JEELS
(Journal of English Education and Linguistics Studies)
P-ISSN: 2407-2575 E-ISSN: 2503-2194
https://jurnalfaktarbiyah.iainkediri.ac.id/index.php/jeels

THE EFFECT OF GROUP-INDIVIDUAL COLLABORATIVE TESTING ON PRIMARY STUDENTS' ACHIEVEMENT IN READING TEST

***Yuliyanto Chandra**[1]
[1]Master's Program in English Education, Widya Mandala Surabaya Catholic University, Indonesia
**yuliyanto.chandra1996@gmail.com*
(*) Corresponding Author

Abstract: Collaborative testing—a two-stage testing modality in which students do tests as a group and individually—has been extensively researched and found to improve learning. However, most collaborative testing research has focused on non-English language learning. Therefore, this study examines the effects of group-individual collaborative testing (GICT) on English reading test achievement and student views of GICT. This sequential explanatory mixed-methods study gathered quantitative data from both treatment and control groups, primarily comprising students' correct answers, followed by qualitative insights derived from students' reflections on their learning experiences. Quantitative data demonstrates that GICT had a significant effect on students' reading achievement tests. While the qualitative data reveals that

[1]**Citation in APA style:**
Chandra, Y. (2024). The effect of group-individual collaborative testing on primary students' achievement in reading test. *JEELS*, *11*(1), 157-184.
DOI: 10.30762/jeels.v11i1.2630

Submission: January 2024, Revision: March 2024, Publication: April 2024

students considered GICT was a good reading testing modality, with some notes on the scoring method. Practice and research implications include 1) using GICT to test other language skills, 2) diversifying group composition, 3) giving more weight to the individual part score because students will retrieve more information during this stage, and 4) instilling collaborative learning values before testing.

Keywords: *collaborative testing, cooperative learning, group-individual, reading achievement, primary students*

INTRODUCTION

The fallacy that language testing merely amounts to providing a grade that students deserve—which somewhat pigeonholes them into certain categories of competence—has made traditional individual written language testing prevalent and a normalized rule in testing across educational levels. However, since such a fallacy does not encapsulate the multiple facets of language testing, reforming the way language testing is conducted is imperative. Hence, this present study considers collaborative testing—a modality that has seen a surge in literature (Caboral-Stevens & Fox, 2020; Cantwell et al., 2017; Eastwood et al., 2020; H. Jang et al., 2017; LoGiudice et al., 2015; Mahoney & Harris-Reeves, 2019; Wissman & Rawson, 2018)—as a reformed way of testing and aims to investigate how such a test can be carried out as learning.

Body of literature has suggested that there are two possible ways of conducting collaborative testing (Eastridge & Benson, 2020). First, collaborative testing can be conducted as an individual-group collaborative test (IGCT) design and second, a group-individual collaborative test (GICT) design. Both designs come with their own benefits and take different considerations when implemented.

IGCT design is when students do the test individually at first and then as a group. In other words, the students do the collaborative testing as a post-test. Collaborative testing as a post-test serves as a preventive effort towards possible academic dishonesty and score

inflation. Jang et al. (2017) argue that collaboration on tests may be beneficial for low-performing students, while remaining not so much towards high-achievers. For that reason, they suggest that IGCT may function as a solution to this concern in collaborative testing. Some previous studies report on the effectiveness of this model towards students' positive perception and increased scores (Eastwood et al., 2020; Jones, 2019). However, from the perspective of students' engagement and participation, this model may not support students' active participation in the collaborative testing. While collaborative testing can be thought of as a form of 'collaborative thinking' which involves "process of two or more people coordinating their thoughts to negotiate shared understanding, solve problems, and accomplish shared goals" (Hard & RaoShah, 2022, p. 2), using collaborative testing as a post-test may not allow students to participate in the collaborative stage due to lack of group brainstorming—which the present study deems as a crucial factor that could possibly lead to better negotiation of meaning, inferencing and successful retrieval. Because this factor assumes that "the ideas expressed by one group member can cue new ideas or categories of ideas another group member would not have generated or surveyed on their own" (Mende et al., 2021, p. 31), the absence of group brainstorming during individual test in IGCT model does not allow students to explore diverse ideas, correct misinformation, and retrieve important information. Hence, it not only hinders them from meaningfully participating in collaborative test but also renders the individual stage rather meaningless as no meaningful negotiation of meaning is performed.

On the other hand, GICT design works when students do a test as a group before doing it individually. This arrangement is derived from the idea that students will improve collaboration with peers and build on their ideas when answering questions (Eastridge & Benson, 2020). Similarly, this design also empowers the students, meaning that it reduces test anxiety as well as stimulates thinking. Different from IGCT, GICT may support longer retention of the learned

materials—as most of the important processes of collaborative thinking transpiring during the collaborative test help students to further retrieve 'enhanced' information in the individual stage. The word 'enhanced' here particularly indicates that the information the students retrieve in the individual stage might be the results of negotiation, correction, and shaping that previously happen in the collaborative stage. Hence, the model not only helps students to perform group brainstorming as a form of *preparation* before the individual test, it also allows students to do the individual test more meaningfully—as much of the collaborative processes become transferrable in the individual test.

As one model of collaborative testing, GICT also offers several benefits for the field of language assessment and testing. Because independent testing does not facilitate students' needs of collaboration—one of the main tenets of 21st century skills—GICT may come across as facilitative of the said skills. Mahoney and Harris-Reeves (2019) argue that collaborative tasks in student testing are effective to enhance students' learning outcomes. The implementation of GICT, thus, is important to promote problem-solving (Eastwood et al., 2020; Levine et al., 2018), sharing of different perspectives (Mahoney & Harris-Reeves, 2019), questioning and understanding different points of view (Mahoney & Harris-Reeves, 2019), and solving disagreements (Bećirović et al., 2022; Mahoney & Harris-Reeves, 2019)

Besides, in relation to the purpose, test generates "an insight into students' learning processes and provide them with feedback that can help them improve their learning, or to identify potential learning problems that need to be addressed through appropriate pedagogical intervention" (De Angelis, 2021, p. 2). From collaborative testing perspective, besides teachers, peers can also assume the role of feedback providers and give equally supportive and valuable feedback for the students as they learn together to reach certain instructional goals. Having peers as feedback givers maintains that collaborative testing provides students with a source of motivation

(Bećirović et al., 2022), as they gain substantial knowledge through the interaction they have with their peers (Caboral-Stevens & Fox, 2020). Collaborative testing is also important due to its role in lessening test anxiety—one factor in traditional individual test which has been described as a negative predictor of language achievement (Sylvia et al., 2023; Zheng & Cheng, 2018). Against this backdrop, a GICT might be facilitative of collaboration as well as improve achievement in language learning.

As a form of active learning and a learning strategy (Levine et al., 2018), collaborative testing supports student-centeredness (H. Jang et al., 2017) and can have several effects on the students' learning and skills. Besides motivation, some other effects of it also include longer retention (LoGiudice et al., 2015; Salomone & Kling, 2017), enhanced problem-solving skills (Levine et al., 2018; Namaziandost et al., 2020, p. 10) and improved creativity (Segundo Marcos et al., 2020). Collaborative testing also accentuates the magnitude of the experiential learning and positive interdependence (Bećirović et al., 2022)—which implies that learners tend to mutually share knowledge and resources for the purpose of learning (Zipp, 2007). Therefore, in cooperative testing, the learners are encouraged to provide mutual help which potentially constitutes offering cognitive assistance and socio-affective support (Koutrouba & Karageorgou, 2013). All of these perceived and research-based benefits of collaborative testing—which includes GICT as one of the models—are derived from the view of summative *assessment as and for learning* (Caboral-Stevens & Fox, 2020; Lam, 2020).

Much of the research body on collaborative testing, however, has tended to focus on three most prevalent variables. For example, collaborative testing has been mostly carried out in the context of higher education (Cantwell et al., 2017; Eastridge & Benson, 2020; Eastwood et al., 2020; Kapitanoff & Pandey, 2018; Levine et al., 2018). Most studies on collaborative testing have also focused on its usage in disciplines such as nursing (Eastwood et al., 2020), math (H. Jang et al., 2017), statistics (Kapitanoff & Pandey, 2018), and health science

(Mahoney & Harris-Reeves, 2019). Little evidence of cooperative test-taking has been generated in the context of EFL. Lastly, literature has also been mostly concerned with the use of IGCT rather than GICT (Eastwood et al., 2020; H. Jang et al., 2017; Jones, 2019; Zipp, 2007).

This present study specifically focuses on applying GICT to test young learners' reading comprehension because reading is defined as meaning-making processes (Khampool & Chumworatayee, 2023; Smith et al., 2021). Khampool and Chumworatayee (2023) also specifically explain that reading is a process where students, as readers, interact with the text, reading strategies and fluency. Because the interaction of all the aforementioned elements potentially creates meaningful reading, GICT, based on the social constructivist take on learning, should prove relevant and beneficial as students scaffold each other in applying appropriate strategies in constructing meaning of the reading materials given to them.

Hence, this study aims to contribute to the growing literature in collaborative testing which specifically focuses on the use of GICT in EFL context, especially to improve reading achievement among primary students. As this modality is applicable for students across levels due to its collaborative nature and benefits, insights from young learners may enhance the applicability of GICT for diverse learners. Understanding young learners' voice about GICT is especially crucial because despite frequently being subjects to positive and negative experiences of assessment or test (Butler et al., 2021), their perceptions and thoughts have frequently been subverted by the immense focus on higher level students. Additionally, by foregrounding their perceptions regarding GICT, the study acknowledges young learners as stakeholders of learning and assessment itself as well as promote student-centeredness.

Given the importance of GICT as a potentially effective language testing modality in testing young learners' reading comprehension and the gap in literature of the said modality, the present study seeks to investigate how GICT improves the achievement of primary students in reading test and how students

perceive the implementation of GICT. For that reason, it behooves this study to answer the following research questions:

1) Does GICT in EFL testing improve primary students' achievement in reading test?
2) Does students' achievement in reading test using GICT significantly differ from traditional individual testing?
3) How do primary students perceive the implementation of GICT in reading test?

METHOD

Research Design

This study employed the explanatory sequential mixed-method, meaning that the quantitative method informed the use of the qualitative method as the latter generated more insightful explanation to and refined the quantitative results (Walker & Baxter, 2019). The idea of using both quantitative and qualitative methods was driven by the need to explore whether GICT had effects on students' reading achievement test or not, and students' perceptions on the implementation of GICT. The quantitative data here were attributed to the needs of providing objective measures to GICT's effects before the qualitative data were analyzed because perceived learning is not synonymous to the actual effects of learning—specifically cognitive effects of GICT (LoGiudice et al., 2015).

At first, quantitative data were collected through implementing a collaborative reading test for Grade 6 students at a private Christian primary school in Surabaya, Indonesia. Then, after the completion of quantitative data collection, the qualitative approach was followed due to the need to delve deeper into the students' thoughts on the implementation of the GICT. The qualitative data were collected through the use of students' reflections. Here, the data from the students' reflective questionnaire gave the explanatory power to the quantitative data.

Participants

Convenience sampling was applied in this study. 42 Grade 6 students participated in the study. They were divided into two groups: 21 for the controlled group and 21 for the treatment group. Each group consisted of an equal number of male (11) and female (10) students. In terms of English proficiency, both groups had similar level of reading comprehension based on the results of the pre-assessment, which will be further discussed in the latter part of this section.

It must be noted, however, that due to the sampling method of this present study, the generated results may not be representative of the population at large. This is in line with Andrade's (2021) idea that the generalizability of the results is only applied to the conveniently accessible population.

Since the participants were minors, parental approvals were necessary. For ethical consideration, online consent forms—which provided information on ensuring confidentiality and anonymity, the objectives of the study, how the tests would be conducted and how the scores would be calculated—were given to the parents. After receiving the consents from all the parents, the study was conducted.

Materials

The GICT design of this study was implemented based on a reason that the reading materials in individual and collaborative phases were made different in terms of topics, but aimed to measure the same competencies. Thus, reading sections of TOEFL Junior, which were available online, were used. Both tests consisted of 20 reading questions.

For validity and reliability measures, the tests were checked and evaluated by the school curriculum consultants before being administered to the students. Then, all the test questions were tried out to another class of the same grade which did not participate in the study. After the try-out, item analysis was conducted and some too-demanding questions were taken out or replaced. The idea of omitting and replacing too demanding questions was based on fair

participation between students. If the questions were too difficult, it was assumed that needs-improvement and satisfactory (see Data Collection section for notes on the group composition) students would not be able to participate in the group discussion.

Data Collection

The instruments used to generate the data for this study were reading test questions from TOEFL Junior, observation and reflective writing. First, the reading test was conducted in two phases. For the treatment group, the students were clustered into groups of three. They did the test collaboratively in the first phase. Then, in the second phase, they did the similar test individually. The underlying idea of implementing GICT was to allow the students to transfer what they had discussed in their respective groups into doing the individual test. On the other hand, the controlled group did the test individually in both phases. The following table shall provide the data collection process.

Table 1.
Data Collection Process

Step	Date	Treatment Group	Control Group
1	26 January 2023	Pre-assessment (used for grouping purpose and getting information about students' current reading comprehension skills)	Pre-assessment (used for getting information about students' current reading comprehension skills)
2	-	Grouping based on pre-assessment results	-
3	3 March 2023	Reading test in groups	Reading test individually
4	10 March 2023	Reading test individually	Reading test individually
5	13 March 2023	Students' reflection	Students' reflection

After acquiring the permission from the school's principal and students' parents, I developed three sets of reading tests; one for pre-assessment, another one for the first phase, and the other one for the

second phase. All of the tests questions were taken from TOEFL Junior questions which were available online.

Pre-assessment was then conducted to gain insights into students' reading ability. The data gathered from this pre-assessment were used to classify the students into three categories: excellent readers, satisfactory readers, and need-improvement readers. The classification adapted the school's grading policy; A for excellent (score 86-100), B for satisfactory (score 70-85), and C for need improvement (score below 70). The results of the pre-assessment revealed that both groups had similar level of reading comprehension. Then, this pre-assessment led to the grouping of the students in the treatment group. Each group consisted of three students; totaling seven groups. The following Table 2 provides the results of the pre-assessment.

Table 2.
Results of pre-assessment for grouping the treatment group

Classification	**Treatment Group**
Excellent readers	8 students
Satisfactory readers	6 students
Need-improvement readers	7 students
Total	**21 students**

In the first phase of the testing—which was done one week before the second phase—both the treatment group and control group did the test on the same day. The treatment group did the test in groups, whilst the control group individually. In the second phase, students from both groups did the last reading test individually. During the administration of the first phase, observation was conducted to record any ongoing activities among the students. Students were observed in terms of how they did the test collaboratively to solve the provided questions.

Following the completion of the test and scoring the test, students had to write up a short reflection. The reflection was used as the source of qualitative data which supposedly delineate the

students' thoughts on the use of collaborative testing to improve reading achievement. This is in line with Mohan (2020) who argues that reflection can be used as a research tool to capture the research participants' thought processes. However, since primary students took part in the study, a more elaborate model of the reflection was given—meaning that some guiding questions were provided to help students reflect on the collaborative testing. The following table provides the overview of the reflective questions that guided the students during their reflection phase.

Table 3.
Reflective questions

Area	Question
Experience in doing GICT	How was your experience while working with your group members during the test?
Perceived strengths of GICT	What do you think were good or the strengths of working together with your group members during the test?
Perceived weaknesses of GICT	What do you think were not so good or the weaknesses of working together with your group members during the test?
Coping skills during GICT	How did you deal with the difficulties?
Suggestion for improvement	What are your suggestions for improving the collaboration?

These reflective questions were designed based on the third research question which specifically aims to explore students' perception on GICT. The questions shall help the students explain the processes in GICT, the strengths and weaknesses, and the suggestions for improving GICT. Furthermore, the reflective activity was given to the students to delve into the perceived benefits of GICT as discussed in the literature review part of this study. For example, by understanding students' experience, perceived strengths and weaknesses, coping strategies and suggestions for improvement, the study may reveal how GICT relates to students' increased motivation,

decreased anxiety, enhanced problem-solving skills, creativity and others. It may also reveal the group processes that happened among students while they negotiated meaning and solved instructional problems.

Similar to test questions, for validity and reliability measures, the reflective questions were tried out with another class of the same grade, which followed the same process of GICT as the treatment group. Amendments of the questions were made in terms of the clarity. Besides, the questions were also checked and evaluated by the school's curriculum consultant.

Data Analysis

After all the data had been gathered from both groups, the analysis was performed. However, a preliminary check was carried out to ascertain normality and identify outliers. JASP software was used to conduct the normality test and all of the other subsequent quantitative tests. The following figure provides the visualization of the data analysis procedures.

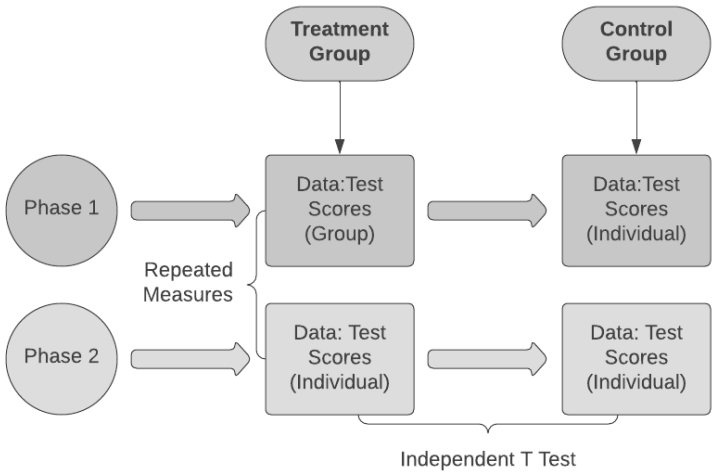

Figure 1. Visuals of Data Analysis Procedures

To analyze the quantitative data, the number of correct answers—which serve as the dependent variable—were tabulated based on the groups and phases. Coding was first assigned to the

students as a way to maintain anonymity. For treatment group, the coding consisted of a letter followed with a number. The letter represented the group, while the number represented students' achievement as determined by the pre-assessment test. The number 1, 2 and 3 indicate high to low achievement respectively. The control group students were coded with the same letter as no grouping was done.

The tabulated data were analyzed twice. As Figure 1 above indicates, first, to answer the first research question, the treatment group's number of correct answers from both phases were analyzed using repeated measures ANOVA. Furthermore, based on the results of the Shapiro-Wilk Test ($\alpha = 0.05$) in the first test, the data followed the normal distribution (p-value was 0.1467). Then, to answer the second research question, the number of correct answers from both groups in phase two were analyzed using independent t-test. Similar to the first normality test, the results of the normality test for both groups also reported that the data were normally distributed (treatment group's p-value was 0.0695 and control group's p-value was 0.1988). While carrying out the test, no outlier was found in the data.

Lastly, to answer the third research question, the data from students' written reflection were retyped and organized in NVIVO, a software for organizing qualitative data. Then, a thematic analysis was carried out. The themes were generated based on the reflective questions. For example, students' coping strategies during group processes, strengths and weaknesses of GICT, and students' suggestion were the central themes. All the sub-themes were also analyzed. Hence, coding was applied in the process of analysis. For instance, CS (coping strategy), St (strength), W (weakness), and Su (suggestion) were used for the main themes and letters A, B, C and so on were used to indicate the generated sub-themes.

However, not all responses from the students were included in the data analysis process. For example, exclusion or omission was applied to students' responses on the question related to suggestion.

Some students provided answers such as, "*I think it's already good*", "*I have no suggestion*", and "*No suggestion*". As these answers did not provide any explanation or insights for the improvement of GICT, such answers were omitted from the analysis.

FINDINGS

In this section, findings of the study are presented in the order of the research question. First, the findings related to the first research question are explored. Then, to compare between the use of GICT and traditional individual test, the results of the second research question are presented. Lastly, to support the quantitative data, the results of qualitative analysis of the students' reflections are presented.

Research Question 1: Does GICT in EFL testing improve primary students' achievement in reading test?

The result of the study revealed that GICT in EFL reading test was fruitful in improving students' reading test achievement. As shown in Table 4, the students in the treatment group had a significant improvement in their individual reading test achievement after collaborating with their peers.

Table 4.
Repeated measures ANOVA results for treatment group's GICT

	df	Sum of squares	Mean square	F	p-value
Within group	1	0.5952	0.5952	0.3852	0.5418

From the results presented in Table 4 above, it can be inferred that there was a statistically significant improvement in the treatment group's reading comprehension test after experiencing the collaborative stage due to the p-value being greater than the significance level ($\alpha = 0.05$). This finding is similar to the finding of Eastridge and Benson (2020) in which students who had GICT performed better.

Research Question 2: Does students' achievement in reading test using GICT significantly differ from traditional individual testing?

Using the independent t-test to compare the two groups, it was found that the average of the treatment group's population is considered to be not equal to the average of the control group's population. Thus, it can be concluded that the difference between the sample average of the treatment and control group is big enough to be statistically significant.

Table 5.
Results of independent t-test of the two groups' reading test in phase two

Group	N	Mean	Std. Deviation	Mean Difference	p-value	Remark
Treatment	21	18.238095	1.480026	2.952381	0.00002366	Significant
Control	21	15.285714	2.411283			

As shown in Table 5 above, the treatment group obtained a bigger mean score (18.24) than the control group (15.29). This demonstrates that students in the treatment group generally performed better on their reading test in phase two after having discussion with their peers in phase one. On the contrary, the lower mean score of the control group indicates that the students generally performed not as well as the treatment group did. However, as the mean difference between treatment and control group did not show significant difference, the results of for research question 2 only partially supported the hypothesis.

Research question 3: How do primary students perceive the implementation of GICT in reading test??

A total of 42 reflections were gathered from both groups. However, since the third research question particularly focuses on the perception of the treatment group who experienced the GICT, only 21 reflections were used for the analysis. This implies that only 50% of the reflections gathered were analyzed.

Given that the students were still primary students, students reflected based on the guiding questions. These questions pertained to several aspects such as: experience, strengths and weaknesses of doing the GICT, coping strategies during group processes, and suggestions for improvement.

In relation to the experience of the learners during GICT, some students described the modality as interesting and less stressful, as opposed to traditional individual testing. The implementation of group and individual tests in subsequent manner was also perceived to be fair.

> *"It was interesting to do it because test is usually done individually."*
>
> *"I feel less stressful. Because I could rely on my friends when I don't understand the reading text. I am bad at reading."*
>
> *"One of the good points is that we did the test twice, in groups and individually. That is fair."*
>
> *"Because the texts in the individual part were different from the ones in the group part, I think it was fair."*
>
> *"I prefer doing this type of test because I felt less worried when I did the test."*

According to the students, the strengths of GICT lied in the way students received help in recalling the materials with the help of their peers (reexposure) and correcting misconceptions. GICT also supported the practice of translanguaging (J. Jang, 2022) and developed a sense of positive interdependence.

> *"I was able to understand how to do some questions in the individual part because I learned the techniques from my friends in the team part. I think [by] changing the reading texts in the individual part, it is more fair for everyone because I could really learn."*
>
> *"I [came to] know that my understanding was not correct after learning from my friends' explanation."*
>
> *"I was assigned as the leader of the group and I learned to take my friends' opinions seriously. Then, evaluated again before we decided on an answer."*

> *"I like it because I didn't have to speak in English to my friends when we discussed."*

However, students also accounted for social loafing or free-riding (McKay & Sridharan, 2023) as the weakness of GICT, especially in the group phase.

> *"Sometimes my members did not help me at all."*
>
> *"One of my friends did not really work well with us. At one point, we had an argument which made the situation awkward."*
>
> *"I think I was the one who handled the questions the most. I don't like [the way] my friends stayed silent during the discussion and I had to do all of the test. But, [I think] the individual test was very useful because it made the test fair."*

Another interesting finding from the students' reflection was the account of the one-week interval between the group and individual test.

> *"I don't like [the fact that] I had to wait for one week to do the individual test. I forgot most of the discussion."*

Students also shared their coping skills towards the problems they experienced during GICT. These coping skills were all related to interpersonal dynamics within the group.

> *"I waited for my friends to respond. If they did not respond, then I went ahead."*
>
> *"I had to use Indonesian when my friends did not understand the reading."*
>
> *"When I had the conflict with my friend, I tried to talk to him carefully because I worried we would not be able to finish the test."*

The students also shared their suggestions to improve GICT in the future. The suggestions mostly addressed the scoring and fair distribution of group work.

> *"I think the weight of the individual test should be bigger as it is the one that shows our true ability."*
>
> *"I think the group test should be distributed to everyone equally because only a few students spoke up during the discussion."*

In short, the results of the students' reflection indicate that the students generally had positive attitude towards the use of GICT, with some commentaries on its power to lessen anxiety and fairness. On the negative note, a few students expressed their concern over social loafing. However, they proposed that it can be solved through fair scoring system.

DISCUSSION

The overarching objective of this study is to investigate if GICT could improve students' reading test. The results of this study demonstrate that GICT might serve as an applicable and beneficial modality in language testing, especially in reading test, as opposed to the traditional individual testing. In accordance with the results of this study, GICT not only helped students to improve their reading test as exemplified by the improved number of correct answers but also promoted social relationships. Therefore, this discussion addresses how GICT works as a primer, promotes the learning of reading strategies, facilitates peer scaffolding as evidenced in the interpersonal dynamics, and increases motivation. Lastly, the present study gives rise to several best practices.

Some studies suggest that collaborative learning be done in an individual-group design as students must be primed by doing the test first before allowing them to engage in a group discussion (LoGiudice et al., 2015). By doing so, students are enabled to retrieve previously attained knowledge unassisted. This shall then allow them to explore further the information during the collaborative stage together. It is worth noting, however, that such a condition applies when the test given in the individual stage bears sameness as the one in the group stage. Taking into account the 'primer' necessary for information

retrieval, the present research suggests that the aforementioned 'primer' could be facilitated in GICT as well. This argument is based on the premise that the test given in the group stage bears only similarities in terms of the targeted competency and reading techniques required to work on the test given in the individual part. In other words, this design demanded the students to retrieve the 'enhanced' information they had obtained from the group discussion to solve reading comprehension questions of similar level in terms of length of text, complexity and required competencies. In other words, the present study gives rise to the need to allow students to explore reading strategies together with their peers before engaging in individual activities.

From a cognitive aspect, students were able to learn from each other about the reading strategies. As one student remarked in their reflection, "*I learned from … about how to find the main idea of the text by finding the key words*." This is in line with previous studies (Amjadi & Talebi, 2021; Babapour et al., 2019; Khampool & Chumworatayee, 2023) which delineate on the magnitude of mastering reading strategies to improve reading comprehension. In their studies, reading strategies should be explicitly instructed to the students, but not necessarily in a teacher-centered or teacher-led manner. Explicit instruction on reading strategies could also benefit from collaborative design where the students exchange ideas about reading strategies. The importance of using collaboration to improve reading comprehension also finds support in the study conducted by Parlindungan et al. (2023) who found that Collaborative Reasoning (CR)—a student-led small discussion approach—has a significant impact on students' reading comprehension. Although this study does not primarily focus on CR, GICT as an approach builds on the concept of CR due to its focus on letting students have a meaningful discussion. It also allows students to seek help from their peers in order to support their comprehension. Seeking help to improve comprehension of a text not only emphasizes the value of collaboration but also self-monitor because seeking help is one of the

steps in Self-Monitoring Approach for Reading and Thinking (SMART) (Syafi'i, 2015). This implies that by allowing students to work on the reading test as a group, students are able to check their own understanding and ask for help when necessary before they do the individual test.

It needs to be noted, however, that the learning of reading strategies in this study may entail several aspects such as reexposure, retrieval of information, cross-cueing, and error pruning. Reexposure here refers to students getting reexposed by their peers to the material or reading strategy previously learned (Crompton et al., 2022). The reexposure the students receive from their peers are expected to assist them in retrieving the information from their memory (Levine et al., 2018). Another possible outcome of reexposure is that the students recall the use of a certain strategy or an item after another group member cues them (Wissman & Rawson, 2018). Lastly, error pruning—referring to "a group rejecting what they collectively consider to be wrong" (LoGiudice et al., 2015, p. 381)—is another one cognitive mechanism that helps students engaging in GICT, or collaborative learning in general, increase learning. These four commonly found cognitive mechanisms in collaborative learning may have been at play during GICT in this study given that students partook in information sharing during the collaborative test. Besides, the present study also demonstrates how students performed better in the individual test after doing the collaborative test—which further suggests that the four cognitive mechanisms were in force.

Body of literature devoted to collaborative testing generally gauges reexposure and cross-cueing in positive light (LoGiudice et al., 2015; Wissman & Rawson, 2018), albeit with some notable commentary on its role in learners' retention. GICT in this present study has shed some lights on how learners may have been reexposed to information by their peers during the group discussion. Such reexposure may have also expedited students' retention of information or reading strategies. However, reexposure may not suffice in ascertaining longer retention of the aforementioned facets.

Hence, as LoGiudice, Pachai and Kim (2015) states, information retrieval leads to better and larger retention benefits than reexposure does. Against this backdrop, GICT was carried out in this study based on the rationale that learners could receive reexposure or cross-cuing from their group members in the first stage as a way to solidify the information they have attained from the teacher during class activities. Then, in the hope that they can retrieve the information unassisted—which could possibly lead to longer retention—students had to do the individual test in the second stage. However, further research is necessary to see if GICT really helps students achieve long-term effects as the present study was designed with only a one-week interval between the group and individual tests.

One observable aspect of GICT is the evidence of interactional scaffolding as part of the group dynamics. Babapour, Ahangari and Ahour (2019) remark that the scaffolding given by a more capable or knowledgeable student is fruitful in helping students retain and develop linguistic ability. This peer interaction scaffolding is related to the input the students receive, because such input—alongside output and internal learner capacities—can be ascertained through effective interaction and negotiation of meaning (Amjadi & Talebi, 2021). The group structure which was determined through the pre-assessment allowed low-performing students to receive scaffolding from more capable students.

Given the scaffolding provided by their peers, the present study also suggests that there were interpersonal dynamics occurring within the collaborative stage. Although in most of the discussions within the group the cream of the crops took the lead, students in this study reported that they learned to be more aware of the other members' feelings when stating opinions. This is an especially valuable point for teachers in applying GICT. When GICT is to be implemented as an alternative reading testing modality, teachers need to explicitly elaborate the values of collaborative testing in order for the students to respect other students' opinions, evaluate their own previous learning by listening to other students' perspectives,

and be encouraged to approach discussions over the texts and their comprehension questions in diplomatic manner. This is in line with the findings of Levine, Borges, Roman, Carchedi, Townsend, Cluver, Frank, Morey, Haidet, and Thompson (2018) who found that the participants in their study negotiated differences, had diplomatic discussions, and recognized when to give in and defend.

Another important aspect of GICT is the increased motivation among students. The qualitative data of this study suggest that the students engaging in the GICT demonstrated their preference towards the testing modality due to their participation in the group discussion as they could use the information they had acquired from group discussion to do the individual test. Most of the students expressed how they had enjoyed the group testing because they could interact with their friends, and hence, the test felt less intimidating to them. This finding aligns with the finding from Eastridge and Benson (2020) which highlights how students rated GICT more highly and positively as they felt less anxiety. Other previous studies have also recognized that motivation—along with saving face (Robinson et al., 2015), personalities (French & Kottke, 2013) and compatibility with group members (Antoniou, 2019)—plays a pivotal role in students' engagement in collaborative testing (Mahoney & Harris-Reeves, 2019). Therefore, this research has suggested that GICT can be used to bolster students' engagement in *assessment for and as learning* (Lam, 2020).

CONCLUSION

To conclude, the present study aims to explore the effect of GICT as an alternative language testing modality on primary students' reading achievement and the perceptions of students on the implementation of GICT. By focusing on reading test and using explanatory sequential mixed methods, this study revealed that GICT had a significant effect on students' reading achievement test but only partially significant if compared to the control group's performance. Furthermore, the present study also revealed that students regarded

the modality in positive light due to its power in lessening anxiety and providing opportunities for collaboration and fair condition.

However, the present study also has several limitations. First, the present study investigated only a small group of students from one specific school in Indonesia. This may restrict the generalizability of the findings. Future studies should therefore investigate the use of GICT in English language testing with more diverse and bigger number of population and in other contexts to ascertain more solid findings. Second, the present study mainly compared between GICT and traditional individual test. This comparison may not tell much about the of individual-group design, as the said design is also considerably under-researched in English language testing. Therefore, future research may explore the comparison between the three testing modalities to generate more insights into the effectiveness of collaborative testing. Third, this research primarily studied the effect of GICT on reading test which may not testify much of the effectiveness of the modality in other language areas such as grammar, vocabulary, listening, writing, and speaking—which may demand different processes. Thus, future research may explore the use of GICT in different language areas.

The study also offers several insights into the best practice to implement GICT to improve the reading achievement test of primary students. First, diversify the group composition. Such diversification promotes the more knowledgeable students as the leader of discussion, which could help low-performing students to understand more about reading strategies and retrieve previously learned information. Second, give more weight to the score of the individual part as students will engage more in the retrieval of information during this stage. Third, instill the values of collaborative learning prior to the testing. In order for the group test in GICT to be effective, students need to be aware of others' feelings, take diplomatic stance in discussion, and be open towards different opinions. By doing so, students will develop a sense of positive interdependence.

Therefore, this study has testified against the prevalent use of traditional individual testing and contributed to scanty research body of GICT in English language testing. English language testing can benefit from the use of GICT as it promotes student-centeredness and positive interdependence. GICT also facilitates priming, scaffolding, and reexposure in the collaborative stage so that students can retrieve information in the individual part.

Overall, from the findings of this present study, combined with other findings related to collaborative testing, English language teachers, educators, and curriculum developers may consider implementing GICT if test is perceived not only *of learning* but also *as* and *for learning*.

ACKNOWLEDGMENTS

I would like to extend my gratitude to Grade 6 students of VITA Elementary School, Surabaya, for their participation in this study and to the school for granting the permission to conduct the study. I also would like to extend my appreciation to Lisa Nugraha, M.Ed., VITA School education consultant, who helped in constructing the test materials; Prof. Anita Lie, Ed.D., who shared her evaluation to improve the research; and Dr. Phil. Maria Teodora Ping, M.Sc., who guided me in improving the use of quantitative method in this study.

REFERENCES

Amjadi, A., & Talebi, S. H. (2024). Extending the efficacy of Collaborative Strategic Reading (CSR) in teaching reading strategies: A socio-cultural emotional perspective. *Language Teaching Research, 28*(2), 714-748. https://doi.org/10.1177/13621688211011501

Andrade, C. (2021). The inconvenient truth about convenience and purposive samples. *Indian Journal of Psychological Medicine*, *43*(1), 86–88. https://doi.org/10.1177/0253717620977000

Antoniou, A. (2019). Compatibility of small team personalities in computer-based tasks. *Challenges*, *10*(1), 1–12. https://doi.org/10.3390/challe10010029

Babapour, M., Ahangari, S., & Ahour, T. (2019). The effect of shadow reading and collaborative strategic reading on EFL learners' reading comprehension across two proficiency levels. *Innovation in Language Learning and Teaching*, *13*(4), 318–330. https://doi.org/10.1080/17501229.2018.1465059

Bećirović, S., Dubravac, V., & Brdarević-Čeljo, A. (2022). Cooperative learning as a pathway to strengthening motivation and improving achievement in an EFL classroom. *SAGE Open*, *12*(1). https://doi.org/10.1177/21582440221078016

Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, *38*(3), 429–455. https://doi.org/10.1177/0265532221992274

Caboral-Stevens, M., & Fox, D. P. (2020). The use of collaborative testing with baccalaureate nursing students. *Teaching and Learning in Nursing*, *15*(1), 37–41. https://doi.org/10.1016/j.teln.2019.09.007

Cantwell, E. R., Sousou, J., Jadotte, Y. T., Pierce, J., & Akioyamen, L. E. (2017). Collaborative testing for improving student learning outcomes and test-taking performance in higher education: a systematic review. *Campbell Systematic Reviews*, *13*(1), 1–18. https://doi.org/10.1002/cl2.186

Crompton, C. J., Wolters, M. K., & MacPherson, S. E. (2022). Learning with friends and strangers: Partner familiarity does not improve collaborative learning performance in younger and older adults. *Memory*, *30*(5), 636–649. https://doi.org/10.1080/09658211.2022.2041038

De Angelis, G. (2021). *Multilingual testing and assessment*.

Eastridge, J. A., & Benson, W. L. (2020). Comparing two models of collaborative testing for teaching statistics. *Teaching of Psychology*, *47*(1), 68–73. https://doi.org/10.1177/0098628319888113

Eastwood, J. L., Kleinberg, K. A., & Rodenbaugh, D. W. (2020). Collaborative testing in medical education: Student perceptions and long-term knowledge retention. *Medical Science Educator*, *30*(2), 737–747. https://doi.org/10.1007/s40670-020-00944-x

French, K. A., & Kottke, J. L. (2013). Teamwork satisfaction: Exploring the multilevel interaction of teamwork interest and group extraversion. *Active Learning in Higher Education*, *14*(3), 189–200. https://doi.org/10.1177/1469787413498034

Hard, B. M., & RaoShah, T. (2022). Developing collaborative thinkers: Rethinking how we define, teach, and assess class participation. *Teaching of Psychology*, *49*(2), 176–184. https://doi.org/10.1177/0098628320986953

Jang, H., Lasry, N., Miller, K., & Mazur, E. (2017). Collaborative exams: cheating? or learning? *American Journal of Physics*, *85*(3), 223–227. https://doi.org/10.1119/1.4974744

Jang, J. (2022). Translanguaging as an agentive action: A longitudinal case study of Uzbek EFL learners in South Korea. *Tesl-Ej*, *26*(3), 1–18. https://doi.org/10.55593/ej.26103a6

Jones, T. A. (2019). Effect of collaborative group testing on dental students' performance and perceived learning in an introductory comprehensive care course. *Journal of Dental Education*, *83*(1), 88–93. https://doi.org/10.21815/jde.019.011

Kapitanoff, S., & Pandey, C. (2018). Collaborative testing in statistics: Group interaction, anxiety, and class performance. *Statistics Education Research Journal*, *17*(2), 51–67. https://doi.org/10.52041/serj.v17i2.158

Khampool, W., & Chumworatayee, T. (2023). Collaborative Strategic Reading (CSR) instruction and its effects on Thai EAP university learners' reading comprehension across two proficiency levels. *LEARN Journal: Language Education and Acquisition Research Network*, *16*(2), 8–24.

Koutrouba, K., & Karageorgou, E. (2013). Cognitive and socio-affective outcomes of project-based learning: Perceptions of Greek Second Chance School students. *Improving Schools*, *16*(3), 244–260. https://doi.org/10.1177/1365480213501062

Lam, R. (2020). Investigating assessment as learning in second language writing: A qualitative research perspective. *International Journal of Qualitative Methods*, *19*, 1–10. https://doi.org/10.1177/1609406920938572

Levine, R. E., Borges, N. J., Roman, B. J. B., Carchedi, L. R., Townsend, M. H., Cluver, J. S., Frank, J., Morey, O., Haidet, P., & Thompson, B. M. (2018). High-stakes collaborative testing: Why not? *Teaching and Learning in Medicine*, *30*(2), 133–140. https://doi.org/10.1080/10401334.2017.1365719

LoGiudice, A. B., Pachai, A. A., & Kim, J. A. (2015). Testing together: When do students learn more through collaborative tests? *Scholarship of Teaching and Learning in Psychology*, *1*(4), 377–389.

https://doi.org/10.1037/stl0000041

Mahoney, J. W., & Harris-Reeves, B. (2019). The effects of collaborative testing on higher order thinking: Do the bright get brighter? *Active Learning in Higher Education*, *20*(1), 25–37. https://doi.org/10.1177/1469787417723243

McKay, J., & Sridharan, B. (2023). Student perceptions of collaborative group work (CGW) in higher education. *Studies in Higher Education*, 1–14. https://doi.org/10.1080/03075079.2023.2227677

Mende, S., Proske, A., & Narciss, S. (2021). Individual preparation for collaborative learning: Systematic review and synthesis. *Educational Psychologist*, *56*(1), 29–53. https://doi.org/10.1080/00461520.2020.1828086

Mohan, M. (2020). Use of critical reflection as a research method: A case of research-induced distress? *Indian Journal of Medical Ethics*, *5*(1), 19–20. https://doi.org/10.20529/IJME.2020.014

Namaziandost, E., Homayouni, M., & Rahmani, P. (2020). The impact of cooperative learning approach on the development of EFL learners' speaking fluency. *Cogent Arts and Humanities*, *7*(1). https://doi.org/10.1080/23311983.2020.1780811

Parlindungan, F., Mursyidin, M., Kurniasih, K., Rahmatillah, R., Nuthihar, R., & Oussou, S. (2023). The effect of collaborative reasoning on Indonesian university students' literacy and argumentation skills. *Journal of English Education and Linguistics Studies*, *10*(9), 97–118. https://doi.org/10.30762/jeels.v10i1.788

Robinson, L., Harris, A., & Burton, R. (2015). Saving face: Managing rapport in a Problem-Based Learning group. *Active Learning in Higher Education*, *16*(1), 11–24. https://doi.org/10.1177/1469787415573355

Salomone, M., & Kling, T. (2017). Required peer-cooperative learning improves retention of STEM majors. *International Journal of STEM Education*, *4*(1). https://doi.org/10.1186/s40594-017-0082-3

Segundo Marcos, R. I., López Ferández, V., Daza González, M. T., & Phillips-Silver, J. (2020). Promoting children's creative thinking through reading and writing in a cooperative learning classroom. *Thinking Skills and Creativity*, *36*, 1–13. https://doi.org/10.1016/j.tsc.2020.100663

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, *42*(3), 214–240.

https://doi.org/https://doi.org/10.1080/02702711.2021.1888348

Syafi'i, M. L. (2015). SMART strategy to boost students' reading comprehension. *Journal of English Education and Linguistics Studies*, *2*(1), 100–115. https://doi.org/https://doi.org/10.30762/jeels.v2i1.30

Sylvia, A. M., Shear, P. K., Jastrowski Mano, K. E., Guerin, J. M., & Mano, Q. R. (2023). Test anxiety and reading comprehension: The key role of fluid reasoning. *Anxiety, Stress, & Coping*, *36*(6), 781–794. https://doi.org/https://doi.org/10.1080/10615806.2023.2169677

Walker, C., & Baxter, J. (2019). Method sequence and dominance in mixed methods research: A case study of the social acceptance of wind energy literature. *International Journal of Qualitative Methods*, *18*, 1–14. https://doi.org/10.1177/1609406919834379

Wissman, K. T., & Rawson, K. A. (2018). Collaborative testing for key-term definitions under representative conditions: Efficiency costs and no learning benefits. *Memory and Cognition*, *46*(1), 148–157. https://doi.org/10.3758/s13421-017-0752-x

Zheng, Y., & Cheng, L. (2018). How does anxiety influence language performance? From the perspectives of foreign language classroom anxiety and cognitive test anxiety. *Language Testing in Asia*, *8*(1). https://doi.org/10.1186/s40468-018-0065-4

Zipp, J. F. (2007). Learning by exams: the impact of two-stage cooperative tests. *Teaching Sociology*, *35*(1), 62–76. https://doi.org/10.1177/0092055X0703500105