Abstract

by No No

ALens: An Adaptive Training System for Academic Abstract Writing

ANONYMOUS AUTHOR(S)

SUBMISSION ID: 1969*

Novice researchers need to master the art of writing abstracts for their papers, but they often encounter difficulties in this process, as has been well documented in the field of higher education. Traditionally, students have been advised to take writing training courses to develop abstract writing skills. However, this process often does not provide students with individual and adaptive feedback on abstract writing. To fill this gap, we first conduct a formative study to derive user requirements for such an abstract writing training tool and propose an adaptive domain-oriented abstract writing training tool, *ALens*, that uses rhetorical structure parsing to identify main ideas, assesses abstract drafts based on linguistic features, and analyzes writing patterns of the reference abstract through visualization. A comparative user study with an alternative tool for abstract writing training demonstrate the effectiveness of our approach.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; *Visualization*; User studies.

Additional Key Words and Phrases: Educational Applications, Writing Support Systems, Automated Feedback, Summarizing Learning

ACM Reference Format:

Anonymous Author(s). 2018. ALens: An Adaptive Training System for Academic Abstract Writing. In . ACM, New York, NY, USA, 24 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Paper writing is an essential skill that junior graduate students or researchers should master [97] because of its importance in learning, understanding, applying, and synthesizing new knowledge [31]. Basically, the structure of a typical research paper follows a pattern called the "*King Model*" [29], which delineates the thematic progression of an article through six sections: *title*, *abstract*, *introduction*, *body*, *discussion*, and *references* [29]. Among the major components of an academic paper, the abstract, which usually consists of a separate paragraph outlining the paper' content [79], has become increasingly important. For example, with the boom in search engines and bibliographic databases, the title and abstract are often the only two parts of a research paper that can be freely viewed by potential readers, while access to the full paper may be subject to charges to the copyright owner [78]. In addition, when researchers conduct systematic investigations of related work, they tend to discard papers with obscure abstracts because reading the full text is too time-consuming for them [3, 79]. In addition, during the blind review process, editors use abstracts to invite appropriate reviewers with expertise in the relevant field to evaluate the paper [3, 78].

However, concerns about the academic abstract writing skills of undergraduate and graduate students in higher education are well documented [23, 33, 77]. From a faculty's perspective, writing well is more than just following writing conventions. It also involves creative inspiration, problem solving, reflection, and editing, culminating in a complete manuscript [23, 46]. In particular, the Organization for Economic Co-operation and Development (OECD) has proposed a learning compass that already emphasizes higher-order thinking skills such as problem solving and reflection as skills needed by further-ready students [34]. From a student's perspective, writing an abstract can be a daunting task, both in terms of getting ideas on paper and mastering writing rules such as *logic*, *summarizing*, *argumentation*, and *grammar* [23, 35]. To help students develop abstract writing skill typically included in paper writing skills, institutions, such as universities, have conventionally suggested students taking thematic writing training courses, such as scientific paper writing and biology essay writing, during which it is important that individual students receive ongoing formative feedback [15]. However, the need to provide optimal formative feedback on individual abstract writing training in traditional large-scale lectures is often hampered by limited financial and instructional resources []. One possible solution for providing individual feedback is to utilize recent advances in Natural Language Processing (NLP) and Machine learning (ML).

We systematically reviewed the literature on abstract writing in the field of educational technology following the rigorous approach suggested in [87]. However, we found that the existing literature is under-researched in terms of **academic abstract writing** training []. In contrast, quite a few tools have been developed to enhance the **summary writing** ability of students. It should be noted that abstracts and summaries are different []. While there are nuances in various accounts of the difference between an abstract and a summary, the general perception is that a summary of an entire article is a more detailed version of an abstract, and that an abstract is usually written in the order of the content of a research paper, while a summary may focus on important aspects of the article [1]. Despite the differences, it has to be acknowledged that both contain important content and require students to have the ability to condense information.

We borrow the ideas from studies on summary writing training from which we can conclude that the systems or methods proposed in computer-assisted summary writing training usually involve a cycle of three stages, i.e., *reading*, *writing*, and *feedback*. First, reading and understanding the main idea of the source text is critical. Previous work [76**?**] used concept maps to help students identify the main ideas and understand their hierarchical structure. While it may be suitable for general summary writing training, it is not appropriate in the scenario of academic abstract writing training because the concept maps in [76**?**] need to be generated by consultants and experts, which is too labor-intensive, especially when it comes to academic papers. In order to annotate concept maps for papers in different fields, a large number of experts are needed, as academic terminology and writing styles vary widely. Even in the same discipline, such as human-computer interaction (HCI), abstracts are written differently due to the different types of papers; for example, an application paper and a survey paper generally have different abstract writing styles. In terms of writing, existing summary training tools provide paradigms and summary strategies to instruct writing, which are good guides for abstract writing training. For the last stage, there are four types of feedback according to literature review, namely: *providing scores* [22, 26, 39, 88], *peer review* [92], *section content coverage* [22, 26, 88] and *summary writing strategy detection* [1, 26, 43]. However, for the same reasons as concept maps, the coverage of chapter content that requires instructor annotation does not apply to academic writing scenario. To the best of our knowledge, there are no principles and proofs in the current literature on how to design automatically adaptive computer-assisted academic abstract writing tools to help a researcher learn abstract writing styles and patterns in his/her field.

To clarify the current status and main concerns of the abstract writing and training process for academic papers, we first systematically reviewed the literature in the field of pedagogy and educational technology [87]. Then, we investigated the pain points of research beginners in writing abstracts through a formative study (a survey of 167 students and semi-structured interviews with 12 students), from which we extracted four major barriers that learners face in writing abstracts, namely: *lack of skills in rephrasing content*, *idea organization*, *main idea identification*, and *writing style recognition*. First, rephrasing is considered to be one of the core skills for paraphrasing key content, which

[1]

is the essence of abstract writing [9]. However, students, especially L2 (second language) learners, resort to copying sentences from other parts of the paper rather than rewriting the main ideas in their own words [9, 35]. Second, when it comes to the organization of ideas in each sector of the paper, most junior students are not skilled at integrating them in a logical and cohesive manner while making the essay fluent and clear [9]. Third, despite the ability of novice researchers to identify the topic of the essay, secondary and irrelevant information is still easily incorporated, which means that they are deficient in their ability to grasp the complete hierarchy of ideas in the text [23]. Fourth, 73% students mentioned in their interviews that it would have been better to show writing style, or at least give them some hints about it.

To address the above issues and fill the gap in abstract writing training, as well as to take advantage of recent advances in NLP technology, we propose a domain-oriented abstract writing training system *ALens* (short for Abstract Lens), an adaptive learning tool that uses rhetorical structure parsing to identify main ideas, evaluates their abstracts in terms of different linguistic features, and uses visualization to analyze the writing patterns of reference abstracts (i.e., ground truth abstracts). Specifically, to address the first challenge and to train users in their paraphrase abilities, we incorporate linguistic features, such as lexical and syntactic complexity, as assessment metrics. To address the idea organization problem, we ran a re-trained sentence classification model that classifies abstract sentences into five genres (i.e. background, objective, methods, results, and conclusions) [19, 35, 44] and displays the results in different colors. Considering the classification feedback, self-regulation [5] regarding the organization of the abstract will be aroused, which will lead the user to discover which parts of the paper need to be included in specific areas and whether ideas are expressed in a logical and cohesive order. To address the third challenge, we use discourse parsing with Rhetorical Structure Theory (RST) [40, 58] to construct RST segments from the perspective of identifying logical relations in the introduction. It separates sentence groups into RST trees with phrases on leaf nodes and logical relations on branches. RST uses rhetorical relations (e.g., elaboration, contrast, etc.) to depict the structure and logic of each part of the text [59]. By parsing different paragraphs or the whole introduction, users can get the hierarchical structure of the text at different granularities and grasp the hierarchy of ideas in the text. Finally, to address the last issue, following the approach utilized in the works about attention [11, 75, 84, 85] an attempt is made to find the relevant tokens in the generated abstract from the source text, applying semantic similarity to align the ideas between the reference abstract and the source text in an attempt to reconstruct the style used by the authors in writing the reference abstract. In addition, about 27% told us in interviews that they might get stuck in, faced with an empty paper, when they have some ideas to write, and they don't know where to start. To facilitate the writing of the first draft, we embedded a summarization model into our system [53, 67, 93] as an option to generate an initial draft as a prompt to start.

We demonstrate the impact of *ALens* on users' abstract writing skills by evaluating our system in a social computing abstract writing training scenario. We quantitatively compared an abstract writing training method with our system. In a user study with 21 students, the results show that with the help of *ALens*, users can successfully write more accurate abstracts with well-organized content than the alternative tool. In addition, we measured the usability, perceived usefulness, and intention to use of both tools using key constructs [82, 83], and the results were encouraging, suggesting that *ALens* motivates students to learn abstract writing patterns in their own domain and to write abstracts in an appropriate style. In summary, the main contributions of this work are:

- We conduct a formative study to understand the problems they encountered in the academic abstract writing process.
- We build *ALens*, an automatic feedback learning tool that first incorporates visualization and interactive features into academic abstract writing training.

- We show the effectiveness of *ALens* by comparing it to an alternative tool for abstract writing training.

2 RELATED WORK

Literature that overlaps with this work can be grouped into four categories, *technology-mediated summary writing assistance, summary evaluation metrics, NLP models in summarization task*, and *self-regulated learning*.

2.1 Technology-Mediated Summary Writing Assistance

We systematically reviewed the literature on abstract writing in the field of educational technology following the rigorous approach suggested in [87]. However, few literature focuses on the development of learning tools for abstract writing, although in the last decade, several tools have been developed to improve students' summary writing skills. The main difference between an abstract and a summary of a whole article is the length and purpose []. Abstracts usually follow the empirical order of content prescribed by the journal or association and cover the major aspects of the research paper []. Summaries may not follow specific guidelines, emphasizing certain important aspects of the paper and providing more detail than the abstract []. Despite the differences, it has to be acknowledged that both are abbreviated versions of the paper and contain important content which requires the ability to understand, express, synthesize and paraphrase [17, 18, 74]. For example, in studies of computer-assisted summary writing training, **concept map** [22, 76] arranges the concepts in the text in a hierarchical manner, with general concepts at a shallower level and specific concepts at a deeper level. It attempts to facilitate students' identification of the main ideas and understanding of the corresponding supporting ideas. Some studies have proposed methods that identify the **summarization strategies**, including deletion, sentence combination, and paraphrasing used by students to help assess teacher summarization processes and target them during training. **Worked examples** [22, 39, 88] are exemplars with worked-out steps and predetermined questions often used as instruction to help students learn to read the original text and summarization strategies. In addition, by comparing multiple worked examples, students can gain the ability to identify patterns of relevant and irrelevant information [22]. **Computer-supported collaborative learning (CSCL) approaches** [92] are embedded in the summary writing training system, where students receive peer feedback through online dialogue and interaction. In the process of digesting peer feedback, students reflect on their summarization process and make further revisions [42].

However, the above methods cannot be directly applied to academic abstract writing training. Specifically, concept maps and worked examples are carefully prepared by the tutor, which need to be annotated article by article as the subject matter and progress of the articles vary. In other words, when it comes to academic papers, the workload of instructors in generating concept maps and worked examples will be very high. To fill the gap in abstract writing training and to take advantage of recent advances in NLP technology, we use rhetorical structure parsing to identify main ideas, evaluate abstracts in terms of different linguistic features, and use visualization to analyze the writing patterns of reference abstracts.

2.2 Summary Evaluation Metrics

In the learning process, it is important to provide individual and adaptive feedback [15], and the same is true for abstract writing training. We consider assessment methods widely used in summary writing training as a potential approach to abstract writing training. For example, assessment scores are a typical method for providing formative feedback in computer-assisted abstract writing training [] and there are three kinds of scores, namely content coverage scores [22, 76, 88], scores given by mathematical methods [26, 48, 54, 76, 94] and scores predicted by pre-trained language models [16, 57, 91]. Specifically, content coverage scores are calculated automatically to measure the degree of coverage of each content in the summary. Although calculated automatically, the exact content to be measured is specified by the instructor on an article-by-article basis. However, this is clearly not appropriate for academic abstract writing, as the differences in disciplines, fields, and paper types result in a significant amount of work for faculty to develop content criteria for each type of article. In contrary, mathematical methods and deep learning methods are the most suitable candidates. Although pre-trained deep learning language models can achieve a high degree of agreement with human estimates, their high performance is highly dependent on the availability of relevant datasets. Due to the unavailability of high-quality datasets, we turn to mathematical methods. Specifically, three approaches are commonly used: metrics used in ML [54, 66, 94], latent semantic analysis [52] and linguistic features [27, 50, 51]. Metrics in ML, such as *ROUGE* [54], *BERTScore* [94] and *Bleu* [66] and LSA all assess the quality of a summary based on semantic overlap with the reference or source text, thus giving an overall score for the summary. However, such single score is not an appropriate feedback [89] to reveal the gap between what is understood and what should be understood [71]. Therefore, we rate summary using different scoring criteria (e.g. lexical complexity and cohesion) based on linguistic features, which is considered more appropriate because it captures different aspects of the summary and thus provides more informative and instructive feedback [13].

2.3 NLP Models in Summarization Tasks

The text processing models behind text summarization tools can be broadly classfied into two categories, namely *extraction* and *abstraction* [32]. Extractive approaches [45, 55, 64, 65, 73, 96] copy salient phrases and sentences from the text and merge them to create summaries [65, 96], thus ensuring that the summary is factually consistent with the source text [20]. However, the extract paradigm is often criticized for being logically inconsistent with the input text [72, 73]. Abstraction methods [36, 53, 90, 93] rearrange the language in the text and add new words/phrases to the abstract as needed [37]. Although abstraction methods perform well in generating fluent human-like summaries [93], factual inconsistency [49] and fidelity [61] issues may lead to a heavier cognitive workload. Since there are no advantages or disadvantages of the above two approaches, in our work we embed both types of summarization models into our system [53, 67, 93] as an option to generate an initial manuscript as a prompt for users to start.

2.4 Self-regulated Learning

It has been hypothesized that providing students with feedback about their writing abilities will enhance their learning experience and facilitate the writing of high quality summaries []. In order to enable self-regulated learning, providing students with formative feedback as well as setting goals is critical [10]. It has been suggested that in order for feedback systems to be effective, learners must be provided with goals, track their progress, and identify actions to help them achieve those goals [38]. However, individuals are unable to track their own progress during the work [14]. Using targeted assessment and feedback is a good way to enhance the learning process [70]. If feedback on students' abilities is provided throughout the intervention, it can increase their chances of achieving better short-term outcomes on specific learning tasks [7, 38, 70]. In this work, we provide students with user-centered adaptive feedback on their abstracts to determine whether they could write and improve organized abstracts.

3 FORMATIVE STUDY

We conducted a systematic literature review on abstract writing in the field of educational technology following the rigorous approach suggested in [87]. However, there are few studies on abstract writing training, whereas there are many works about summary writing training. Because of the similarities between summary writing and abstract writing in terms of the writing process and required competencies, we thought we could borrow ideas from these works. To fill the gap in abstract writing training tools, we use a top-down approach, first distilling possible meta-requirements from the existing literature on summary writing training tools. With this goal in mind, we selected 27 papers for meticulous analysis, from which we distilled a closed loop of summary writing learning. In addition, because abstract writing aids span the fields of education, psychology, and computer science, we focused on these streams of literature. On this basis, we selected 32 papers for further analysis. and learned established pedagogical theories in writing [12] and metacognition [56] in the learning process, which is considered a meta-need for an adaptive learning tool.

Next to derive user requirements for an academic abstract writing training system, we first need to understand the problems students encounter in the academic writing process. Therefore, we design and examine the following four research questions: 1) *RQ1: What are students' current practices when writing abstracts?* 2) *RQ2: What are the challenges students have when writing abstracts?* 3) *RQ3: What kind of support do students need when writing abstracts?* and 4) *RQ4: What do students want to know about the style of the abstracts?* Therefore, a formative survey and follow-up formative interviews were conducted with professors and students who had experience in writing academic abstracts.

3.1 Survey Study

3.1.1 Survey Protocol. The survey was administered on the Microsoft Forms online platform. The survey questions included abstract writing practices, difficulties with abstract writing, help needed when writing abstracts, and demographic questions. The survey contained textual questions and ranking questions about the academic paper writing process. It also contained questions about students' challenges. 5-point Likert scale questions are used to measure students' attitudes towards several potential types of assistance. The survey also contained open-ended questions about student requests. At the end of the survey, respondents can leave their contact information if they would like to be interviewed for a follow-up. Supplemental materials contain detailed survey questions.

3.1.2 Respondents and Recruitment. We recruited 164 respondents (54 female, 106 male, and 4 prefer not to specify) between the ages of 19 and 36 (undergraduate students: 105, master students: 42, PhD. students: 13, others: 4) through advertised posts in online university communities. Of all respondents, 125 with experience in academic writing answered RQ1 – RQ4, the other 39 answered RQ2 – RQ4.

3.2 Interview Study

3.2.1 Interviewees. To gain more insight into students' challenges and requirements when writing their abstracts, we further contacted 11 students (3 female, 8 male; 10 graduate students, 1 PhD. student) who left their email addresses in their survey responses. Their ages ranged from 20 to 26 years old (mean age = 23.09, SD = 1.81). We also contacted a university faculty (male, age = 34) who teaches academic paper writing for 3 years. We refer to the respondents as P1 to P11.

3.2.2 Interview Protocol and Analysis Method. We conducted remote semi-structured interviews using an online communication tool, and audio-recorded the interviews with consent. The interviews consist of three main sections: (1) how students typically write their abstracts; (2) challenges in writing abstracts; and (3) what features students need. We analyzed the results of the interviews based on a content analysis approach [62].

3.3 Findings and Design Goals

We recapitulated the following findings from the survey and interview results. For RQ1, we found that most learners (survey: 139/164, interview: 10/11) wrote the abstract after writing the introduction or body of the paper. They usually first determined the structure of the abstract, i.e., what parts needed to be included and which parts were more important. Then, they purposefully wrote about each part. And most learners (survey: 119/164, interview: 8/11) would refer to the introduction or body of the paper to ensure consistency. The detailed results of the survey are shown in Table 1. The main findings of RQ2 – RQ4 are summarized below.

<table>
<tr><th colspan="5">Had academic writing experience</th><th>Potential assistance</th><th>M</th><th>SD</th></tr>
<tr><td colspan="4">Yes</td><td>No</td><td>Word count</td><td>4.21</td><td>0.35</td></tr>
<tr><td colspan="4">125/164</td><td>39/164</td><td>Online revision</td><td>4.24</td><td>0.88</td></tr>
<tr><td colspan="4">Used writing aids</td><td rowspan="6">/</td><td>Abstract evaluation</td><td>4.09</td><td>0.86</td></tr>
<tr><td colspan="2">Yes</td><td colspan="2">No</td><td>Instructional feedback</td><td>4.03</td><td>0.91</td></tr>
<tr><td colspan="2">67/125</td><td colspan="2">58/125</td><td>Abstract writing style recognition</td><td>4.10</td><td>0.88</td></tr>
<tr><td colspan="4">Wrote abstract referring to</td><td>Present abstract structure</td><td>4.17</td><td>0.84</td></tr>
<tr><td>Intro</td><td>Body</td><td>Full text</td><td>Others</td><td>Present key information of intro</td><td>4.22</td><td>0.69</td></tr>
<tr><td>29/125</td><td>45/125</td><td>47/125</td><td>4/125</td><td>Present logic relation of sentences</td><td>4.06</td><td>0.88</td></tr>
<tr><td colspan="3" rowspan="2">Total respondents number</td><td colspan="2" rowspan="2">164</td><td>Support intro annotation</td><td>4.08</td><td>0.83</td></tr>
<tr><td>Present intro structure</td><td>4.06</td><td>0.88</td></tr>
</table>

TABLE 1. Results of the survey: On the left is the distribution of the number of distinct respondents; on the right is the statistics of 5-point Likert scale questions about potential help (1 – 5: very helpless – very helpful).

3.3.1 Challenges of Academic Abstract Writing. We combined our survey research and interview study to present the following five challenges.

C1: Lacking skills in rephrasing content (N=7). Sometimes students tended to rewrite key sentences from the introduction, especially when writing background and conclusion sentences (P1, P3 – P7). However, rephrasing content is sometimes tricky because students cannot use sentences from the introduction directly. Current summarization techniques produce wording that is still too close to the original, which does not help solve this problem. "*How to precisely express the same meaning in a new way is sometimes an annoying problem(P2, male, age=24).*"

C2: Idea organization (N=9). Having a good insight into the logic of the introduction is crucial for students to write excellent abstracts. According to the survey results, most students (82.96%) wrote their abstracts after writing the body of the paper. However, students may forget the logic flow of the introduction after writing the body of the paper. On the other hand, they cannot remember the logic of the instruction well, especially the parts with complex logical relationships. So students need to reread the introduction to recall and reorganize these logical chains, which wastes time. On the other hand, existing abstract generation techniques tend to break the logic chains, which will mislead students and cause additional problems. For example, "*in my field, the prior experiments section in the introduction includes too many experimental methods. When writing the abstract, I always need to figure out again how they are related (P9, male, age=21).*"

C3: Main idea identification (N=8). According to the survey results, almost half of the respondents (47.6%) answered that it was quite difficult to *summarize all the key points in a limited space*, or to *write them concisely enough.*

C4: Writing pattern and style recognition (N=8). Abstract writing is usually domain-oriented, since different subjects and different types of paper and different journals and conferences generally varies in style and writing pattern. The survey found the 53.0% respondents found that it is time consuming for them to acquire the abstract style by perusing articles in that domain and find the regularity and 65.9% respondents thought it would be better if the regularity is displayed to them.

C5: Requirement for a first draft 3 students in formative interview responded that they don't know where to start with writing. "*I'm used to making changes based on other people's draft, and I can't start with totally blank space (P7, female, age=22).*"

3.3.2 Design Requirements. Based on the identified challenges in academic paper abstract writing and user expectations for satisfactory assistance results and comprehensive functionality, we derive the following design requirements of an adaptive abstract writing training tool.

R1: Provide assistance on rephrasing. Through formative study, we found that how to rewrite key sentences and other information extracted from the introduction is a big challenge for learners (C1). To solve this problem, we should provide guidance on representing information in another way, such as sentence transformation and phrase substitution based on self-regulated learning theory [].

R2: Help learners to have a better understanding of the introduction. From the previous formative study, we found that many learners encountered difficulties in selecting core information in a limited space, i.e., the problem of main idea identification (C3). To effectively address this problem, learners' mastery of the structure and content of the introduction is exceptionally demanding. On the one hand, the introduction is a distillation of the main text, and the relationship between some sentences is difficult to grasp. Therefore, we should provide guidance on identifying the logical relationships of sentences in the introduction. On the other hand, the introduction is long and requires extra time to reread because of forgotten content. We should also help learners to quickly review the structure and content of the introduction.

R3: Assist learners in organizing the main ideas distilled from the introduction. Through the formative study, we found that sometimes learners know what needs to be emphasized but have difficulty organizing these key contents fluently (C2). We should help them to have a better understanding of the information from a new perspective and help them to organize the main ideas in a rational way.

R4: Assistance in understanding domain-specific abstract style. As shown in Table 1, most participants reported that knowledge of how abstracts are written was very useful to them (Mean = 4.10, SD = 0.88) (C4). The underlying style could guide learners to write abstracts with more accurate content and organization. Therefore, in addition to directly presenting the reference abstract, we should also demonstrate its style in a clear and intuitive manner.

R5: Preparation of the first draft version. In the semi-structured interviews, 3 students mentioned the difficulty of writing abstracts from scratch. Despite the relatively low rate, we observed the rise of text summarization platforms such as *TLDR this*[2], *Resoomer*[3], and *Wordtune Read*[4]. Therefore, we believe that there is a trend to harness the power of NLP techniques to facilitate abstract writing, for example, to generate the first-edition draft.

R6: Easy to access and use. From the survey, some potential users were concerned about the complexity of the system's functionality and the difficulty of using it. Therefore, we had to ensure that the tool would not be burdensome and responsive to users, while providing practical features to address the above challenges. For example, users did not need to install additional software or hardware, and we use the typical writing assistance platform interface design

[2]https://tldrthis.com/
[3]https://resoomer.com/en/
[4]https://app.wordtune.com/read

familiar to learners. In addition, the tool needs to have the fundamental features of a writing aid to ensure effectiveness, such as online revision and basic annotation features.

4 APPROACH OVERVIEW

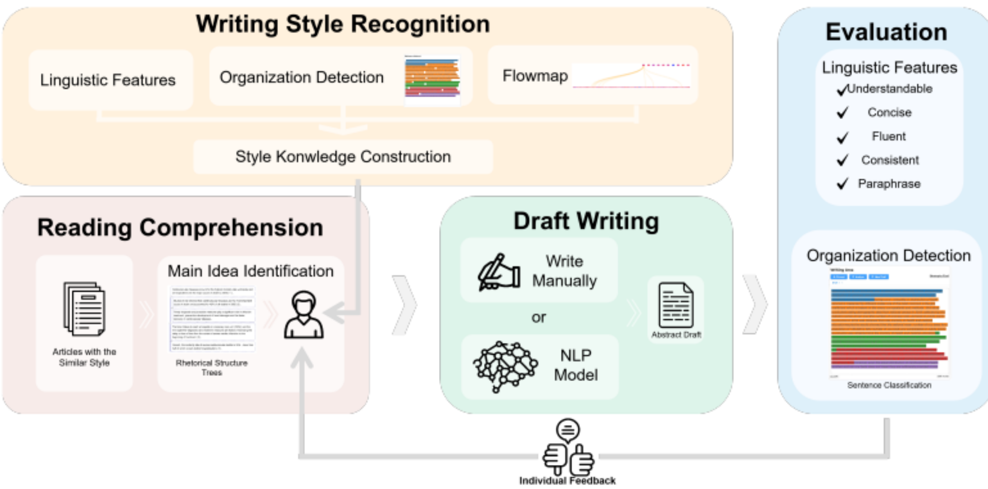

Fig. 1. Pipeline of *IntorLens*: (1) Generating the first manuscript for later revision; (2) Refining the draft with the consideration of organization, main idea and lexicon; (3)Result evaluation based on semantic analysis.

Based on the requirements derived from the formative study, we design an abstract writing training process and incorporate it into a web-based writing assistance platform named *ALens*. It facilitates users to quickly grasp the main ideas of an essay, optionally write using a summarization model in NLP, recognize their deficiencies in abstract writing, and gain knowledge of the style in a specific scenario. To support the writing process in a convenient and user-friendly manner (**R6**), visualizations and interactions are integrated into *ALens* to cater to users' mental habits and to meet the requirements at different granularities. The front-end interface consists of a *Rhetorical Structure View*, a *Writing Area*, an *Evaluation Dashboard*, and a *Reference Abstract with a Flow Map*. Figure 1 describes the general stages of the designed abstract training pipeline. First, the user can select an article that needs to be learned for abstract writing and upload it. Subsequently, the rhetorical structure of the original article is analyzed to help the user quickly identify the main ideas in terms of logical structure (**R2**). Then, the user can choose to write the first draft from scratch or with the help of a summarization model (**R5**). Considering the lack of content organization, sentences in the abstract are divided into several types (e.g. background and conclusion) [4, 25, 63], and the completeness of the abstract is checked according to the domain of the paper, i.e., whether the rough draft properly covers and arranges the domain typically information needed for the abstract and leads the user to reflect on these issues (**R3**). At the same time, *ALens* can automatically analyze the linguistic features of the abstract to check whether it is comprehensible, concise, fluent and consistent with the source text (**R1**). In addition, paraphrase detection is applied to the feedback to guide the user to rephrase sentences instead of copying them. Finally, users can check the writing style of the reference abstract and analyze its linguistic features and organization. Specifically, a flow map is used to align ideas in the reference abstract with the source text, wwhich is to find the most relevant content from the source text. By comparing these features of different articles with similar academic domain "style", users are expected to discover writing patterns and learn writing styles (**R4**).

5 BACK-END ENGINE

The back-end engine first parses the article into a rhetorical tree, and then supports the production of an initial manuscript for later revision. The sentences in the abstract are then classified into several genres and evaluated by different linguistic features, providing personal feedback to stimulate revision according to a self-regulated learning theory [].

5.1 Rhetorical Structure Parsing

To enable learners to quickly grasp the hierarchical structure of a text and to avoid wasting time by repeatedly reviewing the introduction when writing an abstract, we provide a rhetorical structure parsing for each paragraph. In particular, we provide rhetorical relationship recognition for any text with continuous span, i.e., uninterrupted linear intervals. For example, after inputting a text containing five sentences (S1 – S5), the model can output relations for any continuous text span shown in Figure 2. S4 *contrasts* with S5, while the text containing S1 – S3 is the *background* for S4 – S5. All relations and their hierarchical structure build the structure of the whole text and give a deeper understanding of the text structure.

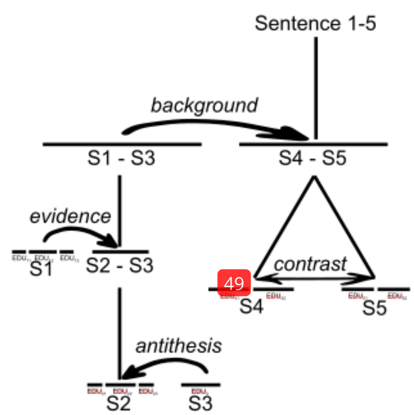

Fig. 2. An example of RST structure tree. There are different rhetorical relations between sentences. Each sentence consists of several elementary discourse units (EDUs).

The identification of rhetorical relations consists of two parts: 1) text segmentation and 2) relation identification. For text segmentation, we need to segment the input text into elementary discourse units (EDUs), i.e., tokens of adjacent text spans that are roughly similar to independent phrases. Each sentence consists of several EDUs. We retrained the model proposed by Heilman et al. [40] to operate the segmentation task with high accuracy. The second relation recognition component was modeled as a classification problem, i.e., classifying the relation of two consecutive EDUs into 16-tuple types such as elaboration, contrast, and joint. For this task, we utilize the ZPar model [95], which predicts rhetorical relationships across text at different levels of granularity. The output of the model is the relationsip of EDUs, i.e., phrase relationships in a sentence, which has no value for abstract writing. Therefore, we modified and retrained the ZPar model to predict rhetorical relations between adjacent sentences with more appropriate granularity.

5.2 Construction of NLP Models

To accomplish the NLP tasks described above, one can directly use existing pre-trained models. Generic models trained on natural language datasets such as news and emails tend to perform poorly on domain-specific tasks. Abstracts of research papers and their sentence classification are such examples, as there are various abstract and writing styles, which require domain-specific models.

5.2.1 Summarization Model. It is important to note that providing a relatively reliable first draft of the abstract is an important foundation for the later revision process. However, one requirement for developing NLP models with the ability to summarize research papers is the availability of relevant datasets. We reviewed the literature on corpora and found that the *arXiv* and *PubMed* datasets released in [25] met our requirements. *PubMed* contains 119*k* pairs of article bodies and abstracts of biomedical literature, while *arXiv* contains 203*k* pairs of articles on different topics. Since different subjects have different terminologies and writing styles, we decided to construct a subset of *arXiv* called *arXiv-cs* that contains only computer science (cs) articles by iterating through the abstracts in *arXiv* and match them in *arXiv-dataset* released in [24], which hosts 1.5*M* metadata of preprinted articles in Physics, Mathematics and Computer Science from 1991 to 2019 to determine whether they belong to the computer science domain.

Based on these two datasets, we developed the following fine-tuning scheme. Since the two datasets belong to the Biomedicine and Computer Science domains, respectively, we developed four summarization models, and for each domain, we developed an extractive model based on *BART-base*[5] [53] and an abstractive model based on *T5-large*[6] [68]. The training parameters follows those described in the original publications and GitHub; in addition, the maximum input token and maximum output token are set to (4096, 512) on *PubMed* and (16384, 512) on *arXiv-cs*, and the model performance for both fine-tunings is shown in Table 2. The results show that these models can be used to generate a relatively reliable abstract for later revision.

	PubMed				arXiv-cs			
	Rouge-1	Rouge-2	Rouge-L	Rouge-Lsum	Rouge-1	Rouge-2	Rouge-L	Rouge-Lsum
BART	48.20	22.35	29.22	44.38	47.67	19.88	27.73	43.16
T5	47.34	22.53	28.74	42.79	46.91	19.67	27.41	41.87

TABLE 2. Results of fine-tuning BART and T5 on two datasets in Biomedicine and Computer Science.

In order to provide users with a fine-grained summarization feature, four fine-grained models are not sufficient. What we have mentioned is that the quality of summaization results is highly domain-dependent. Even in a specific domain, if a user tries to summarize a part of the source text (e.g., a few paragraphs and sentence groups) using a model fine-tuned on a domain-specific corpus, the model does not work well because the output length and input length are shorter than the training data, and the summarization knowledge learned from the whole text cannot be directly applied to the segments of the text, which are derived from the actual practice of our system. Therefore, we use different models for local summarization. We fine-tuned *BART-large* and *T5* on *CiteSum* [60], which contain small paragraph summaries of citation papers of scientific papers, to simulate the summary scenario of scientific paragraphs in domain-specific research papers.

5.2.2 Sentence Classification Model. To facilitate the organization of the ideas distilled from the source text, we detect the organization in the abstract, thus inducing the user to think about the coverage and arrangement of the content. The organization detection task is considered as a multi-class sentence classification task, where each sentence in the abstract is classified into five types, i.e., *background*, *objectives*, *methods*, *results* and *conclusions* []. As for our choice of this classification scheme, the reasons are the absence of domain-specific and annotated sentence classification datasets. For the sentence classification task, we found two datasets, *PubMed 200k RCT* [28], containing 200*k* pairs of type and abstract sentences and *CSAbatract* [6], containing 2*k* pairs corresponding to the Biomedicine and Computer Science domains, respectively. The goal of the classification model is to provide accurate classification to identify sentence

[5]https://github.com/facebookresearch/fairseq/tree/main/examples/bart
[6]https://github.com/google-research/text-to-text-transfer-transformer

Facets of Quality	Linguistic Features
Understandability	Frequency (COCA spoken, AW)
	Word frequency (COCA spoken, AW)
	Frequency (COCA spoken, FW)
	Word frequency (SUBTLEXus, CW)
	Word frequency (SUBTLEXus, AW)
Consistency	Source similarity (word2vec)
Fluency	Adjacent sentence similarity (word2vec)
	Repeated content lemmas and pronouns
	Binary adjacent sentence overlap (FW)
Diversity	Type-token ratio (AW)
	MATTR (FW)
	Number of CW tokens
	MTLD (FW)
	MTLD (AW)
	MTLD (CW)
	Lexical density (Percentage of CWs)
	Type-token ratio (CW)
	SD of dependents per nominal subject
	SD of dependents per clause
	SD of dependents per object of the preposition
Conciseness	Mean length of sentence
	Mean length of clause

TABLE 3. Five facets of the quality for abstracts and the corresponding linguistic features to compute them. AW=All words, CW=content words, FW=Function words, SD=Standard Deviation. For specific definitions of linguistic features, please refer to [27, 50, 51].

intent in the abstract, which can be used to assess the organization of the draft and thus induce self-reflection on how to improve the coverage and arrangement of the content. Following *BERT-base-uncased*[7], a model pretrained on *BookCorpus*[8] and *English Wikipedia*[9] proposed in [30], we do transfer learning on these two datasets separately and train the model with different hyperparameters. For the sentence classification task on *PubMed 200k RCT*, we achieved an F1-score of 83.59%, while for the task on *CSAbstract*, we achieved an F1-score of 86.37%. These results show that we can embed our BERT model into our system to provide users with organizational analysis.

5.3 Evaluation Metric

As with abstract writing training, it is critical to provide individual and adaptive feedback during the learning process [15]. The summaries and abstracts synthesize the main ideas of the article and require comprehension, expression, synthesis, and paraphrasing abilities [17, 18, 74]. Abstract writing training might benefit from the evaluation methods used in summary writing training. Computer-assisted summary writing training typically provides formative feedback in the form of assessment scores. However, this single score is not an appropriate feedback [89], so we tend to use different

[7]https://github.com/google-research/bert
[8]https://github.com/google-research/bert
[9]https://en.wikipedia.org/wiki/English_Wikipedia

scoring criteria for scoring abstracts. From the literature [4, 80, 81] and writing instructions on website[10][11][12], we concluded that a good abstract should be comprehensible, concise, and fluent. In addition, it should be consistent with the source text and use different words and phrases from the source text. Linguistic features are considered to be more appropriate because it captures different aspects of the abstract and thus provides more informative and instructive feedback [13]. Therefore, we use different linguistic features to calculate the five aspects. First, we selected 21 linguistic features that were shown to be related to the quality of abstracts to calculate their quality based on [26]. Then, we clustered the linguistic features to measure the five aspects of abstracts as shown in Table 3. The weighted sum linguistic features are used to measure these facets, where the coefficients are the corresponding correlations calculated in [26].

6 FRONT-END VISUALIZATION

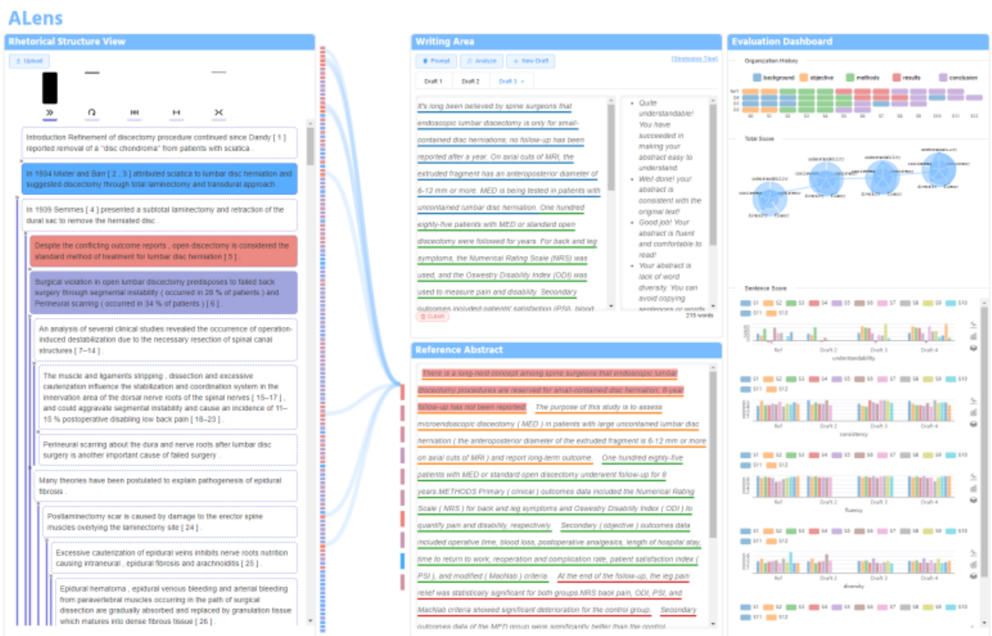

Fig. 3

Following the design principles mentioned in the formative study, we built *ALens* as a responsive web-based application to demonstrate the academic abstract writing training process. The front-end interface includes a *Rhetorical Structure View*, a *Flow Map*, a *Writing Area*, a *Reference Abstract* and an *Evaluation Dashboard*.

6.1 Rhetorical Structure View

[10]https://classroom.synonym.com/list-abstract-qualities-8671549.html
[11]https://www.brandeis.edu/writing-program/resources/students/handouts/features-of-a-good-abstract-handout.pdf
[12]https://www.abstractscorecard.com/uploads/cfp2/images/Abstract_Quality_Standards_Guidelines_13.pdf

The Rhetorical Structure View (Figure 3A) is designed to help the user grasp the hierarchy of ideas and thus quickly identify the main ideas. The rectangles (Figure 3A-a1) at the top of the view indicates the number of respective relations by their lengths. To make the rhetorical relations between sentences more intuitive, these relations are visually encoded with glyphs, as shown in Table 4. Learners can click on these glyphs (Figure 3A-a2) to hide the secondary sentences in a pair of relations. For example, by clicking on the elaboration glyph, all elaborated sentences are preserved and the font color of the supporting sentences is lightened. In this way, learners can quickly capture key information, such as the core sentences in the elaboration relation and the second sentences in a contrast relation, or they can easily check the context by clicking on the glyphs again. Also, we use a flattened tree structure (Figure 3A-a3) to show the hierarchy of ideas in the article, which is compact and makes full use of space. Sentences are wrapped as leaf nodes, logical glyphs are on the inner nodes, and the color depth of the rectangle (Figure 3A-a4) represents the number of corresponding relations in the paragraph, so users can quickly identify the core sentences of each paragraph.

Rhetorical Relation	Glyph
Background	
Contrast	
Elaboration	
Joint	
Sequence	

TABLE 4. Glyphs of Rhetorical Relations

Design Alternative. To represent the hierarchy of ideas, we initially designed the rhetorical structure tree, as shown in Figure 4. The leaf nodes are connected to sentences, and the pop-up tooltip contains the name of the relation. However, during the design iteration, users commented that the left side of the tree has a lot of white space and it is too cumbersome to move the mouse over the internal nodes to see the relations. In addition, they criticized that sentences in the article were placed side-by-side and the paragraph structure was broken, resulting in low readability. Therefore, we chose the current design to visually display the relationships and minimize the differences from the original natural text and ensure the readability.

Fig. 4. Design alternative of rhetorical structure tree.

6.2 Writing Area

The Writing Area (Figure 3B) supports basic edition functions, allowing users to write drafts manually or by clicking on the "prompt" button (Figure 3B-b1), which uses the existing summarization model to predict drafts as prompts. Before writing, a suggested abstract writing strategy (??) is provided by clicking on the "Strategies Tips" (Figure 3B-b4), enabling novices to quickly start abstract writing. After completing a draft, learners can click on the "Analyze" button (Figure 3B-b2) to analyze the abstract draft in terms of six aspects (one for organizational structure and five for linguistic features). The sentences in the writing areas are classified into five types and highlighted in different colors. At the same time, Figure 3B-b5 provides users with guided steps on how to improve the content and style of their abstracts based on the sentence classification results and the scores of linguistic features. Users can implement feedback by creating new drafts (Figure 3B-b5) to gradually improve the quality of their abstracts.

6.3 Evaluation Dashboard

The Evaluation Dashboard (Figure 3C) displays evaluation metrics at different granularities. We design the organization map (Figure 3C-c1) as a row of aligned rectangle tiles – the top row represents the organization scheme of the first draft, and the bottom row represents the latest. Each tile in the row represents a sentence in the draft, and its color encodes the type of that sentence. After writing several drafts, users are expected to find the best organization scheme, and they are anticipated to identify the writing style of a group of essays by analyzing the best organization scheme of each paper in the group. In addition to the organization scheme, the line chart (Figure 3C-c2) records the overall score of the serialized drafts, and the five linguistic features are encoded by the radar plot. However, the whole abstract and its scores for the five aspects may confuse learners [89], as they still need to recognize which parts need to be revised and which parts are already good. Therefore, (Figure 3C-c3) provides a more fine-grained analysis of the abstract. Each row represents a linguistic aspect, and a set of bars in a bar chart represents the sentences in that draft. In this way, users can identify which sentences and which aspects have not been considered and are poorly written. As a result, they can revise their drafts in a more precise and clear direction.

6.4 Reference Abstract with a Flow Map

The Reference Abstract with a Flow Map (Figure 3D&E) are designed to reveal the writing style of the reference abstract. Organization detection is applied to the reference abstract to explicitly reveal its organization scheme. Also, we use a flow map to find the most relevant sentences in the source text for each sentence in the reference abstract. We use the sentence transformer [69] to calculate the semantic similarity score of each sentence in the source text to each sentence in the reference abstract. The square tiles on each side represent a sentence from the source text or the reference. The color depth of each tile on the abstract side is computed by averaging the first k similarity scores, where k can be specified by the user. And the tiles on the source text side represent the similarity score when the user's mouse is placed over a sentence in the abstract. Meanwhile, the top k similar sentences in the source text are highlighted and linked to sentences in the reference. In this way, users can explore the writing patterns of the reference abstract. For example, they may find that the sentence in the reference abstract that describes the background may come from the end of the paragraph introducing the background in the source text. Thus, the knowledge of the abstract writing style can be constructed.

7 EVALUATION

We evaluate the effectiveness of *ALens* in two ways. First, we describe one usage scenario with one target user of *ALens*. Second, we invited 21 participants who had no exposure to our system and conducted a user study to further assess the potency of *ALens*.

7.1 Usage Scenario

In this section, we describe how Anker, a third-year undergraduate student, used *ALens* to train his academic abstract writing skills. Anker comes from the Department of Biomedical Engineering and has been starting his research career for about four months. Prior to using *ALens*, he had no experience writing academic abstracts for journals and conferences, but he did have experience writing abstracts for course essays. We chose a paper from biomedical science and gave him training in writing academic abstracts in a related field. First, he uploaded a prepared text file, and then the article was parsed into a rhetorical tree. We observed that he began to consciously select sentences as he read the article, explaining *"as far as I know, a typical abstract should introduce the purpose of the work, what problems it is trying to solve, the research methods used and the conclusions"*, and then he created a new tab to parse the copied sentences and clicked the "Analyze" button. From the conciseness and comprehensibility indicators (Figure 5) in the evaluation dashboard (Figure 3C-c3), he found that his second draft was not easy to understand and not concise enough, mainly because of the last three sentences. In addition, he found that the last three sentence were too long to read, so he broke them into shorter sentences, added some connectives, and then clicked the "Analyze" button. The result of the analysis showed that the sentences in his draft were more concise and easier to read. He then decided to look at the reference abstract. To his surprise, he found that he had missed the experimental results from the *organization history view*Figure 5. He said, *"this is the first time I know that the experimental results needed to be independent of the conclusions, which I previously thought already contained the results."* Anker concluded with a quick review of the results section of the paper, grasping the description of the combined results under the auxiliary rhetorical relations from the lengthy description of the results.

7.2 User Study

Alternative Training System To evaluate *ALens*, we built a baseline system as shown in Figure ??, We design the baseline to simulate the traditional way of abstract writing and to control the similarities and differences between *ALens* and the baseline. The baseline supports reading the article and writing abstract. We follow the practice of previous summary writing system [39, 88] to provide an overall score of abstract to users as a feedback. The keep *ALens* and the baseline system consistent with each other, both of them follow the same abstract writing and shares many functions. First, both the tools have the "Strategies Tips" button to learn general abstract writing strategies. Furthermore, the new draft button is also implemented to facilitate the abstract iteration process. The " Show Reference " which will disappear when the click it and be present after click the " New Draft " button.

Fig. 5. (a) The understandability of each sentence for the three analyzed drafts and the results for the reference are on the leftmost side. (b)The conciseness of each sentence for the three analyzed drafts and the results for the reference are on the leftmost side. Relatively short bars means sentences are relatively concise. (c) The content organization history of the three analyzed draft and the organization of the reference.

Researches about the self-regulated learning theory suggests that individual feedback can help them learn better [10]. During a learning process, self-reflection is important for effective learning which can trigger the creation of new knowledge through self-regulated learning [98]. Also, writing, as a creation process, highly depends on the engagement [41]. [41] suggests that students can write text in a more cohesive manner when they are concentrated and engaged in the writing process. Therefore, we make the following hypotheses:

- H1: In comparison to the baseline system, individual feedback help users write abstracts with more appropriate style. The style here is defined by content organization and language style.
- H2: In comparison to the baseline system, *ALens* is useful to help users construct knowledge of academic abstract writing.

- H3: In comparison to the baseline system, *ALens* increases the users' satisfaction level with the final draft of the abstract.
- H4: In comparison to the baseline system, *ALens* enables users to be more engaged in the writing process.

To test our hypotheses, we designed a laboratory experiment where participants are required to read and comprehend the given introduction of an article in computer science, write an abstract according to the introduction, learn the writing style of the reference abstract and revise their manuscript at least once. Because academic abstract writing is highly dependent with the domain and students out of the domain usually encounter barriers in reading and comprehending the article, we recruited 21 students from the computer science department in our university through social network. Our participants were randomly assigned in the experimental and control groups. The experimental group used *ALens*, while participants in the control group used the baseline system. After random assignment there are 12 students in the experimental group and 9 students in the control group. We invited them to come to our laboratory and to do the experiment on our devices which are exactly the same. Participants of the experimental group (9 males and 3 females) had an average age of 21.17 (SD=1.64) and they had experienced writing academic abstract 0.33 (SD=0.65) times on average. In the control group, there were 7 males and 2 females. They are average age was 20.89(SD=1.54) and they had written academic abstract for 0.22 (SD=0.67) times on average.

7.2.1 Design and Procedure. We designed an experiment with three stages namely a pre-test stage, a short-term abstract writing training stage and a post-test stage. In the training stage, the experimental group used *ALens* and the the control group used the alternative baseline system to read, write and learn.

Pre-test stage: To test whether our initial random grouping result in a real randomized grouping, we first test our participants from the acceptance to new information technology, feedback seeking, self-confidence and ability of academic abstract writing through a pre-survey with 16 questions. First, we asked four questions about the acceptance to new information technology of participants referring to the methods in [2]. Second, we asked them questions about the ability to to proactively seek feedback according to [8]. Meanwhile, following [8], we also test their self-confidence which aim to control the condition of Mental abilities. Fourth, we test their ability of writing academic abstract by asking their experience in academic abstract writing.

Fig. 6. The baseline system supports users reading the article, write the abstract and inspect the reference. Baseline can provide an overall score on the quality of the abstract

Short-Term abstract writing training stage:
Before the stage started, we provided a brief introduction about the usage of our system and let them play with respective system for about 5 minutes. To test whether our system enable users construct the knowledge about academic abstract writing, we develop a short-term abstract writing training stage as follows. First, participants were asked to read an introduction of computer science paper from IEEE Transactions on Visualization and Computer Graphics (TVCG) since introductions in TVCG usually contains whole information of the article. Our co-author, an experts in the domain of visualization confirmed this fact. They are required to spend at least 5 minutes reading the introduction of about 800 words to ensure that they had a basic comprehension of the article. Then they are asked to spend at least 10 minutes writing the abstract. Subsequently, users can inspect the reference abstract and learn the writing style of that abstract for at least 5 minutes. Then experimental group was using *ALens* to check the organization of the abstract and the locations of core sentences while students in the control group can analyze the abstract based on their knowledge. Then they are required to revised their first draft based on the writing knowledge learned from the reference abstract and according to the evaluation metrics if their are in the experiment group. Finally, they could progressively embellish the abstract until they were satisfied with the draft. All the drafts during the training process were collected by both systems and sent to later assessment.

Post- test stage: In this stage, we first measure user's intention to use our system, usability and usefulness of our system following the technology acceptance test [86]. In addition, we measure users' satisfaction level about their first draft and last draft both group and measure the perceived engagement since writing is creation process where engagement is an important component [41]. Example items for the five constructs are : *"If the system is released, I will use the system to train abstract writing","I can write abstract with appropriate style.","I feel like I can learn to use the system quickly","I am satisfied with the first draft written using the system ,"I was focusing on writing itself, time passed quickly for me"* We use a 1- to 7-point Likert scale (7: very sure, 1:not very sure, with 4 being a neutral statement) for participants to assess.

7.2.2 *Measurement of Abstract Quality.* Technology acceptance, Users' satisfaction level and engagement were used to evaluated the system from the subject perspective of users and test the hypotheses H3 and H4. In addition, we test the hypotheses H1 and H2 by measuring the quality of abstracts from both group. We measured the quality of the drafts from two aspects; 1)the perceived quality and 2) the formal quality.

Perceived quality: Since we adopt as paper in the domain of visualization as a training material for user study, we invite two senior graduate students in that domain to help us assess the abstract on 1- to 7-point Likert scale (7: very good, 1:very bad). Both of them has an research experience of 3 years. Our goal was to subjectively judge the quality of the abstract leveraging the reading experience of experts in that domain. We take the average their scores as the final score of the draft.

Formal quality: We analyze the first drafts and the last drafts for formal quality. We define formal quality of an abstract as following seven aspects: content completeness, content arrangement, understansdability, consistency, fluency, diversity and conciseness. As we have mentioned before, the seven aspects are refined from a bunch of articles and course websites. First, we create a criterion in terms of the seven aspects on 1- to 7-point Likert scale (7: very good, 1:very bad). We annotated the 42 drafts by ourselves and take the average our scores as the final score of the draft on that aspects.

Pre-test Stage					
Group	New technology acceptance	Self-confidence	Feedback seeking	Times for writing abstract	
Mean ALens	5.83	5.42	5.25	0.33	
Mean Baseline	5.89	5.33	5.00	0.22	
SD ALens	0.58	0.90	0.87	0.65	
SD Baseline	0.60	0.87	1.00	0.67	
Double-sided t	-0.21	0.21	0.60	0.38	
p-Value	0.834	0.833	0.557	0.707	
Training Stage					
Group	First Draft Formal quality	First Draft Perceived quality	Second Draft Formal quality	Second Draft Perceived quality	
Mean ALens	4.08	4.50	5.39	5.58	
Mean Baseline	4.22	4.44	4.56	4.78	
SD ALens	0.74	0.90	0.99	0.79	
SD Baseline	1.15	0.73	0.74	0.67	
Double-sided t	-0.32	0.16	2.22	2.52	
p-Value	0.757	0.878	0.039	0.021	
Post-test Stage					
Group	Technology acceptance	First Draft Satisfaction	Seconda Draft Satisfaction	Engagemnt	Knowledge Construction
Mean ALens	4.76	3.50	5.33	4.33	5.33
Mean Baseline	3.78	3.67	4.67	4.44	3.78
SD ALens	0.68	1.09	0.49	0.49	0.98
SD Baseline	0.69	0.71	0.71	0.53	0.67
Double-sided t	3.24	-0.42	2.42	-0.49	4.31
p-Value	0.004	0.676	0.030	0.629	0.000

TABLE 5. Statistic Analysis Results of *ALens* and the baseline system on a 1-7 Likert Scale (1:low, 7:high)

8 EVALUATION AND RESULTS

To evaluate our 4 hypotheses and the technology acceptance, we aim to answer the following research questions (RQ):

RQ1: Did users find *ALens* to be useful for academic abstract writing and want to use it in the future?

RQ2: Did *ALens* help with users to write abstract with more appropriate style?

RQ3: Did users gain some knowledge about academic abstract writing?

RQ4: Are users perceived to be more engaged with the writing process?

RQ5: Are users more satisfied with their last draft compared to the baseline system?

To ensure our random assignment is successful and control the potential effects of the small sample size, we compared the difference of both group from the four aspects as shown in the pre-test stage in Table 5. The p value for the four constructs are all larger than 0.05 which guarantees there are no significant difference between both groups in terms of these four constructs.

Technology acceptance To compare the technology acceptance, we average the Likert score of perceived usefulness, intention to use and the usability. From the column of technology acceptance, we can find the that *ALens* received higher acceptance than the baseline system. Technology acceptance for learning tools is a critical base for users to make further learning on it. The positive technology acceptance provide a promising results to use this tool as an adaptive feddback application.

Writing with more appropriate style To test whether users write abstracts with more appropriate style, we collected 42 drafts of abstracts written by participants and evaluated their formal quality and perceived quality. From Table 5, we can see that there is no significance between the first drafts in two groups. However, there is statistically difference between the second draft of two groups, which demonstrate the effectiveness of *ALens* to train users to write abstracts with more appropriate style. Thus H1 is accepted.

Abstract writing knowledge construction To test whether users whether acquire some knowledge, we asked them three quantitative problems and one qualitative problem. Both results Table 5 shows the promising results that users can acquire some knowledge about abstract writing. Thus H2 is accepted.

Increased satisfaction level The results in post-test stage shows that both *ALens* and alternative tools enable users to increase the satisfaction level about their drafts. Such is human nature that iteration usually makes things better. However, the difference of satisfaction levels between two groups becomes more significant on the second draft which means that *ALens* can greatly increase their satisfaction level. Thus H3 is accepted.

Higher engagement in the writing process The statistic results shows that there is no significance between the two groups in terms of engagement. Thus H4 is rejected. We think that users my feel distracted during the long reading and writing process and the interaction time for learning the writing style and analyze their draft is relatively shorter.

Qualitative feedback We also included open questions in our survey to receive some improvement suggestions. For example, we asked *which part of the system is needed to be improvedWhy?* Overall. most of them hold positive attitude to *ALens*, in particular, the flow map, the evaluation at sentence level, " Strategies Tips", the sentence classification function and the "Prompt" function. However, some participant also give constructive suggestions. Generally, they complain about the mislead caused by rhetorical tree, the wrong sentence classification results and the confused score in evaluation deshboard.

9 DISCUSSION AND LIMITATION

Our user study shows that the adaptive formative feedback on students' abstract draft helps them t write abstracts with more appropriate style including content organization and language style, We verify this from both the formal quality and the perceived quality compared to the alternative tools. The last drafts in both groups are of higher quality than the initial one. We believe that the self-regulated learning theory can explain this effect. However, when comparing the quality level between two groups of the same draft, we found the difference between the drafts in the same batch increases significantly. It is important to keep in mind that the right portion and granularity of feedback is important for learners to absorb and self-regulate [10]. In comparison to the alternative tool, *ALens* provide different levels of individual feedback thus motivate students to change behavior. The proven short-term progression on academic abstract writing in the user study indicates that self-regulation motivates participants to learn the writing style and construct relevant knowledge. In addition, to make our research successfully applied to the actual abstract writing training scenario, we verify the technology acceptance of our system which provides promising results. Hence, our work has several contributions to current research. First, we conduct a formative study to understand the problems they encountered in the academic abstract writing process, on which we derive relevant design requirements. To the best of our knowledge, *ALens* is one of the first research to provide a verified design requirements on the adaptive learning tools for academic abstract writing. We believe that our work can provide some inspirations for those want develop training tools on meta cognition skills. Instructors and tool developers can now borrow our design retirements and findings to develop their own training tools for academic abstract writing.

There are three constructive suggestions from the user study. Fist, 4 participants show concerns about mislead caused by the rhetorical tree owing to its low accuracy on relations like " joint " and " sequence ". Meanwhile, 5 participants complains that sentence classification model can not effectively recognize the "objective" and "background". Here, we forward two options to improve the performance of our RST model and sentence classification model by increasing diversity and quantity of the training data. We will follow the methods describe in [47], first train our RST parser on automatic annotated data and fine tune it on RST-DT [21]. To improve the accuracy of sentence classification, we will use a similar Data Augmentation methods. We first train our bert model on a automatic annotated data and then tuned on domain-specific dataset. In terms of the confused score in evaluation dashboard, we want to improve the understanding of the scores by providing annotated paradigms and explanations to clarify the underlying meaning of the score and compile the the action step guidelines to achieve a higher score. In addition, we consider to try transformer-based models to predict more accurate scores.

In the user study, we verify the short-term positive impact of *ALens* on participants' academic abstract writing. However, the long-term learning effects are still needed to be verified. Therefore, we decide to conduct a field experiment to test the effectiveness and acceptance. We set up two groups and the control group that will not receive any feedback except the guidance from their supervisors and the experimental group that will receive feedbackk from both their supervisors and *ALens*. Finally, an evaluated adaptive learning tool is anticipated to be forwarded for research novices to practice academic abstract writing.

10 CONCLUSION AND FUTURE WORK

REFERENCES

[1] Asad Abdi, Norisma Idris, Rasim M Alguliyev, and Ramiz M Aliguliyev. 2016. An automated summarization assessment algorithm for identifying summarizing strategies. *PloS one* 11, 1 (2016), e0145809.

[2] Ritu Agarwal and Elena Karahanna. 2000. Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS quarterly* (2000), 665–694.

[3] JoAnn Grif Alspach. 2017. Writing for publication 101: Why the abstract is so important. , 12–15 pages.

[4] Chittaranjan Andrade. 2011. How to write a good abstract for a scientific paper or conference presentation. *Indian journal of psychiatry* 53, 2 (2011), 172.

[5] Linda Argote, Paul Ingram, John M Levine, and Richard L Moreland. 2000. Knowledge transfer in organizations: Learning from the experience of others. *Organizational behavior and human decision processes* 82, 1 (2000), 1–8.

[6] Daniel King Bhavana Dalvi Dan Weld Arman Cohan, Iz Beltagy. 2019. Pretrained Language Models for Sequential Sentence Classification. In *EMNLP*.

[7] Natasha Artemeva. 2008. Toward a unified social theory of genre learning. *Journal of business and technical communication* 22, 2 (2008), 160–185.

[8] Susan J Ashford. 1986. Feedback-seeking in individual adaptation: A resource perspective. *Academy of Management journal* 29, 3 (1986), 465–487.

[9] Anis Ashrafzadeh and Vahid Nimehchisalem. 2015. Vocabulary Knowledge: Malaysian Tertiary Level Learners' Major Problem in Summary Writing. *Journal of Language Teaching & Research* 6, 2 (2015).

[10] A Bandura. 1997. Social cognitive theory of self-regulation organizational behavior. (1997).

[11] Jasmijn Bastings and Katja Filippova. 2020. The elephant in the interpretability room: Why use attention as explanation when we have saliency methods? *arXiv preprint arXiv:2010.05607* (2020).

[12] Diane Dewhurst Belcher and George Braine. 1995. *Academic writing in a second language: Essays on research and pedagogy*. Greenwood Publishing Group.

[13] Elizabeth Bernhardt. 2010. *Understanding advanced second-language reading*. Routledge.

[14] Robert A Bjork, John Dunlosky, and Nate Kornell. 2013. Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of psychology* 64 (2013), 417–444.

[15] Paul Black and Dylan Wiliam. 2009. Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)* 21, 1 (2009), 5–31.

[16] Florian Böhm, Yang Gao, Christian M Meyer, Ori Shapira, Ido Dagan, and Iryna Gurevych. 2019. Better rewards yield better summaries: Learning to summarise without references. *arXiv preprint arXiv:1909.01214* (2019).

[17] Ann L Brown, Joseph C Campione, and Jeanne D Day. 1981. Learning to learn: On training students to learn from texts. *Educational researcher* 10, 2 (1981), 14–21.

[18] Ann L Brown and Jeanne D Day. 1983. Macrorules for summarizing texts: The development of expertise. *Journal of verbal learning and verbal behavior* 22, 1 (1983), 1–14.

[19] Jochen WL Cals and Daniel Kotz. 2013. Effective writing and publishing scientific papers, part II: title and abstract. *Journal of clinical epidemiology* 66, 6 (2013), 585.

[20] Ziqiang Cao, Furu Wei, Wenjie Li, and Sujian Li. 2018. Faithful to the original: Fact aware neural abstractive summarization. In *thirty-second AAAI conference on artificial intelligence*.

[21] Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2003. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Current and new directions in discourse and dialogue*. Springer, 85–112.

[22] Chiou Sheng Chew, Wen-Chi Vivian Wu, Norisma Idris, Er Fu Loh, and Yan Piaw Chua. 2020. Enhancing summary writing of ESL learners via a theory-based online tool: system development and Evaluation. *Journal of Educational Computing Research* 58, 2 (2020), 398–432.

[23] Shao Joyce Chin. 2016. Investigating the summary writing performance of university students in Taiwan. In *20th conference of English teaching and learning in the ROC*.

[24] Colin B. Clement, Matthew Bierbaum, Kevin P. O'Keeffe, and Alexander A. Alemi. 2019. On the Use of ArXiv as a Dataset. arXiv:1905.00075 [cs.IR]

[25] Arman Cohan, Franck Dernoncourt, Doo Soon Kim, Trung Bui, Seokhwan Kim, Walter Chang, and Nazli Goharian. 2018. A discourse-aware attention model for abstractive summarization of long documents. *arXiv preprint arXiv:1804.05685* (2018).

[26] Scott A Crossley, Minkyung Kim, Laura Allen, and Danielle McNamara. 2019. Automated summarization evaluation (ASE) using natural language processing tools. In *International Conference on Artificial Intelligence in Education*. Springer, 84–95.

[27] Scott A Crossley, Kristopher Kyle, and Danielle S McNamara. 2016. The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior research methods* 48, 4 (2016), 1227–1237.

[28] Franck Dernoncourt and Ji Young Lee. 2017. Pubmed 200k rct: a dataset for sequential sentence classification in medical abstracts. *arXiv preprint arXiv:1710.06071* (2017).

[29] Michael Derntl. 2014. Basics of research paper writing and publishing. *Int. J. Technology Enhanced Learning* 6, 2 (2014), 105.

[30] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[31] Dylan B Dryer. 2013. Scaling writing ability: A corpus-driven inquiry. *Written Communication* 30, 1 (2013), 3–35.

[32] Wafaa S El-Kassas, Cherif R Salama, Ahmed A Rafea, and Hoda K Mohamed. 2021. Automatic text summarization: A comprehensive survey. *Expert Systems with Applications* 165 (2021), 113679.

[33] Muhammad Fareed, Almas Ashraf, and Muhammad Bilal. 2016. ESL Learners' Writing Skills: Problems, Factors and Suggestions. *Journal of Education and Social Sciences* 4 (10 2016), 81–92. https://doi.org/10.20547/jess0421604201

[34] Organisation for Economic Co-operation and Development (OECD). 2018. The future of education and skills: Education 2030. *OECD Education Working Papers* (2018).

[35] Nancy Frey, Douglas Fisher, and Ted Hernandez. 2003. What's the gist? Summary writing for struggling adolescent writers. *Voices from the Middle* 11, 2 (2003), 43–49.

[36] Mandy Guo, Joshua Ainslie, David Uthus, Santiago Ontanon, Jianmo Ni, Yun-Hsuan Sung, and Yinfei Yang. 2021. Longt5: Efficient text-to-text transformer for long sequences. *arXiv preprint arXiv:2112.07916* (2021).

[37] Som Gupta and Sanjai Kumar Gupta. 2019. Abstractive summarization: An overview of the state of the art. *Expert Systems with Applications* 121 (2019), 49–65.

[38] John Hattie and Helen Timperley. 2007. The power of feedback. *Review of educational research* 77, 1 (2007), 81–112.

[39] Yulan He, Siu Cheung Hui, and Tho Thanh Quan. 2009. Automatic summary assessment for intelligent tutoring systems. *Computers & Education* 53, 3 (2009), 890–899.

[40] Michael Heilman and Kenji Sagae. 2015. Fast rhetorical structure theory discourse parsing. *arXiv preprint arXiv:1505.02425* (2015).

[41] Alan Hirvela and Qian Du. 2013. "Why am I paraphrasing?": Undergraduate ESL writers' engagement with source-based academic writing and reading. *Journal of English for Academic Purposes* 12, 2 (2013), 87–98.

[42] Tasos Hovardas, Olia E Tsivitanidou, and Zacharias C Zacharia. 2014. Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education* 71 (2014), 133–152.

[43] Norisma Idris, Sapiyan Baba, and Rukaini Abdullah. 2011. Identifying students' summary writing strategies using summary sentence decomposition algorithm. *Malaysian Journal of Computer Science* 24, 4 (2011), 180–194.

[44] Mehrdad Jalalian. 2012. Writing an eye-catching and evocative abstract for a research article: A practical approach. *Electronic Physician* 4, 3 (2012), 520–524.

[45] Ruipeng Jia, Yanan Cao, Hengzhu Tang, Fang Fang, Cong Cao, and Shi Wang. 2020. Neural extractive summarization with hierarchical attentive heterogeneous graph network. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 3622–3631.

[46] Margaret R Kirkland and Mary Anne P Saunders. 1991. Maximizing student performance in summary writing: Managing cognitive load. *Tesol Quarterly* 25, 1 (1991), 105–121.

[47] Naoki Kobayashi, Tsutomu Hirao, Hidetaka Kamigaito, Manabu Okumura, and Masaaki Nagata. 2021. Improving neural RST parsing model with silver agreement subtrees. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 1600–1612.

[48] Grzegorz Kondrak. 2005. N-gram similarity and distance. In *International symposium on string processing and information retrieval*. Springer, 115–126.

[49] Wojciech Kryściński, Bryan McCann, Caiming Xiong, and Richard Socher. 2019. Evaluating the factual consistency of abstractive text summarization. *arXiv preprint arXiv:1910.12840* (2019).

[50] Kristopher Kyle. 2016. Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication. (2016).

[51] Kristopher Kyle, Scott Crossley, and Cynthia Berger. 2018. The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior research methods* 50, 3 (2018), 1030–1046.

[52] Thomas K Landauer, Danielle S McNamara, Simon Dennis, and Walter Kintsch. 2013. *Handbook of latent semantic analysis*. Psychology Press.

[53] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461* (2019).

[54] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*. 74–81.

[55] Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. *arXiv preprint arXiv:1908.08345* (2019).

[56] Jennifer A Livingston. 2003. Metacognition: An Overview. (2003).

[57] Annie Louis and Ani Nenkova. 2013. Automatically assessing machine summary content without a gold standard. *Computational Linguistics* 39, 2 (2013), 267–300.

[58] William C Mann and Sandra A Thompson. 1987. *Rhetorical structure theory: A theory of text organization*. University of Southern California, Information Sciences Institute Los Angeles.

[59] William C Mann and Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse* 8, 3 (1988), 243–281.

[60] Yuning Mao, Ming Zhong, and Jiawei Han. 2022. CiteSum: Citation Text-guided Scientific Extreme Summarization and Low-resource Domain Adaptation. *arXiv preprint arXiv:2205.06207* (2022).

[61] Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. *arXiv preprint arXiv:2005.00661* (2020).

[62] Philipp Mayring. 2015. Qualitative content analysis: Theoretical background and procedures. In *Approaches to qualitative research in mathematics education*. Springer, 365–380.

[63] Rui Meng, Khushboo Thaker, Lei Zhang, Yue Dong, Xingdi Yuan, Tong Wang, and Daqing He. 2021. Bringing structure into summaries: a faceted summarization dataset for long scientific documents. *arXiv preprint arXiv:2106.00130* (2021).

[64] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*. 404–411.

[65] Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Thirty-first AAAI conference on artificial intelligence*.

[66] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 311–318.

[67] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2019. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *arXiv e-prints* (2019). arXiv:1910.10683

[68] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* 21, 140 (2020), 1–67.

[69] Nils Reimers and Iryna Gurevych. 2020. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://arxiv.org/abs/2004.09813

[70] Henry L Roediger III and Jeffrey D Karpicke. 2006. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science* 17, 3 (2006), 249–255.

[71] D Royce Sadler. 1989. Formative assessment and the design of instructional systems. *Instructional science* 18, 2 (1989), 119–144.

[72] Abigail See. 2021. *Neural Generation of Open-Ended Text and Dialogue*. Stanford University.

[73] Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368* (2017).

[74] Arie S Spirgel and Peter F Delaney. 2016. Does writing summaries improve memory for text? *Educational Psychology Review* 28, 1 (2016), 171–196.

[75] Hendrik Strobelt, Sebastian Gehrmann, Michael Behrisch, Adam Perer, Hanspeter Pfister, and Alexander M Rush. 2018. S eq 2s eq-v is: A visual debugging tool for sequence-to-sequence models. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 353–363.

[76] Yao-Ting Sung, Chia-Ning Liao, Tao-Hsing Chang, Chia-Lin Chen, and Kuo-En Chang. 2016. The effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. *Computers & Education* 95 (2016), 1–18.

[77] Thanatkun Tangpermpoon. 2008. Integrated approaches to improve students writing skills for English major students. *ABAC journal* 28, 2 (2008).

[78] MS Tullu and S Karande. 2017. Writing a model research paper: A roadmap. *Journal of postgraduate medicine* 63, 3 (2017), 143.

[79] Milind S Tullu. 2019. Writing the title and abstract for a research paper: Being concise, precise, and meticulous is the key. *Saudi journal of anaesthesia* 13, Suppl 1 (2019), S12.

[80] Oleg Vasilyev and John Bohannon. 2020. Is human scoring the best criteria for summary evaluation? *arXiv preprint arXiv:2012.14602* (2020).

[81] Oleg Vasilyev, Vedant Dharnidharka, and John Bohannon. 2020. Fill in the BLANC: Human-free quality estimation of document summaries. *arXiv preprint arXiv:2002.09836* (2020).

[82] Viswanath Venkatesh and Hillol Bala. 2008. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences* 39, 2 (2008), 273–315.

[83] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS quarterly* (2003), 425–478.

[84] Jesse Vig. 2019. A multiscale visualization of attention in the transformer model. *arXiv preprint arXiv:1906.05714* (2019).

[85] Jesse Vig and Yonatan Belinkov. 2019. Analyzing the structure of attention in a transformer language model. *arXiv preprint arXiv:1906.04284* (2019).

[86] Jan Vom Brocke, Wolfgang Maaß, Peter Buxmann, Alexander Maedche, Jan Marco Leimeister, and Günter Pecht. 2018. Future work and enterprise systems. *Business & Information Systems Engineering* 60, 4 (2018), 357–366.

[87] 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 (2015), 9.

[88] David Wade-Stein and Eileen Kintsch. 2004. Summary Street: Interactive computer support for writing. *Cognition and instruction* 22, 3 (2004), 333–362.

[89] Sara Cushing Weigle. 2002. *Assessing writing*. Cambridge University Press.

[90] Lijun Wu, Juntao Li, Yue Wang, Qi Meng, Tao Qin, Wei Chen, Min Zhang, Tie-Yan Liu, et al. 2021. R-drop: Regularized dropout for neural networks. *Advances in Neural Information Processing Systems* 34 (2021), 10890–10905.

[91] Stratos Xenouleas, Prodromos Malakasiotis, Marianna Apidianaki, and Ion Androutsopoulos. 2019. Sumqe: a bert-based summary quality estimation model. *arXiv preprint arXiv:1909.00578* (2019).

[92] Yu-Fen Yang. 2016. Transforming and constructing academic knowledge through online peer feedback in summary writing. *Computer Assisted Language Learning* 29, 4 (2016), 683–702.

[93] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International Conference on Machine Learning*. PMLR, 11328–11339.

[94] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675* (2019).

[95] Yue Zhang and Stephen Clark. 2011. Syntactic processing using the generalized perceptron and beam search. *Computational linguistics* 37, 1 (2011), 105–151.

[96] Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. 2020. Extractive summarization as text matching. *arXiv preprint arXiv:2004.08795* (2020).

[97] Wei Zhu. 2004. Faculty views on the importance of writing, the nature of academic writing, and teaching and responding to writing in the disciplines. *Journal of second language Writing* 13, 1 (2004), 29–48.

[98] Barry J Zimmerman and Dale H Schunk. 2001. *Self-regulated learning and academic achievement: Theoretical perspectives*. Routledge.

Abstract

ORIGINALITY REPORT

9%	7%	7%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	arxiv.org Internet Source	1%
2	Submitted to Free University of Bolzano Student Paper	1%
3	www.alexandria.unisg.ch Internet Source	1%
4	Jeongyeon Kim, Yubin Choi, Meng Xia, Juho Kim. "Mobile-Friendly Content Design for MOOCs: Challenges, Requirements, and Design Opportunities", CHI Conference on Human Factors in Computing Systems, 2022 Publication	<1%
5	Laixin Xie, Ziming Wu, Peng Xu, Wei Li, Xiaojuan Ma, Quan Li. "RoleSeer: Understanding Informal Social Role Changes in MMORPGs via Visual Analytics", CHI Conference on Human Factors in Computing Systems, 2022 Publication	<1%
6	link.springer.com Internet Source	<1%

7	"ECAI 2020", IOS Press, 2020 Publication	<1%
8	Makiko Kato. "Examining the Dependability and Practicality of Analytic Rubric of Summary Writing Using Multivariate Generalizability Theory: Focusing on Japanese University Students with Lower-Intermediate Proficiency in English", English Language Teaching, 2022 Publication	<1%
9	aclanthology.org Internet Source	<1%
10	cidecuador.org Internet Source	<1%
11	"Natural Language Processing and Chinese Computing", Springer Science and Business Media LLC, 2020 Publication	<1%
12	Submitted to Bath Spa University College Student Paper	<1%
13	"Intelligent Computing", Springer Science and Business Media LLC, 2022 Publication	<1%
14	Nagsen Komwad, Paras Tiwari, Banoth Praveen, C. Ravindranath Chowdary. "A survey on review summarization and	<1%

sentiment classification", Knowledge and Information Systems, 2022
Publication

15	Thiemo Wambsganss, Christina Niklaus, Matthias Cetto, Matthias Söllner, Siegfried Handschuh, Jan Marco Leimeister. "AL: An Adaptive Learning Support System for Argumentation Skills", Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020 Publication	<1%
16	academypublication.com Internet Source	<1%
17	edoc.pub Internet Source	<1%
18	saudija.org Internet Source	<1%
19	Chiou Sheng Chew, Wen-Chi Vivian Wu, Norisma Idris, Er Fu Loh, Yan Piaw Chua. "Enhancing Summary Writing of ESL Learners via a Theory-Based Online Tool: System Development and Evaluation", Journal of Educational Computing Research, 2019 Publication	<1%
20	Md Tahmid Rahman Laskar, Enamul Hoque, Jimmy Xiangji Huang. "Domain Adaptation with Pre-trained Transformers for Query	<1%

Focused Abstractive Text Summarization", Computational Linguistics, 2022
Publication

21 Ning Ye, Fan Zhi-Ping. "An Empirical Study on the Context of Tacit Knowledge Transfer in Knowledge Alliances", 2006 International Conference on Service Systems and Service Management, 2006 <1%
Publication

22 Som Gupta, S. K Gupta. "Abstractive summarization: An overview of the state of the art", Expert Systems with Applications, 2019 <1%
Publication

23 Www.Journalcra.Com <1%
Internet Source

24 "Impacts of highly automated vehicles on urban passenger transport and tourism", Corvinus University of Budapest, 2022 <1%
Publication

25 ebin.pub <1%
Internet Source

26 icls2020.org <1%
Internet Source

27 ir.canterbury.ac.nz <1%
Internet Source

28	pure.uva.nl Internet Source	<1 %
29	researcharchive.vuw.ac.nz Internet Source	<1 %
30	studentsrepo.um.edu.my Internet Source	<1 %
31	vgupta123.github.io Internet Source	<1 %
32	www.ifets.info Internet Source	<1 %
33	Ameena Hassan, Akhtar Abbas. "Chapter 13 The Impact of Written Corrective Feedback on Second Language Composition in English", Springer Science and Business Media LLC, 2022 Publication	<1 %
34	Charlene Polio, Hyung‐Jo Yoon. "The reliability and validity of automated tools for examining variation in syntactic complexity across genres", International Journal of Applied Linguistics, 2018 Publication	<1 %
35	Fath U Min Ullah, Mohammad S. Obaidat, Amin Ullah, Khan Muhammad, Mohammad Hijji, Sung Wook Baik. "A Comprehensive Review on Vision-based Violence Detection in	<1 %

Surveillance Videos", ACM Computing Surveys, 2022
Publication

36	Ilia Kuznetsov, Jan Buchmann, Max Eichler, Iryna Gurevych. "Revise and Resubmit: An Intertextual Model of Text-based Collaboration in Peer Review", Computational Linguistics, 2022 Publication	<1 %
37	Satyajit Ghosh, Mousumi Dutta, Tanaya Das. "Indian Legal Text Summarization: A Text Normalisation-based Approach", Institute of Electrical and Electronics Engineers (IEEE), 2022 Publication	<1 %
38	Wenchao Li, Chun Zhang, Yang yang Wang, Lifei Xiao, Yan Feng, Xianhao Huo, Chaofan Wang, Yu Sun, Feng Wang, Tao Sun. "Alterations of RNAs in the insula related to cocaine-induced condition place preference in adolescent mice", Biochemical and Biophysical Research Communications, 2022 Publication	<1 %
39	classroom.synonym.com Internet Source	<1 %
40	core.ac.uk Internet Source	<1 %

41	deepai.org Internet Source	<1 %
42	digital.lib.washington.edu Internet Source	<1 %
43	e-space.mmu.ac.uk Internet Source	<1 %
44	espace.curtin.edu.au Internet Source	<1 %
45	export.arxiv.org Internet Source	<1 %
46	mjltm.org Internet Source	<1 %
47	perer.org Internet Source	<1 %
48	researchspace.auckland.ac.nz Internet Source	<1 %
49	scholarsmine.mst.edu Internet Source	<1 %
50	so04.tci-thaijo.org Internet Source	<1 %
51	www.aclweb.org Internet Source	<1 %
52	www.addblueoff-tr.com Internet Source	<1 %

53	www.statmt.org Internet Source	<1%
54	"Natural Language Processing and Information Systems", Springer Science and Business Media LLC, 2019 Publication	<1%
55	Krisztina Károly. "Aspects of Cohesion and Coherence in Translation", John Benjamins Publishing Company, 2017 Publication	<1%
56	Thiemo Wambsganss. "Designing Adaptive Argumentation Learning Systems Based on Artificial Intelligence", Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, 2021 Publication	<1%
57	"Artificial Intelligence in Education", Springer Science and Business Media LLC, 2019 Publication	<1%
58	Asad Abdi, Siti Mariyam Shamsuddin, Shafaatunnur Hasan, Jalil Piran. "Machine learning-based multi-documents sentiment-oriented summarization using linguistic treatment", Expert Systems with Applications, 2018 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		