Standardized Evaluation on the Effectiveness of eKidz 2024

Evaluation by Dr. Bernd Nuss **July 4, 2024**

eKidz was born in 2017 out of a direct need for children in professional families, relocated internationally, to adapt and access the education systems in their new homelands. The Ukrainian founder, Nataliya Tetruyeva, was desperate to help her own children assimilate the languages of their new schools for them to progress their learning and at the same time keep their national identity and their heritage language. She came up with a solution eKidz.

eKidz has grown to benefit other migrant and refugee families either facing challenges of integration into a new society or operating from mobile classrooms in war-torn environments.

Since its foundation, eKidz has been scientifically evaluated in 22 German schools under a Research and Development project, ALFAII. It was supported by the federal government of Germany under an initiative called, “New Educational Platform”. This research places particular emphasis on the performance of at-risk segments of the student population, such as German as a Second Language learners, students with low reading proficiency, and students with low motivation.

From the 1,050 students who were initially engaged in the study, 364 elementary school students in the 2nd grade were selected to build 3 experimental groups and 1 control group that established a baseline for comparison. The design of the study used two similar standardized tests to estimate the entry and exit level in reading speed and comprehension. The time frame between the pre-test and post-test was six weeks. Further research will be conducted over the summer and include AI automatic scoring of oral reading fluency.

Evidence-based Outcomes:

1. Intensive app usage results in **10.4% to 12% higher reading comprehension complexity**, compared to only 5.8% with weak or no usage.
2. Intensive app usage results in **3% to 5% higher reading speed** compared to weak or no usage.
3. Intensive app usage results in **1% to 3% higher accuracy of content reading comprehension** compared to weak or no usage.
4. Intensive app usage results in the **highest growth rates (12.0%)** in comprehension of complex content for readers at the **upper end of the performance spectrum**.
5. Intensive app usage results in **11.4% to 15.0%** higher overall growth rates in the comprehension of complex texts for **non-German-speaking learners** at the lower and upper end of the performance spectrum compared to the overall group.
6. Intensive app usage results in an average of **3.4 to 4.0 points growth** in text comprehension complexity for **learners with low reading interest** compared to learners with high reading interest. App usage clearly compensates for the disadvantages of poor reading motivation.
7. **Learners with low scores in self-assessment suffer a growth disadvantage of up to 4.8 points growth** due to **lack of app usage** compared to a comparison group with similar benchmark results.

1. Participants:

n: 364 total; 91 per quartile
The control group was eliminated for the following reasons:

1. the control group of 64 was too small to serve as a control group.
2. the baseline data of the control group was unequally higher to achieve actual comparative data.

Reduction:
The initial test group size of 1048 was reduced to the final number of 364 due to the following parameters:

1. missing pre-test
2. missing or incomplete (1) post-test
3. missing information on app usage
4. participant (1) without declaration of consent

2. Objects of investigation:

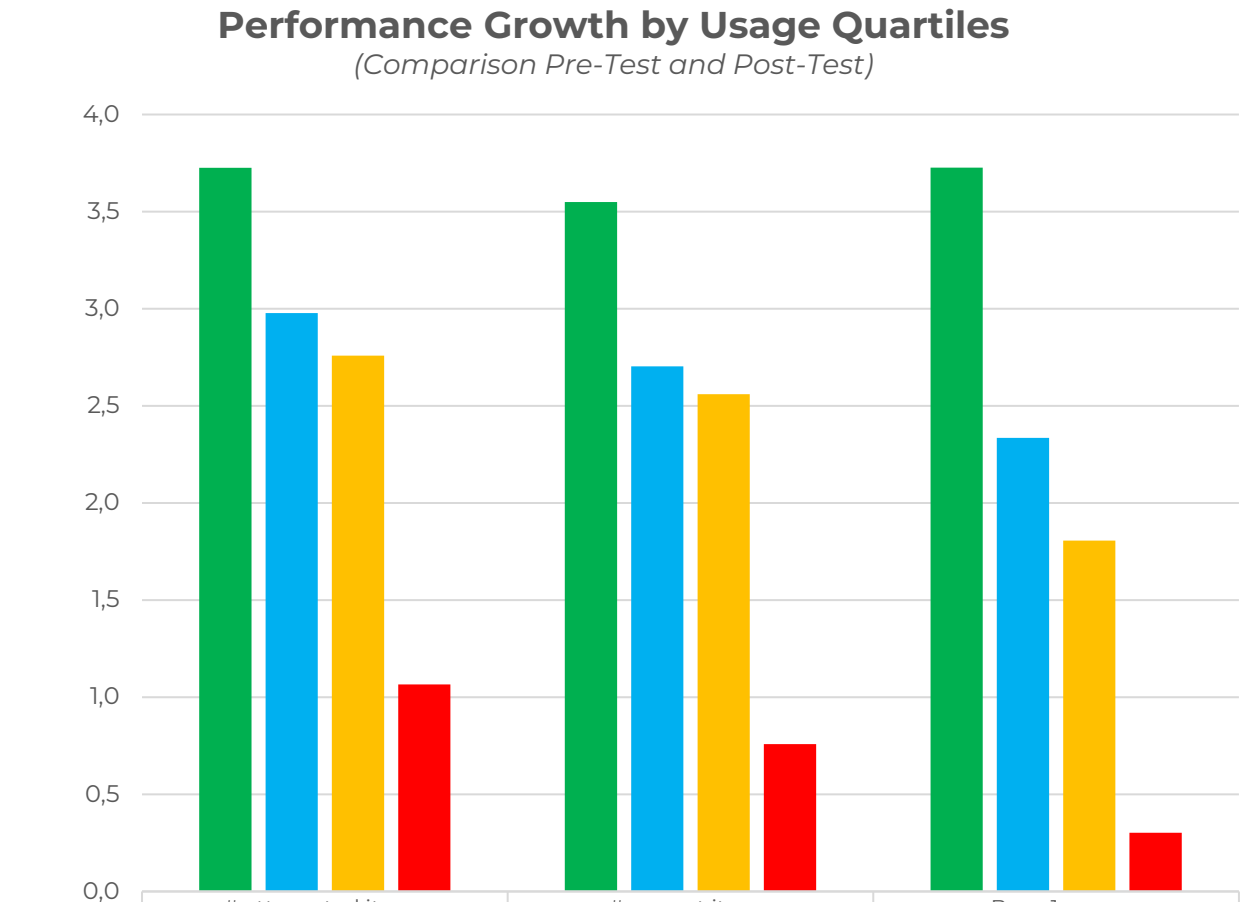
1. Text-related:
   Increase in working speed, increase in accuracy of content comprehension (represented by the absolute number of items that have been correctly identified), and increase of reading comprehension of progressively complex texts (represented by the accumulated summation of the base 1 indices of the individual statements respectively).
   a. Working speed: number of statements processed in a given time.
   b. Accuracy of reading comprehension: Number of correctly completed statements in a given time.
   c. Reading comprehension complexity: Accumulated base 1 values of statements that have been processed correctly.
2. Reader-related:
   Relation between text-related determinants and the intensity of eKidz usage, defined by the number of "activities" during the total usage period of the app; representation in four quartiles.
   i. Overall test group
   ii. Comparison between the entire test group with the subgroup that did not specify German as a spoken language in the demographic survey
   iii. Comparison between the entire test group with subgroups who described themselves as either very uninterested or very interested in their reading behavior survey
   iv. Comparison between the entire test group with subgroups who rated themselves as very weak or very strong in terms of their own reading ability.
   v. Comparison between the entire test group divided into quartiles based on the intensity of app use (determined by the number of "activities"} with the entire test group divided into quartiles based on reading motivation (determined by a motivation factor resulting from a self-evaluative survey before the start of the test series).

3. Tests:

SLS A2 (pre-test) and SLS B1 (post-test), all statements scored according to the integrated base-1 method.

4. Results with discussion:

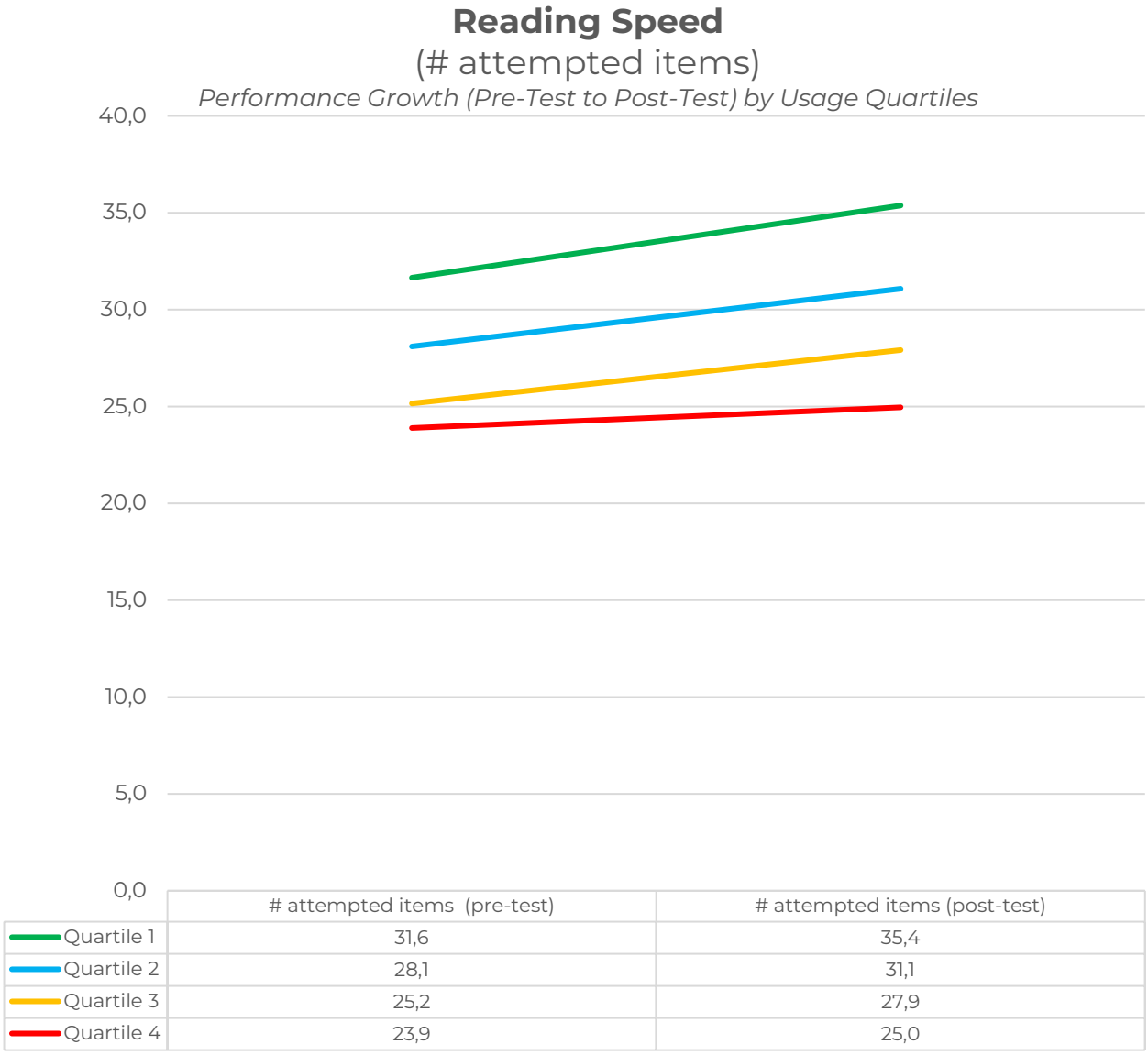
a. Overall test group



	# attempted items	# correct items	Base-1 score
Quartile 1	3,7	3,5	3,7
Quartile 2	3,0	2,7	2,3
Quartile 3	2,8	2,6	1,8
Quartile 4	1,1	0,8	0,3

Fig. 1: Specific performance gains for four participant quartiles (based on app usage intensity) in the areas of working speed (# attempted items), reading comprehension content accuracy (# correct items) and reading comprehension complexity (Base-1 score) calculated from the difference between the values of the post-test and the values of the pre-test.

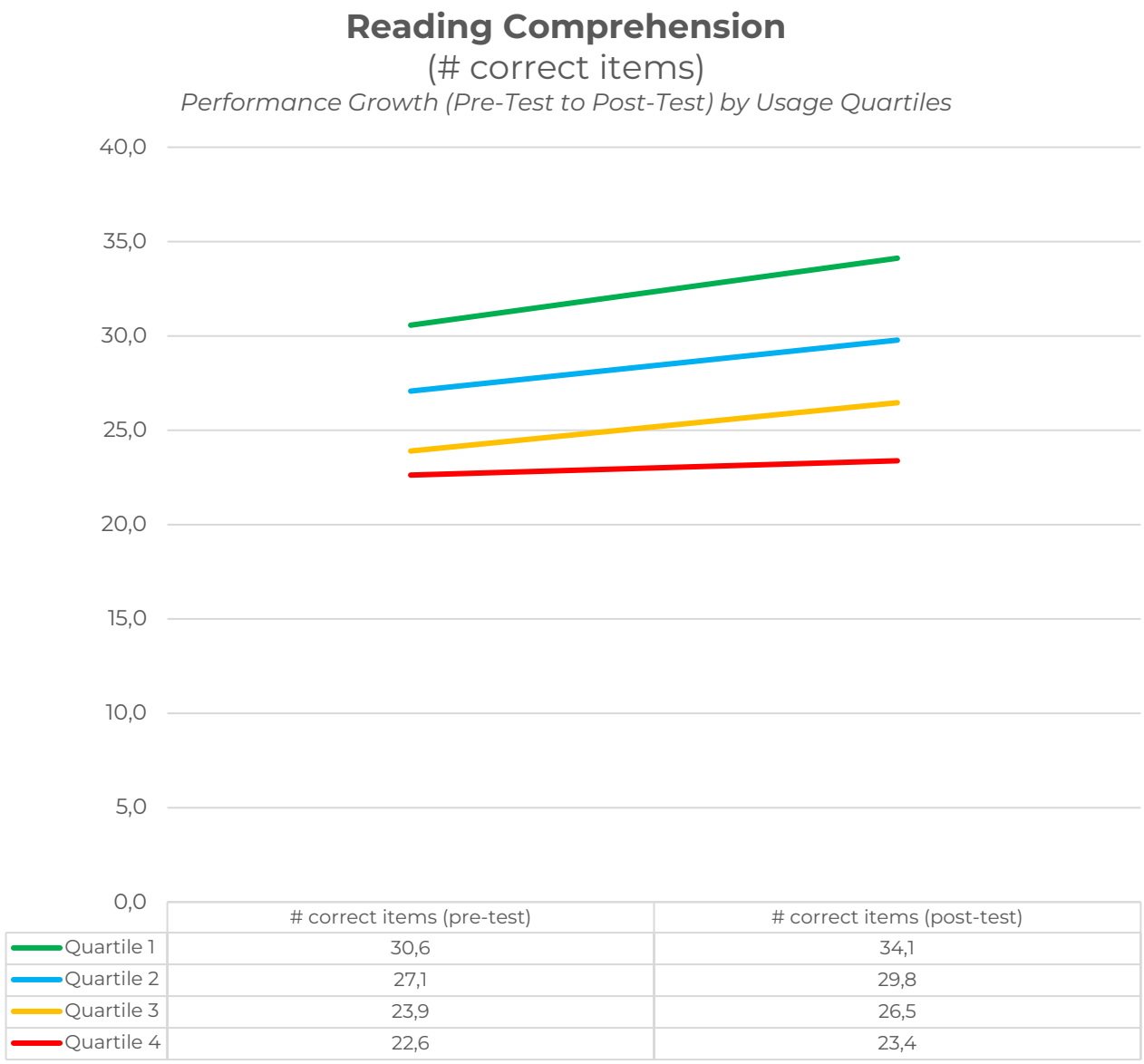
With intensive app usage, the reading results increase evenly at higher levels between 10.4% and 12% according to Base-1, with the two middle quartiles showing relative proximity in all text-related areas. With a lack of use, the results improve between the pre- and post-test with only marginal increases in performance in the individual text-related areas (5.8% according to Base1).



	# attempted items (pre-test)	# attempted items (post-test)
Quartile 1	31,6	35,4
Quartile 2	28,1	31,1
Quartile 3	25,2	27,9
Quartile 4	23,9	25,0

Fig. 2: Specific performance gains for four participant quartiles (based on app usage intensity) in the areas of working speed (# attempted items) calculated from the difference between the values of the post-test and the values of the pre-test.

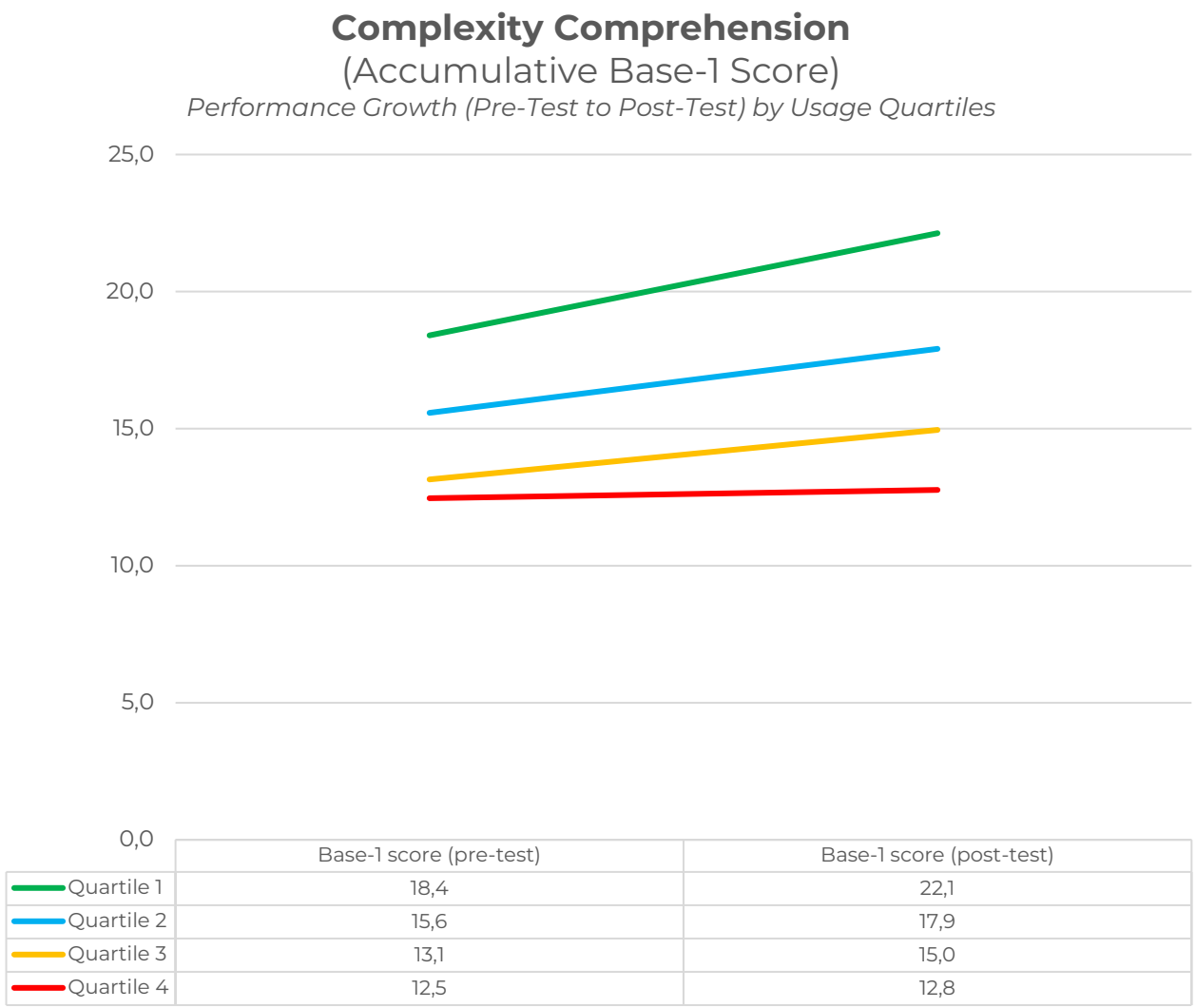
Weak students in particular benefit from intensive app usage in terms of reading speed. Although the pre-test results of the groups in quartiles 1 and 2 only differ by 1.3 points, only the participants with intensive activity rates (quartiles 2, 3 and 4) achieve comparable increase rates of 3.2 points on average. In contrast, the increase rate of the group with low to no use (quartile 4) only reached 1.1 points. In conclusion, intensive app usage achieves between 3% and 5% higher reading speed compared to weak or absent usage.



	# correct items (pre-test)	# correct items (post-test)
Quartile 1	30,6	34,1
Quartile 2	27,1	29,8
Quartile 3	23,9	26,5
Quartile 4	22,6	23,4

Fig. 3: Specific performance gains for four participant quartiles (based on app usage intensity) in the areas of reading comprehension content accuracy (# correct items) calculated from the difference between the values of the post-test and the values of the pre-test.

Particularly weak students benefit from app usage in terms of the accuracy of content-related reading comprehension, as the benchmark scores of the groups in quartiles 1 and 2 differ by only 1.2 points. However, after the pretest, only the participants with higher activity rates (quartiles 2, 3 and 4) achieved comparable growth rates of 2.9 points on average. In contrast, the rate of increase in the group with low to no use (quartile 4) was only 0.8 points. In conclusion, app usage can be seen as causal for increased accuracy of reading comprehension.



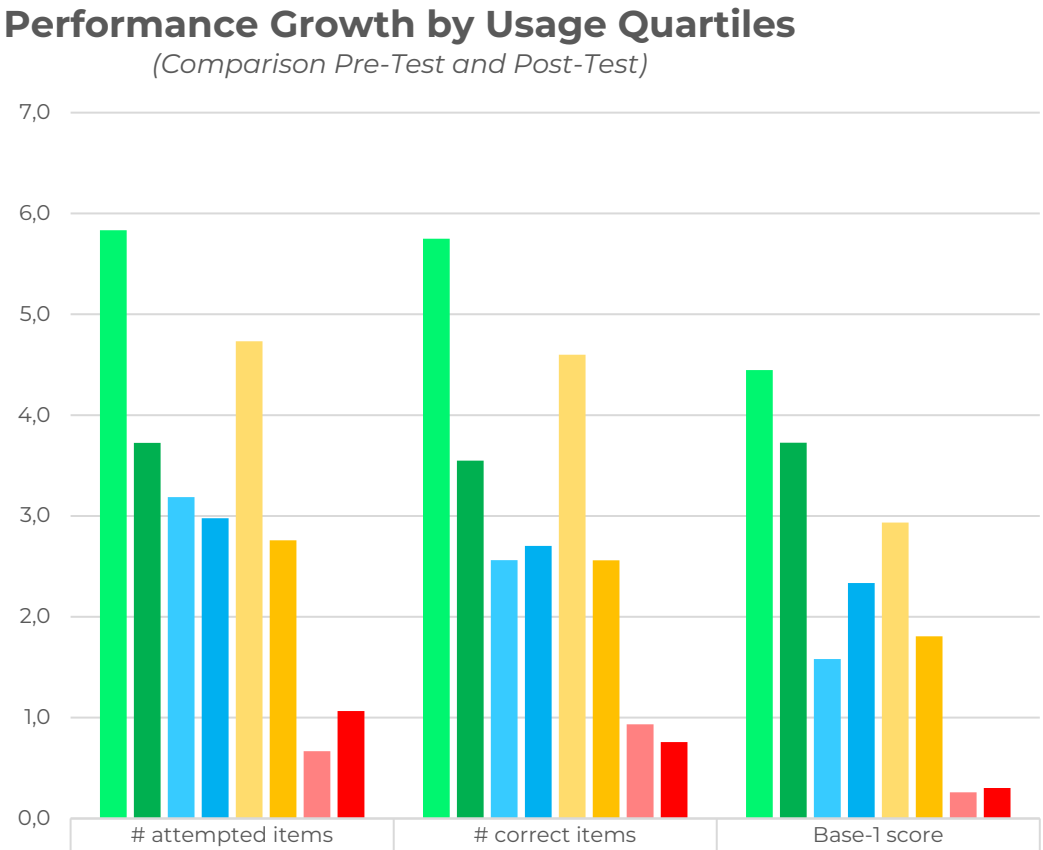
	Base-1 score (pre-test)	Base-1 score (post-test)
Quartile 1	18,4	22,1
Quartile 2	15,6	17,9
Quartile 3	13,1	15,0
Quartile 4	12,5	12,8

Fig. 4: Specific performance gains for four participant quartiles (based on app usage intensity) in the area of reading comprehension complexity (Base-1 score) calculated from the difference between the values of the post-test and the values of the pre-test.

Particularly weak students benefit from intensive app usage in terms of reading comprehension of complex reading content. Although the pre-test results of the groups in quartiles 1 and 2 only differ by 0.6 points, only the participants with intensive activity rates (quartiles 2, 3 and 4) achieve average increase rates of 2.6 points (1.8, 2.3 and 3.7). In contrast, the growth rate of the group with low to no use (quartile 4) only reached 0.3 points. In conclusion, intensive app usage achieves between 4.6% and 6.2% higher accuracy of content reading comprehension compared to weak or absent usage.

Furthermore, participants with high activity rates in the 1st quartile achieved higher gains (3.7) than even the participants in the 2nd and 3rd quartiles (2.7, and 2.6). This corresponds to a growth rate according to the base-1 method of 12.0% (2nd quartile: 11.6%, 3rd quartile: 10.4%, 4th quartile: 5.8%). In conclusion, intensive app usage achieves the highest growth rates in the understanding of complex content for readers at the upper end of the performance spectrum.

b. Comparison between the entire test group with the subgroup that did not specify German as a spoken language in the demographic survey

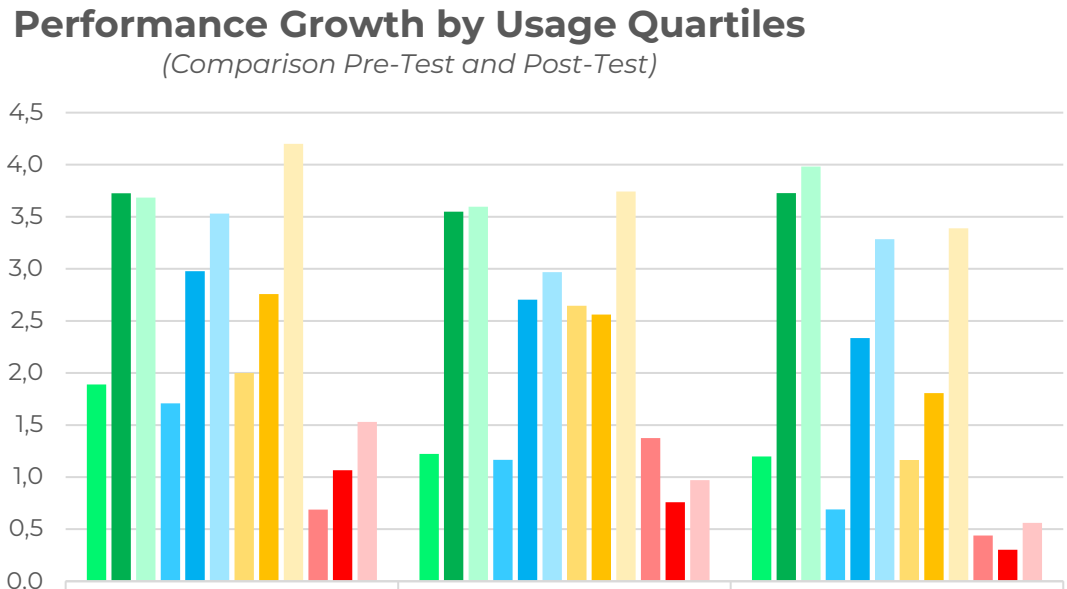


	# attempted items	# correct items	Base-1 score
Quartile 1 (non-German)	5,8	5,8	4,4
Quartile 1 (overall)	3,7	3,5	3,7
Quartile 2 (non-German)	3,2	2,6	1,6
Quartile 2 (overall)	3,0	2,7	2,3
Quartile 3 (non-German)	4,7	4,6	2,9
Quartile 3 (overall)	2,8	2,6	1,8
Quartile 4 (non-German)	0,7	0,9	0,3
Quartile 4 (overall	1,1	0,8	0,3

Fig. 5: Specific performance gains for four participant quartiles (based on app usage intensity) in the areas of working speed (# attempted items), reading comprehension content accuracy (# correct items) and reading comprehension complexity (Base-1 score) comparing the overall group and the subgroup of participants who did not specify German as their spoken language, calculated from the difference between the values of the post-test and the values of the pre-tests.

Non-German-speaking students show well above average gains in working speed, reading comprehension content accuracy and reading comprehension complexity in quartile 1 (57%, 66%, 19%) and quartile 3 (67%, 77%, 61%) compared to the overall group. In conclusion, intensive app usage achieves 15.0%, and 11.4% respectively, higher overall growth rates in comprehension of complex texts for non-German speaking learners in the upper and lower performance spectrum respectively compared to the overall group.

c. Comparison between the entire test group with subgroups who described themselves as either very uninterested or very interested in their reading behavior survey

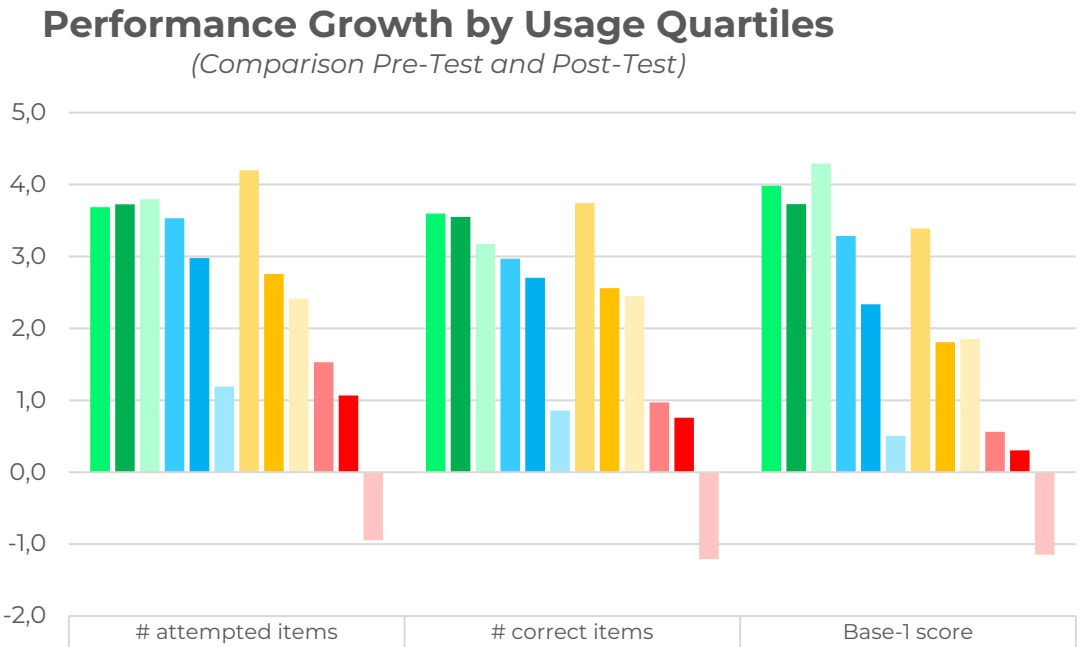


	# attempted items	# correct items	Base-1 score
Quartile 1 (high interest)	1,9	1,2	1,2
Quartile 1 (overall)	3,7	3,5	3,7
Quartile 1 (low interest)	3,7	3,6	4,0
Quartile 2 (high interest)	1,7	1,2	0,7
Quartile 2 (overall)	3,0	2,7	2,3
Quartile 2 (low interest)	3,5	3,0	3,3
Quartile 3 (high)	2,0	2,6	1,2
Quartile 3 (overall)	2,8	2,6	1,8
Quartile 3 (low interest)	4,2	3,7	3,4
Quartile 4 (high interest)	0,7	1,4	0,4
Quartile 4 (overall	1,1	0,8	0,3
Quartile 4 (low interest)	1,5	1,0	0,6

Fig. 6: Specific performance gains for four participant quartiles (based on app usage intensity) in the areas of working speed (# attempted items), reading comprehension content accuracy (# correct items) and reading comprehension complexity (Base-1 score) compared between the overall group and the subgroup of participants who indicated a low or increased interest respectively in reading, calculated from the difference between the values of the post-test and the values of the pre-tests.

The data analysis with regard to self-declared reading interest shows a consistent picture in favor of learners with low reading interest. Learners with low interest in reading achieve higher growth rates in the area of reading comprehension complexity between 6.3% in the 4th quartile and 18.1% in the 1st quartile than readers with a high interest in reading. It can be assumed that the motivating character of the app has a high influence on reading performance, which decreases relatively with increasing reading interest. Intensive app usage achieves on average 10.7% higher growth rates in text comprehension complexity for learners with low reading interest compared to learners with high reading interest.

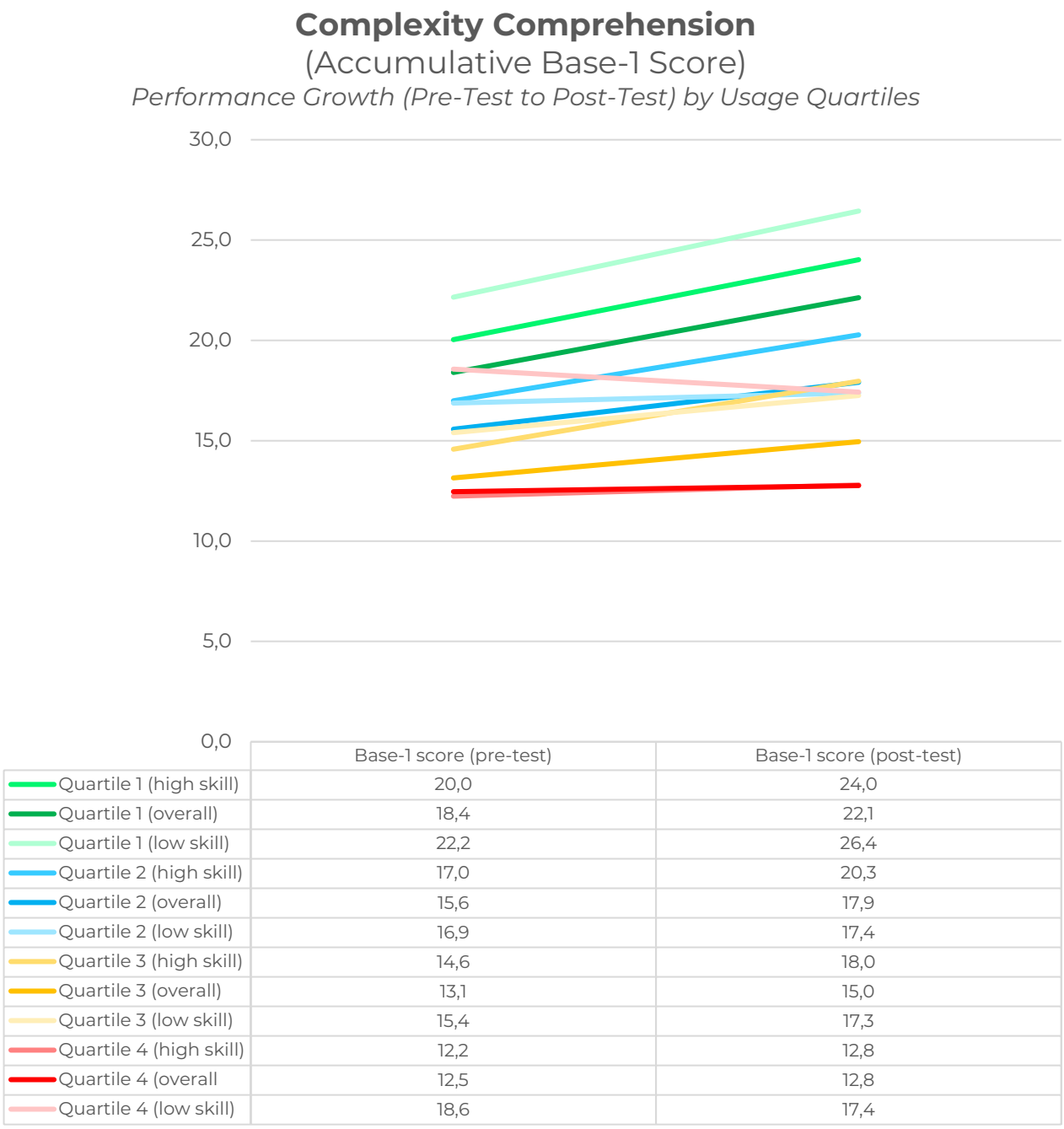
d. Comparison between the entire test group with subgroups who rated themselves as very weak or very strong in terms of their own reading ability.



	# attempted items	# correct items	Base-1 score
Quartile 1 (high skill)	3,7	3,6	4,0
Quartile 1 (overall)	3,7	3,5	3,7
Quartile 1 (low skill)	3,8	3,2	4,3
Quartile 2 (high skill)	3,5	3,0	3,3
Quartile 2 (overall)	3,0	2,7	2,3
Quartile 2 (low skill)	1,2	0,9	0,5
Quartile 3 (high skill)	4,2	3,7	3,4
Quartile 3 (overall)	2,8	2,6	1,8
Quartile 3 (low skill)	2,4	2,4	1,9
Quartile 4 (high skill)	1,5	1,0	0,6
Quartile 4 (overall	1,1	0,8	0,3
Quartile 4 (low skill)	-0,9	-1,2	-1,1

Fig. 7: Specific performance gains for four participant quartiles (based on app usage intensity) in the areas of working speed (# attempted items), reading comprehension content accuracy (# correct items) and reading comprehension complexity (Base-1 score) comparing the overall group and the subgroup of participants who reported low or increased reading skills, calculated from the difference between the values of the post-test and the values of the pre-test.

Overall, the learners' self-assessment does not appear to be indicative of their actual reading behavior. For example, the subgroup of learners who rated themselves as weak achieved the highest increases in the highest quartile in the areas of reading speed and reading comprehension of complex texts. In the second and third quartiles, the self-assessments appear to be more accurate. Even if the learners in the lowest quartile who rate themselves as weak show a negative increase in all areas, their initial scores in the pre-test always achieve the third highest value of all subgroups in all reading areas. This leads to the reverse conclusion that this subgroup would also have benefited from higher app usage.

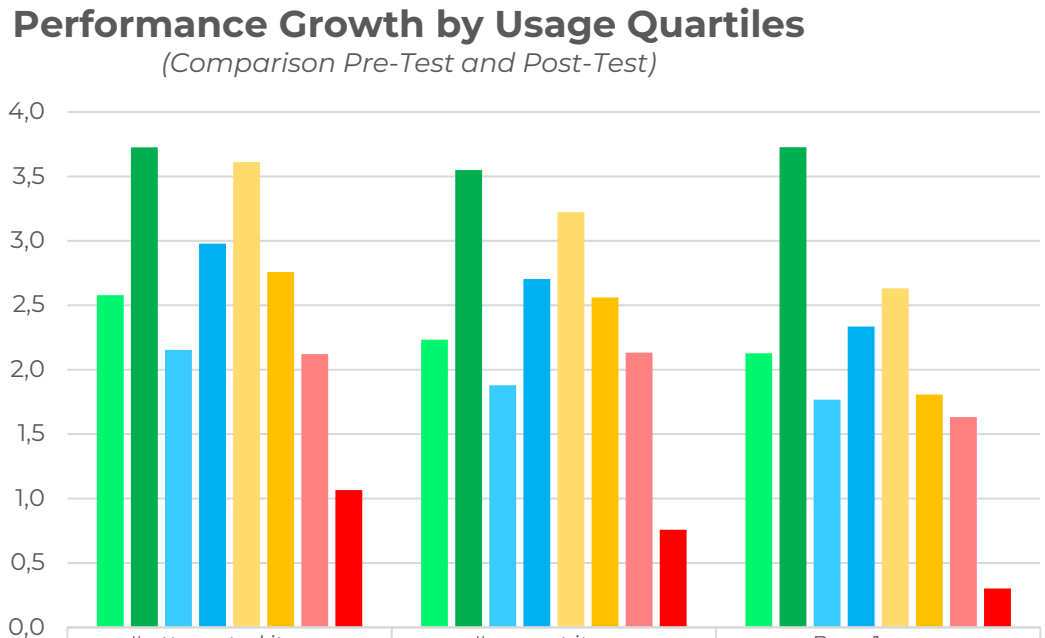


	Base-1 score (pre-test)	Base-1 score (post-test)
Quartile 1 (high skill)	20,0	24,0
Quartile 1 (overall)	18,4	22,1
Quartile 1 (low skill)	22,2	26,4
Quartile 2 (high skill)	17,0	20,3
Quartile 2 (overall)	15,6	17,9
Quartile 2 (low skill)	16,9	17,4
Quartile 3 (high skill)	14,6	18,0
Quartile 3 (overall)	13,1	15,0
Quartile 3 (low skill)	15,4	17,3
Quartile 4 (high skill)	12,2	12,8
Quartile 4 (overall	12,5	12,8
Quartile 4 (low skill)	18,6	17,4

Fig. 8: Specific performance gains for four participant quartiles (based on app usage intensity) in the area of reading comprehension complexity (Base-1 score) calculated from the difference between the values of the post-test and the values of the pre-tests, comparing the overall group and the subgroup of participants who reported low or increased reading skills.

At the same time, however, it is precisely this subgroup that shows the third highest reading comprehension score according to Base-1 of 18.6 after the pre-test, which is only exceeded by subgroups in the 1st quartile, for example, with the highest app usage. Through the comparison with similar performance subgroups, this constellation allows the reversal of the postulate that learners with low self-assessment suffer a growth disadvantage of -19% due to a lack of app usage.

e. Comparison between the entire test group divided into quartiles based on the intensity of app use - determined by the number of "activities" - with the entire test group divided into quartiles based on reading motivation - determined by a motivation factor resulting from a self-evaluative survey before the start of the test series.



	# attempted items	# correct items	Base-1 score
Quartile 1 (motivation)	2,6	2,2	2,1
Quartile 1 (usage)	3,7	3,5	3,7
Quartile 2 (motivation)	2,2	1,9	1,8
Quartile 2 (usage)	3,0	2,7	2,3
Quartile 3 (motivation)	3,6	3,2	2,6
Quartile 3 (usage)	2,8	2,6	1,8
Quartile 4 (motivation)	2,1	2,1	1,6
Quartile 4 (usage)	1,1	0,8	0,3

Fig. 8: Specific performance gains for four participant quartiles in the areas of working speed (# attempted items), reading comprehension content accuracy (# correct items) and reading comprehension complexity (Base-1 score) comparing the overall group based on app usage intensity and the overall group based on the motivation factor, calculated from the difference between the values of the post-test and the values of the pre-tests.

The comparison of usage and motivation shows consistently for all text-related areas that learners with low reading motivation benefit from using the app and achieve rates of increase that in some cases double the average rates of increase. This is partly because learners of all app usage intensities are in the lowest motivation quartile. However, it should also be noted that the difference between the increase in all text-related areas of highly motivated learners and the increase of low-motivated learners only achieves a maximum difference of 0.5. In conclusion, app usage clearly compensates for the disadvantages of poor reading motivation.