

From Cognition to Writing—Planning Activities in Digital Argumentative Texts of Ninth Graders in Two Different School Types

Winnie-Karen Giera, Lucas Deutzmann* and Subhan Sheikh Muhammad

University of Potsdam, Germany

*Corresponding Author

Lucas Deutzmann, University of Potsdam, Germany.

Submitted: 2025, Oct 28; Accepted: 2025, Nov 26; Published: 2025, Dec 08

Citation: Giera, W. K., Deutzmann, L., Muhammad, S. S. (2025). From Cognition to Writing—Planning Activities in Digital Argumentative Texts of Ninth Graders in Two Different School Types. *J App Lang Lea*, 2(2), 01-17.

Abstract

This pilot study investigates the role of planning activities in digital argumentative texts of ninth graders in two different school types—in shaping the quality of argumentative writing among 357 ninth-grade students in Germany. This is a sub-study that is part of a bigger research project. It aims - building on cognitive models of writing and executive functioning- to investigate how time spent on writing, time spent on planning, and switching behaviour between a writing plan and text field predict writing quality. Using a quasi-experimental, longitudinal design, data was collected digitally via Gorilla Experiment Builder at four measurement points in six secondary schools representing both higher and lower academic tracks. The digital writing environment allowed for the detailed logging of student activity, including writing time, planning time, and frequency of between writing field and writing plan. The results from the correlation and regression analyses show that writing time and switching behaviour are significant predictors of text quality, whereas planning time had nonsignificant or weak association with switching activity. The students in lower academic tracks spent more time planning and writing but produced lower-quality texts, suggesting a differential reliance on external scaffolding. The findings underscore the importance of cognitive flexibility in writing and highlight the potential of digital planning tools to support writing development, especially among students with fewer prior skills. The implications for writing instruction and assessment are also discussed.

Keywords: Argumentative Writing, Cognitive Flexibility, Planning, Digital Writing, Secondary Education

1. Introduction

1.1. Background and Rationale

Writing is considered a complex and essential form of human communication. It includes, at the base level, the simple expression of language and continues to engage a number of complex domains, including cognitive, social, emotional, cultural, and contextual processes [1]. In educational and professional environments, writing is considered the distribution of knowledge, a way of expressing an individual's identity, and the initiation of critical thinking [2]. Its significance can be noticed by observing different educational policies, curricula, and pedagogical principles that focus on mastering writing skills from early childhood to adulthood [3,4].

The nature of writing being a cognitive activity makes it demanding for both researchers and educators [5,6]. Writing seems to be just a simple activity when it comes to its appearance, simply putting words on paper, but at the same time it involves the coordination of a number of processes simultaneously, including ideation, planning, the translation of thoughts into written language, revising, editing, etc. [7]. Moreover, these domains are connected with affective factors such as motivation for writing, self-efficacy, anxiety, social influences, including cultural expectations as well as peer and teacher perceptions, and the societal values that are meant to be placed on written expression [1,8].

Cognitive flexibility is the ability to adapt thinking while performing tasks. As Kuhn explained, writers who are able to adapt their style to different genres and communicate in a variety of

ways can demonstrate their versatility and a deeper understanding of rhetorical strategies [9]. It is evident that expert writers employ a variety of genre-specific strategies, which are tailored to different audiences and purposes. These competencies encompass the ability to recognise genre features, adapt tone (formal or informal), and employ suitable organisational structures. The cultivation of such adaptability is of paramount importance in lifelong learning [10,11].

Meta-cognitive awareness (which means knowing what to do and when) is also connected with improved writing results. McNamara & Kendeou demonstrate the significance of implementing meta-cognitive strategy training, which involves the goal of guiding learners in the planning, monitoring, and evaluation of writing [12].

In particular, argumentative writing tasks are highlighted as highly demanding for secondary school students [13,14]. In the majority of cases, students are required either to imagine their dialogue partners within the written argumentation or are given a fictitious argumentation context. According to secondary school students, this is one of the primary reasons that they usually have more difficulties with argumentative writing than with oral argumentation. In addition, it is a particularly demanding cognitive activity in which the writers must invest cognitive resources in representing their arguments and those of their imagined dialogue partners [13]. Challenges for students arise because they also have to consider arguments of the opposing position for text composition.

Ninth-grade students are also expected to be proficient in writing different genres of texts and to use planning and revision strategies. In German examinations at the end of the tenth grade, for example, students are expected to write an argumentative text that considers pro and con arguments and in the end leads to a personal statement [15]. In the federal states of Berlin and Brandenburg, this exam requires the completion of a writing plan that prepares students to write pro and con arguments. However, this writing plan is not understood as a tool to assist, but rather its completion is assessed with points, which in turn has an influence on the assessment of the examination. However, the question must be asked of whether this mandatory assessment corresponds to the state of research described above and whether writing plans should not be seen primarily as an element in the task environment to support the writing process of students. In addition, students in Germany still have to handwrite their final exams at the end of the tenth grade. This means that both the text and the writing plan must be handwritten. This makes it difficult to integrate notes from the writing plan into the draft text.

The present study addresses the aspects of planning and digital writing tools through its research design, firstly by measuring the planning activities of students while writing an argumentative text, and secondly by using a digital writing plan (see Chapter 1.3).

1.2. Objective of the Research

The objective of this sub-study is to reanalyse learners' writing processes digitally to recommend evidence-based practices for teachers in their classrooms. This reanalysis is part of a larger pilot study (see Chapter 3.1) [16]. The following information about the German school system is important for understanding the study: In Germany, two pathways exist in secondary education that provide access to higher education. After a common primary education phase, the system becomes stratified, with each of the 16 federal states establishing its own school types. Despite these structural differences, all systems follow a similar pattern: beginning in the fifth or seventh grade (around ages 10 or 12), students are typically enrolled in either a higher academic track or a lower academic track, with inclusive classes for students with special educational needs in learning, language, or socio-emotional development [17]. These two pathways are guided by distinct curriculum frameworks, especially concerning the subject of German. These peculiarities of the German education system also influence how argumentation is taught in classrooms. At higher academic-track schools, students begin practicing controversial written argumentation as early as the eighth grade [18]. In contrast, in lower-academic-track schools, this typically begins one year later, in the ninth grade. The information about German school system is discussed more in Chapter 3.

It has already been emphasized that good writers are characterised by cognitive flexibility and meta-cognitive skills. This is also the case for planning activities, which must be set in correspondence with the resulting text. To this end, this long-term study focuses on investigating precisely this aspect: it aims to contribute to the understanding of how students' planning activities using a writing plan can contribute to the improvement of their text quality. In order to achieve this goal, this sub-study uses a research design that utilises a digital tool. This digital tool (*Gorilla Experiment Builder*) offers students a digital task environment that combines a digital writing plan with a digital writing task including a text-input field [7]. The aforementioned task environment is used as part of a survey at four measurement points in order to record the ninth grade students' planning and writing activity and to compare this with the quality of the written argumentative texts. This raises central questions for the present study: firstly, how variables of writing and planning activity (e.g., time spent on the writing plan) are related to text quality, and, secondly, how these variables differ between students in the higher and lower academic tracks (see Chapter 3.3.). The objective of analyzing learners' writing processes digitally in order to recommend evidence-based practices for teachers in their classrooms is closely related to the research questions. The focus of this study is well defined, as outlined in the research questions and hypotheses presented in Chapter 3.3. These questions are being investigated as part of long-term study, which also includes two interventions (lesson series debating and written argumentation based on SRSD approach). These interventions, however, are not the central subject of this paper; they will be discussed in more detail in Giera et al. (2025). [16]. The focus of the current study is on the use of a digital writing plan and its effect on text quality. First, an overview of important studies on the

topic of planning activities in writing is provided. Chapter 3 then describes the research design, including the data collection using the digital tool *Gorilla Experiment Builder*. The results for the central research questions are presented in Chapter 4 with regard to various variables and student groups, and are then discussed in Chapter 5.

To better understand the underlying cognitive demands of writing and the role of planning in student performance, the following chapter introduces established models of the writing process.

2. Theoretical Framework

Given the challenges discussed in Chapter 1, cognitive models such as that of Hayes offer a useful lens through which to examine how planning processes affect writing performance in school contexts [7]. The vast scope of writing, including a lot of variation due to individuals' cognitive abilities and environmental or social contexts, has encouraged many studies in the domain of writing development [7,19]. The fields being engaged in the inquiry include cognitive science, educational psychology, and linguistics, mainly focussing on how writing competency appears, how it can be promoted, and what the factors or themes are that affect its success [5,10].

Historical theoretical models on writing are more focused on the cognitive dimensions. One of the promoted and influential models is that of Hayes and Flower, which is focused on planning, translating, and revising as core components with recursive feedback loops, and Berninger and Swanson's work notes executive functioning, especially emphasizing the role of working memory, which supports these processes [5,6]. These are the bases for the development of pedagogical strategies that are focused on some specific cognitive skills. Studies explain that clear instructions in planning and revising are powerful in fostering text coherence and argument strength [1,20,21]. Furthermore, longitudinal studies show that children's growth in planning and revising skills is correlated with higher-quality texts over time [23].

Writing education research has been more focused on obtaining insights about effective classroom practices. Graham and Harris proposed systematic reviews in which they summarized the results of many intervention studies, mentioning that clear and direct instructions in the context of planning, organising, and self-monitoring can significantly increase students' writing performance [1].

One of the crucial constructs related to writing research is working memory, which involves the short-term storage of information and adapting it accordingly for performing complex tasks. According to Berninger and Swanson, the extent of working memory works as a restriction on the speed and fluency of the transcription and translation of all abstract ideas into meaningful concrete information, such as written text [5]. This concept is also supported by contemporary meta-analyses, which demonstrate that individuals possessing elevated working memory capacity tend to exhibit more logical, well-structured texts [24]. The significance of other executive functions involving planning, selfmonitoring, and

cognitive flexibility has also been acknowledged in the literature. Hayes proposed a version of a writing process model that includes emotional and motivational parts of the thinking process [7,25]. This allows us to better understand how writers concentrate, manage ideas, and control how much effort they expend. This model shows that skilled writers can easily switch between different parts of the writing process, such as planning, writing, and revising, depending on the task. It shows how important it is to be able to control what you are doing and to plan carefully if you want to write something that makes sense and has an effect [7].

Furthermore, the recent researches are focused on self-regulation in secondary school writing. Self-regulated writing is based on increased planning, monitoring and revision techniques that are meant to alter as the academic and cognitive abilities of the students ascends [26,27].

The writing process is a miscellaneous activity composed of multiple interlinked stages, including planning, drafting, revising, and editing [7]. In the existing literature, planning is considered to be a crucial stage that has a significant impact on the overall quality and coherence of final texts [19,23]. The stages involved in successful planning comprise idea generation, organisation, goal setting, and structural outlining, which acts as a cognitive roadmap that guides the whole writing task [20,28].

It has been indicated by studies that good writers spend an excessive amount of time planning before and during writing [5,22]. During this stage, writers often brainstorm, order their ideas, create graphic organisers, or write detailed outlines [6,28]. These strategies help to make thinking processes clearer, organize ideas better, and stop texts from being disorganised or superficial [29]. Planning effectively helps in writing better because it provides a structure that helps ideas be developed in a logical way [20,30].

Skilled writers often show that they can adapt their writing process. They can change, expand, or reorganize their plans as they write [31]. This ability shows that a writer understands how to think about their writing, which allows them to change their plan when it does not work and make important adjustments [32]. It has been conceptualised from the study that these adaptations are capable of making writing more fluent and coherent [29]. It has been shown by the studies that people who plan and revise texts daily are more motivated in the tasks and generate better work [4,22].

With regard to teaching, explicitly teaching students how to plan has been shown to have a large positive effect on their writing [21]. Studies have shown that strategies such as idea mapping, outlining, goal-setting, and reflective planning can help students to organize their thoughts and stay focused during the writing process [33]. In addition, students who think carefully about their writing and change it when needed have been shown to develop important skills for writing well on their own.

It has been shown that developing good planning skills can improve writing performance and promote long-term writing development

and self-regulation [7,29]. By the time students become more proficient in the planning process, they feel elevated confidence and autonomy that allow them to engage with a range of genres, audiences, and purposes with greater efficacy [29]. Consequently, the integration of explicit instructions on planning strategies into writing curricula is imperative for skilled, autonomous writers who can competently manage complex and diverse writing tasks [20,33].

The results from meta-analyses demonstrate that interventions based on Self-Regulated Strategy Development (SRSD) yielded significant improvements in writing accuracy, organisation, and motivation. For instance, studies indicate that using a writing plan improves writing quality [22,34,35]. The current study is built upon a foundation of methodological precision via effect size calculation and meta-analyses, and it is also argued that the present study highlights the significance of aligning intervention design with ongoing evaluation to ensure both effectiveness and scalability. This approach ensures clarification in diversified educational contexts.

In connection with this study, the increasing importance of keyboard typing in schools must also be mentioned, as the texts in this study were also written on a computer keyboard. Keystroke logging has been studied particularly intensively in recent years. It allows every keystroke and mouse click made by the writer to be recorded [36]. For example, the *Inputlog* programme records keystrokes, pauses, corrections, mouse/window changes in millisecond resolution [37]. Time stamps for these activities are also reported to indicate when each event occurs and how long it lasts. In their study, Tian and Cushing summarize findings from previous years that indicate a positive influence of keystroke logging on various aspects: From the students' perspective, this approach offers the opportunity to focus more on the writing process and less on the final product, which can have a positive effect on their self-regulation skills [38]. From the teachers' perspective, keystroke logging provides a chance to reflect on the writing process together with the students and to teach writing strategies. It also gives teachers insight into their students' writing behaviour [38]. The method presented in this study for measuring writing process data using the *Gorilla Experiment Builder* can be seen as an alternative to the keystroke logging method. More details about this programme will be given in Chapter 3.2.1. Both methods have in common that they relate to cognitive writing process data and provide insight into the complex and dynamic processes that occur during writing.

3. Materials and Methods

3.1. Research Design

Based on the theoretical assumption that cognitive flexibility and recursive planning are crucial for successful writing (see Chapter 2), this study investigates how these components manifest in students' interactions with a digital writing tool. The central goal of the sub-study aims to focus on the influence of planning activities on the text quality of students of different groups in ninth grade at six German higher- and lower-academic-track schools.

The sub-study is embedded in a broader study and research project. Detailed information on the research design and interventions of this research project can be found in Giera et al. (2025) [16]. However, the focus of this paper is not on investigating the effect of a treatment, but on the measured writing process data of the students. Therefore, only the most important information on the research design of the project is provided.

In this study, a quasi-experimental design was chosen to maintain an authentic learning context. In Germany, the school system is differentiated into different types after elementary school, which enables different educational pathways. Secondary schools of the higher academic track (HSS) provide a higher education entrance qualification in the twelfth grade. Secondary schools of the lower academic track (LSS) provide an intermediate school-leaving certificate in the tenth grade.

The participating schools were recruited via a survey of HSS and LSS schools in Brandenburg (Germany) that wanted to learn more about students' writing skills for argumentative texts. Argumentative texts are relevant in secondary schools in Germany for achieving educational qualifications for vocational and academic track and represent a cognitively demanding task for students [13,16]. When recruiting schools, the aim was to achieve a balanced ratio of HSS and LSS schools, which was also achieved: Three HSS and three LSS schools were recruited for the study. The responses from the schools resulted in the present sample of 6 schools with 357 students. HSS students wrote 614 texts and LSS students 447 texts (see Tab. 2).

The participating classes were further differentiated into intervention and control groups. To understand the background of this study, the following aspect is particularly relevant: unlike the control group, the intervention on argumentative writing included, for example, explicit work with the writing plan, which was also provided to the students in the subsequent surveys as an aid for writing the texts (see Chapter 3.2.1). The allocation of a class to an intervention or control group was carried out in consultation with the cooperating teachers. When allocating students to the intervention and control groups, care was taken to ensure a balanced ratio between the two groups. In addition, it was agreed with the teachers of the control groups that no writing plans would be used in their classes. The intervention groups were taught by members of the research team during the study. All lessons kept to their regular schedule.

Surveys were conducted at four different measurement points (t1-t4) as part of the study. T1 took place before the intervention, t2 during the intervention, t3 immediately after the intervention, and t4 as maintenance eight weeks after the intervention. The current sub-study employed a quasi-experimental design that was longitudinal without the involvement of random assignment. The main purpose was to observe naturally occurring differences in students' switching behaviour, writing and planning time across existing classroom groups rather than to manipulate variables experimentally. There was no possibility of individual

randomization within authentic classroom setting, however the involvement of treatment and control groups gave rise to structured group-based comparison across several schools. The four measurement points (pre-test, intermediate, post-test and follow-up) in time from repeated measures resulted in longitudinal element of the design. During the study course, the total sample size (357 students, 1,061 texts) made it possible for sufficient coverage for reliable group-level analyses.

Except for the treatments and type of school, further differentiation of the sample (e.g., in terms of age or gender) is not possible due to the anonymized data collection, a requirement of the Ministry of Education to carry out this study. To ensure the exclusion of any reference to individual students and their data, the students created a five-letter code consisting of two capital letters, two digits, and a unique character before the first data collection (pre-test). Neither the research team nor teachers knew about these identification numbers during the process. The students are, therefore, the only people who know their respective codes. At this point, it is important to emphasize the statement from Chapter 1.2. that all students, regardless of gender and other personal data, must take exams at the end of the tenth grade. Therefore, this limitation is not essential for the overall objective of this study.

3.2. Methods

3.2.1. Gorilla Experiment Builder

The participating students, used the software *Gorilla Experiment Builder*, this software was chosen for the study because the independent survey design could be implemented free of charge without extensive programming skills. In addition, the survey used *Gorilla Experiment Builder*, which is compliant with EU data protection directives [39]. The reference to *Gorilla Experiment Builder* as a digital online platform is documented in numerous international studies between 2018 and 2023 [39,40]. For data collection, the writing tasks (see Appendix A.1 to A.4) were integrated into the back end of Gorilla Experiment Builder. A task environment was created for the students to work on this writing task, which enables them to enter their argumentative text digitally. This task environment consists of the following two components:

Firstly, the students had the writing task and a field in front of them in which they could enter their text. The students always had the task in front of them during the survey while they were entering the text.

Secondly, the students could also switch to a digital writing plan. This shows the structure of pro and con argumentation (see Table 1). The students could fill in individual fields of this writing plan according to their needs (e.g., in sentences or bullet points).

Introduction	
Occasion	
Introduction to the topic	
Main Part	
Claim	•
1st Argument (pro or con)	•
Example/ evidence	•
2nd Argument (pro or con)	•
Example/ evidence	•
3rd Argument (pro or con)	•
Example/ evidence	•
Counterclaim	•
1st Argument (pro or con)	•
Example/ evidence	•
2nd Argument (pro or con)	•
Example/ evidence	•
3rd Argument (pro or con)	•
Example/ evidence	•
Conclusion	
Opinion	•
Summary	•
Recommendation	•

Table 1: Structure of Writing Plan

In this way, an analogue task environment, which is common for school exams, was translated into a digital one with the help of *Gorilla*. In *Gorilla*, writing tasks can be easily uploaded and integrated into a survey. These can then be linked to supporting scaffolds such as writing plans on a digital interface, with *Gorilla* counting the number of times users switch between these different windows during the survey (see below). For the present study, the writing plan (see Table 1) and the writing tasks (see Appendix A) were integrated into *Gorilla's* backend. Thus, the students were able to switch between the two parts of the task environment (the writing task with text field on the one hand and the writing plan on the other) as often as they wished during working time. The previous notes in both the text field and the writing plan were saved.

This structure of the task environment makes it possible to record students' writing and planning activities with the help of *Gorilla*. Firstly, *Gorilla* measures the time the students spent in the writing field, i.e., how much time they invested in the actual writing of the text (variable: time on writing). Technically speaking, *Gorilla* measures how long the cursor of a student's computer remained in the writing field during the survey. Secondly, *Gorilla* can be used to measure for how long the students used the writing plan (variable: time on planning). Similar to the writing field, the length of time the cursor remains in the writing plan during the survey is also measured. In addition, the number of switches is measured. how often the students switched between the two windows, the text field with the writing task and the writing plan. Regarding this variable, *Gorilla* counts every switch between the writing field window, including the writing task, and the writing plan window.

Gorilla provides the data as Excel files from which the values for the relevant variables (time on writing and planning and number of switches) must be extracted and prepared for data analysis (e.g. with SPSS). While *Gorilla* provides the number of switches directly, the software initially displays the measured time in the writing field and in the writing plan in milliseconds. These must be converted into minutes for a better understanding of the data (see Table 3).

To conduct the survey with *Gorilla* in the manner described above, only a few basic requirements are necessary: an investment of one Dollar per participant for a token and a digital device with internet access so that the students can submit their text. In addition, researchers must provide a text box for entering text via *Gorilla*, as well as text boxes for filling out the writing plan for the students.

Both the time spent on the writing plan and the number of switches provide information about the students' planning activities and the extent to which they used the writing plan to write their own texts. The number of switches indicates a predictor for cognitive flexibility, mentioned in Chapters 1 and 2.

3.2.2. Measuring Text Quality

In general, text quality was measured in the course of a longitudinal study at four different measurement points using four different

writing tasks (see Appendix A.1–A.4). In all cases, the writing tasks are standardised writing exercises that have been used in this form in authentic examinations in German at the end of the tenth grade. All of these argumentative writing tasks had a similar structure: the students were required to write both pro and con arguments on a particular issue. One argument for each position is given in a speech bubble (see Appendix A.1–A.4).

Moreover, the text quality at all four measurement points was assessed at the following levels: holistic and analytical text quality based on language pragmatics (e.g., text content organisation and structure). All anonymized pro and con arguments ($n = 1,061$) were assessed by two raters through double-blind peer review using and adapting IMOSS (“Integrated Model of School Writing”) encoding according to the writing task (pro and con argument essay) [41]. This instrument consists of, firstly, one item to assess global text quality and, secondly, five items for measuring language pragmatics. The items for language pragmatics include 1. content organisation (implementation of the task), 2. text structure (e.g., use of paragraphs), 3. coherence of arguments and persuasiveness, 4. addressing the reader and style, and 5. vocabulary. This coding method was based on a 5-point scale (0 = low, 5 = high text quality). For the language pragmatics, the scores of the five individual items for each participant were summarised to form a mean score. The development of the IMOSS coding procedure is based on the experience of the DESI “German English Student Performance International” study, a German large-scale study, in which the quality of 10,056 texts (t1) and 9,844 texts (t2) of ninth-grade students in German and English was assessed both holistically and analytically [42].

The reliability for this coding method is at an appropriate level with Cronbach's $\alpha = .90$. Cohen's Kappa measures the inter-rater reliability. It is in a satisfactory range ($\kappa = .727$) for holistic text quality. However, it does not consider texts declared to be missing data because they do not fulfil the task or are too short (less than two sentences). Regarding these texts, however, it is unclear whether the participating students were not motivated or did not have the necessary skills to write an argumentative text. In order to avoid completely excluding students who failed their writing assignments from data analysis, thereby gaining more statistical power, the IMOSS scale was expanded to a 6-point scale for the present study (1 = failed text, 6 = high text quality). In the analysis, task text scores of 1 (failed text) were both included and excluded to demonstrate their effect on the overall evaluation; the inclusion of scores of 1 was considered to be crucial to ensure the integrity of the dataset. With this approach, the comprehensive view of the data distribution and variability was analyzed. The inclusion of scores of 1 was also significant in terms of making a comparison between the ranges of students' performance, especially for struggling ones. Finally, scores of 1 were included to provide baseline knowledge about the effectiveness of interventions; the analysis carried out that included scores of 1 allowed for more selective data interpretation.

From a psychological point of view, using a 6-point scale lessens the central tendency bias, where extreme rating categories are avoided. It aids in clearly representing raters' judgement, thus helping to produce richer data with potentially varied responses and allowing a more nuanced assessment of attitudes and perceptions.

Two raters were trained for the rating procedure, and they were same throughout the study period, they were two experienced German teachers. Rijlaarsdam et al. emphasize, the independent training of raters with a reliable coding procedure is a prerequisite for a rating procedure that causes the fewest errors possible [43]. To meet this requirement, the raters were first trained independently by the research team to rate the argumentative texts. This involved explaining and applying the rating criteria described above using 50 sample texts from measurement point t1. During this, the ratings of both raters were disclosed after this initial rating and compared

with each other. The ratings of texts that differed greatly between the two raters (more than one level) were discussed and adjusted, resulting in benchmark texts for each quality level that served as a guide for the raters in the subsequent double-blind rating. Further criteria for the rating procedure were taken into account by defining benchmark texts prior to the rating for higher inter-rater reliability and by both raters coding the texts both holistically and analytically in the same procedure. The raters followed the instructions from the IMOSS coding procedure and started with holistic text rating, followed by the rating of language pragmatics. According to this rating procedure, all texts were double-blind-rated [41].

With the aim of better adapting the rating procedure, Table 2 below summarises the rating criteria.

Overall Impression	Language Pragmatics	Language Systematics
Evaluate the holistic quality of the text. It can be classified as a normative rating	Sub-criteria for rating procedure	
	Content Design Assess to what extent the text fulfills the task requirements. Text Structure Analyze paragraph division into introduction, main body, and conclusion, which can be realized through the insertion of paragraphs and the use of text procedures. Coherence/Persuasiveness Examine how the introduced topic of the task unfolds in the text, the text's semantic-thematic quality within its deep structure, and the linking of arguments and examples. Audience orientation/style Examine the linguistic style of the respective text types (formal vs. informal context). One of the decisive factors here is whether the text is intended for a specific audience (who is the text aimed at?). Vocabulary Assess how varied the vocabulary is within the text.	Sentence Construction Structure of the sentences. Spelling Examine spelling errors in the text. Grammar Refers to errors in conjugation, declension, tense, case, mode, and prepositional structure. Punctuation Appropriate usage of punctuation.

Table 2: Overview of The Dependent Variables for Text Quality

3.3. Research Questions

The following key points emerge from the considerations in the first three chapters for this study: To improve students' writing skills, it appears worthwhile to encourage their planning activities while writing. This can be achieved, for example, by providing them with a writing plan that supports them in writing a text. This study aims to demonstrate how this goal can be achieved by allowing students to use a digital writing plan when writing argumentative texts. This plan is integrated into a digital task

environment that also contains the writing task and a text-input field. The study aims to gain insights into which factors of writing and planning activities are particularly relevant for predicting text quality and whether differences can be identified between students of different school types. These objectives lead to the following central research questions:

RQ1: Do time on writing, time on planning, and number of switches predict text quality among ninth-grade students writing

argumentative texts?

H1: It is assumed that all three variables predict text quality because they indicate the level of cognitive flexibility of students during the writing process.

RQ2: Are there significant differences in time on writing, time on planning, and number of switches across school types (higher and lower academic tracks)?

H2: It is assumed that students in the higher academic track spend more time on planning and switch more often between the editor field and writing plan than their counterparts in lower-academic-track schools. This assumption is due to the higher level of planning activities displayed by more skilled writers.

4. Results

4.1. Data Analysis

This section will be focused on the testing of the hypotheses and research questions mentioned (in Chapter 3) by using different statistical analyses, starting from the RQ1, followed by correlation and regression analysis, and then moving towards the RQ2, that was proceeded with descriptive statistics (see Table 3), independent samples t-test, and One-Way ANOVA.

The amount of variance for school, class and participant levels was determined by conducting preliminary multilevel analyses. It was indicated by the findings that 38% of the total variance in text quality was found between the participants, 16% at the class level, and only about 6% at the school level. The focus of the study was on between-group differences rather than hierarchical dependency, the between-school variance was minimal, the analyses used conventional parametric tests (t-tests, ANOVAs, and multiple regression). This approach conserved analytical transparency while also preserving the statistical validity.

SPSS 28 was used for descriptive statistics, being followed by multi-level models used primarily for variance decomposition and exploratory insight, not for hierarchical dependency modelling. Despite the presence of a hierarchical or nested structure in the data (for example, students nested within schools), the primary focus of this study was on conducting between-group comparisons rather than explicitly modelling the multi-level dependence. The experimental design comprized different groups that were not repeatedly measured within nested structures in the way being requiring multi-level modelling.

The analysis revealed that most of the variance in the outcome variable, text quality, is observed between participants and learning groups. However, the minimal variance was observed to be consistent with the schools. Notably, the total participant variance was 38%: 16% was attributed to class and only 6% to school, which showed that school could be excluded as a level if needed. In a comparison using only the learning group levels, the model with three levels (participants as level 2 and class as level 3) explained significantly more variance. The context

variable accounted for 36% of the total residual variance, while other individual differences, including those caused by the training measures, accounted for 64% of the error variance.

The primary objective of this study was to examine variations in the development of argument text quality with respect to differences between times on writing and planning in terms of the number of switches being hypothesized to be observed differently among LSSs and HSSs. The analyses of independent sample t-tests and one-way ANOVA were deemed appropriate for the observation of the differences in all three perspectives (time on writing, time on planning, and number of switches) in terms of specified groups.

Furthermore, the evaluations from the assessment highlighted that the variance attributed to the nested structures (i.e., difference among schools) was comparatively small and has no impact on the outcome variables. Thus, while taking into consideration analytical simplicity and interpret-ability the use of simplified parametric tests, such as ANOVA and t-test, provided a concise and statistically valid and acceptable ways to observe the group differences and developmental changes. In order to test the hypotheses and answer the research questions, multi-level analytical technique was taken into consideration. To handle missing data (<5%) and allowing all available data to be taken into account without considerably impacting the results, full information maximum likelihood was employed.

4.2. RQ1—Do time on writing, time on planning, and number of switches predict text quality among ninth-grade students writing argumentative texts?

4.2.1. Correlational Analyses

Pearson's correlation coefficients provided initial insights into the relationships among the key variables. The number of switches were significantly positively correlated with active writing time ($r = 0.213$, $p < 0.001$). This moderate correlation suggests that students who engaged in more frequent switching tended to spend longer durations actively writing, possibly indicating a more dynamic or engaged approach to composing.

More importantly, both the number of switches and time on writing showed strong positive correlations with holistic text quality (holistic text quality assumed to be a composite measure of writing quality), with correlation coefficients of $r = 0.529$ and $r = 0.513$, respectively (both $p < 0.001$). These findings suggest that students who switched more frequently and spent more active time on their writing tended to produce higher-quality texts, potentially reflecting a link between cognitive flexibility, sustained effort, and successful writing outcomes. In order to explore group differences more explicitly, both one-way ANOVA and independent sample t-tests were used to compare the active writing time across the different groups.

4.2.2. Regression Analysis

A hierarchical multiple regression analysis was carried out in order to examine the extent to which students' time on writing, number of switches and other contextual factors predicted the overall text

quality. The detailed summary of the analysis can be observed in Table 4 that gives detailed model summary for each analysis. In model 1, Time on writing and number of switches were entered as predictors in the first model. The variance explained in this model for text quality was 44.7%, $R^2 = 0.45$, $F(2, 1058) = 428.44$, $p < .001$. The number of switches ($B = .051$, $SE = .003$, $\beta = .439$, $t = 18.79$, $p < .001$, 95 % CI [.046, .057]) and time on writing ($B = .051$, $SE = .003$, $\beta = .420$, $t = 17.94$, $p < .001$, 95 % CI [.046, .057]) both were statistically significant. Time in planning was not included in this analysis and because of its weak and non-significant association with the text quality rather was explained well with the previous analyses.

When treatment group and school type were included in the second model, the explained variance rose to 50.7 % ($\Delta R^2 = .06$, $F(4, 1056) = 271.43$, $p < .001$, adjusted $R^2 = .505$). School type was a positive predictor of text quality ($B = .680$, $SE = .060$, $\beta = .243$, $t = 11.25$, $p < .001$), on the other hand, there was non-significant pattern observed for the treatment group ($B = -.018$, $SE = .036$, $\beta = -.011$, $t = -0.50$, $p = .617$). All the tolerance values were surpassed. The lake of multicollinearity was confirmed by variance inflation values of less than 1.1 and 0.9. It has been suggested by the findings that a minor contribution from the school type, greater text quality is connected to longer writing times and more frequent switching.

S. No	Predictor	Model 1 B	Model 1 β	Model 1 p	Model 2 B	Model 2 β	Model 2 p
1.	Constant	1.59	—	< .001	0.57	—	< .001
2.	Number of switches	0.05	.44	< .001	0.05	.43	< .001
3.	Writing time (min)	0.05	.42	< .001	0.05	.42	< .001
4.	School type	—	—	—	0.68	.24	< .001
5.	Treatment group	—	—	—	-0.02	-.01	.617
Model Summary							
S. No	Model	R	R ²	Adjusted R ²	F(df)	p	
1.	Model 1	.669	.447	.446	(2, 428.44 1058)	< .001	
2.	Model 2	.712	.507	.505	(4, 271.43 1056)	< .001	
Note. Dependent variable = MW_B001 (text quality). Model 1 includes writing time and number of switches. Model 2 adds school type and treatment group. B = unstandardized coefficient; β = standardized coefficient; ΔR^2 = change in R^2 .							

Table 3: Hierarchical Multiple Regression Predicting Overall Text Quality

4.3. RQ2—Are there significant differences in time on writing, time on planning, and number of switches across school types (higher and lower academic tracks)?

4.3.1. Descriptive Statistics

The analysis is focused on providing insights into individual constructs being observed in the sample, monitoring the differences in the mean scores among different groups (see Table 3).

The behaviour and performance measures of the participants during the tasks are elaborated on with descriptive statistics derived from the various groups (see Table 3). In total, 357 students from ninth grade produced 1,061 texts across the four time points. Each cell in Table 3 presents one learning group nested within either higher-academic-track schools or lower-academic-track schools. The smaller numbers presented in Table 3 (e.g., 38–60 cases per cell) reflect the number of texts collected within each subgroup and testing phase, not the total participant count. Thus, the dataset comprised multiple texts per participant, explaining the difference between the number of students and the number of observations.

The distributions for the main observable variables (time in writing, time in planning, and the number of switches) were checked for normality. The values of skewness and kurtosis indicated moderate right-skewness but remained within acceptable limits ($|\text{skew}| < 2$, $|\text{kurtosis}| < 3$) for parametric analyses. Approximate normality was

confirmed by visual inspection of histograms; therefore, no further transformations were required.

In the pre-test phase, participants in lower-academic-track schools (LSS) spent an average of 49.05 minutes on writing ($SD = 11.13$), while those in higher-academic-track schools (HSS) spent 46.90 minutes ($SD = 12.05$). The skewness values (0.37 and 0.31) and kurtosis values (0.74 and 0.68) indicate slight right-skewness with minimal outliers.

During the intermediate test, the mean writing time for LSS students was 48.92 minutes ($SD = 10.90$), compared to 45.70 minutes ($SD = 12.11$) for HSS students. The post-test revealed average writing times of 46.56 minutes ($SD = 13.14$) for LSS and 46.90 minutes ($SD = 13.10$) for HSS participants, showing relatively stable engagement across groups. In the follow-up phase, LSS students spent an average of 48.30 minutes ($SD = 12.30$) on writing, while HSS students averaged 47.45 minutes ($SD = 12.40$), with distributions remaining within acceptable skewness and kurtosis ranges.

For planning time, the LSS group showed an average of 36.14 minutes ($SD = 14.15$) at pre-test and 42.49 minutes ($SD = 12.48$) at the intermediate stage, whereas the HSS group showed corresponding averages of 37.45 minutes ($SD = 13.10$) and 41.30

minutes (SD = 12.87). At the post-test, mean planning times were 41.74 minutes (SD = 16.08) for LSS and 41.42 minutes (SD = 19.35) for HSS students. During the follow-up, averages were 39.29 minutes (SD = 14.13) for LSS and 41.65 minutes (SD = 15.67) for HSS, suggesting consistent engagement with planning activities across time points.

Regarding switching behaviour, LSS students averaged 8.17 switches (SD = 8.48) at pre-test and 12.84 switches (SD = 8.73) at the intermediate phase, while HSS students averaged 10.18 (SD = 8.95) and 11.73 (SD = 8.52), respectively. At the posttest, LSS

students showed an average of 14.13 switches (SD = 7.72) and HSS students 13.02 switches (SD = 7.79). During the follow-up, switching increased to 17.45 (SD = 9.10) in the LSS group and 15.67 (SD = 8.64) in the HSS group.

These results indicate that switching frequency increased across all phases, with HSS students generally performing slightly more switches than their LSS counterparts, particularly in earlier phases. However, LSS students showed a stronger progression over time, indicating increased engagement and cognitive flexibility during the later writing stages.

	School Type	Time Point	<i>N</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Time in writing (min)	LSS	t1	38	49.05	11.13	0.37	0.74
		t2	43	48.92	10.90	0.36	0.70
		t3	36	46.56	13.14	0.32	0.79
		t4	48	48.30	12.30	0.29	0.88
	HSS	t1	60	46.90	12.05	0.31	0.68
		t2	54	45.70	12.11	0.32	0.71
		t3	50	46.90	13.10	0.33	0.78
		t4	56	47.45	12.40	0.29	0.84
Time in planning (min)	LSS	t1	38	36.14	14.15	0.30	0.74
		t2	43	42.49	12.48	0.26	0.71
		t3	36	41.74	16.08	0.34	0.73
		t4	48	39.29	14.13	0.33	0.89
	HSS	t1	60	37.45	13.10	0.32	0.68
		t2	54	41.30	12.87	0.30	0.70
		t3	50	41.42	19.35	0.33	0.76
		t4	56	41.65	15.67	0.29	0.83
Number of switches	LSS	t1	38	8.17	8.48	0.46	0.83
		t2	43	12.84	8.73	0.48	0.76
		t3	36	14.13	7.72	0.43	0.80
		t4	48	17.45	9.10	0.41	0.89
	HSS	t1	60	10.18	8.95	0.40	0.68
		t2	54	11.73	8.52	0.41	0.71
		t3	50	13.02	7.79	0.38	0.78
		t4	56	15.67	8.64	0.39	0.82

Note. LSS = lower-academic-track schools; HSS = higher-academic-track schools; t1 = Pre-test, t2 = Intermediate, t3 = Post-test, t4 = Follow-up. *M* = mean; *SD* = standard deviation. Values represent aggregated means across treatment groups. Skewness and kurtosis values confirm approximately normal distributions ($|\text{skew}| < 1.5$, $|\text{kurtosis}| < 2.0$).

Table 4: Descriptive statistics for time on writing, time on planning, and number of switches across school types and time points

4.3.2. One-Way ANOVA

The analysis was carried out for the following three conditions: DW (Debate Writing), WD (Writing Debate) and Control were compared for the number of switches and writing time using a one-way analysis of variance. There was no observable group difference in the analysis of the number of switches ($F(2, 1058) = 1.49$, $p = .225$, $\eta^2 = .003$). There was observable significant

difference for the writing time, $F(2, 1058) = 3.62$, $p = .027$, $\eta^2 = .007$. Students belonging to WD condition ($M = 17.06$ min, $SD = 11.44$) spent considerably more time writing than those in the Control condition ($M = 14.70$ min, $SD = 11.06$), according to post hoc comparisons (Tukey HSD), with a p -value of 0.02. DW did not differ significantly from the other two groups ($M = 15.91$ min, $SD = 11.36$).

Significant changes in the amount of time spent on planning were found across treatment groups using a one-way ANOVA ($F(2, 1058) = 8.12, p < .001, \eta^2 = .015$), suggesting a minor but statistically significant impact of the intervention. Students who utilized the writing plan during writing spent more time planning than those in the control group, according to post hoc testing (Tukey HSD) ($p < .01$).

4.3.3. Independent Sample T-Test

Independent-samples t-tests were performed to examine whether students from lower-academic-track (LSS) and higher-academic-track (HSS) schools differed in their number of switches and time on writing.

For *number of switches*, Levene's test indicated equal variances ($F(1, 1059) = 1.50, p = .22$), and the group difference was not significant, $t(1059) = -1.34, p = .18$, Cohen's $d = -0.08$ (95 % CI $[-0.21, 0.04]$). Students from HSS ($M = 14.96, SD = 11.54$) switched slightly more often than those from LSS ($M = 13.98, SD = 12.09$), but the effect was negligible.

For *time on writing*, the assumption of equal variances was violated ($F(1, 1059) = 13.81, p < .001$); therefore, Welch's test was used. The difference between LSS ($M = 15.95$ min, $SD = 12.31$) and HSS ($M = 15.82$ min, $SD = 10.54$) was not significant, $t(869) = 0.18, p = .86$, Cohen's $d = 0.01$ (95 % CI $[-0.11, 0.13]$).

In the independent-samples t-test, there was a slight but significant difference between the amount of time students in lower-academic-track schools ($M = 22.77$ min, $SD = 15.33$) and those in higher-academic-track schools ($M = 19.50$ min, $SD = 12.19$) spent planning ($t(1059) = 3.87, p < .001$, Cohen's $d = 0.24$).

The current findings suggested that the students belonging to both school types depicted similar patterns of engagement in terms of time in writing and switching behavior. The effect sizes from the results also suggested that the type of school had small real time impact on these behavioral measures.

5. Discussion

The current study focused on investigating the cognitive and behavioural factors that affect the overall text quality among students, with more focus on specific research questions: RQ1 whether time on writing and planning as well as the number of switches are predictors of text quality and RQ2 the differences among these variables across different school types and groups. The results provided insights into the interaction between engagement, flexibility, and contextual factors in writing performance.

This study set out to explore how cognitive flexibility and planning activities, as defined in the theoretical models of Hayes & Flower and further developed by Hayes, influence the quality of argumentative writing among ninth-grade students [6,7]. Building on the assumption that writing is not only a cognitive but also a socially situated and emotionally influenced activity, our findings provide important insights into how students engage with digital

tools that support planning during writing [1,8].

5.1. RQ1: The Roles of Writing Time, Planning Time, and Switching in Predicting Text Quality

The results of RQ1 confirm the foundational claims made in Chapters 1 and 2: students who exhibit cognitive flexibility through frequent switching between the writing plan and the text editor tend to produce higher-quality texts. This supports the presumption that good writing is deeply connected to metacognitive awareness and self-regulation [12]. In contrast, the mere duration of planning time seems to be not a significant predictor of text quality unless accompanied by frequent switching. This nuance highlights that it is not time alone but how students use the writing plan that matters—a critical distinction echoed in earlier work on executive functioning and working memory [5,24].

Overall, the findings align with the theoretical models presented in Chapter 2, particularly regarding the role of planning in writing and the value of adaptive task environments that encourage active engagement with writing processes [7,20].

It has been proven with both correlational and regression analyses that the writing time and number of switches are significant positive predictors of the overall text quality. This also aligns with the hypothesis stating that more engaged and flexible cognitive processes during writing contribute to higher-quality outcomes [23]. The findings from the study give rise to the assumption that students who spend more time engaged with writing and at the same time depict more cognitive flexibility in the form of frequent switching are more likely to produce texts that are more effective [24].

Findings were consistently revealed by the regression analysis: Increased duration on writing and greater switching behaviour both predict higher text quality. Moreover, time spent on planning did not appear to be the key predictor in the model, indicating that the amount of time spent on planning alone could not directly predict better writing outcomes. This supports Hayes' view that effective planning is not a discrete phase but a recursive process embedded within the act of writing itself [7]. It shows the importance of active engagement and flexibility during the writing process rather than just dedicating time to planning [29].

One reason for this observation could be that students who work extensively on a writing plan ultimately lack the time to transfer these insights to the final argumentative text. The number of switches, on the other hand, indicates the extent to which students manage to integrate their notes from a writing plan into the finished text. Switching between the two components of the task environment can be seen as an indicator of the commitment students invest in writing texts. It seems that the number of switches is a prerequisite for whether planning activities can actually be used effectively to improve text quality. In other words, the number of switches could be seen as an indicator of cognitive flexibility, which characterises good writers [9].

Future studies should investigate whether the number of switches alone is meaningful enough to predict an improvement in text quality. It would be interesting, for example, to examine how the number of switches develops over the course of students' writing processes—for example, whether they increase or decrease more at the beginning or end of the writing process. Another aspect that must be considered in this discussion is that students may well formulate their entire texts in writing plans and then copy them entirely into the text-input field. While this implies a low number of switches, it does not necessarily indicate poor writing effort or poor text quality.

Given the results, H1 can only be partially confirmed. While writing time and the number of switches can certainly predict text quality, this only applies to a limited extent to planning time. Planning time only has a positive influence on text quality if it is also associated with a high number of switches.

5.2. RQ2: Group Differences and Contextual Factors

RQ2 revealed surprising group differences. While students from lower academic-track schools spent more time on planning and writing, their texts were generally of lower quality. This suggests that scaffolding tools such as digital writing plans may be particularly valuable for this group, who likely have less prior experience with argumentative writing (see Chapter 1.2.). Their increased time investment may reflect effort spent and a reliance on external structures, while higher-track students—more accustomed to the genre—may internalise such structures and thus require fewer external aids. These findings reinforce the notion that instructional tools must be adaptable to student needs, especially when educational pathways diverge, as is the case in the German school system [17,18].

This study highlighted the significance of the environment in terms of writing behaviour by exploring the differences between school types. Students belonging to lower-academic-track schools seemed to spend more time on writing and planning than higher-academic-track students. This is surprising at first, given that LSS students produce lower-quality writing, and planning activities are generally associated with high writing skills. However, it is evident that students, who, due to curricular requirements, have less knowledge and skills in writing argumentative texts than their counterparts from higher-academic-track schools, accept this scaffold in the form of a writing plan. The writing plan appears to support the writing process; it visualizes the structures of pro and con arguments, which students can use as an initial basis for developing their texts (see Table 1). Students on the higher academic track appear to use this support less due to their greater experience with the pro and con argumentation genre, instead writing more directly in the text-input field.

The switching behaviour highlighted significant group differences but with small effect sizes, underlining the role of the intervention and school environment is modest in affecting students' cognitive flexibility while writing. The contextual variations seems to emphasize the consideration of environmental and instructional

factors while designing strategies to improve writing quality.

The results from descriptive and inferential statistics suggested visible differences in both constructs, i.e., time on writing and time on planning across school types. It was indicated by the findings that there is observed variability in the measures in some specific groups, such as the follow-up group, which depicted greater average times and a bigger spread. This may be an indicator of elevated cognitive load and greater engagement over time. Statistical findings suggested that some specific groups, especially the follow-up group, saw a prominent increase in the duration of time allocated to both planning and writing, indicating that the progression through study phases exhibited a significant influence on the time allocated to these tasks.

Furthermore, the inconsistencies among school types were quite obvious, with participants from specific school types showing longer or shorter durations, which may be linked to various factors such as differences among the instructional approaches, familiarity with the writing tasks, or prior exposure. These findings mirror the systemic disparities outlined in Chapter 1.2. and suggest that students in lower academic tracks may rely more heavily on external scaffolding to compensate for limited prior writing experience.

6. Conclusion

This research explored the complex relationship between teaching methods, students' behaviour, and their writing outcomes on the basis of different writing tasks in different school types. The results indicate that teaching strategies and school types both have a large impact on writing quality, especially the amount of time spent on writing and planning. The fluctuations in the writing and planning times across different groups of participants suggest how specific instructions can make a valuable difference to the way people are thinking and perceiving while writing. The analyses also shows that switching between ideas or sections can also have a significant impact on the quality of texts that students write. It seems to depend directly or indirectly on the time spent on planning. More switching behaviour was linked to higher time on planning leading to lower text quality. The differences among various schools in terms of time spent on planning and writing reflect the significance of school environments in affecting the way students write. This study also suggest that sometimes students require some extra help and guidance to perform better in the classroom setting.

It shows that they need a task environment that enables them to effectively use planning activities to improve the quality of their writing. The study demonstrated an exemplary setting in which students were able to use a digital writing plan to integrate notes into their draft texts. This showed that it is not just the time spent in the writing plan that is crucial, but the extent to which students are able to link the text and writing plan and to transfer notes from the writing plan into the text. This ability was highlighted in the first chapter as cognitive flexibility, and over the course of the study it became clear that providing a digital writing plan helps students, especially those on the lower academic track, to better use this

ability to improve their texts.

6.1. Limitations

A different method of collecting data may have been needed if the data were to be collected by a researcher instead of a teacher. The results would more likely be different if there was a larger sample size, and the generalizability of the data would be increased.

This study mainly focused on text quality (no other personal data were collected), because the text was assessed based on this domain. However, mean language systematics and pragmatics, which focused, e.g., on sentence construction, spelling, punctuation, grammar, text structure, and content quality, were excluded from this analysis. They were not the centre of attention for this study.

Furthermore, it was only possible to identify 44 participants who took part at all 4 measurement points with the same anonymized code (of which only 7 were from lower secondary schools; see Chapter 3.1). For this reason, individualized progress diagnostics are not possible for all other students. However, during the project, all the field participants remained stable, and the same learning groups were maintained throughout, so analysing data in Chapter 4 was possible.

Finally, it should be noted that this pilot project was carried out in 2021 and 2022 during the COVID-19 pandemic. With a few exceptions, all lessons could be held in the classrooms of the schools on site. Sometimes, lessons had to be held online due to school requirements. This being the case, there is a need for further research with a bigger sample to increase the data power and combine it with a quality research approach to analyse the individual learning paths of learners.

6.2. Educational Implications

This study contributes to the growing body of literature that emphasises the importance of meta-cognitive strategies and cognitive flexibility in writing instruction. In line with the theoretical background outlined in Chapters 1 and 2, the results suggest that successful writing instruction—particularly in genres such as argumentative writing—requires more time spent on the task. Additionally, it requires intentional support structures that help students actively navigate between planning and composing.

The use of a digital writing plan emerged as a particularly promising approach. It seems to create a task environment that enabled students to integrate their planning into the writing process effectively—something not easily achievable in handwritten exam formats. This integration was apparently most successful when students switched frequently between the planning and writing interfaces, which we interpret as an indicator of cognitive flexibility, a trait closely tied to writing quality in prior research [9].

Educationally, these findings highlight the need to view writing plans not as an assessable product but as a dynamic tool for supporting cognitive processes during writing. Especially for

students in lower academic tracks, structured digital environments offer an important means of compensating for prior deficits in genre knowledge and writing fluency.

In conclusion, it seems advisable to support argumentative writing not only by increasing the planning time alone but also by encouraging the strategic and integrated use of planning tools. This aligns directly with the theoretical and empirical claims discussed in Chapters 1 and 2 and indicates a strong basis for further research and pedagogical innovation in writing education.

Considering the educational challenges identified in Chapter 1, these findings advocate for a systematic integration of digital planning tools into writing instruction, particularly in schools serving academically diverse student populations. The current study indicates that time on planning alone is not enough to predict text quality [43-45].

References

1. Graham, S., & Harris, K. R. (2021). Evidence-based practices in writing instruction: A meta-analysis and synthesis. *Journal of Educational Psychology*, 113(6), 859–878.
2. Applebee, A. N., & Langer, J. A. (2011). EJ Extra: A snapshot of writing instruction in middle schools and high schools [free access]. *English journal*, 100(6), 14-27.
3. UNESCO (2023). The futures we build: abilities and competencies for the future of education and work. In *UNESDOC Digital Library*. UNESCO.
4. Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of writing research*, 1(1), 1-26.
5. Berninger, V., & Swanson, H. L. (1994). Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writing. In E. C. Butterfield (Ed.), *Children's writing: Toward a process theory of the development of skilled writing*. JAI Press.
6. Hayes, J. R., & Flower, L. S. (2016). Identifying the organization of writing processes. In *Cognitive processes in writing* (pp. 3-30). Routledge.
7. Hayes, J. R. (2012). Modeling and remodeling writing. *Written communication*, 29(3), 369-388. <https://doi.org/10.1177/0741088312451260>
8. Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational psychology review*, 18(4), 315-341.
9. Kuhn, D. (2022). Metacognition matters in many ways. *Educational Psychologist*, 57(2), 73-86. <https://doi.org/10.1080/00461520.2021.1988603>
10. OECD (2019). OECD future of education and skills 2030. Conceptual learning framework. *Attitudes and values 2030*. https://www.oecd.org/content/dam/oecd/en/about/projects/edu/education-2040/concept-notes/Attitudes_and_Values_for_2030_concept_note.pdf
11. European Union. (2018). *Key competences for lifelong learning*. Publications of European Union. <https://>

- op.europa. eu/en/publication-detail/-/publication/297a33c8-a1f3-11e9-9d01-01aa75ed71a1/language-en/format-PDF/source-107520914
12. McNamara, D. S., & Kendeou, P. (2022). The early automated writing evaluation (eAWE) framework. *Assessment in Education: Principles, Policy & Practice*, 29(2), 150-182
 13. Ferretti, R. P., & Graham, S. (2019). Argumentative writing: Theory, assessment, and instruction. *Reading and Writing: An Interdisciplinary Journal*, 32(6), 1345–1357. <https://doi.org/10.1007/s11145-019-09950-x>
 14. Leitão, S. (2003). Evaluating and selecting counterarguments: Studies of children's rhetorical awareness. *Written Communication*, 20(3), 269-306.
 15. Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs (2022). Bildungsstandards für das Fach Deutsch. Erster Schulabschluss (ESA) und Mittlerer Schulabschluss (MSA).
 16. Giera, W.-K., Deutzmann, L., & Sheikh Muhammad, S. (2025). Merging Oral and Written Argumentation: Supporting Student Writing Through Debate and SRSD in Inclusive Classrooms. *Education Sciences*, 15(11), 1471. <https://doi.org/10.3390/educsci15111471>
 17. Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs (2023). *Basic Structure of the Education System in the Federal Republic of Germany*. Diagram.
 18. Müller, N., & Busse, V. (2023). Herausforderungen beim Verfassen von Texten in der Sekundarstufe—Eine differenzielle Untersuchung nach Migrationshintergrund und Familiensprachen. *Zeitschrift für Erziehungswissenschaft*, 26(4), 921-947.
 19. Graham, S., & Perin, D. (2007). Writing next-effective strategies to improve writing of adolescents in middle and high schools.
 20. Graham, S., & Harris, K. R. (2009). Almost 30 years of writing research: Making sense of it all with The Wrath of Khan. *Learning disabilities research & practice*, 24(2), 58-68.
 21. Graham, S., & Harris, K. R. (2005). Improving the writing performance of young struggling writers: Theoretical and programmatic research from the center on accelerating student learning. *The journal of special education*, 39(1), 19-33.
 22. Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary educational psychology*, 30(2), 207-241.
 23. Limpo, T., Alves, R. A., & Fidalgo, R. (2014). Children's high-level writing skills: Development of planning and revising and their contribution to writing quality. *British Journal of Educational Psychology*, 84(2), 177-193.
 24. Sun, T., Wang, C., & Wang, Y. (2022). The effectiveness of self-regulated strategy development on improving English writing: Evidence from the last decade. *Reading and Writing*, 35(10), 2497-2522.
 25. Mason, L. H., Harris, K. R., & Graham, S. (2011). Self-regulated strategy development for students with writing difficulties. *Theory into practice*, 50(1), 20-27.
 26. Zimmerman, B. J., & Kitsantas, A. (2014). Comparing students' self-discipline and self-regulation measures and their prediction of academic achievement. *Contemporary educational psychology*, 39(2), 145-155.
 27. Bruning, R., & Horn, C. (2000). Developing motivation to write. *Educational psychologist*, 35(1), 25-37.
 28. McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of educational psychology*, 86(2), 256.
 29. Graham, S., & Harris, K. R. (2013). Designing an effective writing program. *Best practices in writing instruction*, 2, 3-25..
 30. Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2), 199-218.
 31. Britton, J. (1970). *Language and learning* (pp. 11-18). London: Allen Lane.
 32. Bransford, J. D., & Stein, B. S. (1993). The IDEAL problem solver.
 33. Harris, K. R., & Graham, S. (1996). Making the writing process work: Strategies for composition and self-regulation. (No Title).
 34. Washburn, E., Sielaff, C., & Golden, K. (2016). The use of a cognitive strategy to support argument-based writing in a ninth grade social studies classroom. *Literacy Research and Instruction*, 55(4), 353-374.
 35. Chai, C. (2006). Writing plan quality: Relevance to writing scores. *Assessing Writing*, 11(3), 198-223.
 36. Vandermeulen, N., Leijten, M., & Van Waes, L. (2020). Reporting writing process feedback in the classroom using keystroke logging data to reflect on writing processes. *Journal of Writing Research*, 12(1), 109-139.
 37. Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, 30(3), 358-392.
 38. Tian, Y., & Cushing, S. T. (2025). Exploring the application of keystroke logging techniques to research in second language (L2) writing. *Research Methods in Applied Linguistics*, 4(1), 100179.
 39. Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior research methods*, 53(4), 1407-1425.
 40. Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior research methods*, 53(4), 1407-1425.
 41. Neumann, A., & Matthiesen, F. (2011). *Indikatorenmodell des schulischen Schreibens: Test-dokumentation 2009, 2010 (DidaktikDiskurse; Vol. 5)*. Leuphana Universität Lüneburg.
 42. Neumann, A. (2012). Advantages and disadvantages of different text coding procedures for research and practice in

- a school context. In *Measuring writing: Recent insights into theory, methodology and practice* (pp. 33-54). Brill.
43. Rijlaarsdam, G., Van den Bergh, H., Couzijn, M., Janssen, T., Braaksma, M., Tillema, M., Van Steendam, E., & Raedts, M. (2012). Writing. In K. Harris, S. Graham, T. Urdan, A. G., Bus, S. Major, & H. L. Swanson (Hrsg.), *APA educational psychology handbook*, Vol. 3. Application to learning and teaching. (pp. 189–227). American Psychological Association.
44. Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 1–27). Lawrence Erlbaum. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780203811122-2/new-framework-understanding-cognition-affect-writing-john-hayes>
45. McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational psychology review*, 8(3), 299-325.
46. Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional science*, 37(4), 375-401.

Appendix A

A.1 Writing Task A (Pre-Test, t1)

Screen or paper – How do you read today?

The student representatives of your school propose that the entire book stock of your school to be converted to e-books. At the school conference, concerns are expressed about this.

You take this as an opportunity to discuss the problem in more detail in an article for the school newspaper.

First read the following opinions!



Now write the article in the form of a written argumentation.

Fill in the writing plan on the next page!

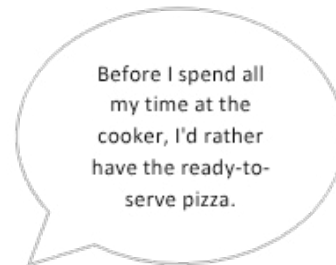
Source: MBJS & SenBJF (2017). *Schriftliche Prüfungsarbeit zur erweiterten Berufsbildungsreife und zum mittleren Schulabschluss 2017 im Fach Deutsch. Donnerstag, 4. Mai 2017, 28.*

A.2. Writing Task B (Intermediate Test, t2)

Fast Food or cooking by yourself?

There is a project at your school about healthy eating. The magazine "The Little Gourmet" is very interested in your results. You take this as an opportunity to write an article, in which you weigh the advantages of home-cooked food and fast food against each other.

First read the following opinions!



Now write the article in the form of a written argumentation.

Fill in the writing plan on the next page!

Source: MBJS & SenBJF (2017). *Schriftliche Prüfungsarbeit zum mittleren Schulabschluss 2013 im Fach Deutsch*. Dienstag, 16. April 2013, 24.

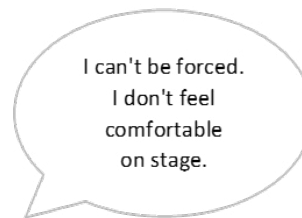
A.3. Writing Task C (Post-Test, t3)

Scenic Play compulsory for all students?

In some states, there is a debate about whether the subject of performing (theatre lessons) should be made compulsory for all students from grade 7 onwards. The magazine "Curtain up!" has set up a discussion forum for which you should write a contribution to the discussion.

To prepare the article, you have collected opinions on the topic in your school.

First read the following opinions!



Now write the article in the form of a written argumentation.

Fill in the writing plan on the next page!

Should everybody hold an honorary office?

On your last hiking day, you observed a rescue mission by the volunteer fire brigade. They work for free. Not only in the fire brigade, but also in sport, health care health care, nature conservation and childcare people are active on a voluntary basis. Since they do this work in their free time and without pay, this is a controversial topic in your class.

You take this as an opportunity to discuss the different positions in an article for the school newspaper.



Now write the article in the form of a written argumentation.

Fill in the writing plan on the next page!

Source: MBJS & SenBJF (2017). *Schriftliche Prüfungsarbeit zur erweiterten Berufsbildungsreife und zum mittleren Schulabschluss 2017 im Fach Deutsch. Mittwoch, 29. Mai 2017. Nachschreibetermin, 28.*

Copyright: ©2025 Lucas Deutzmann, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.