École Polytechnique Fédérale de Lausanne

Understanding Revision Behavior in Adaptive Writing Support Systems for Education

Machine Learning For Education Lab

Luca Mouchel

6th January 2023

Abstract

Revision behavior in adaptive writing support systems is an important and relatively new area of research that can help improve the design and effectiveness of these tools for students' self-regulated learning. Understanding how these tools are used is key to improving them to better support learners in their writing and learning processes. In this paper, we present a novel pipeline and insights to analyze revision behavior of students at scale. We leverage a data set of two students groups using an adaptive writing support tool in an educational setting. With our novel pipeline, we show that the tool was effective in promoting revision among the learners. Depending on the writing feedback, we were able to analyze different strategies of learners when revising their texts. However, we also found that users of the exemplary case are far less engaged over time, regardless of the group. Our research contributes a pipeline for measuring self-regulated learning behaviors at scale in writing tasks (i.e., engagement or revision behavior) and informs the design of future adaptive writing support systems for education, with the goal of enhancing their effectiveness in supporting student writing.

This semester project is in collaboration with Thiemo Wambsganss, with whom we plan to submit our findings at the Educational Data Mining Conference or the Artificial Intelligence In Education Conference.

Chapter 1

Introduction

Intelligent writing support tools (e.g., Grammarly, WordTune or Quilbot) offer new ways for learners to receive feedback and thus revise their texts (Nova 2018). These and other writing support systems bear the potential to provide learners with needed adaptive feedback on their writing exercises when educators are not present, (e.g., on grammatical mistakes (Yuan 2017), argumentation (Wambsganss, Janson and Leimeister 2022), or general persuasive writing (Wambsganss, Caines and Buttery 2022)). They can help students in their self-regulated learning process (Fuente et al. 2022; Zimmerman 2002), e.g, to organize their thoughts and ideas, reflect on their learnings, or simply to receive feedback on frequently occurring grammar or argumentation mistakes.

The backbone of these tools is Large Language Models such as GPT-3 (M. Lee, Liang and Yang 2022a) or ChatGPT that can generate a wide variety of prose and dialogues with an unprecedented level of fluency out of the box. Through fine-tuning, these models can further become specialized at particular tasks, such as composing emails (Buschek, Zürn and Eiband 2021) or providing health consultations (L. Wang et al. 2021). While most of these tools and systems are driven by researchers from NLP and ML, especially from an educational perspective, it is important to understand how these tools are used by learners in educational settings and how they improve the effectiveness of educational scenarios (Chen, Breslow and DeBoer 2018; Mejia-Domenzain et al. 2022). In fact, automatic insights into the self-regulated learning behavior of students when using writing support tools for learning tasks and insights into their usage can bear the foundation to improve the effectiveness of the design and embedding of such tools in educational scenarios (Quintana et al. 2006). However, research in educational data mining on the writing and revision behavior of learners with intelligent writing tools is rather scarce. To the best of our knowledge, no publicly available pipeline exists that focuses on processing the keystroke behavior of learners and helps to analyze self-regulated learning characteristics such as engagement, revision, or visualizing the learning path.

In the context of writing support systems, self-regulated learning can be supported through features such as setting writing goals, tracking progress towards those goals, and providing tools and resources to help writers identify and address areas for improvement in their writing (Zhu, Liu and H.-S. Lee 2020). A writing support system might allow writers to set goals for the number of words they aim to write each day and provide tools for tracking progress towards those goals. It might also provide resources such as grammar and style guides, or writing prompts and exercises to help writers improve their skills (Wong et al. 2021). By providing students with the tools and resources they need to take control of their own learning and development, writing support systems can help students become more self-regulated learners and improve their writing skills over time. Students who lack self-regulated learning/writing capacity tend to struggle more at the initial stages of writing tasks (Bowen, Thomas and Vandermeulen 2022).

Present research is largely focused on designing and building writing support systems[1] (Coenen et al. 2021; Padmakumar and He 2021). However, there are not many insights into the effects of the usage of these tools and their impact on students' self-regulated learning processes (Bjork, Dunlosky and Kornell 2013; Vandermeulen, Leijten and Van Waes 2020), which is why we contribute a novel pipeline analysing and visualising revision behavior to better understand how we can design, develop and improve existing systems to better support students.

A solution to understand revision behavior and explain self-regulated learning bears techniques from the field of data mining. One such technique is keystroke logging. Keystroke logging allows us to use educational data mining to analyze user behavior in writing tasks (Leijten and Van Waes 2013; Zhu, Mo Zhang and Deane 2019; Sinharay, Mo Zhang and Deane 2019). In fact, tools such as InputLog (Leijten and Van Waes 2013) or ETS (Zhu, Mo Zhang and Deane 2019) are used to track user activity. One advantage of keystroke logging compared to other data collection methods is that the mechanism is less intrusive. The data is collected in the background without interfering with the writer's performance or thinking process (Leijten and Van Waes 2013). With this study, we apply data mining techniques to model, inspect, and analyze quantitative data in learners' writing interaction through keystroke logging. In recent years, observing writers via keystroke logging has become more and more popular and the possibilities to analyze the resulting logging data have increased rapidly (Vandermeulen, Leijten and Van Waes 2020). With inspiration from Zhu, Mo Zhang and Deane 2019, we contribute to this research space by providing a novel pipeline analysing keystroke logs.

The aim of this study is to create and provide a pipeline to analyze and visualize the revision and writing behavior of users in a writing task. Past research has evaluated the differences in self-regulated learning and revision behavior between users in different groups (Mejia-Domenzain et al. 2022; Bowen, Thomas and Vandermeulen 2022; Zhu, Mo Zhang and Deane 2019) with different criteria. We use a keystroke log from an experiment, where users were divided into two groups. The first one was given adaptive feedback and the second one was not. We intend to first identify and visualize the differences between these two groups in their revision process and compare different user profiles and measure their engagement over time. We use an exemplary data set to build this pipeline and apply data mining in order to gain insights into the underlying process of this writing activity[2].

[1] e.g *In2Writing*, a workshop dedicated to enhancing and understanding writing assistants

[2] The source code for this research will be made publicly available upon publication of the paper. The code will be hosted on a publicly accessible repository on GitHub and will be released under an open-source license to facilitate reproducibility and further research.

Chapter 2

Related Works

2.1 Research on Automatic Data Mining for Writing Behaviour

Research in writing process analysis can be traced to the 1970s (Emig 1971; Stallard 1974). However, only more recently have studies been focusing theoretically on behavioral and cognitive processes of writing (John R Hayes 2012; McCutchen 1996). In fact, Flower and Hayes (Flower and John R. Hayes 1981) laid the groundwork for research on the psychology of writing. They propose that the act of writing is propelled by goals, which are created by the writer and grow in number as the writing progresses. Today, writing support tools need to support this cognitive process as it emphasizes writers' intentions, rather than their actions (Gero et al. 2022). It is important to understand what these tools help with, and how we may design new ones (Gero et al. 2022). Feedback features in previous studies address the timing of the feedback and the content of the feedback. In terms of timing, feedback can be immediate or delayed. In classroom settings, immediate feedback was usually found to be more effective than delayed feedback (Shute 2008) because immediate feedback can catch and correct conceptual or procedural errors before students fully assimilated misconceptions or internalized procedural mistakes (Zhu, Liu and H.-S. Lee 2020). With digital writing support systems, the advantage of immediate feedback is magnified (Chen, Breslow and DeBoer 2018).

Nowadays, with the advance in technology, digitally based writing has become mainstream (Sinharay, Mo Zhang and Deane 2019) and writing support systems are widely considered essential to complete writing tasks. As such, it is easier to collect, observe and analyze the writing process for a given task. While prior works on text revision (Coenen et al. 2021; Padmakumar and He 2021; M. Lee, Liang and Yang 2022b; Wambsganss, Niklaus et al. 2020) have proposed machine collaborative writing interfaces, they focus on collecting human-machine interaction data to better train neural models, rather than understanding the underlying processes of text revision. However, some tools are becoming more and more elaborate and attempt to tackle these problems. When revising, performing multiple high-quality edits at once is very challenging, which is why (Du et al. 2022) uses a Human in the Loop text revision system to make helpful editing suggestions by interacting with users iteratively. This effectively shows why understanding revision behavior is essential to better improve these tools to maximise their effectiveness. This study focuses on comparing two different groups and their writing strategies in a writing task.

Many studies in the past have used keystroke logging as a technique to study revision (Leijten and Van Waes 2013; Zhu, Mo Zhang and Deane 2019; Sinharay, Mo Zhang and Deane 2019) in different settings. Several studies have been conducted to understand and evaluate keystroke log features in a writing task context. However, until now, keystroke logging has been scarcely used in the classroom

(Vandermeulen, Leijten and Van Waes 2020). One reason is the technical complexity of the logging tools and the logged data. Since keystroke logging is developed as a research tool, the output is quite technical, very fine-grained and often hard to grasp for laymen (Vandermeulen, Leijten and Van Waes 2020). Finally, another issue that complicates the use of keystroke loggers is their invasive nature. These tools record every typing action, so they might register private information (Vandermeulen, Leijten and Van Waes 2020).

Previous research has suggested that writing time and number of keystrokes, which are indicative of general writing fluency and efforts, are related to writing quality (Mo Zhang et al. 2016; Allen et al. 2016). Another feature of interest are pause times. Zhang et al. 2017 found that, under a certain timed-writing test condition, a shorter pre-writing pause is preferred as that indicates an adequate understanding of the task requirements, more familiarity with the writing topic, and better task planning (Sinharay, Mo Zhang and Deane 2019).

Another technique to capture the underlying process of a writing task, to better understand revision behavior, is process mining. Process Mining for education is still a rather nascent field. Only since around 2009 have researchers aimed to apply process mining to raw educational data on event logs of students' learning processes (Wambsganß et al. 2021). For educational process mining, three basic types of process mining can be distinguished (Bogarín, Cerezo and Romero 2018): Process Discovery, Conformance Checking and Process Enhancement. Because of the nature of this study, we focus on process discovery to uncover any significant differences between the two study groups. Process discovery consists of modelling and visualizing the learning process of students, in order to monitor a student's learning journey or path (Wambsganß et al. 2021). In the context of writing tasks, process mining can be used to understand the underlying processes involved in creating written content. This could include understanding the steps involved in researching, writing, editing, and publishing written materials. One particular way of applying Discovery Process Mining is using Petri Nets (Howard, Johnson and Neitzel 2010; Berti, Zelst and Aalst 2019; Southavilay, Yacef and Calvo 2010). In this study, however, we focus on Directly-Follows Graphs (DFGs) (Berti, Zelst and Aalst 2019). DFGs are simple representations of process models. Each node represents an activity and the edges describe the relationship between various activities (e.g, average time between activities or number of users from one activity to another). Because our contribution will be made public, it is easy to use the same event logs and apply any other process mining techniques using the `pm4py` (Berti, Zelst and Aalst 2019) library.

2.2 SELF-REGULATED LEARNING

To analyze revision behavior, we rely on the lens of self-regulated learning (SRL). SRL refers to the proactive process that learners engage in to optimize their learning outcome (Zimmerman 2002). According to Zimmerman's model of SRL (Zimmerman 2002), there are three major phases: forethought, performance and self-reflection. The forethought phase includes task analysis, such as goal setting and strategic planning and self-motivational beliefs. The performance phase includes self-control processes, such as task and attention-focusing strategies. The self-reflection phase includes processes involving self-judgment and self-reaction (Wong et al. 2021). SRL is essential in the context of studying revision behavior in writing support systems as it allows writers to take an active role in identifying and addressing their own writing weaknesses, rather than simply relying on the writing support system to automatically detect and correct errors. This can lead to a deeper understanding of the writing process.

One way to promote SRL is through reflective prompts (Wong et al. 2021) or via example-learning (Renkl, Hilbert and Schworm 2009). According to the self-explanation effect, learners benefit more from the examples if they can actively explain the examples to themselves. Naturally, the quality of the self-explanations determines what is learned from the examples (Schworm and Renkl 2007). The application of prompts is a possible intervention to increase the quality and depth of the explanations. These prompts should stimulate active processing of learning materials and direct attention to the central aspects (Wong et al. 2021; Renkl, Hilbert and Schworm 2009; Schworm and Renkl 2007).

With the pandemic, educators widely shifted to Massive Online Open Courses (MOOCs), which require learners to self-regulate their learning far more than in traditional settings (Mejia-Domenzain et al. 2022). However, learners' ability to self-regulate their learning varies and not all learners are highly capable of SRL (Bjork, Dunlosky and Kornell 2013). Therefore, it is important to examine ways to support and enhance SRL to increase learners' likelihood of being successful in MOOCs (Wong et al. 2021). Given the digital nature of MOOCs, it is easier to collect data and understand how certain processes are used by students. Then, it becomes easier for teachers to adapt to students and improve their SRL. In fact, (Wong et al. 2021) suggests that SRL support in the form of videos or at least a format that is integrated in the MOOCs can be effective. This study contributes to a better understanding of how users self-regulate by studying revision behavior and understanding how they use the system when they receive feedback.

2.3 GENDER COMPARISON

Research in writing has often shown similar results in terms of comparing genders. What research finds is that men tend to be more impaired in composing text in comparison with women. A study showed girls were judged superior writers in elementary school, but there were no gender differences in writing self-efficacy (Pajares and Valiante 1999). However, research comparing genders in a digital writing task is scarce. Literature offers little evidence regarding differences in the composition processes that might lead to the observed score differences (Zhu, Mo Zhang and Deane 2019). The study in (Zhu, Mo Zhang and Deane 2019) found that female students performed better than male students on a number of levels. Women had higher scores, revised more and were more efficient (e.g, they revised more per unit of time, exhibiting greater writing fluency).

CHAPTER 3

METHODS

To investigate student revision behavior in writing processes, we propose a pipeline for the automatic analysis of the self-regulated learning behavior of users during their writing tasks.

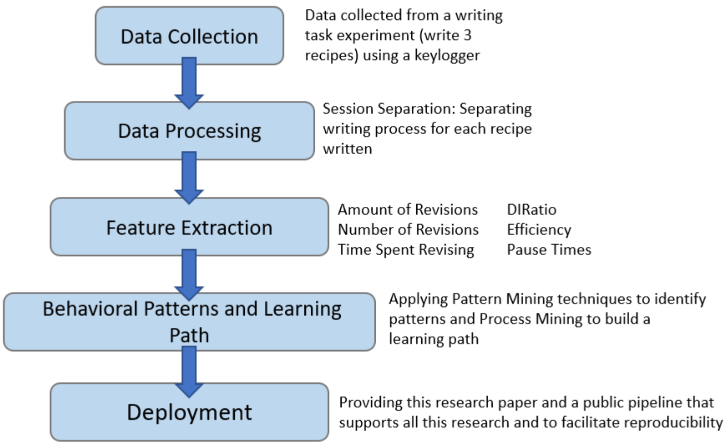

FIGURE 3.1
Overview of the pipeline and methodology

3.1 DATA COLLECTION

To conduct this research we collect data from a writing experiment. The experiment consists of 73 users divided into two groups as follows.

	No. Participants	Age		Gender		
		Mean	Std	% Female	% Male	% Other
With Adaptive Feedback (G1)	34	26.8	3.3	43	51	6
Without Adaptive Feedback (G2)	39	26.3	2.8	51	46	3

TABLE 3.1
Demographics of participants per group

Each user was given a writing task, to write three instructional texts (cooking recipes). The first group (G1) consists of 34 users. When submitting their recipes, the platform analyzed the text and provided adaptive feedback in the form of structure and clarity suggestions as well as more detailed descriptions of what the user wrote. The second group (G2) consists of 39 users and they were not provided with any feedback. The experiment ran as follows for G1: When submitting their texts, the platform provided feedback. The user could then decide to reset and start the next recipe or modify and revise their texts to improve it. The protocol for G2 is the same except they did not receive feedback.

Our analysis is based on the log data collected from the experiment that tracked users' activities using a keylogger[1]. The log entries report the user, the timestamp, a JSON file with the keystroke log and the submitted recipe in text format (e.g, 2022-09-04, 12:31:03.343, user1, [{'time': 1, 'character': '1'}, ... }], "1) Cook ...").

Additionally, we captured perception scores consisting of users' answers to a survey. The survey consists of questions asking users what they modified for each recipe, as well as how they rate the platform in different ways. This data also maps users to their groups and their genders.

3.2 Data Processing

To build an effective pipeline to understand revision behavior in adaptive writing support systems, it is crucial to properly process the data set. Given that the logs consist of the users' first attempts and revision phases, it is important to separate them in order to focus only on the revision sessions. We define sessions for a user as all the data that is collected from them for **one** recipe. For each recipe, users write, many revise at least once, which is why it is not trivial to separate sessions. As such, we use `scipy` and use cosine distance to separate sessions. Cosine distance can be used to measure the similarity between the different recipes by representing each recipe's content as a vector in a high-dimensional space. One advantage of using cosine distance for text comparison is that it is relatively insensitive to the length of the strings. This is because the distance is calculated based on the angle between the vectors, which is not affected by the length of the vectors. In contrast, other measures of distance such as Euclidean distance are sensitive to the length of the vectors and can be affected by the presence of common words that do not contribute significantly to the meaning of the strings. We load a GloVe model (Pennington, Socher and Manning 2014) using `gensim`, which is already trained on Wikipedia, where each word is represented as a 50-dimensional vector. We use this model to vectorize the recipes and run a recursive algorithm on the recipes submitted, computing the cosine distance between recipes at indices n, $n+1$, ..., until we find $k > n$ such that:

$$1 - \text{COSINE}_{dist}(n, k) < 0.995$$

When we do, we define k as the index of a new recipe in our data. This way we collect the indices of all the start of recipes in our dataset. However, this algorithm is only 91% accurate as some indices are misclassified as being new recipes. Most of the time, this is because, for a same recipe, users' revisions lead to the algorithm detecting it as a new recipe (e.g, when a user adds several steps to their recipe).

3.3 Feature Extraction

Different aspects of SRL have been researched extensively (Mejia-Domenzain et al. 2022). In a meta-analysis on online education, Broadbent and Poon 2015 found significant associations with academic achievement for five sub-scales of SRL: effort regulation (persistence in learning), time management (ability to plan study time), metacognition (awareness and control of thoughts), critical thinking (ability to carefully examine material), and help-seeking (obtaining assistance if needed). Based on these findings, we use the following dimensions to represent student behavior: effort regulation (Number of Edits, Number of Revisions), time management (Time Spent Revising), metacognition (Efficiency, Pause Time

[1]The data set includes the entirety of the writing task, not only the revision phases.

Feature Variables	Description
Number Of Revisions	For each user, we count the amount of times they revise each time they write a recipe. This gives a sense of the effort put into the revision phase of the writing task.
Number of edits	The total number of insertions and deletions during a revision step. Insertions are counted as any characters that are typed including whitespaces, and deletions are counted as the number of times the user presses any of the Backspace or Delete buttons. There are other keystroke actions in the dataset such as 'Enter' or "Tab" but the number of insertions and deletions account for over 95% of all the actions so we decide to focus on these only.
Time Spent Revising in seconds	Most users revise at least once. We compute the average time users spend revising for each group, for each recipe. This allows us to compare the two groups and to estimate the effort put in by both groups.
DIRatio (Delete-Insert Ratio)	The average deletions over insertions ratio, which approximately captures the extent of editing and revision of any kind (Zhu, Mo Zhang and Deane 2019).
Efficiency	Estimated by the number of insertions per second, which indicates a general writing speed. This feature is arguably an indicator of writing fluency (Zhu, Mo Zhang and Deane 2019).
Pause Time during Revision in seconds	For each user, we collect the inter-key time interval and compute the mean of these intervals. This captures the average lag time between two adjacent keystroke actions (Zhu, Mo Zhang and Deane 2019). The pause time for a group is the mean of the collected means for users belonging to that group. This feature gives a sense of the amount of reflection put into revising a text. It also captures the effort and persistence level of users.

TABLE 3.2
Feature Variables included in the study

During Revision), and critical thinking (DIRatio). The nature of our log data does not allow to represent help-seeking. A description of these feature variables can be found in Table 3.2.

3.4 IDENTIFICATION OF BEHAVIORAL PATTERNS

To understand revision behavior, it is essential to identify and categorise patterns in between groups to see how the writing support system is used, based on the support provided by the system. As such, our pipeline has a section dedicated to Pattern Mining (Y. Wang et al. 2019; Fournier-Viger et al. 2017). We apply the `PrefixSpan` algorithm on insert-delete sequences to identify frequent patterns in writing for different users but because of the numerous amount of sequences, the results are not conclusive. However, because the code is public, we encourage readers to contribute to making the pipeline more robust and conduct further analysis using pattern mining. An approach could be to cluster different revisions into different categories based on their content or characteristics derived from the pattern mining approach. This can be useful to identify any trends or patterns that can help in differentiating the groups. One can also use association rule mining to identify patterns in sequences of revisions made to the texts.

3.5 BUILDING THE LEARNING PATH

Understanding revision behavior implies understanding the underlying process in the writing task (e.g, how long do users in a group take to revise on average or how many users revise). In order to understand this better, process discovery can help us detect the writing process for users in a group and design a learning path when using adaptive writing support systems. This is useful for the field of SRL as it provides a way to visualize and analyze the steps involved in a process, especially revision. This is why we use process mining, to uncover underlying differences in the writing task between the two groups.

Process mining, which focuses on extracting process-related knowledge from event logs recorded by an information system, can be used to extract underlying behavior in this writing task (Southavilay, Yacef and Calvo 2010). We aim to apply discovery process mining and to do this, we create an event log (Southavilay, Yacef and Calvo 2010) for each group and use the `pm4py` library for visualisation purposes (Berti, Zelst and Aalst 2019). The event logs focus on the activity of each user in the group, tracking their activity with each recipe they write and their revisions. In this study, we focus on DFGs (directly-follows graphs) as they give an overall idea of the underlying writing process of this task. Because the event-logs are already created for the two groups, it is easy to use the `pm4py` library to adapt and apply other process mining techniques such as *Alpha Miner*, *Heuristic Miner* or *Inductive Miner* (Berti, Zelst and Aalst 2019).

3.6 PIPELINE

In this study, we design and deploy a pipeline to analyze quantitative data (e.g., time spent or the number of edits) over qualitative data (e.g, what did users revise). As such, we consciously overlook what users write in the keystroke logs, as long as we manage to perform accurate analysis. We decide to index each user and map this index to the indices in the dataset where this user has written (e.g, user 1 writes at indices [0,1,2], user 2 at indices [3, 4, 5, 6], ...). We also index recipes (section 3.2), which allows us to perform data analysis on users' revisions of every recipe they write (e.g, if user 2 writes a new recipe at indices 3 and 5, then indices 4 and 6 are revisions.

Our pipeline is carefully structured to be easily understandable by other researchers. We separated different major parts of our study into various Jupyter Notebooks, the most important one being `RevisionsStudy.ipynb`. We have dedicated notebooks to pattern mining and process mining and another one for extracting the same feature variables (section 3.3) we study from a similar data set as the one we use.

Chapter 4

Results

With this study, we find that users in different groups revise their texts differently. Recall that G1 is given adaptive feedback and G2 is not. By providing insightful feedback on what a user can change in their writing, users tend to have more revision steps with less edits at each step (figure 4.1). However, users not receiving feedback follow the opposite trend, they have fewer revision steps, with a larger number of revisions at each step 4.1.

	With Adaptive Feedback (G1)		Without Adaptive Feedback (G2)		
Feature Variables	Mean	Std	Mean	Std	p-values
Number of Revisions	1.882	2.166	0.846	0.735	0.0071
Number of Edits	75.734	95.114	222.727	338.574	0.24
Time Spent Revising (seconds)	224.48	237.37	264.01	530.47	0.694
Pause Time in Revision (seconds)	2.594	0.641	2.344	0.183	0.629

Table 4.1
First Recipe Data

	With Adaptive Feedback (G1)		Without Adaptive Feedback (G2)		
Feature Variables	Mean	Std	Mean	Std	p-values
Number of Revisions	1.147	1.033	0.692	0.722	0.033
Number of Edits	95.051	192.92	57.52	61.31	0.26
Time Spent Revising (seconds)	121.19	148	70.15	120.9	0.11
Pause Time in Revision (seconds)	2.25	0.39	1.72	0.24	0.29

Table 4.2
Second Recipe Data

	With Adaptive Feedback (G1)		Without Adaptive Feedback (G2)		
Feature Variables	Mean	Std	Mean	Std	p-values
Number of Revisions	1.12	1.05	0.737	0.676	0.073
Number of Edits	87.61	270.75	57.714	71.7	0.456
Time Spent Revising (seconds)	81.2	115.8	86.5	231.9	0.905
Pause Time in Revision (seconds)	1.76	0.358	1.94	0.97	0.71

Table 4.3
Third Recipe Data

To better visualize the data, our pipeline has designed bubble plots, plotting the revision quantity at each step for the two groups (figure 4.1). This takes into account all insertions and deletions at each revision step. Please note that revision step 0 is not a revision, but rather the moment users first submit their recipes. Revision starts when the revision step is 1. The bubble plots allow us to visualize this phenomenon, clearly allowing us to differentiate the writing process for G1 and G2.

The bubble plots confirm the fact that users in G1 have more revision steps but fewer edits at each step in general in comparison with G2.

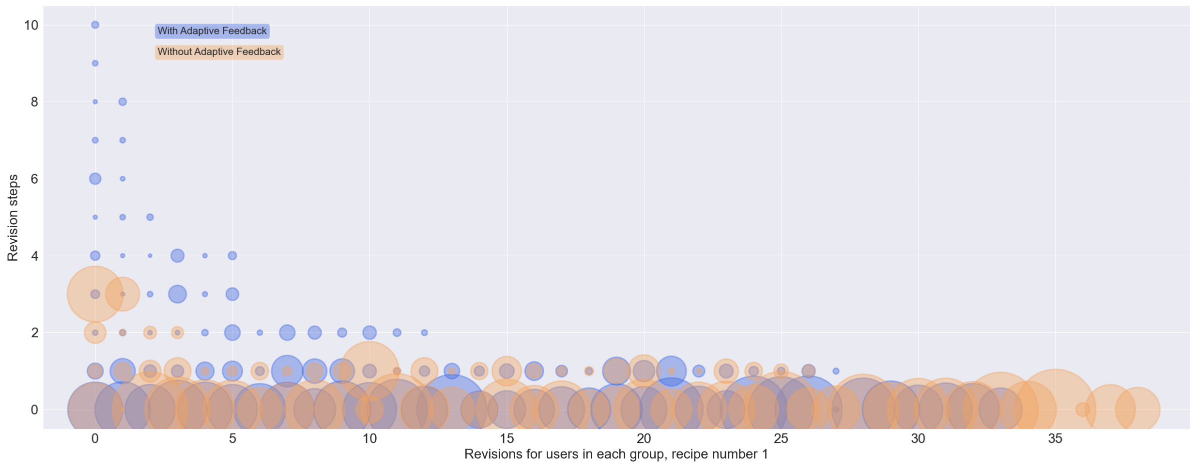

FIGURE 4.1
Bubble plot for the first recipe (size of the bubbles corresponds to the average number of revisions for a group)

4.1 DIFFERENT WRITING STRATEGIES

The strategies outlined in the previous section show a clear distinction between G1 and G2 in the writing process. We have highlighted the fact users tend to revise more often with fewer edits at each revision step when they receive feedback, and the opposite when they do not. In fact, for the first and second recipes (Tables 4.1, 4.2), we find p-values below the usual significance level (0.05), which indicates a clear distinction in the number of revisions users have which outlines the strategies mentioned. This is also underlined by the mean number of revisions and edits. On average, users in G1 tend to revise their texts more often, with fewer edits at each step (tables 4.1, 4.2, 4.3).

Regarding time, for G2, the average pause time varies by less than 0.5 seconds between the first and third recipe (tables 4.1, 4.3) but the time spent revising decreases by over 65% (figure 4.3). Because users do not receive feedback, users reflect equally over time, but they spend less time revising overall. This arouses the question of quality over time. When users write the first draft and revise, they tend to revise less once they are deep into a writing session. Because of the lack of reflection and effort put into revising their texts, the quality decreases. This can be confirmed if we perform a qualitative analysis of the recipes in this writing task, which we encourage readers to do. For the second recipe, we find a very small p-value (table 4.2) for time spent revising which highlights the difference in the distribution when users write the second recipe. However, the p-value for the third recipe shows a close correlation in the distributions for time spent, meaning users from both groups approximately revise equally in terms of time spent.

The directly-follows graphs (figure 4.2) confirm the strategy outline established earlier. From the graphs, we see that users spend approximately the same amount of time writing recipes and revising. However, we see that users in G2 revise much longer when having consecutive revision sessions (6 min) compared to G1 (56 s)(figure 4.2). This confirms the fact users in G1 have shorter revision sessions but more in quantity, whereas G2 has fewer revision steps, but more edits at each step. We notice users in both groups revise almost equally in terms of time spent revising on the first revision step after each recipe is submitted, the differences come from the consecutive revision steps.

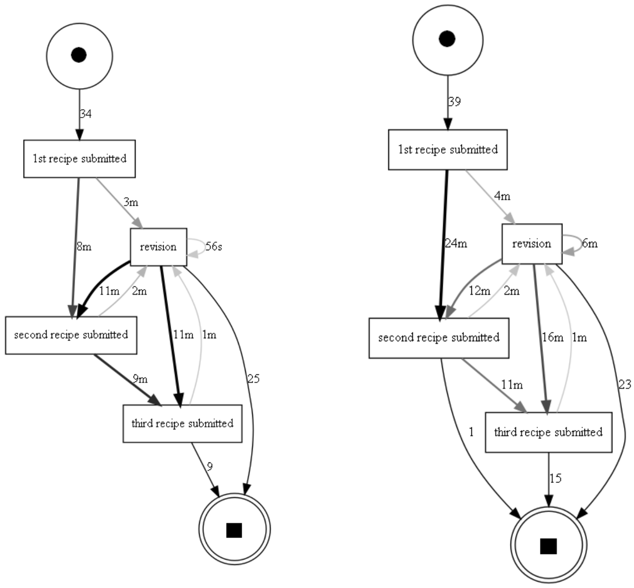

FIGURE 4.2
Directly-Follows Graphs for G1 (left) and G2 (right)

4.2 ENGAGEMENT

From the second recipe onwards, we find that users are far less engaged in their task (figure 4.3). This result is intuitive and follows the pattern from the study in Shi 2021. Young users tend to be less concentrated over time and less engaged with the task at hand and the data confirms it. The feature pause time gives a sense of the amount of reflection put into the writing. Users with higher pause times think more about the next thing to write (Zhang et al. 2017). The fact both groups have a decreasing pause time by the time the experiment ends shows that users reflect less when reaching the end of their task. For G1, because the average pause time for users decreases by almost 1 second between the first and third recipe may inform us about how the adaptive feedback is used when users are deep into a writing task. We can speculate users in G1 rely on adaptive feedback when revising to finish the writing task. In fact, When writing the third recipe, users in G1 spend on average only 80 seconds revising (table 4.3) with a smaller pause time. The fact users are less engaged makes them more inclined to follow the feedback they are given without reflecting upon it. This is also confirmed by the average number of revisions which increases by 27.14% for G1 between the first and third recipe (figure 4.3). This means, on average, users in G1 have more edits on the third recipe. However, users spend 63.85% less time revising and have a much smaller pause time as well (figure 4.3). As users are less focused, they are more likely to trust the feedback and use it to finish their tasks faster. Writing support systems are especially helpful to quickly finish tasks, but over time, one can claim it inhibits writing creativity and correctness if users rely on it to finish their tasks (Parkin et al. 2012). On the other hand, we see a stronger decline in user engagement for G2. They spend much less time in the revision phase (67% less) and make much fewer edits when they write the third recipe in comparison to the first (figure 4.3).

Furthermore, in terms of engagement, although we have seen users spend less time revising and reflecting on what to write next, 73.5% of users in G1 make the effort of revising the last recipe (25/34 users) and

59% of users in G2 revise their last recipe (23/39 users) (figure 4.2). As such, although writing support tools may not be used in the way in which they were conceived, which was to assist students in their writing, they do promote revision in the dying embers of the writing task.

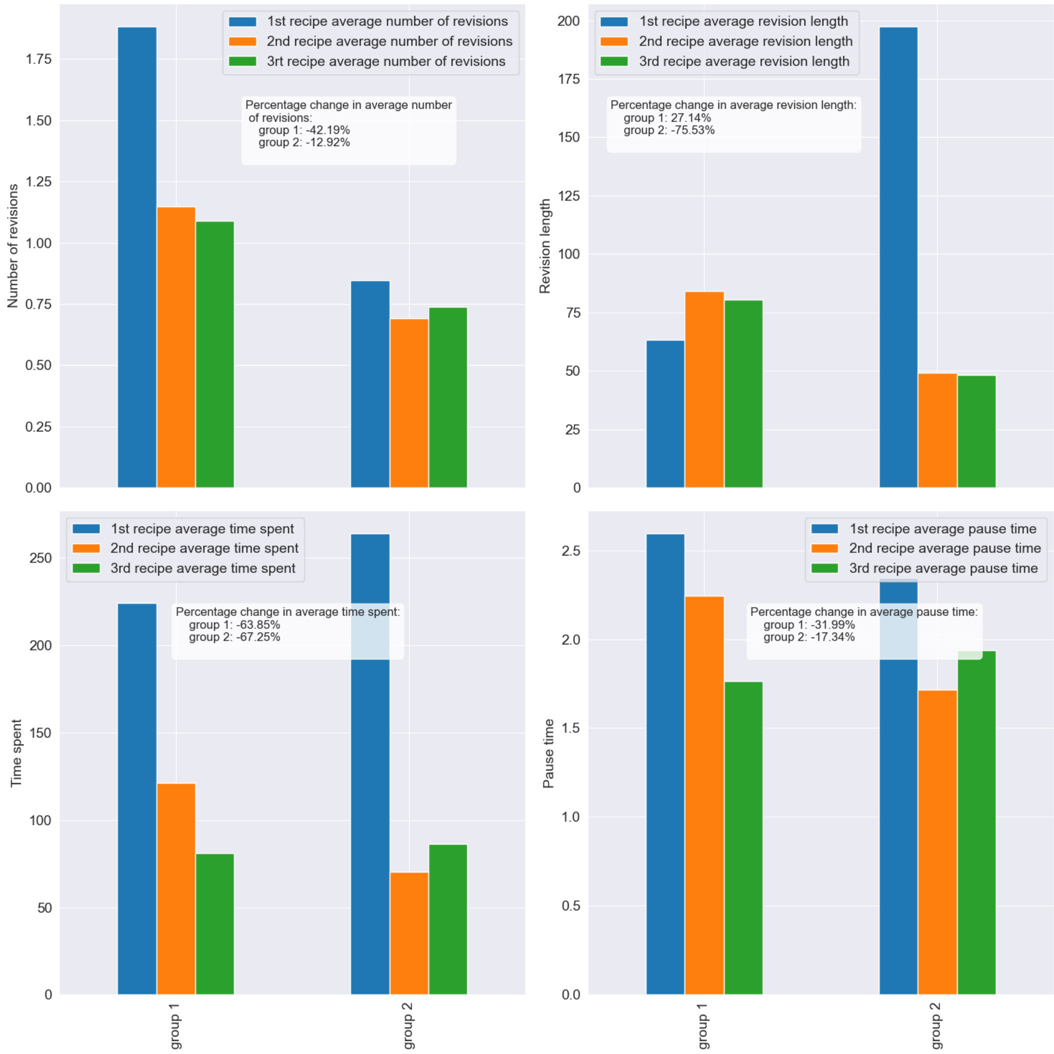

FIGURE 4.3
Percentage Change between the first and third recipe of different feature variables

4.3 GENDER COMPARISON

When studying revision behavior, it is crucial to profile students in order to provide useful tools for them to succeed. This entails studying the differences between genders. We find that there is a clear distinction in the writing capabilities between men and women. In figure 4.4, we focus on all of the writing tasks, not on sessions. We find that overall, women are more confident in their writing and use their time more efficiently. Men consistently spend more time revising than women, and regardless of the number of times revised, men revise more often than women and have more edits as well, i.e, have more insertions and/or deletions than women. Efficiency, as defined earlier, is an indicator of writing fluency. We see that women are far more efficient (Zhu, Mo Zhang and Deane 2019) in this writing task compared to men and this

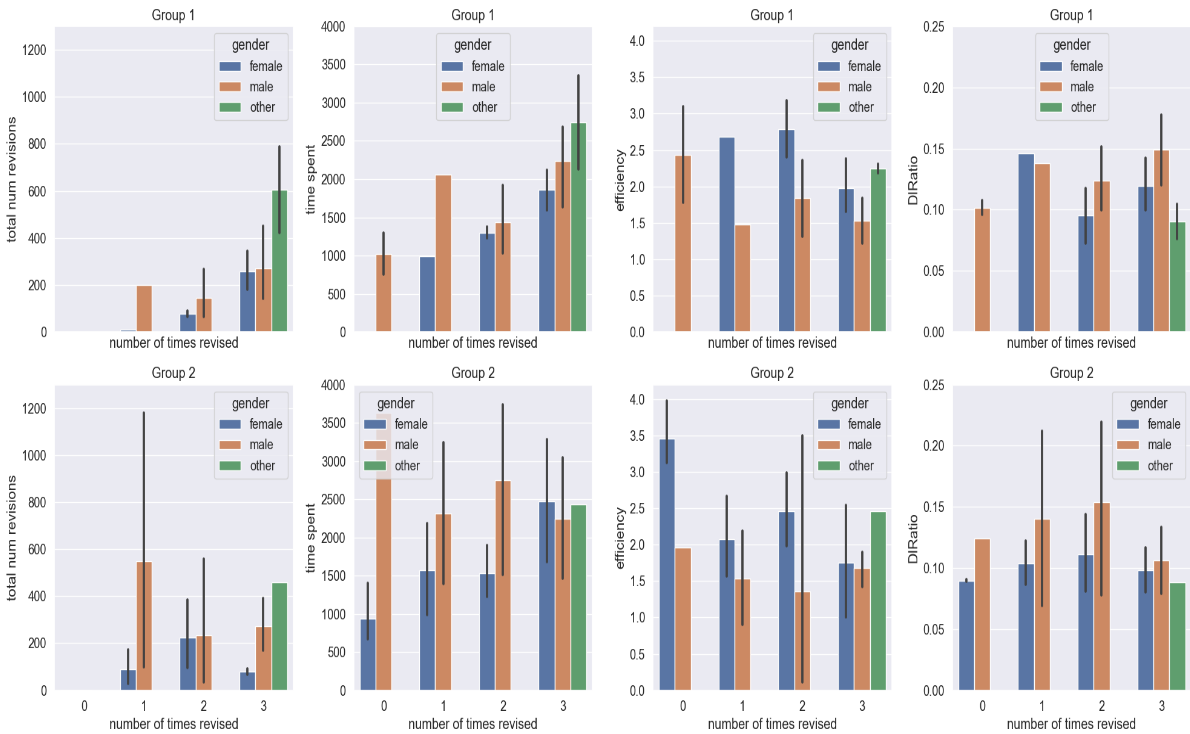

FIGURE 4.4
Gender comparison over 4 feature variables

confirms women write more confidently on average. [1].

Furthermore, regarding the Delete-Insert ratio (DIRatio), we find that men in G2 have higher scores. Having higher DIRatio scores indicates a certain lack of confidence in writing, meaning users delete a large portion of their texts (over 20% for some users in G2 (figure 4.4)). Looking back at self-regulated learning, especially on the critical thinking part Broadbent and Poon 2015, which is defined as the ability to examine material, we can see men are more self-critical and delete a larger portion of their texts compared to women. To sum up, women spend less time and have higher efficiency scores on average, exhibiting greater writing fluency (Zhu, Mo Zhang and Deane 2019).

[1]Please note the plots remove several outliers which corrupt the results

CHAPTER 5

DISCUSSION

5.1 OUR CONTRIBUTION

In this research, we focused on understanding revision behavior in adaptive writing support systems. To do this, we developed a pipeline to collect and analyze data on user writing and revision activity. Our analysis revealed that users in different groups revise using different strategies. Users who were equipped with adaptive feedback revised more often, with fewer edits at each revision step and users without adaptive feedback followed the opposite trend. This suggests that the support provided by the system may influence revision behavior. Additionally, we found that overall user engagement decreases over time. These findings have important implications for the design and evaluation of adaptive writing support systems. By understanding how revision behavior changes in these systems, designers can create more effective support that caters to the needs of different user groups. Our results also suggest that it may be important to consider how engagement changes over time and to design strategies to sustain user engagement in adaptive writing support systems. Overall, our pipeline and findings contribute to a deeper understanding of revision behavior and user engagement in adaptive writing support systems, and they provide valuable insights for future research in this area.

5.2 LIMITATIONS & FUTURE DIRECTIONS

In this section, we will discuss the limitations of our study and suggest directions for future research in this area. The small sample size of this study (73 users) certainly is a limitation and could affect the results of this study. A larger sample size may have allowed for a more precise measurement of the study variables, as the larger number of participants could have provided a more representative sample of the population. This could have resulted in more accurate and reliable findings. As such, we only form an intuition of revision behavior in writing support systems and future research must contain a larger sample size to increase the accuracy of the studies. Moreover, given the sample choice is rather small and we have several outliers, this further affects the results and interpretations of the study.

The decision to focus on quantitative data in this study may have resulted in some important limitations in our understanding of the topic. While quantitative data can be a powerful tool for analyzing relationships and trends, it may not capture the full depth and richness of qualitative data, such as different types of revisions in each group. This limitation was a conscious choice, as we wanted to examine the numerical relationships between the variables and the outcomes of interest. However, it is important to recognize that this may have resulted in some limitations in the generalizability of our findings. Studying qualitative data can inform us about the writing fluency of users in different groups and underline the effectiveness of the writing support system.

Another limitation is in the data itself. The lack of scores for the quality of the recipes or other relevant data for the samples in this study may have resulted in some limitations. Without this information, we were unable to control for certain factors that could have influenced the outcomes of interest, which may have impacted the accuracy and validity of our results. This limitation was due to the nature of the data that was available to us, and we were unable to address it in our analysis.

Future research should focus on addressing the limitations of this analysis. This study focuses on quantitative data over qualitative. We find that users revise differently, but we do not compare what users revise. Future directions consist of analysing what users have revised for each recipe and comparing the quality difference of revisions. This can be done using the same data set this study uses, so we encourage readers to engage in qualitative analysis. This can be done in part by using pattern mining 3.4 to understand different underlying patterns in the writing process. Moreover, we've already prepared some tools for qualitative analysis, by summarizing what users answered to the survey on what they edited at each recipe so that clustering techniques can be used to group similar revisions together and to get a better picture of how users revise.

Another metric that can be interesting is providing scores for the quality of the texts. This way, we can clearly see if users equipped with adaptive feedback score better, meaning they write more qualitatively than users without adaptive feedback. If so, this would easily underline the utility of writing support systems for writing processes.

Finally, because we measure a steep decline in engagement, this can inform us about how to design future writing support tools to better support engagement. Keeping users engaged over a long period of time is challenging but managing to do so would bring out more quality in students' writing. In fact, some attempts have been made to counter this issue, such as (Du et al. 2022) which interacts with the user iteratively.

Chapter 6

Conclusion

This study outlines different writing strategies when revising texts, depending on the feedback received. We find users receiving feedback perform more revision sessions, with fewer edits at each step. On the other hand, users who do not receive feedback have fewer revision sessions, with a greater number of revisions. These strategies underline the utility of a writing support tool. Because users received feedback only when they submitted their recipes and not during the writing procedure, they tend to submit their recipes more often, look at the feedback and adapt it to their texts to make it more qualitative.

We also find engagement strategies are something writing support tools must implement to keep users more focused on the task at hand. In fact, when performing a repetitive task, like writing recipes, users tend to lose focus over time. They revise less in terms of time spent, number of revision steps and number of edits. By the time users write their third recipe, they tend to care less about the quality of their texts. Users receiving feedback tend to reflect less and rely on the platforms' suggestions to finish the task.

In conclusion, our research on revision behavior in adaptive writing support systems has shed light on how users in different groups approach revision and how overall user engagement evolves over time. The development of a pipeline to study this topic has allowed us to collect and analyze data on user writing and revision activity, leading to the discovery of important patterns and trends. By understanding how revision behavior and engagement change in adaptive writing support systems, we can design and evaluate these systems in a way that better meets the needs of different user groups and sustains user engagement over time. Our findings have important implications for the design and evaluation of adaptive writing support systems and for future research in this area. Overall, our study has made a significant contribution to the field by providing a deeper understanding of revision behavior and user engagement in these systems and by offering valuable insights for future research in this area.

ACKNOWLEDGEMENTS

In this section, I want to extend a heartfelt thank-you to the entire ML4ED team for providing a very friendly and inclusive place, always being open to discussion and including semester project students in multiple lab meetings. I had an amazing experience working in this lab, and I am so grateful to have had Thiemo as my mentor and colleague. His tireless efforts to challenge me and help me produce the best possible results were invaluable, and I can't thank him enough for his guidance and support. I dedicated a lot of my time to work on this project and I am truly excited about having written a paper that we will submit to a conference. When I first started this project, I could never have dreamed that I would be in the position to submit a paper to a conference, especially as a Bachelor's student. It's been an incredible journey, and I am so grateful to have had the opportunity to work on such a meaningful and rewarding project.

I also want to give a special shout-out to Paola, who I had the pleasure of meeting with just once, but who left a lasting impression with her valuable suggestions and fresh perspective on the data. Her insights ended up being helpful as I continued to work on the project.

Bibliography

Nova, Muhamad (2018). 'Utilizing Grammarly in evaluating academic writing: A narrative research on EFL students' experience'. In: *Premise: Journal of English Education and Applied Linguistics* 7.1, pp. 80–96.

Yuan, Zheng (Mar. 2017). *Grammatical error correction in non-native English*. Tech. rep. UCAM-CL-TR-904. University of Cambridge, Computer Laboratory. DOI: 10.48456/tr-904. URL: https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-904.pdf.

Wambsganss, Thiemo, Andreas Janson and Jan Marco Leimeister (2022). 'Enhancing argumentative writing with automated feedback and social comparison nudging'. In: *Computers Education* 191, p. 104644. ISSN: 0360-1315. DOI: https://doi.org/10.1016/j.compedu.2022.104644. URL: https://www.sciencedirect.com/science/article/pii/S0360131522002159.

Wambsganss, Thiemo, Andrew Caines and Paula Buttery (2022). 'ALEN App: Persuasive Writing Support To Foster English Language Learning'. In: *BEA 2022*, p. 134.

Fuente, Jesús de la et al. (2022). 'Advances on Self-Regulation Models: A New Research Agenda Through the SR vs ER Behavior Theory in Different Psychology Contexts'. In: *Frontiers in Psychology* 13. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2022.861493. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.861493.

Zimmerman, Barry (June 2002). 'Becoming a Self-Regulated Learner: An Overview'. In: *Theory Into Practice* 41, pp. 64–70. DOI: 10.1207/s15430421tip4102_2.

Lee, Mina, Percy Liang and Qian Yang (2022a). 'CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities'. In: *CoRR* abs/2201.06796. arXiv: 2201.06796. URL: https://arxiv.org/abs/2201.06796.

Buschek, Daniel, Martin Zürn and Malin Eiband (2021). 'The Impact of Multiple Parallel Phrase Suggestions on Email Input and Composition Behaviour of Native and Non-Native English Writers'. In: *CoRR* abs/2101.09157. arXiv: 2101.09157. URL: https://arxiv.org/abs/2101.09157.

Wang, Lu et al. (2021). 'An Evaluation of Generative Pre-Training Model-based Therapy Chatbot for Caregivers'. In: *CoRR* abs/2107.13115. arXiv: 2107.13115. URL: https://arxiv.org/abs/2107.13115.

Chen, Xin, Lori Breslow and Jennifer DeBoer (2018). 'Analyzing productive learning behaviors for students using immediate corrective feedback in a blended learning environment'. In: *Computers Education* 117, pp. 59–74. ISSN: 0360-1315. DOI: https://doi.org/10.1016/j.compedu.2017.09.013. URL: https://www.sciencedirect.com/science/article/pii/S0360131517302178.

Mejia-Domenzain, Paola et al. (2022). 'Identifying and Comparing Multi-Dimensional Student Profiles Across Flipped Classrooms'. In: *Artificial Intelligence in Education: 23rd International Conference, AIED 2022, Durham, UK, July 27–31, 2022, Proceedings, Part I*. Durham, United Kingdom: Springer-Verlag, pp. 90–102. ISBN: 978-3-031-11643-8. DOI: 10.1007/978-3-031-11644-5_8. URL: https://doi.org/10.1007/978-3-031-11644-5_8.

Quintana, C. et al. (Jan. 2006). ‘Learner-centered design’. In: *Cambridge Handbook of the Learning Sciences*, pp. 119–134.

Zhu, Mengxiao, Ou Lydia Liu and Hee-Sun Lee (Jan. 2020). ‘The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing’. en. In: *Computers & Education* 143, p. 103668. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2019.103668. URL: https://www.sciencedirect.com/science/article/pii/S0360131519302210 (visited on 3rd Oct. 2022).

Wong, Jacqueline et al. (2021). ‘Examining the use of prompts to facilitate self-regulated learning in Massive Open Online Courses’. In: *Computers in Human Behavior* 115, p. 106596. ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2020.106596. URL: https://www.sciencedirect.com/science/article/pii/S0747563220303435.

Bowen, Neil Evan Jon Anthony, Nathan Thomas and Nina Vandermeulen (Mar. 2022). ‘Exploring feedback and regulation in online writing classes with keystroke logging’. en. In: *Computers and Composition* 63, p. 102692. ISSN: 8755-4615. DOI: 10.1016/j.compcom.2022.102692. URL: https://www.sciencedirect.com/science/article/pii/S8755461522000019 (visited on 24th Dec. 2022).

Coenen, Andy et al. (2021). ‘Wordcraft: a Human-AI Collaborative Editor for Story Writing’. In: *CoRR* abs/2107.07430. arXiv: 2107.07430. URL: https://arxiv.org/abs/2107.07430.

Padmakumar, Vishakh and He He (2021). ‘Machine-in-the-Loop Rewriting for Creative Image Captioning’. In: *CoRR* abs/2111.04193. arXiv: 2111.04193. URL: https://arxiv.org/abs/2111.04193.

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

Vandermeulen, Nina, Mariëlle Leijten and Luuk Van Waes (2020). ‘Reporting writing process feedback in the classroom using keystroke logging data to reflect on writing processes’. In: *Journal of Writing Research* 12.1, pp. 109–139.

Leijten, Mariëlle and Luuk Van Waes (July 2013). ‘Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes’. en. In: *Written Communication* 30.3. Publisher: SAGE Publications Inc, pp. 358–392. ISSN: 0741-0883. DOI: 10.1177/0741088313491692. URL: https://doi.org/10.1177/0741088313491692 (visited on 5th Oct. 2022).

Zhu, Mengxiao, Mo Zhang and Paul Deane (2019). ‘Analysis of Keystroke Sequences in Writing Logs’. en. In: *ETS Research Report Series* 2019.1. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ets2.12247, pp. 1–16. ISSN: 2330-8516. DOI: 10.1002/ets2.12247. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ets2.12247 (visited on 24th Dec. 2022).

Sinharay, Sandip, Mo Zhang and Paul Deane (Apr. 2019). ‘Prediction of Essay Scores From Writing Process and Product Features Using Data Mining Methods’. In: *Applied Measurement in Education* 32.2. Publisher: Routledge _eprint: https://doi.org/10.1080/08957347.2019.1577245, pp. 116–137. ISSN: 0895-7347. DOI: 10.1080/08957347.2019.1577245. URL: https://doi.org/10.1080/08957347.2019.1577245 (visited on 24th Dec. 2022).

Emig, Janet (1971). ‘The composing processes of twelfth graders.’ In.

Stallard, Charles K (1974). ‘An analysis of the writing behavior of good student writers’. In: *Research in the Teaching of English* 8.2, pp. 206–218.

Hayes, John R (2012). ‘Modeling and remodeling writing’. In: *Written communication* 29.3, pp. 369–388.

McCutchen, Deborah (1996). ‘A capacity theory of writing: Working memory in composition’. In: *Educational psychology review* 8.3, pp. 299–325.

Flower, Linda and John R. Hayes (1981). ‘A Cognitive Process Theory of Writing’. In: *College Composition and Communication* 32.4, pp. 365–387. ISSN: 0010096X. URL: http://www.jstor.org/stable/356600 (visited on 31st Dec. 2022).

Gero, Katy et al. (May 2022). 'A Design Space for Writing Support Tools Using a Cognitive Process Model of Writing'. In: *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*. Dublin, Ireland: Association for Computational Linguistics, pp. 11–24. DOI: 10.18653/v1/2022.in2writing-1.2. URL: https://aclanthology.org/2022.in2writing-1.2.

Shute, Valerie J. (2008). 'Focus on Formative Feedback'. In: *Review of Educational Research* 78.1, pp. 153–189. DOI: 10.3102/0034654307313795. eprint: https://doi.org/10.3102/0034654307313795. URL: https://doi.org/10.3102/0034654307313795.

Lee, Mina, Percy Liang and Qian Yang (Jan. 2022b). 'CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities'. In: *arXiv e-prints*, arXiv:2201.06796, arXiv:2201.06796. arXiv: 2201.06796 [cs.HC].

Wambsganss, Thiemo, Christina Niklaus et al. (2020). 'AL: an adaptive learning support system for argumentation skills'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.

Du, Wanyu et al. (2022). 'Read, Revise, Repeat: A System Demonstration for Human-in-the-loop Iterative Text Revision'. In: *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*, pp. 96–108. DOI: 10.18653/v1/2022.in2writing-1.14. arXiv: 2204.03685[cs]. URL: http://arxiv.org/abs/2204.03685 (visited on 27th Dec. 2022).

Zhang, Mo et al. (2016). 'Classification of writing patterns using keystroke logs'. In: *Quantitative psychology research*. Springer, pp. 299–314.

Allen, Laura K et al. (2016). '{ENTER} ing the Time Series {SPACE}: Uncovering the Writing Process through Keystroke Analyses.' In: *International Educational Data Mining Society*.

Zhang, M et al. (2017). 'An investigation of the writing processes in timed task condition using keystrokes'. In: *Understanding and investigating response processes in validation research*, pp. 321–339.

Wambsganß, Thiemo et al. (Oct. 2021). 'The Potential of Technology-Mediated Learning Processes: A Taxonomy and Research Agenda for Educational Process Mining'. In.

Bogarín, Alejandro, Rebeca Cerezo and Cristóbal Romero (2018). 'A survey on educational process mining'. In: *WIREs Data Mining and Knowledge Discovery* 8.1, e1230. DOI: https://doi.org/10.1002/widm.1230. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1230. URL: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1230.

Howard, Larry, Julie Johnson and Carin Neitzel (Oct. 2010). 'Examining Learner Control in a Structured Inquiry Cycle Using Process Mining'. In: pp. 71–80.

Berti, Alessandro, Sebastiaan J. van Zelst and Wil van der Aalst (15th May 2019). *Process Mining for Python (PM4Py): Bridging the Gap Between Process- and Data Science*. DOI: 10.48550/arXiv.1905.06169. arXiv: 1905.06169[cs]. URL: http://arxiv.org/abs/1905.06169 (visited on 25th Dec. 2022).

Southavilay, Vilaythong, Kalina Yacef and Rafael Calvo (Oct. 2010). 'Process Mining to Support Students' Collaborative Writing.' In: pp. 257–266.

Renkl, Alexander, Tatjana Hilbert and Silke Schworm (Mar. 2009). 'Example-Based Learning in Heuristic Domains: A Cognitive Load Theory Account'. In: *Educational Psychology Review* 21, pp. 67–78. DOI: 10.1007/s10648-008-9093-4.

Schworm, Silke and Alexander Renkl (May 2007). 'Learning argumentation skills through the use of prompts for self-explaining examples'. In: *Journal of Educational Psychology* 99, pp. 285–296. DOI: 10.1037/0022-0663.99.2.285.

Pajares, Frank and Giovanni Valiante (1999). 'Grade Level and Gender Differences in the Writing Self-Beliefs of Middle School Students'. In: *Contemporary Educational Psychology* 24.4, pp. 390–405. ISSN: 0361-476X. DOI: https://doi.org/10.1006/ceps.1998.0995. URL: https://www.sciencedirect.com/science/article/pii/S0361476X9890995X.

Pennington, Jeffrey, Richard Socher and Christopher D Manning (2014). 'Glove: Global vectors for word representation'. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Broadbent, Jaclyn and Walter L Poon (2015). 'Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review'. In: *The Internet and Higher Education* 27, pp. 1–13.

Wang, Yu et al. (Dec. 2019). 'Recognizing patterns of student's modeling behaviour patterns via process mining'. In: *Smart Learning Environments* 6.1, p. 26. ISSN: 2196-7091. DOI: 10.1186/s40561-019-0097-y. URL: https://doi.org/10.1186/s40561-019-0097-y (visited on 29th Sept. 2022).

Fournier-Viger, Philippe et al. (2017). 'A survey of sequential pattern mining'. In: *Data Science and Pattern Recognition* 1.1, pp. 54–77.

Shi, Yali (15th Oct. 2021). 'Exploring Learner Engagement With Multiple Sources of Feedback on L2 Writing Across Genres'. In: *Frontiers in Psychology* 12, p. 758867. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2021.758867. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8554298/ (visited on 25th Dec. 2022).

Parkin, Helen J et al. (2012). 'A role for technology in enhancing students' engagement with feedback'. In: *Assessment & Evaluation in Higher Education* 37.8, pp. 963–973. DOI: 10.1080/02602938.2011.592934. eprint: https://doi.org/10.1080/02602938.2011.592934. URL: https://doi.org/10.1080/02602938.2011.592934.