12] Cansu Canca. 2020. Operationalizing Ai Ethics Principles *. *Commun. ACM* 63, 12 (2020), 3–6. <https://doi.org/10.1145/3430368>
- [13] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A review of technology-mediated nudging in human-computer interaction. In *Conference on Human Factors in Computing Systems - Proceedings*. 1–15. <https://doi.org/10.1145/3290605.3300733>
- [14] Ana Caraban, Loukas Konstantinou, and Evangelos Karapanos. 2020. The nudge deck: A design support tool for technology-mediated nudging. In *DIS 2020 - Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 395–406. <https://doi.org/10.1145/3357236.3395485>
- [15] Winston Carlile, Nishant Gurrapadi, Zixuan Ke, and Vincent Ng. 2018. Give me more feedback: Annotating argument persuasiveness and related attributes in student essays. *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)* 1 (2018), 621–631. <https://doi.org/10.18653/v1/p18-1058>
- [16] Artem Chernodub, Oleksiy Oliynyk, Philipp Heidenreich, Alexander Bondarenko, Matthias Hagen, Chris Biemann, and Alexander Panchenko. 2019. TARGER: Neural Argument Mining at Your Fingertips. (2019), 195–200. <https://doi.org/10.18653/v1/p19-3031>
- [17] Robert B. Cialdini and Noah J. Goldstein. 2004. Social influence: Compliance and conformity. *Annual Review of Psychology* 55 (2004), 591–621. <https://doi.org/10.1146/annurev.psych.55.090902.142015>
- [18] Harris M. Cooper. 1988. Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society* 1, 1 (1988), 104–126. <https://doi.org/10.1007/BF03177550>
- [19] Peter Daly and Dennis Davy. 2016. Structural, linguistic and rhetorical features of the entrepreneurial pitch: Lessons from Dragons' Den. *Journal of Management Development* 35, 1 (2 2016), 120–132. <https://doi.org/10.1108/JMD-05-2014-0049>

- [20] R De Groot, R Drachman, R Hever, B Schwartz, U Hoppe, A Harrer, M De Laat, R Wegerif, B M McLaren, and B Baurens. 2007. *Computer Supported Moderation of E-Discussions: the ARGUNAUT Approach*. Technical Report. <http://www.argunaut.org>
- [21] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (2018). <http://arxiv.org/abs/1810.04805>
- [22] Nicholas Diana. 2018. Leveraging educational technology to improve the quality of civil discourse. In *International Conference on Artificial Intelligence in Education*, Vol. 10948 LNAI. Springer Verlag, 517–520. https://doi.org/10.1007/978-3-319-93846-2_97
- [23] Rosalind Driver, Paul Newton, and Jonathan Osborne. 2000. Establishing the norms of scientific argumentation in classrooms. *Science Education* 84, 3 (2000), 287–312. [https://doi.org/10.1002/\(SICI\)1098-237X\(200005\)84:3<287::AID-SCE1>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A)
- [24] Ana Maria Ducasse. 2020. Evidence-based persuasion: A cross-cultural analysis of entrepreneurial pitch in English and Spanish. *Journal of International Entrepreneurship* 18, 4 (12 2020), 492–510. <https://doi.org/10.1007/s10843-020-00278-0>
- [25] Richard A. Duschl and Jonathan Osborne. 2002. Supporting and Promoting Argumentation Discourse in Science Education. *Studies in Science Education* 38, 1 (1 2002), 39–72. <https://doi.org/10.1080/03057260208560187>
- [26] Frans H van Eemeren, Rob Grootendorst, Ralph H Johnson, Christian Plantin, Charles A Willard, Rob Grootendorst, Ralph H Johnson, Christian Plantin, and Charles A Willard. 1996. *Fundamentals of Argumentation Theory*. Routledge. <https://doi.org/10.4324/9780203811306>
- [27] Sean B. Eom, H. Joseph Wen, and Nicholas Ashill. 2006. The Determinants of Students' Perceived Learning Outcomes and Satisfaction in University Online Education: An Empirical Investigation*. *Decision Sciences Journal of Innovative Education* 4, 2 (7 2006), 215–235. <https://doi.org/10.1111/j.1540-4609.2006.00114.x>
- [28] Lauren Eskreis-Winkler, Katherine L Milkman, Dena M Gromet, and Angela L Duckworth. 2019. A large-scale field experiment shows giving advice improves academic outcomes for the advisor. *Proceedings of the National Academy of Sciences of the United States of America* 116, 30 (2019), 14808–14810. <https://doi.org/10.1073/pnas.1908779116>
- [29] Charles Fadel, Maya Bialik, and Bernie Trilling. 2015. *Four-dimensional education : the competencies learners need to succeed*. 177 pages.
- [30] José Santiago Fernández-Vázquez and Roberto Carlos Álvarez-Delgado. 2019. The interaction between rational arguments and emotional appeals in the entrepreneurial pitch. *International Journal of Entrepreneurial Behaviour and Research* 26, 3 (10 2019), 503–520. <https://doi.org/10.1108/IJEBR-06-2019-0334>
- [31] Frank Fischer, Ingo Kollar, Karsten Stegmann, and Christof Wecker. 2013. Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational Psychologist* 48, 1 (1 2013), 56–66. <https://doi.org/10.1080/00461520.2012.748005>
- [32] Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76, 5 (11 1971), 378–382. <https://doi.org/10.1037/h0031619>
- [33] R Flesch. 1943. Marks of readable style; a study in adult education. *Teachers College Contributions to Education* 897 (1943).
- [34] Eirini Florou, Stasinos Konstantopoulos, Antonis Kukurikos, and Pythagoras Karampiperis. 2013. Argument extraction for supporting public policy formulation. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. ACL, 49–54.
- [35] María Paz García-Villalba and Patrick Saint-Dizier. 2012. A Framework to Extract Arguments in Opinion Texts. *International Journal of Cognitive Informatics and Natural Intelligence* 6, 3 (7 2012), 62–87. <https://doi.org/10.4018/jcini.2012070104>
- [36] Shirley Gregor, Leona Chandra Kruse, and Stefan Seidel. 2020. Research perspectives: The anatomy of a design principle. *Journal of the Association for Information Systems* 21, 6 (2020), 1622–1652. <https://doi.org/10.17705/1jais.00649>
- [37] Katrin Hartwig and Christian Reuter. 2021. Nudging users towards better security decisions in password creation using whitebox-based multidimensional visualisations. *Behaviour and Information Technology* (2021). <https://doi.org/10.1080/0144929X.2021.1876167>
- [38] John Hattie and Helen Timperley. 2007. The power of feedback. , 81–112 pages. <https://doi.org/10.3102/003465430298487>
- [39] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (11 1997), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [40] Kate S Hone and Ghada R El Said. 2016. Exploring the factors affecting MOOC retention: A survey study. *Computers and Education* 98 (2016), 157–168. <https://doi.org/10.1016/j.compedu.2016.03.016>
- [41] Chenn Jung Huang, Shun Chih Chang, Heng Ming Chen, Jhe Hao Tseng, and Sheng Yuan Chien. 2016. A group intelligence-based asynchronous argumentation learning-assistance platform. *Interactive Learning Environments* 24, 7 (2016), 1408–1427. <https://doi.org/10.1080/10494820.2015.1016533>

- [42] David H. Jonassen and Bosung Kim. 2010. Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology Research and Development* 58, 4 (2010), 439–457. <https://doi.org/10.1007/s11423-009-9143-8>
- [43] Zixuan Ke, Hrishikesh Inamdar, Hui Lin, and Vincent Ng. 2019. Give Me More Feedback II: Annotating Thesis Strength and Related Attributes in Student Essays. (2019), 3994–4004. <https://doi.org/10.18653/v1/p19-1390>
- [44] Timothy Koschmann. 1996. *Paradigm Shifts and Instructional Technology*. Technical Report. 1–23 pages. http://opensiuc.lib.siu.edu/meded_books/4
- [45] Klaus Krippendorff. 1980. Content analysis : an introduction to its methodology. <https://lib.ugent.be/en/catalog/rug01:000547037>
- [46] Klaus Krippendorff. 2004. Measuring the Reliability of Qualitative Text Analysis Data. *Departmental Papers (ASC)* 38 (2004). <https://doi.org/10.1007/s11135-004-8107-7>
- [47] Deanna Kuhn. 1992. Thinking as Argument. *Harvard Educational Review* 62, 2 (7 1992), 155–179. <https://doi.org/10.17763/haer.62.2.9r424r0113t670l1>
- [48] Deanna Kuhn. 1993. Science as argument: Implications for teaching and learning scientific thinking. *Science Education* 77, 3 (6 1993), 319–337. <https://doi.org/10.1002/sce.3730770306>
- [49] Deanna. Kuhn. 2005. *Education for thinking*. Harvard University Press. 209 pages. <http://www.hup.harvard.edu/catalog.php?isbn=9780674027459>
- [50] John Lawrence and Chris Reed. 2019. Argument mining: A survey. *Computational Linguistics* 45, 4 (2019), 765–818. <https://doi.org/10.1162/COLLa00364>
- [51] Doug Lederman. 2018. New data: Online enrollments grow, and share of overall enrollment grows faster. , 11–12 pages. <https://www.insidehighered.com/digital-learning/article/2018/11/07/new-data-online-enrollments-grow-and-share-overall-enrollment>
- [52] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2011. Mining behavioral economics to design persuasive technology for healthy choices. In *Conference on Human Factors in Computing Systems - Proceedings*. 325–334. <https://doi.org/10.1145/1978942.1978989>
- [53] Katja Lehmann, Matthias Söllner, and Jan Marco Leimeister. 2016. Design and Evaluation of an IT-based Peer Assessment to Increase Learner Performance in Large-Scale Lectures. *ICIS 2016 Proceedings* (12 2016). <https://aisel.laisnet.org/icis2016/ISCurriculum/Presentations/8>
- [54] Marco Lippi and Paolo Torroni. 2015. Argumentation Mining: State of the Art and Emerging Trends. *IJCAI International Joint Conference on Artificial Intelligence 2015-Janua*, 2 (2015), 4207–4211. <https://doi.org/10.1145/2850417>
- [55] Marco Lippi and Paolo Torroni. 2016. MARGOT: A web server for argumentation mining. *Expert Systems with Applications* 65 (2016), 292–303. <https://doi.org/10.1016/j.eswa.2016.08.050>
- [56] Sandra L. Marcus and Lance J. Rips. 1979. Conditional reasoning. *Journal of Verbal Learning and Verbal Behavior* 18, 2 (1979), 199–223. [https://doi.org/10.1016/S0022-5371\(79\)90127-0](https://doi.org/10.1016/S0022-5371(79)90127-0)
- [57] Xia Meng, Zhu Qian, Xingbo Wang, Fei Nie, Huamin Qu, and Ma Xiaojuan. 2022. Persua : A Visual Interactive System to Enhance the Persuasiveness of Arguments in Online Discussion. *Proceedings of the ACM on Human-Computer Interaction* 1, 1 (2022), 1–30. arXiv:4264657 [submit]
- [58] Tobias Mirsch, Christiane Lehrer, and Reinhard Jung. 2017. Digital Nudging: Altering User Behavior in Digital Environments. In *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)*.
- [59] Tobias Mirsch, Christiane Lehrer, and Reinhard Jung. 2018. Making digital nudging applicable: The digital nudge design method. In *International Conference on Information Systems 2018, ICIS 2018*. Association for Information Systems. AIS Electronic Library (AISeL). <https://research.cbs.dk/en/publications/making-digital-nudging-applicable-the-digital-nudge-design-method>
- [60] Raquel Mochales and Marie-Francine Moens. 2011. Argumentation mining. *Artificial Intelligence and Law* 19, 1 (3 2011), 1–22. <https://doi.org/10.1007/s10506-010-9104-x>
- [61] Benjamin Motz, Matthew G Mallon, and Joshua D Quick. 2021. Automated Educative Nudges to Reduce Missed Assignments in College. *IEEE Transactions on Learning Technologies* (2021). <https://doi.org/10.1109/TLT.2021.3064613>
- [62] E. Michael Nussbaum, Denise L. Winsor, Yvette M. Aqi, and Anne M. Poliquin. 2007. Putting the pieces together: Online argumentation vee diagrams enhance thinking during discussions. *International Journal of Computer-Supported Collaborative Learning* 2, 4 (11 2007), 479–500. <https://doi.org/10.1007/s11412-007-9025-1>
- [63] Stephen D. O’Connell and Guido Lang. 2018. Can Personalized Nudges Improve Learning in Hybrid Classes? Experimental Evidence From an Introductory Undergraduate Course. *Journal of Research on Technology in Education* 50, 2 (4 2018), 105–119. <https://doi.org/10.1080/15391523.2017.1408438>
- [64] OECD. 2018. The Future of Education and Skills - Education 2030. <https://doi.org/2018-06-15>
- [65] OECD. 2019. OECD learning compass 2030: In brief. (2019).
- [66] Jonathan F. Osborne, J. Bryan Henderson, Anna MacPherson, Evan Szu, Andrew Wild, and Shi Ying Yao. 2016. The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching* 53, 6 (2016), 821–846. <https://doi.org/10.1002/tea.21316>

- [67] Alexander Osterwalder, Yves Pigneur, Greg Bernarda, and Alan Smith. 2014. *Value proposition design : how to create products and services customers want : get started with ...* <https://books.google.de/books?id=jQ5yBgAAQBAJ&dq=osterwalder+business+value+proposition+design&hl=de&lr=>
- [68] Raquel Mochales Palau and Marie-Francine Moens. 2009. Argumentation Mining: The Detection, Classification and Structure of Arguments in Text. *Proceedings of the 12th international conference on artificial intelligence and law* (2009), 98–107. <https://doi.org/10.1145/1568234.1568246>
- [69] Niels Pinkwart, Kevin Ashley, Collin Lynch, and Vincent Alevén. 2009. *Evaluating an Intelligent Tutoring System for Making Legal Arguments with Hypotheticals*. Technical Report. 401–424 pages. http://iaiedsoc.org/pub/1302/file/19_4_05_Pinkwart.pdf
- [70] George E Raptis and Christina Katsini. 2021. Beter, funner, stronger: A gameful approach to nudge people into making less predictable graphical password choices. In *Conference on Human Factors in Computing Systems - Proceedings*. ACM, New York, NY, USA, 17. <https://doi.org/10.1145/3411764.3445658>
- [71] Carolyn Rosé, Yi Chia Wang, Yue Cui, Jaime Arguello, Karsten Stegmann, Armin Weinberger, and Frank Fischer. 2008. Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning* 3, 3 (2008), 237–271. <https://doi.org/10.1007/s11412-007-9034-0>
- [72] Omar Sabaj, Paula Cabezas, Germán Varas, Carlos González-Vergara, and Álvaro Pina-Stranger. 2020. Empirical Literature on the Business Pitch: Classes, Critiques and Future Trends. *Journal of technology management {I&I}* 15, 1 (2020), 55–63. <https://doi.org/10.4067/s0718-27242020000100055>
- [73] Oliver Scheuer. 2015. *Towards adaptive argumentation learning systems*. <https://www.researchgate.net/publication/298087259>
- [74] Oliver Scheuer, Frank Loll, Niels Pinkwart, and Bruce M. McLaren. 2010. Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning* 5, 1 (2010), 43–102. <https://doi.org/10.1007/s11412-009-9080-x>
- [75] Oliver Scheuer, Bruce M McLaren, Frank Loll, and Niels Pinkwart. 2012. Automated Analysis and Feedback Techniques to Support and Teach Argumentation: A Survey. *Educational Technologies for Teaching Argumentation Skills* 2 (2012), 71–124. <https://doi.org/10.2174/978160805015411201010071>
- [76] Christoph Schneider, Markus Weinmann, and Jan Vom Brocke. 2018. Digital nudging: Guiding online user choices through interface design. *Commun. ACM* 61, 7 (2018), 67–73. <https://doi.org/10.1145/3213765>
- [77] Sofia Schöbel, Torben Barev, Andreas Janson, Felix Hupfeld, and Jan Marco Leimeister. 2020. Understanding User Preferences of Digital Privacy Nudges – A Best-Worst Scaling Approach. *HICSS 2020 Proceedings* (2020), 3918–3927.
- [78] Julia E. Seaman, I. E. Allen, and Jeff Seaman. 2018. *Higher Education Reports - Babson Survey Research Group*. Technical Report. <http://www.onlinelearningsurvey.com/highered.html><https://www.onlinelearningsurvey.com/highered.html>
- [79] Paul Smolensky, Barbara Fox, Roger King, and Clayton Lewis. 1988. Computer-Aided Reasoned Discourse or, How to Argue with a Computer | Taylor & Francis Group. In *R. Guindon (Ed.), Cognitive Engineering in the Design of Human-Computer Interaction and Expert Systems*. Amsterdam, 109–162. <https://www.taylorfrancis.com/chapters/computer-aided-reasoned-discourse-argue-computer-paul-smolensky-barbara-fox-roger-king-clayton-lewis/e/10.4324/9780203782040-8>
- [80] Elliot Soloway, Mark Guzdial, and Kenneth E Hay. 1994. Learner-Centered Design The Challenge For WC1 In The Xst Century. *Interactions* (1994), 36–48.
- [81] Yi Song, Michael Heilman, Beata Beigman Klebanov, and Paul Deane. 2014. Applying argumentation schemes for essay scoring. In *Proceedings of the First Workshop on Argumentation Mining*. 69–78.
- [82] Christian Sonnenberg and Jan vom Brocke. 2012. Evaluation Patterns for Design Science Research Artefacts. Springer, Berlin, Heidelberg, 71–83. https://doi.org/10.1007/978-3-642-33681-2_7
- [83] Christian Stab and Iryna Gurevych. 2014. Annotating Argument Components and Relations in Persuasive Essays. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* ., 1501–1510. <http://www.ukp.tu-darmstadt.de>
- [84] Christian Stab and Iryna Gurevych. 2014. Annotating Argument Components and Relations in Persuasive Essays. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*. 1501–1510.
- [85] Christian Stab and Iryna Gurevych. 2014. Identifying Argumentative Discourse Structures in Persuasive Essays. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)(Oct. 2014), Association for Computational Linguistics, p.(to appear)*. 46–56. www.ukp.tu-darmstadt.de
- [86] Christian Stab and Iryna Gurevych. 2017. Parsing Argumentation Structures in Persuasive Essays. *Computational Linguistics* 43, 3 (2017), 619–659. <https://doi.org/10.1162/COLI-1.1.16271>
- [87] Christian Stab and Iryna Gurevych. 2017. *Recognizing Insufficiently Supported Arguments in Argumentative Essays*. Technical Report. 980–990 pages. www.ukp.tu-darmstadt.de

- [88] Karsten Stegmann, Christof Wecker, Armin Weinberger, and Frank Fischer. 2012. Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. *Instructional Science* 40, 2 (2012), 297–323. <https://doi.org/10.1007/s11251-011-9174-5>
- [89] Daniel D Suthers. 2003. Representational Guidance for Collaborative Inquiry. In *Arguing to Learn*. 27–46. https://doi.org/10.1007/978-94-017-0781-7_{T1\textbackslash}\textbackslash}_{\{\}\}2
- [90] Daniel D Suthers and Christopher D Hundhausen. 2001. *European Perspectives on Computer-Supported Collaborative Learning*. Technical Report. 577–584 pages. <http://ilt.ics.hawaii.edu/papers/2001/Suthers-Hundhausen-Euro-CSCL-2001.pdf>
- [91] Richard Thaler and Cass Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness*.
- [92] Richard H Thaler and Cass R Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press. 1–293 pages. [https://doi.org/10.1016/s1477-3880\(15\)30073-6](https://doi.org/10.1016/s1477-3880(15)30073-6)
- [93] Su Hie Ting. 2018. Ethos, logos and pathos in university students’ informal requests. *GEMA Online Journal of Language Studies* 18, 1 (2018), 234–251. <https://doi.org/10.17576/gema-2018-1801-14>
- [94] Heikki Topi. 2018. Using competencies for specifying outcome expectations for degree programs in computing: Lessons learned from other disciplines. *2018 SIGED International Conference on Information Systems Education and Research* (2018).
- [95] Stephen E. Toulmin. 1984. *Introduction to Reasoning*.
- [96] Stephen E. Toulmin. 2003. *The uses of argument: Updated edition*. 1–247 pages. <https://doi.org/10.1017/CBO9780511840005>
- [97] Crystal Tse and Katie Roof. 2021. Coursera Prices IPO at Top of Range to Raise \$519 Million - Bloomberg. <https://www.bloomberg.com/news/articles/2021-03-31/online-educator-coursera-s-top-of-range-ipo-raises-519-million>
- [98] Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *science* 185, 4157 (1974), 1124–1131.
- [99] John Venable, Jan Pries-Heje, and Richard Baskerville. 2016. FEDS: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems* 25, 1 (2016), 77–89. <https://doi.org/10.1057/ejis.2014.36>
- [100] Viswanath Venkatesh and Hillol Bala. 2008. Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences* 39, 2 (5 2008), 273–315. <https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- [101] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 27, 3 (2003), 425–478.
- [102] Jan vom Brocke, Wolfgang Maaß, Peter Buxmann, Alexander Maedche, Jan Marco Leimeister, and Günter Pecht. 2018. Future Work and Enterprise Systems. *Business and Information Systems Engineering* 60, 4 (2018), 357–366. <https://doi.org/10.1007/s12599-018-0544-2>
- [103] Jan vom Brocke, Alexander Simons, Kai Riemer, Bjoern Niehaves, Ralf Plattfaut, and Anne Cleven. 2015. Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems* 37, 1 (8 2015), 205–224. <https://doi.org/10.17705/1cais.03709>
- [104] Claudia von Aufschnaiter, Sibel Erduran, Jonathan Osborne, and Shirley Simon. 2008. Arguing to learn and learning to argue: Case studies of how students’ argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching* 45, 1 (1 2008), 101–131. <https://doi.org/10.1002/tea.20213>
- [105] Lev Semenovich Vygotsky. 1980. *Mind in society: The development of higher psychological processes*. Harvard university press.
- [106] Douglas Walton, Christopher Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511802034>
- [107] Thiemo Wambsganss, Sebastian Guggisberg, and Matthias Söllner. 2021. ArgueBot: A Conversational Agent for Adaptive Argumentation Feedback. In *16th International Conference on Wirtschaftsinformatik*. Essen, Germany.
- [108] Thiemo Wambsganss, Tobias Kueng, Matthias Soellner, and Jan Marco Leimeister. 2021. ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411764.3445781>
- [109] Thiemo Wambsganss and Christina Niklaus. 2022. Modeling Persuasive Discourse to Adaptively Support Students’ Argumentative Writing. *60th Annual Meeting of the Association for Computational Linguistics, ACL 2022 - Proceedings of the Conference* (2022).
- [110] Thiemo Wambsganss, Christina Niklaus, Matthias Cetto, Matthias Söllner, Jan Marco Leimeister, and Siegfried Handschuh. 2020. AL : An Adaptive Learning Support System for Argumentation Skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14.
- [111] Thiemo Wambsganss, Christina Niklaus, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. 2020. A Corpus for Argumentative Writing Support in German. In *28th International Conference on Computational Linguistics*

- (Coling). Barcelona, Spain. <https://doi.org/10.18653/v1/2020.coling-main.74>
- [112] Thiemo Wambsganss and Roman Rietsche. 2019. Towards designing an adaptive argumentation learning tool. In *40th International Conference on Information Systems, ICIS 2019*, Munich, Germany.
 - [113] Wenting Wang, Deeksha Arya, Nicole Novielli, Jinghui Cheng, and Jin L C Guo. 2020. ArguLens: Anatomy of Community Opinions on Usability Issues Using Argumentation Models. *Conference on Human Factors in Computing Systems - Proceedings* (2020). <https://doi.org/10.1145/3313831.3376218>
 - [114] World Economic Forum WEF. 2018. *The Future of Jobs Report 2018*. Technical Report. <https://doi.org/10.1177/0891242417690604>
 - [115] Robert J Weijers, Björn B de Koning, and Fred Paas. 2020. Nudging in education: from theory towards guidelines for successful implementation. *European Journal of Psychology of Education* (2020), 1–20. <https://doi.org/10.1007/s10212-020-00495-0>
 - [116] Armin Weinberger and Frank Fischer. 2006. A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers and Education* 46, 1 (2006), 71–95. <https://doi.org/10.1016/j.compedu.2005.04.003>
 - [117] Markus Weinmann, Christoph Schneider, and Jan vom Brocke. 2016. Digital Nudging. *Business and Information Systems Engineering* 58, 6 (2016), 433–436. <https://doi.org/10.1007/s12599-016-0453-1>
 - [118] Rainer Winkler, Matthias Söllner, and Jan Marco Leimeister. 2021. Enhancing problem-solving skills with smart personal assistant technology. *Computers and Education* 165 (2021). <https://doi.org/10.1016/j.compedu.2021.104148>
 - [119] Wei Xu, Marvin J Dainoff, Liezhong Ge, and Zaifeng Gao. 2021. From Human-Computer Interaction to Human-AI Interaction: New Challenges and Opportunities for Enabling Human-Centered AI. (2021), 1–73. <http://arxiv.org/abs/2105.05424>
 - [120] Barry J Zimmerman and Dale H Schunk. 2001. *Self-regulated learning and academic achievement: Theoretical perspectives*. Routledge.
 - [121] Verena Zimmermann and Karen Renaud. 2021. The nudge puzzle: Matching nudge interventions to cybersecurity decisions. *ACM Transactions on Computer-Human Interaction* 28, 1 (2021). <https://doi.org/10.1145/3429888>

Received January 2022; revised April 2022; accepted August 2022