

Comparing Student High and low Reading Performance With Differentiated Digital Reading Materials

Lisa Paleczek^{1,3}, Daniela Ender^{1,2,3}, Andrea Kogler¹ and Susanne Seifert^{1,3}

¹University of Graz, Institute of Education Research and Teacher Education, Department for Inclusive Education, Austria

²Private University College of Teacher Education Augustinum, Department of Educational Science, Graz, Austria

³Research Center for Inclusive Education, Graz, Austria

lisa.paleczek@uni-graz.at

daniela.ender@pph-augustinum.at

andrea.kogler@edu.uni-graz.at

susanne.seifert@uni-graz.at

Abstract: The present study investigates eight Grade-4 classes (9- to 12-year-olds, 52.1% girls) who worked on at least 10 lessons using project RegioDiff material. The study focuses on one of these lessons (including nine text passages and corresponding tasks) and on students with low (19 students, percentile <15) and high reading skills (29 students, percentile > 70). While students were working with the material, screencasts were recorded (30h). The construct “task performance” (processing time, response accuracy, and task engagement) was then analysed using the screencasts. The analysis revealed that the two groups differed significantly in the processing time of two tasks, but not in the total time spent on all nine tasks. Significant differences were revealed also for general task engagement. Task engagement was highly correlated with processing time. Participants with higher reading skills spent more time on the tasks and were more engaged than participants with lower reading skills. However, we did not find any differences in terms of answer accuracy. This indicates that task difficulty and student reading skills were well matched. The study also shows how tasks may be adapted or augmented in order to match the learning environment more closely to student learning needs.

Keywords: differentiation, digital materials, reading skills, screencast analysis, primary school

1. Introduction

In today’s classrooms, teachers encounter many different forms of student diversity (Messiou et al., 2020). Students also often differ in terms of their reading skills, even though they are in the same grade level (OECD, 2019). Differentiation and digitalization allow for greater inclusion when teaching in a diverse classroom setting.

In the project RegioDiff, we developed digital materials for primary school content lessons (Paleczek, 2020; Paleczek, Ender, Berger, Prinz & Seifert, 2022). These materials combine factual texts with tasks containing elements that foster reading comprehension (reading strategies, cooperative learning, vocabulary work, etc.). The texts and tasks are differentiated into four different difficulty levels. While all students work on the same topic within a comparable time frame (i.e., one content lesson), the text they read is matched to their reading skills, thus aiding greater inclusion during the lesson (Feyerer, 2012). The digital materials used in the present study are described in the following section. In subsequent sections we then focus on addressing whether students with above average reading skills, and students with strongly below average reading skills, differ in terms of their processing time, response accuracy and task engagement, and how these three constructs correlate with one another. The question of how or whether material might need to be adapted to make it more inclusive was also addressed.

2. The digital learning environment

The RegioDiff digital learning environment is designed to cover different topic areas. Each topic consists of various text passages, where, usually, each passage is linked to a corresponding task right after or before reading the text passage. The different types of tasks provided are all known to support students’ reading comprehension.

Students first enter a code in order to access the digital learning environment. They then see the topics prepared by the teacher on their tablets. Students choose their assigned level by clicking on the appropriate fruit symbol (see Figure 1) and start working on the topic individually. At certain points, students cooperate

across reading levels in order to solve tasks. Working in such a digital learning environment requires access to an internet connection. The differentiated texts and tasks are now described in more detail below.





percentile rank	reading performance	fruits	approximate number of words per differentiation level
>70	above-average	apple 	750
31-70	average	cherry 	575
15-30	below-average	pear 	350
<15	severely below-average	grapes 	250

Figure 1: Reading levels

2.1 Differentiated texts with glossary words

Each topic consists of various text passages that are differentiated into four reading levels (see Figure 1). On the one hand, the differentiation is based on word count, with the number of words decreasing as one moves from the highest to the lowest reading level. On the other hand, the words, sentences, and texts also vary in terms of their complexity (using fewer complicated synonyms, fewer nested sentences, etc.). Table 1 shows two different levels (the highest differentiation level (hdl) and the lowest differentiation level (ldl)) for the topic used in the present investigation.

Table 1: ldl and hdl reading text information

Passage	Title	Word-count (hdl/ldl)	Sentence-count (hdl/ldl)	Passive sentences (hdl/ldl)	gSmog (hdl/ldl)	RIX (hdl/ldl)	Glossary Words (hdl/ldl)
1	The city Graz	58/17	6/2	0/0	5.42/4.71	2.89/2.72	5/1
2	Districts of Graz	35/14	4/2	0/0	4.71/3.48	2.74/2.41	1/1
3	The coat of arms of Graz	49/28	5/5	0/0	4.48/1.46	3.90/2.15	3/3
4	Center of Graz	110/21	10/1	1/0	6.12/10.25	4.10/4.36	12/4
5	Eggenberg Palace	142/80	13/10	0/0	5.29/3.72	4.81/3.49	8/4
6	Kunsthau Graz	98/21	8/2	0/0	5.50/3.48	5.29/3.02	7/3
7	Island in the Mur	73/29	7/3	2/0	6.02/3.48	4.42/2.87	5/2
8	The castle Hill and the clock tower	159/66	14/8	2/0	5.02/3.81	5.39/3.65	4/2
9	Graz Cathedral	46/18	4/2	2/0	5.25/3.48	4.17/2.77	6/2
Graz hdl	Whole Text incl. headlines (6 words)	775	70	7	5.49	7.69	51
Graz ldl	Whole Text incl. headline (6 words)	300	36	0	3.85	4.90	22

Difficult words are marked in the texts as glossary words. These words can be specific content-related vocabulary, places (countries, cities, rivers, etc.), persons etc. that students might be unfamiliar with. Students obtain explanations of such terms by clicking on them. A pop-up-window then opens and offers a written explanation, a picture (if possible) and an audio of the explanation in order to ease understanding. The latter feature is especially helpful for weak readers. The explanations are the same for all reading levels. Figure 2 shows a sample pop-up window for a glossary word. Figure 3 shows the same passage in the learning environment, comparing hdl and ldl reading level.

At the beginning of each text passage, students are asked to read the passage carefully and then to click on the button that leads them to the text-related task.

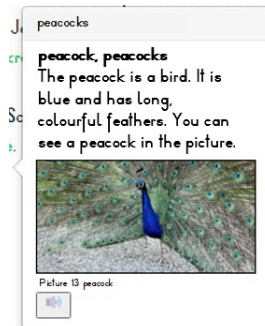


Figure 2: Example for a glossary word pop-up-window

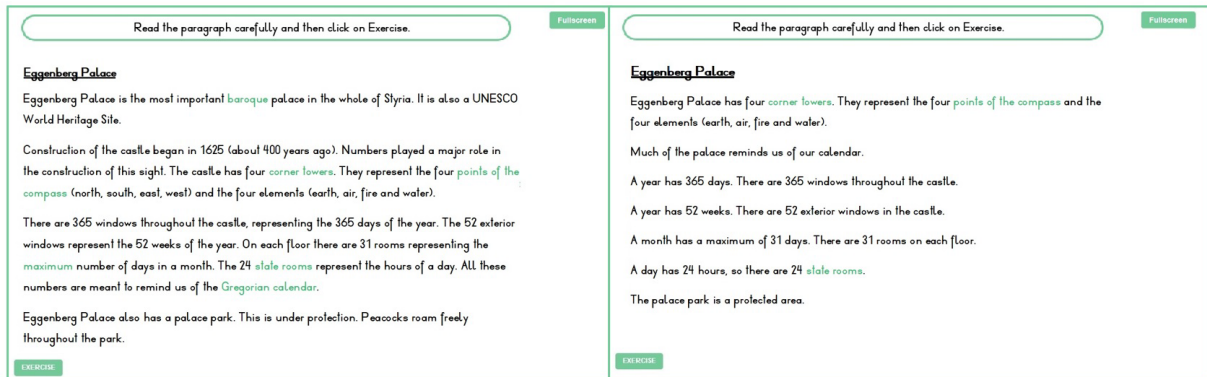


Figure 3: Example for a text passage hdl (left) and ldl (right)

2.2 Tasks

There are different task formats following the text passages. In general, the tasks are either reading strategy tasks (e.g., Duke et al., 2021), cooperative learning tasks (e.g., Hattie & Zierer, 2019), or reading comprehension tasks (e.g., cloze texts, judging whether statements are true or false; Duke et al., 2021). All such tasks are known to foster reading comprehension. Usually, the tasks are also differentiated into four difficulty levels. Only the first and the last task are the same for all differentiation levels. The first task always involves a pre-reading strategy (predicting content by the heading) paired with a cooperative learning exercise (think-pair-share). The last task is the balloon game (see Figure 4) in which students have to answer a text-related question by clicking on the correct letters in order to fill the boxes provided. Clicking on wrong letters makes the balloons burst. As soon as the task is solved, the balloons fly into the air.

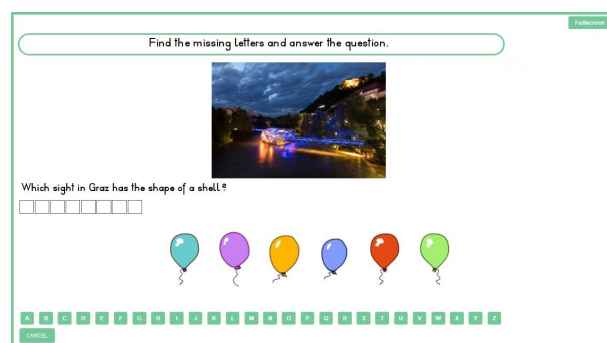


Figure 4: Balloon exercise

2.3 Reading levels and differentiation

Reading abilities within a classroom differ even though children are taught at the same grade level (Paleczek, 2020). The RegioDiff reading material offers four differentiation levels. To decide on the appropriate differentiation level, prior to beginning the study, student reading comprehension was measured using the GraLeV Reading Comprehension Test (Paleczek, Seifert, Franz, Wohlhart, & Riedl, in prep.; Seifert & Paleczek, 2021). The GraLeV was developed to assess the reading comprehension of learners in Grades 3 and 4 and it

consists of four subtests (one at word level, one at sentence level, and two at the text level). The number of correctly solved tasks in a certain amount of time is converted into percentile ranks, which are then used to determine the appropriate differentiation level of the text for each child. For example, if a child achieved a percentile rank of 60, 60 percent of participants performed worse or the same and 40 percent performed better (Paleczek & Seifert, 2021). Children with a percentile rank of 70 or above were classified as achieving above-average reading performance (hdl), and those at a percentile rank below 15 were placed in the severely below-average reading level (ldl) (Paleczek, 2020). Further information on the reading levels is provided in Figure 1.

Differentiating the texts makes it possible for all children in a class, regardless of their reading level, to work on the same topic and thus to acquire knowledge together (Seifert, Schwab & Gasteiger-Klicpera, 2014). Providing the text at different levels of difficulty is intended to make it possible for all students to read the text in approximately the same amount of time (i.e., in one lesson for both the text and the tasks). Moreover, in terms of Vygotski's zone of proximal development (Vygotski, 1987), student reading of a text matched to their own reading abilities facilitates the development of reading comprehension.

2.4 Lessons

Once teachers hand out the tablets, the students can start work immediately. They start with the first reading task (as described above). This activates prior knowledge and facilitates pre-reading conversation with the whole classroom. Students then work individually on the tablets focussing on their own texts and tasks. Each time they get to a cooperative task, they have a look around to see who of their classmates is available and pair up to work on the cooperative tasks. This is not dependent on differentiation levels. It is recommended that the whole class finish the lesson together. This provides an opportunity for post-reading conversation on the topic and thus helps to deepen knowledge acquisition.

3. Aims and research questions

The aim of the present paper is to investigate how students of different reading levels interact in the learning environment and whether it is possible to offer inclusive lessons with such material. To investigate this, we chose one topic (Graz) and analysed screencasts from hdl and ldl readers.

Student task performance (consisting of processing time, response accuracy and task engagement) was thus evaluated. The focus was on readers with above-average (working with the hdl) and severely below-average (working with the ldl) reading skills (Paleczek, 2020). It was thought that differences in processing time between and within groups would provide insight into how materials might be adapted or improved. It was thus of particular interest to investigate whether there was a relationship between processing time and response accuracy, and how students from the ldl and the hdl differed in this regard.

This led to the following research questions:

- 1. What are the respective differences, in terms of task performance, when hdl and ldl readers interact in the learning environment?
- 2. What should be adapted in the digital learning environment to enable students from these two reading levels to work together on a topic?

4. Methods

4.1 Sample

Eight Grade 4 classrooms (N = 131 students) used the digital material on their tablets after parents' consent to take part in the study. In the present study, we concentrated on those students assigned to the ldl (n = 24) and to the hdl (n = 32). After eliminating incomplete, faulty, or inadequate data (e.g., arising from students turning off the screencast prematurely, technical problems, student absence), 48 screencasts remained for analysis (ldl 19 (9 female) and hdl 29 (16 female) students), providing a total of approximately 30 hours of video material.

The students attended Grade 4 primary school and were between nine and twelve years old. Regarding the children's first language (L1), 70.8% were L1 learners (ldl: 52.6%; hdl: 82.8%).

4.2 Design and procedure

For the screencasts, the program "AZ Screen Recorder - No Root" was installed on the tablets to record the screen while students were working in the digital learning environment. The screencast program was started before the lesson began. Prior to the recording, students had already worked on between three and five digital texts. This was to ensure that everyone was able to orient themselves within the learning environment and to reduce any bias that might occur due to lack of familiarity.

The construct *task performance* was defined in advance as being a proxy for student behaviour in the learning environment. Task performance consists of three categories: 1) processing time, 2) response accuracy, and 3) task engagement. These were all determined by means of the screencasts.

For the *processing time*, the start and end time of a text passage task was noted and its duration was calculated. This procedure was based on Baum et al. (2019) and Fadljević et al. (2020).

Additionally, we added the student scores. Since the two differentiation levels differed in the total number of scores students could achieve (28 and 46 for ldl and hdl, respectively), we calculated percentages (*response accuracy*). For Exercises 2 and 5, which involved solving the task in the booklet instead of in the digital learning environment itself, students automatically received 100% as soon as the task was completed. For example, in Exercise 5, learners had to write sentences and read them to another child.

For the category *task engagement*, we captured the students' response behavior. For this purpose, we used an ordinal scale: "3" meant that the task had been solved independently and completely; "2" meant that the task had been solved partially or with help (for example, from a classmate); "1" was selected when children had not solved the task or when someone else had solved it for them. The audio data from the screencasts were the main source for determining task engagement scores. Children's statements were used to track whether and how a child had solved the tasks. This category seems particularly significant with regard to Exercise 2 and Exercise 5, as these are not digitally solvable tasks. For statistical analyses, a sum score was formed (maximum 27: three points for each of the nine exercises).

4.3 Analyses

In addition to one of the authors of the present paper, seven interns of the RegioDiff project, who were trained in advance by the author, were responsible for analysing the screencasts. In order to calculate interrater reliability, all interns evaluated the same screencast after their training. Interrater reliability as indicated by Cohen's Kappa coefficient, shows how much the evaluators assigned the same value to the measured variables (Grouven et al., 2007; Hemmerich, 2019). For the evaluated data, the level of agreement between an evaluator and the author was then checked. Since the available evaluation data are partially interval-scaled, the linearly weighted kappa was calculated. According to Altman (1991), a kappa value of .60 or higher is good; .80 or higher is very good (Hemmerich, 2019). For the screencast evaluated, it was between .873 and .952 (see Table 2), which is why no further training or adaptation of the procedure was required and the evaluators then assessed further screencasts independently.

Table 2: Linearly weighted Cohen's kappa to test interrater reliability.

Rater	Cohen's Kappa (linear weighting)
1	.927
2	.945
3	.91
4	.891
5	.952
6	.873

Descriptive analyses, mean comparisons and correlations were conducted using SPSS. Since the samples were smaller than 30, a test for normality was performed (Shapiro-Wilk test: Shapiro & Wilk, 1965, cited in Hemmerich, 2021). The categories *processing time* and *answer accuracy* were not normally distributed ($p < .001$), so that a non-parametric test (Mann-Whitney U test) was used (Bortz & Döring, 2016). Mann-Whitney U tests were also used for the ordinal data on *task engagement* (Pospeschill & Siegel, 2018). Additionally, a Bonferroni correction was made.

5. Results

Table 3 gives descriptive data on task performance, operationalised by processing time, answer accuracy and task engagement for both differentiation levels. The results for these indicators are described separately in the following sections.

Table 3: Descriptives for ldl and hdl

Differentiation Level		Time spent in learning environment in min	Processing time per task in sec	Mean Percentage of scores achieved per task	Task engagement
ldl (n=19)	Mean (SD)	25.91 (12.57)	70.41 (45.49)	87.32 (17.84)	23.90 (2.99)
	Minimum	4.97	13.00	42.86	17.00
	Maximum	48.48	169.44	100.00	27.00
hdl (n=29)	Mean (SD)	28.54 (8.33)	100.40 (34.91)	90.46 (10.46)	26.21 (1.31)
	Minimum	15.43	46.44	65.29	23.00
	Maximum	49.28	161.67	100.00	27.00

5.1 Differences between ldl and hdl interaction in the digital learning environment

5.1.1 Processing time

As can be seen in Table 3, the range of processing time was very large in both groups. Thus, the students' processing time differs not only between but also within groups. The largest range was seen for the time taken to complete tasks among those in ldl (13.00 to 169.44 seconds per task).

An overview of the individual tasks helps provide insight into how the learning environment may be adapted and into the differences arising concerning student interaction with it. Figure 5 shows the processing time per task for both groups. For each task and level, there is a boxplot showing the range, the median and the first and third quartile. The points above and below are outliers, i.e., the particularly low and high values, respectively (Kohn, 2005).

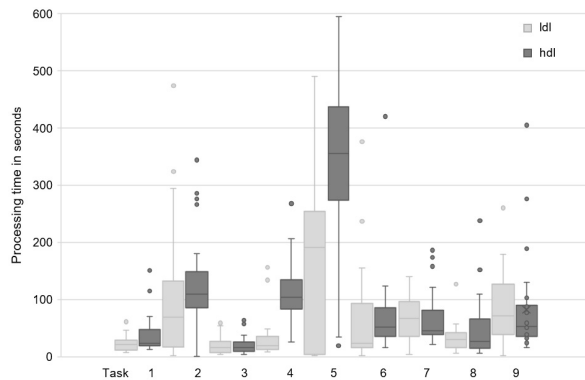


Figure 5: Boxplots of the processing time per task in ldl and hdl readers

As can be seen in Figure 5, there are clear differences between the groups in the tasks 2, 4 and 5. However, after applying a Bonferroni correction of the alpha-level (.005), the Mann-Whitney U tests showed a statistically significant difference only between the two groups in Task 4 (U = 51.0, Z = -4.74, p < .005) and Task 5 (U = 110.0, Z = -3.49, p < .005). No difference was detected in the total time the two groups spent reading and solving tasks in the learning environment.

Task 4 was a cloze text (see Figure 6) and in Task 5, the participants had to write an info box about a famous Graz landmark described in the text. In Task 5, about 75 percent of ldl finished within 250 seconds, while 75 percent of level hdl finished within 435 seconds.

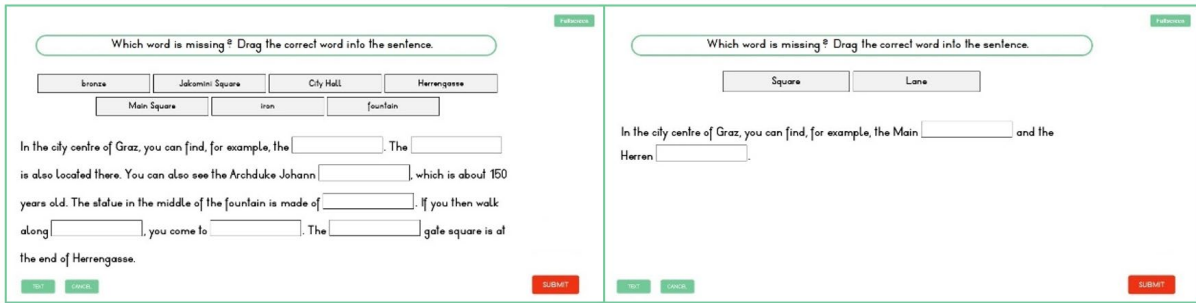


Figure 6: Cloze text in hdl and ldl

5.1.2 Response accuracy

With regard to average response accuracy, there is a difference of 3.14 percentage points between the groups. Table 4 shows the response accuracy for each task. As response accuracy has no validity for Tasks 2 and 5, they are excluded from the table. Both groups scored the lowest in Task 7, in which the correct word had to be selected from three in a cloze text (see Figure 7). In none of the exercises did the two groups differ significantly in terms of response accuracy.

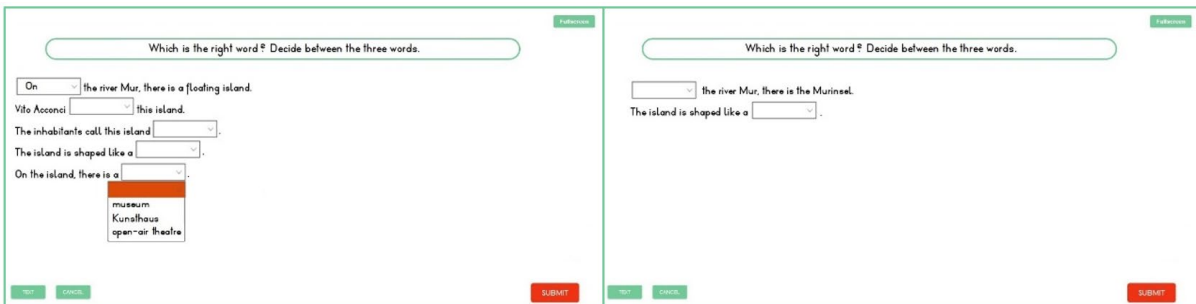


Figure 7: Cloze text with word selection in hdl and ldl

Table 4: Descriptives for percentage of response accuracy per task

task	<i>ldl</i>	<i>hdl</i>
	Mean (SD)	Mean (SD)
1	97.37 (11.47)	94.83 (13.98)
3	94.74 (22.94)	96.55 (18.57)
4	89.47 (31.53)	86.03 (18.70)
6	84.21 (37.46)	94.48 (14.04)
7	76.63 (38.62)	84.14 (19.55)
8	84.21 (23.88)	86.62 (22.63)
9	84.89 (21.65)	90.59 (16.26)

5.1.3 Task engagement

Task engagement shows the students' response behavior. A maximum task engagement score of 27 points was possible (9 tasks, each a maximum of 3 points). A Mann-Whitney U test revealed a significant difference in overall task engagement between the groups ($U = 142.5$, $Z = -3.035$, $p = .002$), with students in ldl scoring 2.31 points lower than those in hdl (see Table 4).

Frequency analyses of task engagement among those who achieved the maximum number of points provide an overview of the maximum engagement per task (see Figure 8). It was found that only 52.6% in ldl (hdl: 86.2%) were capable of solving Task 2 fully and independently. In Task 6, all hdl readers achieved full points for task engagement.

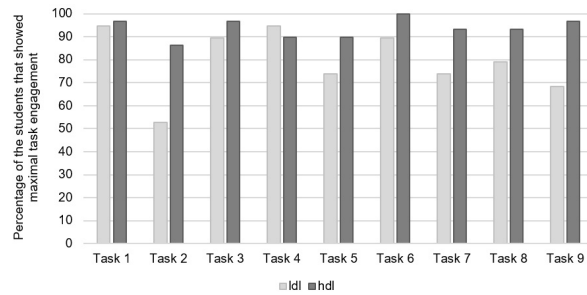


Figure 8: Proportion of participants achieving the maximum number of points in task engagement, separately for both differentiation levels

5.2 Correlations

Task engagement correlated highly significantly with the response accuracy (Spearman's $\rho = .48$, $p < .001$). Thus, students scored higher when they were more engaged. There is a medium correlation (Cohen, 1988) between processing time and task engagement (Spearman's $\rho = .35$, $p = .015$). The correlation between processing time and response accuracy was $r = .34$ ($p = .018$), which also indicates a medium level of correlation (Bortz & Döring, 2016; Cohen, 1988).

Figure 9 shows the time required for the exercises on the X-axis, and the achieved scores in percent on the Y-axis. The two groups are similarly distributed, although there are two ldl readers who clearly spent less time and solved fewer tasks correctly than others. Most students finished the last task between minute 20 and minute 30. The graph also shows that children who took at least 40 minutes solved at least 80 percent of the tasks correctly.

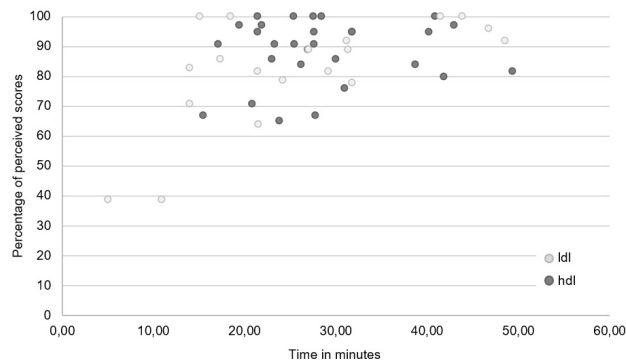


Figure 9: Required time and percentage of perceived scores

6. Discussion

The screencast analysis conducted on one project RegioDiff text (about the city of Graz) provides hints on how the digital learning environment can be adapted so that students of the two reading levels can work together on the materials. The construct task performance was used to compare the work of the below-average readers (the ldl group) with that of those with high reading skills (the hdl group). Task performance consisted of three categories, namely, processing time, response accuracy, and task engagement. The group comparisons showed that the groups did not differ significantly in terms of total processing time spent in the digital learning environment. In general, this thus supports the belief that the material facilitates inclusive lessons where all children learn about the same topic in a similar amount of time, albeit with materials fitting their own reading level. However, there were differences between the groups on specific tasks. The biggest difference in processing time was in Task 5, in which an information box on a Graz landmark had to be created. The ldl readers had to write two sentences and the hdl readers five sentences. It is thus understandable that ldl students completed the task faster. In order to equalize the time needed to complete the task, the number of sentences required for the ldl group could therefore be increased to three. A similar solution could be arranged for Task 4 (cloze text). By adding a sentence to the ldl cloze text, the two groups could perhaps work on the exercise for approximately the same amount of time.

The fact that students in both groups showed the same response accuracy shows that the difficulty level of the tasks was matched to the students' needs in the respective groups. On average, participant task accuracy was between 76% and 97%. Thus, according to Klieme et al. (2010), it can be concluded that the tasks were not too difficult, and that self-directed reading was successful. The most difficult exercise was Task 7. Nevertheless, 76.63% (ldl) and 84.14% (hdl) of the participants were able to solve it correctly.

For task engagement, which represents the response behaviour, it can be concluded that the students in hdl were slightly more engaged and were thus able to solve the tasks more independently compared to ldl readers. One of the reasons for the lower task engagement of weaker readers could be a lack of reading motivation (Brügelmann, 2020). It might be possible to compensate for this by adding more motivational elements to the digital learning environment. The task settings could also be specifically adapted to address the ldl readers, for example, in Task 2. In this task, students had to write down the names of districts of Graz that begin with a G. During the evaluation, we noticed that many children clicked on "submit" in this exercise without actually solving the task. To simplify solving the task, the district map of Graz, which was originally placed beside the text, could also be added to the task itself. Adding the possibility for students to change the font and the font size, as Bosse (2017) suggests, could also be introduced. This would make the learning environment even more individualized. Additional elements could also be added to stimulate task completion among the less motivated students, e.g., by providing direct feedback through star ratings (Grünberger, 2014).

The correlations found between task engagement and response accuracy showed that response behaviour and scores achieved were related. As children who do not engage in a task, or solve it only incompletely, receive no or few points, such correlation is to be expected. As task engagement correlated highly with score achievement and moderately with processing time, it can be concluded that hdl readers needed slightly more time, but scored higher than ldl readers. Fadljević et al. (2020) also concluded that more points were achieved when learners spent more time on a task. However, using RegioDiff as an example, it is difficult to say whether the shorter processing time is due to low task engagement or to the given level of the task or text (text length, complexity). How long students spend on a text and its corresponding task always depends on individual characteristics (interest, motivation, prior knowledge). Therefore, the level of cognitive activity or task engagement cannot always be inferred from the measured time (Christmann, 2002).

Task performance was deliberately operationalized in this paper to capture several characteristics that might be important in assessing student interaction with the learning environment. Distinguishing between processing time, answer accuracy, and task engagement, made it possible to identify several characteristics that are significant in such interaction. Adding further (assessment) categories is likely to provide even more insight. For example, Ozuru et al. (2007) found that accessing the text during the reading task resulted in less activation of prior knowledge. Schaffner and Schiefele (2013) reached a similar conclusion when they found that children scored lower on reading comprehension checks when they were able to read up on the text. Despite this, in RegioDiff, the decision was made to allow access to the text. In this way, the children have the opportunity to look up passages in the text if they are unclear or while working on the reading tasks. Clicking back to the text was noted in the evaluation, but not considered further in the construct task performance. A more detailed consideration of this aspect would be interesting in more in-depth research.

Acknowledgements

We would like to thank the federal state of Styria for providing financial support. Additionally, we thank the wider project team, including our cooperation partner Wohlhart Lernsoftware for the technical implementation of the material. We are grateful to the teachers and children for their help, and to the students and interns from the University of Graz.

References

- Baum, K., Kirsch, N., Reese, K., Schmidt, P., Wachter, L. and Wolf, V. (2019) "Informatikunterricht in der Grundschule? – Erprobung und Auswertung eines Unterrichtsmoduls mit Calliope mini", In: A. Pasternak, ed., *Informatik für alle, Lecture Notes in Informatics (LNI)*, pp.49-58, Gesellschaft für Informatik, Bonn.
- Bortz, J. and Döring, N. (2016) *Forschungsmethoden und Evaluation: In den Sozial- und Humanwissenschaften* (5th ed.), Springer-Verlag, Berlin.
- Bosse, I. (2017) "Gestaltungsprinzipien für digitale Lernmittel im Gemeinsamen Unterricht. Eine explorative Studie am Beispiel der Lernplattform Planet Schule", in K. Mayrberger, J. Fromme, P. Grell & T. Hug (Eds.), *Jahrbuch*

- Medienpädagogik 13: Vernetzt und entgrenzt – Gestaltung von Lernumgebungen mit digitalen Medien (pp 133-149). Springer Fachmedien Wiesbaden.
- Brügelmann, H. (2020) "Leises und funktionales Lesen stärken! Empirische Studien zeigen: Leseförderung lebt von einem reichen, didaktisch strukturierten Methoden-Repertoire - braucht aber auch eine klare pädagogische Haltung" (2nd ed.) https://www.pedocs.de/volltexte/2020/20356/pdf/Bruegelmann_2020_Leises_und_funktionales_Lesen_II.pdf
- Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., Routledge, New York.
- Christmann, U. (2002) "Methoden der Verstehens- und Verständlichkeitserhebung", *Zeitschrift für Literaturwissenschaft und Linguistik* (128), pp 76–97.
- Duke, N. K., Ward, A. E. and Pearson, P. D. (2021) "The science of reading comprehension instruction", *The Reading Teacher*, 74(6), pp 663–672. <https://doi.org/10.1002/trtr.1993>
- Fadljević, L., Maitz, K., Kowald, D., Pammer-Schindler, V. and Gasteiger-Klicpera, B. (2020) "Slow is Good: The Effect of Diligence on Student Performance in the Case of an Adaptive Learning System for Health Literacy", LAK '20: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, pp 112–117. <https://doi.org/10.1145/3375462.3375502>
- Feyerer, E. (2012) "Allgemeine Qualitätskriterien inklusiver Pädagogik und Didaktik", *Zeitschrift für Inklusion-online.net*, 3, [online], <http://www.inklusion-online.net/index.php/inklusion-online/article/view/51/51> [25.05.2022].
- Grouven, U., Bender, R., Ziegler, A. and Lange, S. (2007) "Der Kappa-Koeffizient [The kappa coefficient]", *Deutsche medizinische Wochenschrift*, 132, Art. 23, pp 65-68. <https://doi.org/10.1055/s-2007-959046>
- Grünberger, N. (2014), "Räume zum Flanieren, Spielen und Lernen: Überlegungen zur Gestaltung von Bildungs- und Lernräumen im Kontext kultureller Entwicklungen", in K. Rummel (Ed.), *Medien in der Wissenschaft: Bd. 67. Lernräume gestalten - Bildungskontexte vielfältig denken*, (pp 56-67). Waxmann, Münster.
- Hattie, J., and Zierer, K. (2019) *Visible learning insights*, 1st ed., Routledge, London. <https://doi.org/10.4324/9781351002226>
- Hemmerich, W. (2019) "StatistikGuru: Cohen's Kappa für zwei Rater berechnen", [online], <https://statistikguru.de/rechner/cohens-kappa-zwei-rater-berechnen.html>
- Hemmerich, W. (2021) "Normalverteilung in SPSS prüfen: Shapiro-Wilk-Test", [online], <https://statistikguru.de/spss/gepaarter-t-test/normalverteilung-ueberpruefen-3.html>
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W. and Stanat, P. (Eds.) (2010) *PISA 2009: Bilanz nach einem Jahrzehnt*, Waxmann Verlag, Münster.
- Kohn, W. (2005) *Statistik: Datenanalyse und Wahrscheinlichkeitsrechnung. Statistik und ihre Anwendungen*, Springer-Verlag Berlin Heidelberg.
- Messiou, K., Bui, L. T., Ainscow, M., Gasteiger-Klicpera, B., Bešić, E., Paleczek, L., Hedegaard-Sørensen, L., Ulvseth, H., Vitorino, T., Santos, J., Simon, C., Sandoval, M. and Echeita, G. (2020) "Student diversity and student voice conceptualisations in five European countries: Implications for including all students in schools", *European Educational Research Journal*, 21(2), pp 355–376. <https://doi.org/10.1177/1474904120953241>
- OECD. (2019) *PISA 2018: Was Schülerinnen und Schüler wissen und können* (1st ed.), wbv Media. <https://elibrary.utb.de/doi/book/10.3278/9783763962983>
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A. and McNamara, D. S. (2007), "Influence of Question Format and Text Availability on the Assessment of Expository Text Comprehension", *Cognition and Instruction*, 25(4), pp 399–438. <https://doi.org/10.1080/07370000701632371>
- Paleczek, L. (2020) "How to Produce and Acquire Regional Knowledge Digitally and in Print: Conceptualisation of the RegioDiff-Project", in C. Busch, M. Steinicke, and T. Wendler (Eds.), *Proceedings of the 19th European Conference on e-Learning*, (pp 611-614). Academic Conferences International Limited, Reading.
- Paleczek, L., Ender, D., Berger, J., Prinz, K., and Seifert, S. (2022) „A feasibility study of digital content use in inclusive, Austrian primary school practice“, *International Journal of Educational Research*, 112. <https://doi.org/10.1016/j.ijer.2022.101938>
- Paleczek, L., and Seifert, S. (2021) „Diagnostik als Basis für Förderung: der Grazer Leseverständnistest“, *Sprachtherapie aktuell: Forschung – Wissen – Transfer*, 8(2), e2021-36.
- Paleczek, L., Seifert, S., Franz, A., Wohlhart, D., and Riedl, S. (in prep.) „GraLeV: Grazer Leseverständnistest [Graz reading comprehension test]“, open access.
- Pospeschill, M. and Siegel, R. (2018) „Methoden für die klinische Forschung und diagnostische Praxis“, Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-54726-7>
- Schaffner, E. and Schiefele, U. (2013) "The prediction of reading comprehension by cognitive and motivational factors: Does text accessibility during comprehension testing make a difference?", *Learning and Individual Differences*, 26, pp 42–54. <https://doi.org/10.1016/j.lindif.2013.04.003>
- Seifert, S., and Paleczek, L. (2021a) "Digitally Assessing Text Comprehension in Grades 3-4: Test Development and Validation", *Electronic Journal of e-Learning*, 19(5), pp 336-348.
- Seifert, S., Schwab, S., and Gasteiger-Klicpera, B. (2014) „Ein Konzept für differenzierten Leseunterricht in heterogenen Klassen“, *Sprache Stimme Gehör*, 38, pp e21-e27. <https://doi.org/10.1055/s-0034-1370988>.
- Wygotski, L. (1987) *Ausgewählte Schriften. Arbeiten zur psychischen Entwicklung der Persönlichkeit: Bd. 2*, Pahl-Rugenstein, Köln.