Slow is Good: The Effect of Diligence on Student Performance in the Case of an Adaptive Learning System for Health Literacy

Leon Fadljević*
Know-Center GmbH
Graz, Austria
lfadljevic@know-center.at

Katharina Maitz*
Know-Center GmbH
Graz, Austria
kmaitz@know-center.at

Dominik Kowald
Know-Center GmbH
Graz, Austria
dkowald@know-center.at

Viktoria Pammer-Schindler
Know-Center GmbH
Graz, Austria
vpammer@know-center.at

Barbara Gasteiger-Klicpera
University of Graz
Graz, Austria
barbara.gasteiger@uni-graz.at

ABSTRACT

This paper describes the analysis of temporal behavior of 11-15 year old students in a heavily instructionally designed adaptive e-learning environment. The e-learning system is designed to support student's acquisition of health literacy. The system adapts text difficulty depending on students' reading competence, grouping students into four competence levels. Content for the four levels of reading competence was created by clinical psychologists, pedagogues and medicine students. The e-learning system consists of an initial reading competence assessment, texts about health issues, and learning tasks related to these texts. The research question we investigate in this work is whether temporal behavior is a differentiator between students despite the system's adaptation to students' reading competence, and despite students having comparatively little freedom of action within the system. Further, we also investigated the correlation of temporal behaviour with performance. Unsupervised clustering clearly separates students into slow and fast students with respect to the time they take to complete tasks. Furthermore, topic completion time is linearly correlated with performance in the tasks. This means that we interpret working slowly in this case as diligence, which leads to more correct answers, even though the level of text difficulty matches student's reading competence. This result also points to the design opportunity to integrate advice on overarching learning strategies, such as working diligently instead of rushing through, into the student's overall learning activity. This can be done either by teachers, or via additional adaptive learning guidance within the system.

KEYWORDS

reading competence, health literacy, differentiation, diversity, adaptive e-learning system, clustering, learning analytics

*Both authors contributed equally to this work.

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

LAK '20, March 23–27, 2020, Frankfurt, Germany

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7712-6/20/03…$15.00
https://doi.org/10.1145/3375462.3375502

ACM Reference Format:
Leon Fadljević, Katharina Maitz, Dominik Kowald, Viktoria Pammer-Schindler, and Barbara Gasteiger-Klicpera. 2020. Slow is Good: The Effect of Diligence on Student Performance in the Case of an Adaptive Learning System for Health Literacy. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK '20), March 23–27, 2020, Frankfurt, Germany*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3375462.3375502

1 INTRODUCTION

In school-settings, teachers strive to give every student the opportunity to develop to her or his full potential. This can be a challenging task due to the high diversity in students' fundamental abilities such as reading competences or language proficiency, or metacognitive skills such as adequate learning strategies. In secondary school, especially differences in reading abilities (e.g., due to linguistic or socio-economic background or disabilities) are an important issue [17]. Students with weak reading abilities are disadvantaged not only in language classes but also in the content areas, because learning in the content areas involves to a considerable degree text-based materials with content-specific vocabulary.

However, content area teachers often cannot react to this issue properly due to a lack of time because they have to deliver their topics' content to meet the requirements of the curriculum [12]. Differentiation of learning materials regarding their linguistic difficulty, and thus matching the reading ability of a student, can be a possible solution. This simultaneously fosters reading competences [14] and facilitates content learning. Providing linguistically differentiated materials which teach the same topic and content is therefore a strategy to teach diverse classes because it enables every student to learn the same, while minimizing disadvantages resulting from different initial reading competence. However, there are several challenges for teachers to implement this strategy. One major challenge is that teachers have to assign a difficulty level to every student in their class [2]. Doing that properly requires suitable and validated assessment instruments with cut-off values corresponding to the difficulty levels of the materials. Furthermore, as students achieve different learning gains during one school year, assessments to determine the reading level of a student should be performed several times during a school year in order to examine whether the assigned difficulty level of the material still fits the student's reading competence. Adaptive learning systems that automatically provide such assessments and differentiated content can be a a possible solution to these issues.

Background. Research on MOOCs and Intelligent Tutoring Systems has shown that reading behavior is a predictor for student performance and that reading ability is an effective basis for system adaptation [6, 18, 19]. Nonetheless, e-learning systems adapting to reading behavior are sparse, and we could not find any systems that aim to support learning content via adapting text difficulty. Usually adaptive learning systems that adapt to text difficulty focus on improving reading abilities, and thus provide content that is different based on student's prior content knowledge, affective state, or interests [21]. For example, iSTART (Interactive Strategy Training for Active Reading and Thinking) [10, 11], an adaptive Web-based program to teach reading strategies for science texts, adapts to the performance of students by varying the amount of support in the program. Stairstepper [15], an iSTART module, is a system that adapts text difficulty based on in-program performance and was explicitly designed to foster reading comprehension and help weak readers in improving in reading. However, the purpose of Stairstepper is not to mediate the same knowledge via texts of differing difficulty as is the goal of differentiated content teaching as investigated in our work.

The present work. In this paper, we present results of our investigations on the temporal behavior of students and the relationship between temporal behavior and performance in the given system. Existing literature on different student behaviors in e-learning environments based on temporal behavior analysis [3, 9, 13] shows that students differ significantly in their temporal behavior. Furthermore, differences in temporal behavior have been shown to significantly correlate with performance, such that students who are consistently more active throughout a course, and overall invest more time for learning typically perform better [1, 4]. However, this research has been carried out in learning environments that give considerable freedom of action and self-direction to learners, and with university students as learners.

Complementing this, our present research has 11-15 year old students as learners, and our adaptive learning system significantly constrains users' freedom of actions within it. This means that students can only indirectly influence the level of difficulty they will get, and cannot choose it completely freely. They also navigate through a predefined sequence of topics, and need to do predefined tasks in order to "unlock" subsequent tasks and content. In particular, students use the system all at the same time (within their lessons), such that overall time spent, and broad patterns of how often and with which spacing students access the system is the same for all students.

The temporal behavior that we therefore investigated was not on overall time spent in the system as it was in related work, but it was on topic completion time. Also topic completion time is to some extent constrained in our study, as overall time available was predefined by the school lessons in which the system was used. Consequently, the remaining differences in topic completion time can be understood to be related to the match of student's reading competence and the text difficulty, and students' metacognitive strategy to work more or less diligently.

Research questions and hypothesis. Thus, in this work, we investigated the following research questions:

- RQ1: Is temporal behavior a differentiator between students?
- RQ2: Is temporal behavior correlated with performance?

Our hypothesis was that we would see four groups of students: fast students with low performance because they were too sloppy, fast students with high performance for whom text difficulty was too low (and who therefore should move to the next higher level of difficulty in the upcoming topic or are already at the highest difficulty level), slow students with low performance for whom text difficulty was too high (and who should move to the next lower level of difficulty in the upcoming topic), and slow students with high performance because they were diligently working at an adequate level of text difficulty.

2 STUDY

This section describes the setup of our study including the study environment (i.e., our adaptive learning system) and the adaptation mechanism, the study participants and our data analysis methods.

2.1 Study Environment

Our adaptive learning system for health literacy consists of five modules that focus on different health-related topics such as injuries and hemophilia (the bleeding disease), vaccination, allergies, resistance to antibiotics, and breast cancer. These medical topics are embedded into stories (see Figure 1a). For each module, texts and tasks were co-developed by clinical psychologists, pedagogues and medicine students in four difficulty levels, differing in text length and linguistic complexity but representing the same content.

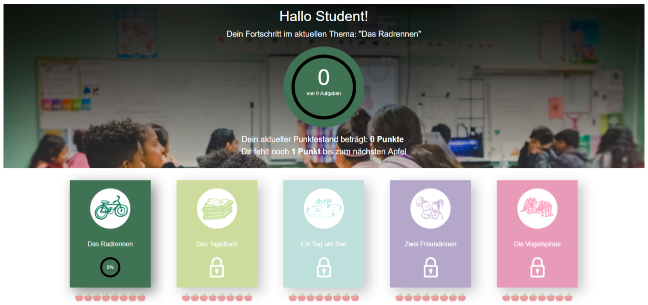
On the technical side, we used Moodle[1], which is a free and open-source learning management system (LMS), as the core of our system since it provides adaptable out-of-the box assessment systems and functionality for learning process navigation.

The dashboard of our system is shown in Figure 1a where students can overview the status of active and finished topics together with all related scores. The scores are displayed as "apples" students can earn when they answer a certain percentage of questions within a topic correctly and a badge for each finished topic based on the amount of "apples" earned in the topic (bronze, silver or gold). Before students can enter a topic, they have to do a reading competence assessment. For subsequent topics, the system decides on the text difficulty of the next topic depending on i) the reading competence assessment ii) the task performance within the topic, and iii) the self-assessment of the student (see Table 1). Each subsequent topic is locked until all prior topics are finished.

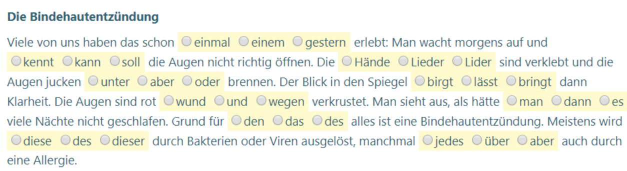
2.2 Adaptation Mechanism

Algorithmically, the above three elements for adaptation translate to i) the reading competence assessment result - the score of every student in the instructionally designed reading competence assessment (r_n), ii) performance score - a student's performance for the current topic (p_n), and iii) self-assessment result - students can rate

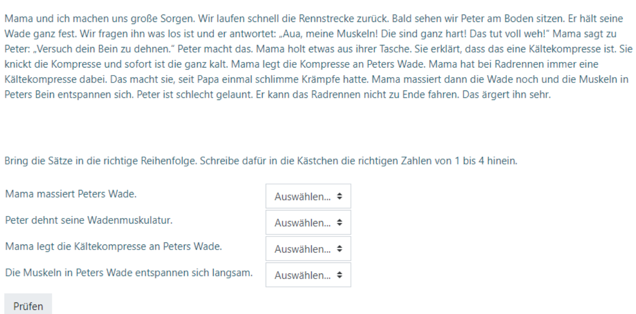
[1]https://download.moodle.org/



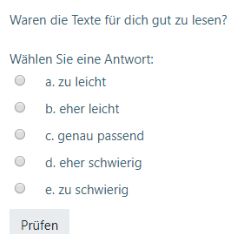
(a) System dashboard (in German).



(b) Reading competence assessment (in German).



(c) Topic assessment (in German).



(d) Self-assessment (in German).

Figure 1: Adaptive learning system overview. Figure (a) shows the dashboard of the system, Figure (b) illustrates the reading competence assessment at the start of each topic, Figure (c) shows one of the reading texts with its assessment, and Figure (d) shows the self-assessment.

at the end of a topic whether they think the text difficulty was too easy, adequate, or too difficult (s_n).

With respect to the text difficulty, level 1 corresponds to the highest level of difficulty, and 4 to the lowest level of difficulty. The difficulty levels are unevenly distributed over the assessment scores,

Table 1: Overview of notations we use throughout this paper.

Symbol	Description
r_n	Reading competence assessment score
p_n	Performance score
s_n	Self-assessment score
res_n	Result for the current topic
d_{n+1}	Next text difficulty level

in order to provide more differentiation, and via this more support, to students with lower reading competence. Each topic starts with a *reading competence assessment* (see Figure 1b), which consists of a speed test (time-limit: 4 minutes) with maze selection tasks. The outcome of this reading competence assessment is given by:

$$r_n(x) = \begin{cases} 1, \text{ if } x > 49\% \ correct\ tasks \\ 2, \text{ if } 35\% < x \leq 49\% \ correct\ tasks \\ 3, \text{ if } 29\% < x \leq 35\% \ correct\ tasks \\ 4, \text{ if } x \leq 29\% \ correct\ tasks \end{cases} \tag{1}$$

For the first topic, the assigned difficulty level d_1 directly corresponds to $r_1(x)$:

$$d_1 = r_1(x) \tag{2}$$

The result of a topic n (res_n) is given by the sum of the reading competence assessment score before the text and tasks r_n, the performance score p_n based on the tasks of topic n (see Figure 1c) and the student's self assessment s_n after the text and tasks (see Figure 1d). Here, the student states whether he or she found the text too easy (-1), suitable (0), or too difficult (1). Thus, formally res_n is given by:

$$res_n = r_n + p_n + s_n \quad \begin{cases} r_n, Res_n \epsilon \{1, 2, 3, 4\} \quad p_n \epsilon \{0, 1\} \\ s_n \epsilon \{-1, 0, 1\} \end{cases} \tag{3}$$

Finally, the next difficulty level d_{n+1} is calculated as the sum of current topic's result res_n and the reading competence assessment score r_{n+1} that precedes the next topic:

$$d_{n+1} = \frac{1}{3}[res_n + 2r_{n+1}(x)] \quad \{ \ d_{n+1} \epsilon \{1, 2, 3, 4\} \tag{4}$$

Summed up, the decision is influenced by one third with the result of the current topic res_n and by two thirds with the result of the reading competence assessment, which precedes the upcoming topic (i.e., r_{n+1}).

2.3 Study Participants

Two lower-secondary schools participated in the study. School 1 was in a rural area of the Austrian federal state of Styria and school 2 was in an urban area of Styria. Our sample consisted of 196 students from grades 6 to 8 (111 students in school 1 and 85 students in school 2). 83 students (42.3%) were sixth graders, 66 students (33.7%) were seventh graders and 47 students (24%) were in grade eight. 191 students provided demographic data. The students were between 11 and 15 years old (M=12.8, SD=0.94). 48.69% were female and 51.31% were male. 93% of the students were born in Austria and

Table 2: Set of input features we use for our k-means clustering approach. With these features, we get the best results with respect to silhouette score.

Input features	
s_n	Self-Assessment
p_n	Performance score
completion time	Avg. topic completion time

80.65% spoke German as their first language. The students worked with the system over the course of four weeks (approx. 3 lessons per week). All statistical analysis and results, which we present in this work are based on the cumulative data from both schools where students produced almost 1/2 million events ($4.53 \cdot 10^5$).

2.4 Data Analysis Methodology

Beside grading levels in the learning program, which were described previously, topic completion time has been measured. Motivation for time recording has been found in [8], where positive correlations have been determined between reading speed and knowledge state when the reader is familiar with a topic. Further on, negative correlations were found between reading speed and quiz performance, which encouraged us to pursue our goal to relate the topic completion time with the overall students' performance [20].

Data cleaning. After the initial data exploration phase and the data cleanup process, we focused on finding groups of students who share similar behavior patterns. During the cleanup process, student attempts which completion time was longer than an hour were left out since those students apparently could not finish their topics during the class in school and therefore would represent noise in the data with respect of analyzing the total topic completion time.

Student clustering. For clustering students, we use the computationally inexpensive and commonly used k-means clustering algorithm [5, 7]. Here, the parameter k (i.e., the number of clusters) can easily be found using a grid-search approach in order to find the k with the highest silhouette score. This score measures the distance between each data point relative to the cluster centroids [16]. In order to get the best clustering result, we tested the algorithm with multiple variations of input features where features from Table 2 produced best clustering results.

Correlation analysis. Another advantage of k-means clustering is that it allows for the inspection of the generated cluster centroids. Here, we noticed a possible linear relationship between the students' topic completion time and their overall performance (i.e. performance score p_n). Trying to solve the linear regression problem between these two variables, where we aimed to predict the students' performance using solely the given completion time, we found a strong correlation between these two features. These clearly pointed out the direct influence of students' diligence on their performance. Furthermore, we examined the relationship between topic completion time and final topic level using a One-Way ANOVA in order to see if there are any statistically significant differences noticeable between the four difficulty levels.

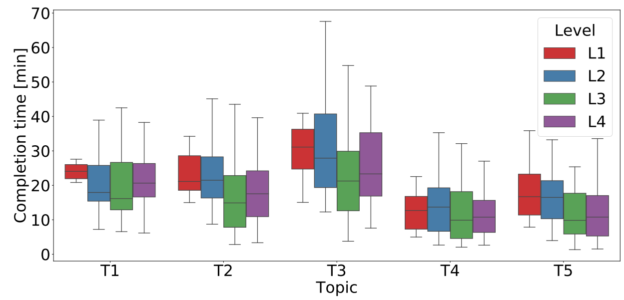


Figure 2: Topic completion time distribution across all topics and related text difficulty levels. We see that students placed in level 3 have the lowest completion time median value in each topic.

Table 3: Characteristics of the four clusters: arithmetic mean and standard deviation for avg. performance score (*range* = 0.28 − 0.88), avg. topic completion time in minutes (*range* = 5.15 − 46.8), and avg. self-assessment (*range* = −1 − 1).

	Performance Score	Completion Time	Self-Assessment
C1 (n=8)	M=0.74 (SD=0.07)	M=41.5 (SD=3.20)	M=-0.38 (SD=0.52)
C2 (n=46)	M=0.67 (SD=0.12)	M=29.4 (SD=2.46)	M=-0.33 (SD=0.52)
C3 (n=85)	M=0.61 (SD=0.12)	M=20.5 (SD=2.47)	M=-0.39 (SD=0.54)
C4 (n=57)	M=0.49 (SD=0.10)	M=11.1 (SD=2.79)	M=-0.12 (SD=0.57)
Total (n=196)	M=0.59 (SD=0.13)	M=20.7 (SD=8.32)	M=-0.30 (SD=0.55)

Table 4: Adaptivity mechanism outcome overview across clusters and in general.

	Improved	Aggravated	Constant	Varied
C1 (n=8)	62.5%	0%	12.5%	25%
C2 (n=46)	58.7%	6.5%	15.2%	19.6%
C3 (n=85)	57.6%	12.9%	16.5%	12.9%
C4 (n=57)	40.4%	15.8%	12.3%	31.6%
Total (n=196)	53,1%	11.7%	14.8%	20.4%

3 RESULTS

To sum up the efforts from the experimental phase, in this section, we present all the results of our work. We tried to find the relationship between students' diligence and performance using the k-means unsupervised clustering algorithm to group the students with similar behavior patterns. On top of that, we extended our research findings with a regression analysis where we tried to predict the student performance using solely the topic completion time.

3.1 Descriptive Statistics

In order to get a general overview of the topic completion time, we investigated the completion time distribution of students from all difficulty levels across all topics (Figure 2). Independent of the difficulty level, the amount of time needed to complete a topic was different for the five topics, because the topics were not equally large. Additionally, Figure 2 shows that there is no linear correlation between completion time and difficulty level. Students in the second lowest difficulty level (level 3) provided the lowest medians (i.e., were the fastest) in all five topics.

In Table 3, the characteristics of the four clusters are displayed. It can be seen that students in clusters one and two achieved higher average performance scores than students in clusters three and four. Students in cluster four achieved on average less than 50% of the total number of points. In terms of average topic completion time, students in cluster one were on average the slowest and students in cluster four the fastest with a difference of more than thirty minutes. Regarding the self-assessment scores, the high mean value for students in cluster four shows, that they were more likely to indicate that the text difficulty was too high than students in the other three clusters.

The general overview of students' performance after the complete learning process is shown in Table 4. Here, 53.1% of the students improved during the process (were at a better level in the system at the end than in the beginning), 11.7% of the students aggravated (were at a lower level in the system at the end than in the beginning), 14.8% of the students were constantly at the same level, and 20.4% of the students ended up at the same level but varied throughout the learning process.

However, we do not assume that half of the students improved in reading during the four weeks they worked with the system. It is more likely that the improvement regarding the difficulty level is due to the fact that the students needed to get used to working with the program, the reading competence assessments and the tasks within a topic. Furthermore, it is possible that some students did not discover the achievement system (i.e., "apples") until they completed the first topic. After they discovered it, their motivation was raised and they put more effort into completing tasks and reading competence assessments.

3.2 The Effect of Diligence on Student Performance

In order to answer our research questions i) whether topic completion time is a differentiator between students, and ii) if there is a correlation between performance and topic completion time, we further analysed the relationship between these two variables. Figure 3 shows the positive correlation ($r = 0.59$, $p < 0.001$) between average topic completion time and average performance score. A linear regression analysis with average topic completion time as predictor for overall performance score ($F = 105.7_{(1,194)}$, $p < 0.001$) supported our clustering results and revealed that 35% of the variance in overall performance score can be explained by topic completion time ($R^2 = 0.35$). The more time students gave themselves to read and perform tasks the better their overall performance score was.

Our clustering result reflects this behavioral pattern as shown in Table 3. Students in cluster one can be described as diligent - working slowly, but accurately, whereas students in cluster four showed an opposite behavior - fast and less accurate. Although the distribution of students across clusters and difficulty levels (Table 5) shows that diligent behavior seems to be more likely in students in level one and level two, the comparison of cluster-membership and final text difficulty level revealed that also some students in level three and four show diligent behavior.

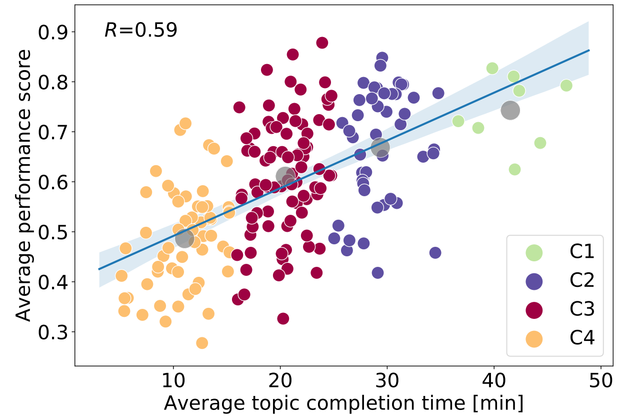


Figure 3: Our four student clusters according to k-means visualized using the average performance score over the average topic completion time (i.e., diligence). We see that there is a positive correlation ($r = 0.59, p < 0.01$) between these two dimensions, which indicates a positive effect of diligence on student performance.

In order to further investigate the relationship between final text difficulty level and average topic completion time, we performed a One-Way ANOVA with post-hoc comparisons (we chose TukeyHSD because the variances were homogeneous). Differences in average topic completion time between the four text difficulty levels are displayed in Table 6. The result of the ANOVA revealed that the four groups differ significantly ($F = 7.65_{(3,\, 192)}$, $p < 0.001$). The post-hoc comparisons showed that this significant result is due to significant differences between level three and level one ($p < 0.01$), level four and level one ($p < 0.01$) and level three and level two ($p < 0.01$). All other differences were not significant.

Students with the highest text difficulty level at the end were significantly slower than students with one of the two lowest text difficulty levels. Students with the second lowest text difficulty level were significantly faster than students with one of the two highest text difficulty levels. This is reflected by the clustering results, that showed that from the students with a final difficulty level three only six out of sixty were assigned to cluster one or cluster two (diligent behavior). Interestingly students with the lowest text difficulty level did not differ significantly from students with the second highest text difficulty level. This is mirrored by the results displayed in Table 5 that show a relatively similar distribution across the four clusters for level two and level four compared to level one and level three.

4 DISCUSSION AND CONCLUSION

Our research questions were, i) whether we could find different temporal behaviors, and ii) whether we could relate these behaviors to performance despite the instructionally designed adaptive system, and despite the comparatively little freedom in action given to the students. The results of unsupervised clustering showed that at each difficulty level, students can be clearly separated into a class of slow and a class of fast students. Subsequent linear regression

Table 5: Student distribution across clusters vs. difficulty level of the final topic.

	Final difficulty level			
Cluster	L1 (n=20)	L2 (n=52)	L3 (n=60)	L4 (n=64)
C1 (n=8)	5% (n=1)	7.5% (n=3)	3.3% (n=2)	3.1% (n=2)
C2 (n=46)	50% (n=10)	30.2% (n=16)	6.6% (n=4)	25% (n=16)
C3 (n=85)	45% (n=9)	43.4% (n=23)	50.8% (n=30)	35.9% (n=23)
C4 (n=57)	0.0%	18.9% (n=10)	39.3% (n=24)	35.9% (n=23)

Table 6: Differences in average topic completion time between the four final difficulty levels.

Final difficulty level	N	Mean	SD	min	max
L1	20	26.2	6.82	17.6	46.8
L2	52	23.0	8.21	5.43	44.3
L3	60	18.0	7.94	5.15	42.4
L4	64	19.5	8.05	7.46	41.9

analyses showed that this temporal behaviour is a predictor of performance, such that slow students have better performance than fast students. In contrast to our initial hypotheses, we did not find groups of students who were fast or slow because the text difficulty was unsuitable (too easy or too difficult) for them. This indicates that the overall assignment of difficulty levels to students was suitable, and also that the differentiation into four, and particularly into four unevenly distributed groups (with respect to reading competence) was suitable. Furthermore, our analysis showed that final text difficulty level and cluster membership were related, but not in a linear way. This again points to a suitable adaptation mechanism, which leaves space for individual differences beyond reading competence to appear.

On a wider level, these results highlight the necessity of teaching strategies for learning and performance even in the presence of personalised learning systems. Such teaching could be given either by teachers, in which case the design opportunity and necessity lies in instructional design, and in designing for embedding technologies suitably for teaching. Such teaching could also be integrated into the technological environment by giving additional and adaptive support for strategically approaching learning tasks with diligence. In this case the design opportunity and necessity lies in technology design. With respect to earlier research on temporal behaviour of students, we highlight that the novelty of our results lies in showing differences of temporal behaviour not in terms of temporal patterns of accessing tasks (e.g., frequently throughout the semester) or time freely invested in learning, as found e.g., in [1, 4]. Rather, the differences are evident in the more fine-granular temporal behaviour on single tasks.

Future work. An ANOVA showed significant differences in average topic completion time between the four text difficulty levels, but not between all four levels. An interesting result of our analyses was, that one particular group of students (final text difficulty level three) was less likely to show diligent behavior than the other groups. Further investigation of the characteristics of this group of students is necessary in order to get better insights regarding the reasons for the different behavior of students in this group, as we have to date no suitable hypothesis for explanation. One possibility for further investigation are qualitative methods; another possibility is to check whether the same outlier behaviour of this group can be replicated in follow-up studies.

Acknowledgements. The authors would like to thank the participating schools for their collaboration. This work was supported by the Gesundheitsfonds Steiermark project HELI-D and the Know-Center GmbH (FFG COMET program).

REFERENCES

[1] Nora'ayu Ahmad Uzir, Dragan Gašević, Wannisa Matcha, Jelena Jovanović, Abelardo Pardo, Lisa-Angelique Lim, and Sheridan Gentili. 2019. Discovering Time Management Strategies in Learning Processes Using Process Mining Techniques. In *Transforming Learning with Meaningful Technologies*, Maren Scheffel, Julien Broisin, Viktoria Pammer-Schindler, Andri Ioannou, and Jan Schneider (Eds.). Springer International Publishing, Cham, 555–569.

[2] Julie W Ankrum and Rita M Bean. 2008. Differentiated reading instruction: What and how. *Reading Horizons: A Journal of Literacy and Language Arts* 48, 2 (2008).

[3] Mina Shirvani Boroujeni and Pierre Dillenbourg. 2018. Discovery and temporal analysis of latent study patterns in MOOC interaction sequences. In *Proc. of LAK'18*. ACM, 206–215.

[4] Analía Cicchinelli, Eduardo Veas, Abelardo Pardo, Viktoria Pammer-Schindler, Angela Fessl, Carla Barreiros, and Stefanie Lindstädt. 2018. Finding Traces of Self-regulated Learning in Activity Streams. In *Proc. of LAK'18 (LAK '18)*. ACM.

[5] Chris Ding and Xiaofeng He. 2004. K-means clustering via principal component analysis. In *Proc. of ML'04*. ACM, 29.

[6] Michael Eagle, Albert Corbett, John Stamper, Bruce M McLaren, Angela Wagner, Benjamin MacLaren, and Aaron Mitchell. 2016. Estimating individual differences for student modeling in intelligent tutors from reading and pretest data. In *ITS'16*.

[7] Greg Hamerly and Charles Elkan. 2004. Learning the k in k-means. In *Advances in neural information processing systems*. 281–288.

[8] Yun Huang, Michael Yudelson, Shuguang Han, Daqing He, and Peter Brusilovsky. 2016. A framework for dynamic knowledge modeling in textbook-based learning. In *Proc. of UMAP'16*. ACM, 141–150.

[9] Beata Beigman Klebanov, Anastassia Loukina, Nitin Madnani, John Sabatini, and Jennifer Lentini. 2019. Would you?: Could you? On a tablet? Analytics of Children's eBook Reading. In *Proc. of LAK'19*. ACM, 106–110.

[10] Danielle S McNamara, Irwin B Levinstein, and Chutima Boonthum. 2004. iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers* 36, 2 (2004), 222–233.

[11] Danielle S McNamara, Tenaha P O'Reilly, Rachel M Best, and Yasuhiro Ozuru. 2006. Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research* 34, 2 (2006), 147–171.

[12] Molly K Ness. 2009. Reading comprehension strategies in secondary content area classrooms: Teacher use of and attitudes towards reading comprehension instruction. *Reading Horizons* 49, 2 (2009), 143–166.

[13] Sahba Akhavan Niaki, Clint P George, George Michailidis, and Carole R Beal. 2019. The Impact of an Online Tutoring Program for Algebra Readiness on Mathematics Achievements; Results of a Randomized Experiment. In *Proc. of LAK'19*. ACM, 363–372.

[14] Rollanda E O'Connor, Kathryn M Bell, Kristin R Harty, Louise K Larkin, Sharry M Sackor, and Naomi Zigmond. 2002. Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology* 94, 3 (2002), 474.

[15] Cecile A Perret, Amy M Johnson, Kathryn S McCarthy, Tricia A Guerrero, Jianmin Dai, and Danielle S McNamara. 2017. StairStepper: An adaptive remedial iSTART module. In *AIED'17*. Springer, 557–560.

[16] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.

[17] Margaret E Shippen, David E Houchins, Candace Steventon, and Donya Sartor. 2005. A comparison of two direct instruction reading programs for urban middle school students. *Remedial and Special Education* 26, 3 (2005), 175–182.

[18] Valerie J Shute and Diego Zapata-Rivera. 2012. Adaptive educational systems. *Adaptive technologies for training and education* 7, 27 (2012), 1–35.

[19] Khushboo Thaker, Paulo Carvalho, and Kenneth Koedinger. 2019. Comprehension Factor Analysis: Modeling student's reading behaviour: Accounting for reading practice in predicting students' learning in MOOCs. In *Proc. of LAK'19*. ACM.

[20] Khushboo Thaker, Yun Huang, Peter Brusilovsky, and He Daqing. 2018. Dynamic knowledge modeling with heterogeneous activities for adaptive textbooks. In *EDM'18*. 592–595.

[21] Mieke Vandewaetere, Piet Desmet, and Geraldine Clarebout. 2011. The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior* 27, 1 (2011), 118–130.