

CHAPTER V

CONCLUSION AND SUGGESTION

5.1 Conclusion

Drawing on the data analysis and discussion in the previous chapter, the researcher arrives at the following conclusion. This study set out to determine whether students' reading comprehension improvement on descriptive texts differed significantly between those assessed through the gamified digital assessment mode (Wayground) and those assessed through the conventional digital assessment mode (ProProfs), among tenth-grade students in the TKR program at SMKN 1 Mojokerto during the 2025/2026 academic year.

Addressing the research question, whether students' reading comprehension improvement from pre-test to post-test differs significantly between the gamified (Wayground) and conventional (ProProfs) digital assessment modes, the findings showed no statistically significant difference between the two groups. Because the two groups' pre-test scores were not equivalent at baseline (Wayground: $M = 87.41$; ProProfs: $M = 82.93$), the main comparison relied on each student's gain score (post-test minus pre-test) rather than the raw post-test score. The Mann-Whitney U test on these gain scores produced $U = 389.500$, $Z = -0.493$, $p = .622$ ($p > .05$), with a negligible effect size of $r = 0.065$, so H_0 was retained and the study did not gather sufficient evidence to support H_1 . Descriptively, ProProfs edged out Wayground in mean gain ($M = 10.17$, $SD = 11.763$ versus $M = 7.93$, $SD = 9.114$), though this gap fell short of statistical significance. Within each group, the Wilcoxon Signed-Rank Test confirmed significant pre-test-to-post-test improvement for both Wayground ($Z = -3.645$, $p < .001$, $r = 0.677$) and ProProfs ($Z = -4.026$, $p < .001$, $r = 0.748$). Although the two groups' improvement patterns looked different descriptively, that gap in gain scores was not statistically significant, so the study lacks sufficient evidence that either digital assessment mode produced greater reading comprehension improvement than the other.

These results suggest that, within this study's scope, a digital assessment platform's level of gamification was not the deciding factor behind how much students' reading comprehension improved. Both groups improved significantly from pre-test to post-test with large effect sizes, and no statistically significant gap emerged between the two modes in the size of that improvement, meaning instructional factors such as text quality, question design, and post-assessment feedback may matter just as much for reading comprehension outcomes as whether gamification was present. The control group's slightly higher descriptive gain

should be read cautiously, since it was not statistically significant and may partly reflect that group's lower starting point.

This study adds to the EFL digital assessment literature by offering empirical evidence that gamification level does not significantly moderate reading comprehension improvement, pushing back on the assumption that gamified platforms are inherently the stronger assessment tool. For EFL reading pedagogy specifically, it extends gamification research, which has largely centered on vocabulary and grammar, into extended-text reading comprehension, showing that the sustained, focused attention descriptive text processing demands may not gain from, and could even be hampered by, competitive gamification features. These findings give vocational EFL teachers and digital assessment platform developers evidence-based guidance for choosing assessment tools based on pedagogical fit rather than assumed motivational superiority.

These findings point to both gamified and conventional digital assessment modes as equally valid options for assessing reading comprehension, with the choice between them reasonably resting on instructional objectives, classroom conditions, and teacher preference rather than on any statistically significant improvement difference.

5.2 Limitations of the Study

Several limitations should be kept in mind when interpreting these findings. First, the two groups' pre-test scores were not equivalent at baseline; although gain scores helped offset this initial imbalance, the non-equivalent control group design means some residual influence of prior ability cannot be fully excluded. Second, the sample was capped at 58 students across two intact classes, limiting statistical power and generalizability. Third, the study took place at a single vocational school (SMKN 1 Mojokerto), so its results may not extend to other schools, educational levels, or regions. Fourth, the treatment ran for a relatively short span, only a few sessions, which may not have been enough to capture the assessment modes' longer-term effects on reading comprehension. Fifth, these findings are tied specifically to the two platforms examined, Wayground and ProProfs, and may not generalize to other gamified or conventional digital assessment platforms with different features. Future researchers should factor in these limitations, as the next section outlines.

5.3 Suggestion

Based on the findings and conclusions of this study, the researcher proposes the following suggestions for relevant parties.

5.3.1 For English Teachers

Based on these findings, the following recommendations are offered to English teachers at vocational high schools. (1) Since this study found no statistically significant difference between the two digital assessment modes in reading comprehension improvement, this should not be read as proof that Wayground (gamified) and ProProfs (conventional) work equally well; it simply means that, within this study's scope and limitations, no significant improvement gap between the two platforms was detected. Teachers can therefore choose either platform for reading comprehension assessment based on practical factors, school conditions, internet access, class size, teacher readiness, familiarity with the platform, and the specific learning objectives at hand. (2) When using Wayground (gamified mode), teachers should give students a brief platform orientation in the first session so they can navigate gamified features (timers, leaderboards, answer animations) without confusion, and for lower-proficiency classes, consider turning off the public leaderboard to ease competitive anxiety that might distract from reading. (3) When using ProProfs (conventional mode), teachers should follow up the assessment with an active post-quiz discussion, reviewing wrong answers in detail and having students locate supporting evidence in the text, since the platform's minimal interactivity means the teacher needs to supply motivational and reflective scaffolding directly. (4) Regardless of platform, teachers should favor descriptive text passages that connect to students' vocational field (automotive-related texts for TKR students, for instance), since familiar content lowers the vocabulary burden and frees up more cognitive resources for comprehension. (5) Gamified platforms such as Wayground suit short, low-stakes formative checks, vocabulary or grammar review, for example, better than extended reading comprehension assessment, where sustained focus outweighs competitive engagement; when they are used for reading assessment, exposing students to the platform across at least three to four sessions lets the novelty effect settle before drawing conclusions about its impact, and features like strict countdown timers should be eased or dropped for lower-proficiency classes to cut down on time-pressure distraction.

5.3.2 For Students

Students are encouraged to approach reading comprehension assessment seriously and carefully no matter which digital platform is in use. This study's findings suggest that strong reading comprehension scores come from focused attention to the text itself rather than from leaning on an assessment platform's motivational features. Students would also do well to actively build their vocabulary, especially vocabulary tied to their vocational field, since this

will strengthen their ability to comprehend descriptive texts across different assessment settings.

5.3.3 For Future Researchers

Building on these findings and limitations, future researchers might consider the following. (1) Replicate this study with larger samples (at least $n = 60$ per group) drawn from multiple schools or regions to boost statistical power and generalizability. (2) Extend the treatment to at least four to six weeks so gamification's novelty effect has time to settle, since a two-session treatment may not be enough to capture its full motivational and achievement effects. (3) Use matching or random assignment to secure pre-test equivalence between groups before treatment, or apply ANCOVA to statistically adjust for baseline differences and obtain more internally valid treatment effect estimates. (4) Pair achievement tests with process-level measures, such as validated motivation questionnaires (IMI or AMS, for example), engagement observation checklists, cognitive load scales, or think-aloud protocols, to capture how different assessment modes shape the reading process itself and not just its outcome. (5) Examine whether proficiency level moderates the relationship between assessment mode and achievement, particularly whether lower-proficiency students systematically gain more from conventional digital platforms than from gamified ones, as the convergence pattern in this study hints.