

Creating Process Indicators from Log Data to analyze Primary School Students' Multiple Text Comprehension

Norbert Erdmann
nwmerd@utu.fi

Kasper Rautio
Kasper.h.rautio@utu.fi

Department of Teacher Education
University of Turku, Finland

Abstract

In today's world learning often takes place on the Internet, even in primary education (Florit et al., 2020; Mason, 2018). The aim of this study is to create indicators from log data to investigate the complex process of multiple text comprehension. Multiple text comprehension consists of the following phases: reading the assignment, searching for relevant pages, reading the pages, organizing the relevant text segments, and writing a synthesis based on multiple texts. A closed web environment, where students practice all multiple text comprehension skills, was designed for this study. A software application was also created to collect log data in this project. Some basic process indicators like time on task and frequencies concerning how often the phases were visited (Greiff et al., 2016) were used to make students' multiple text comprehension processes visible. The preliminary results indicate that primary school students' processes vary greatly and they used versatile methods to perform the task.

Keywords: e-learning; log data; multiple text comprehension; primary education; learning analytics; process indicators

Introduction

The development of technology, such as the World Wide Web (WWW), has initiated the need for additional skills to learn at school, like multiple text comprehension (MTC) (e.g. Magliano et al., 2017). MTC refers to the adaptive process in which the learner goes through different phases, such as searching for texts, evaluating knowledge, and selecting relevant texts and text segments (Kaakinen & Hyönä, 2014) and writing synthesis based on multiple texts (Rouet & Britt 2011). In current world learning often takes place within the Internet even in primary education (Florit et al. 2020; Mason 2018). Often during multiple txt comprehension

nor students neither their teachers are aware of the learning process. The learner can select different paths to proceed during multiple text comprehension. Some paths can better support student learning than the other ones. Technological solutions like learning analytics can enable us to get insight into the individual learning process and difficulties via logfiles. (Knight & Sullivan 2022; Knight & Shum 2017).

Previous research has shown that students need support in learning MTC skills (Wiley et al., 2018). A big challenge seems to be, how to practice MTC skills in an authentic classroom situation and how to use log data for analyzing behavioral patterns to provide indicators to support students' effective and flexible MTC strategies. The aim of this study is to create indicators based on log data from primary students' MTC for further development. A prototype called KidNet was used as a digital learning environment where students practiced all MCT skills initiated by inquiry task. KidNet is a closed learning environment which offers all the phases for MTC in a flexible manner like reading the inquiry task, searching for information with a search engine, bookmarking and evaluation of the webpages, marking of the main ideas in the bookmarked web pages, preparing the structure of the answer by organizing the main ideas, and writing a synthesis by using the structured main ideas. Depending on the inquiry task, different configurations in KidNet are possible, i.e. the number and content of the texts or web pages are made available to the students for this task. In this respect, the system knows both which pages and which text sections in these relevant pages are relevant for the task. The MTC process consists of several phases which can be analyzed for example in terms of time used in differences and frequency of visited pages and different phases.

Through learning analytics, we can analyze the learning process and make it visible for the students and for their teachers. The study attempts to answer the following question: What process indicators can be found in the log data related to multiple text comprehension process?

Method

In our study, 258 fifth and sixth graders from 12 primary schools took part. After a short introduction explaining the purpose of the study and introducing the www learning environment by showing a video, all students performed an online inquiry that lasted approximately 45 minutes. The closed web-based learning environment, called KidNet, simulates www and enables learners to go through all MTC phases and use individual and flexible strategies. The learning environment KidNet has six main phases. In the phase "assignment", the student can read the inquiry task. In our study, the task consisted of several sentences with the main task to explain "how does rain occur?" with some contextual information and some instructions. The phase "search" was realized by common search engine. The phase "reading pages" showed the selected page from the SERP (Search Engine Result Page) and additionally the student could evaluate the page for reliability and usefulness, bookmark the page, and select some text passages for later usage. In the phase "bookmark" the student received an overview in tabular form of the websites that had been bookmarked and the personal evaluation of the pages, for example, in terms of credibility and usefulness. In the phase "snippets" the student could organize the selected text

passages, named snippets, for the writing phase. During the “synthesis” writing phase the student could make use of the organized snippets. For this purpose, the student had a word processor application.

Analysis

The environment collected log data from all phases. Log data was generated through the usage of the environment. In the first phase of the analysis, selection, summarization, and presentation of log data is essential. Log data can be divided into two types, the *action* recorded at a certain point in time (e.g. activate one phase, i.e. the “search” with the search engine to generate a SERP), and the state as a *state* for a duration (e.g. the time in the “search” phase, in which the student among others read the SERP as the result from the search) (Goldhammer et al. 2021). Process indicators are based on these two types of log data and accordingly processed. For example, the total time on search is a summation of all state types for the MTC-phase searching.

Process performance was analyzed using time indicators and frequencies. For example, the time spent in individual phases was recorded, as well as the time at which the individual phase was visited and how often the different MTC phases were visited. Additionally, the number of switches between phases and the phase to which the user switches from each phase was recorded.

Results

Our preliminary results indicate that students spent an average of 36 minutes on the task, with time ranging from a minimum of 10 minutes to a maximum of 50 minutes (see Table 1). Students mainly used their time to write the synthesis (on average about 14 minutes) and to read the pages (on average about 8 minutes). They spend an average of 5 minutes searching, 3 minutes arranging the snippets, and considerably less time reading the assignment (1.5 minutes) and the bookmark list (0.5 minutes).

	Minimum	Maximum	Mean	Std. Deviation
Total time (Tt)	1.19	49.65	36.23	7.74
Tt assignment	0.01	14.52	1.57	1.82
Tt search	0.45	19.81	5.23	3.15
Tt reading pages	0	28.82	11.31	4.49
Tt bookmarks	0	6.45	0.54	0.69
Tt snippets	0	24.10	3.26	3.39
Tt synthesis	0.17	30.79	13.89	6.20

Table1. Descriptive Statistics for time indication in minutes (N=258)

Table 2 shows that students visited the phases “search” and “pages” phases fairly often (13 and 12 times, respectively), but the phases “bookmark”, “preparing snippets”, and “writing synthesis” phases almost half as often. With an average of around four times, the “assignment” phase is the least visited.

	Minimum	Maximum	Mean	Std. Deviation
Frequency (Fr) total	10	188	52.41	25.60
Tt assignment	1	16	4.27	2.95
Tt search	1	43	12.90	6.63
Tt reading pages	0	44	11.80	6.34
Tt bookmarks	0	40	6.70	6.41
Tt snippets	0	30	6.57	5.21
Tt synthesis	1	44	6.98	5.90

Table 2. Descriptive Statistics for frequency indicators (N=258)

Table 3 reports the statistical values for when students visited the different MTC phased. The timeslot indicator has values between 1 and 3. Values between 1 and 2 indicate a high frequency of use, particularly in the first half of the total individual time, with smaller values indicating earlier visits to the phase. Values between 2.0 and 3.0 indicate very frequent use in the second half of the total individual time. While the phases “search”, “pages”, “bookmarks” and “snippets” used most frequently in the first two timeslots, is phase “synthesis” is used most frequently in the middle and at the end of the process.

	Minimum	Maximum	Mean	Std. Deviation
Timeslot assignment	1	3	1.60	0.47
Timeslot search	1	2.45	1.58	0.33
Timeslot reading pages	0	3	1.71	0.34
Timeslot bookmarks	0	3	1.73	0.64
Timeslot snippets	0	3	1.86	0.51
Timeslot synthesis	1	3	2.27	0.41

Table 3. Descriptive Statistics for timeslot indicators (N=258)

Table 4 shows how often students switch between phases. For instance, they transitioned directly from the assignment phase to the search phase 699 times, but only 290 times from the search phase to the assignment phase. The assignment phase as starting point is the only mandatory requirement. The students start with the assignment phase. From the assignment phase the students mainly switch directly to the search phase (699 times). The most frequently used combinations are changing from the from the “search” phase to the “reading pages” phase, 2,195 times, and vice versa from the “reading pages” phase to “search” phase, 1393 times. This result was to be expected, as a search engine search is often followed by clicking on a page in the SERP that seems particularly suitable for the task in order to read it further. The combination of “snippets” and “synthesis” cannot be analyzed to the same extent with 674 and 429 instances respectively. Other combinations tend to be

used unilaterally; from “reading pages” to “synthesis” 619 times in one direction and 118 times in the other, and from “reading pages” to “snippets” 567 times and 71 times in the opposite direction.

FROM / TO	assignment	search	red.pages	bookmarks	snippets	synthesis
assignment		699	1	75	67	142
search	290		2195	398	159	206
reading pages	63	1393		409	567	619
bookmarks	60	448	763		498	150
snippets	78	362	71	456		674
synthesis	135	438	118	363	429	

Table 4. From phase to phase frequencies for all main phases for all students (N=258)

Discussion

This study aimed to create process indicators to improve understanding of how primary school children perform in multiple text comprehension (MTC) tasks. Four indicators were developed using log data from 258 primary school children on KidNet. The time indicator describes the total time used in each of the six MTC phases. The frequency indicator is a click rate for each phase, showing how often the children used each phase. The timeslot indicator shows which phase the children visited most frequently, i.e. at the beginning or at the end of the process. The “from phase to phase” indicator, in turn, shows how often the children used the combination of phases and supplements the other three indicators with this very simple path information. These four indicators provide a basic overview of how children use the MTC application and can therefore form an important basis for further analysis. They can be linked to the result data, such as synthesis quality or the success of bookmarking or selecting snippets from texts. Successful and unsuccessful users can be profiled. However, differences in the values of these four indicators suggest that there is probably no single optimal way to perform the MCT particularly successfully, but rather different effective approaches. However, there are probably also some ineffective ways of performing this task. These four indicators can, of course, be adapted to the needs of the further developments. For example, it is obvious that the “from phase to phase” indicator could be extended to a path indicator that analyses more than just the combination of two phases. Alternatively, the timeslot indicator could be designed in much more detail.

References

Florit, E.; Cain, K.; Mason, L.: Going beyond children's single-text comprehension: The role of fundamental and higher-level skills in 4th graders' multiple-document comprehension. In:

British Journal of Educational Psychology, Vol. 90, 2020, No. 2, pp. 449–472.

<https://doi.org/10.1111/bjep.12288> (last check 2026-07-23)

Goldhammer, F.; Hahnel, C.; Kroehne, U.; Zehner, F.: From byproduct to design factor: On validating the interpretation of process indicators based on log data: Large-scale Assessments in Education. In: Large-Scale Assessments in Education, Vol. 9, 2021, No. 1, pp. 1–25. <https://doi.org/10.1186/s40536-021-00113-5> (last check 2026-07-23)

Greiff, S.; Niepel, C.; Scherer, R.; Martin, R.: Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. In: Computers in Human Behavior, Vol. 61, 2016, pp. 36–46. <https://doi.org/10.1016/j.chb.2016.02.095> (last check 2026-07-23)

Kaakinen, J. K.; Hyönä, J.: Task Relevance Induces Momentary Changes in the Functional Visual Field During Reading. In: Psychological Science, Vol. 25, 2014, No. 2, pp. 626–632. <https://doi.org/10.1177/0956797613512332> (last check 2026-07-23)

Knight, R.; Sullivan, S.: Developing a framework for initial teacher education: Route map or reflective tool? In: Journal of Education for Teaching, Vol. 48, 2022, No. 2, pp. 146–158. <https://doi.org/10.1080/02607476.2021.1985938> (last check 2026-07-23)

Knight, S.; Shum, S. B.: Theory and learning analytics. In: Lang, C.; Siemnes, G.; Wise, A.; Gasevic (Eds.): Handbook of Learning Analytics. SoLAR 2017, Vol. 1, pp. 17–22. DOI: 10.18608/hla17 (last check 2026-07-23)

Magliano, J. P.; McCrudden, M. T.; Rouet, J.-F.; Sabatini, J.: The Modern Reader. In: Schober, M. F.; Rapp, D. N.; Britt, M. A. (Eds.): The Routledge Handbook of Discourse Processes. Routledge, 2nd ed., 2017, pp. 343–361. <https://doi.org/10.4324/9781315687384-18> (last check 2026-07-23)

Mason, L.: Multiplicity in the digital era: Processing and learning from multiple sources and modalities of instructional presentations. In: Learning and Instruction, Vol. 57, 2018, pp.76–81. <https://doi.org/10.1016/j.learninstruc.2018.03.004> (last check 2026-07-23)

Rouet, J.-F.; Britt, M. A.: Relevance processes in multiple document comprehension. In: Text relevance and learning from text. IAP Information Age Publishing, Emerald Publishing, Leeds, 2011, pp. 19–52

Wiley, J.; Jaeger, A. J.; Griffin, T. D.: Effects Of Instructional Conditions On Comprehension From Multiple Sources. In: Braasch, J. L. G.; Bråten, I.; McCrudden, M. T. (Eds.): History And Science. 1st