

FOREIGN LANGUAGE EDUCATION: JAPANESE LANGUAGE INSTRUCTION IN NON-NATIVE ENVIRONMENTS WITH A FOCUS ON GRAMMATICAL COMPETENCE AND LANGUAGE PRODUCTION SKILLS

Hana Bogdanova^{1*}, Lena Kralikova Hashimoto²

¹Faculty of Arts, Constantine the Philosopher University in Nitra, Štefánikova 67, 949 01 Nitra, Slovakia

²Faculty of Arts, Comenius University, Gondova ulica 2, 811 02 Bratislava, Slovakia

Abstract

This study examines the effects of two instructional approaches applied to two student groups. The first group followed instruction methods, which included explicit grammar instruction focused on introducing specific morphosyntactic structures supplemented by practice focused on situational accuracy and meaning and output practice. The second group followed modified instruction approach, which placed greater emphasis on implicit methods of instruction with greater prominence on output practice and lesser prominence on explicit grammar instructions.

The students were observed over the course of one academic year. To evaluate the impact of changes in instruction method, scores from tests aligned with the expected language proficiency level of the participants were compared and analyzed. The assessment focused on three key areas: formal accuracy, referring to the correct production and use of grammatical forms in controlled setting; contextual accuracy, or the ability to apply grammar and vocabulary based on context; and output, or the ability to produce language in real -world tasks, like writing and conversation.

Keywords: Japanese language instruction, Grammatical competence, Language production skills

1. INTRODUCTION

In structured group-based education, while the importance of curriculum is significant, the teaching style also plays a crucial role. The instructor is responsible for delivering the content in a clear and comprehensible manner. Furthermore, their responsibilities extend to ensuring that all students are actively engaged and encouraging active participation in both instruction and practice. Additionally, the instructor must ensure that enrolled students' progress at an appropriate pace and acquire the intended competencies, thereby preventing dissatisfaction, frustration, and ultimately the risk of students leaving or dropping the course or their studies. In this context, it becomes evident that the approach to teaching and instruction style must take multiple factors into account and represents a highly complex issue.

2. THEORETICAL FRAMEWORK

The knowledge of the target language's morphosyntactic system is undeniably important both for the study and the acquisition of a foreign language. For this reason, to this day, explicit grammar-centric instruction continues to play a key role in formal second language education. However, as Ellis ([1]) notes, conscious knowledge of grammatical structures alone does not guarantee a learner's ability to apply this knowledge in productive language use. And since the ease or the quality of language production is often the motivation for studying second language, the role of explicit grammar-centric instruction is increasingly being reconsidered through the lens of its communicative function ([2], [3], [4]). In pedagogical practice, this has led to a gradual shift from rigid, form drilling focused instruction towards a more integrated approach that combines form with meaning and communication. This shift reflects a broader approach to instruction itself, and to the need to understand its nature and effect on learners' performance ([5], [6]). This perspective is supported by recent research, including a study by Woyomo et al. ([7]), which supports the position that instruction combining explicit and implicit methods with communicative, task-based activities enhances both grammatical knowledge and spoken language

accuracy. This only further highlights current tendency of combining grammatical instruction with meaningful language use in communicative contexts.

This need becomes even more pressing in instructional contexts where direct exposure to the target language is significantly limited, as is the case with Japanese language study in most parts of Europe. Due to the increased accessibility of audio-visual media and social networks, the exposure to the target language has become easier, though, it is still very input-dominant. Despite the fact, that contemporary learners' target language input is no longer restricted to classroom environments and teacher-led instruction, but has dominantly moved towards easily accessible natural input, the overall exposure itself is often limited to the said input alone. In other words, today's environment encourages passive consumption of language rather than active, productive use.

Therefore, the need to explicitly address and cultivate learners' ability to produce language, particularly through structured opportunities for output, ought to become a crucial aspect of formal language education, which is usually input dominant as well ([8]). This highlights the pedagogical importance of not only focusing on input-rich instruction, but also ensuring that learners are systematically supported in developing their productive skills within instructional settings. Our study seeks to address this need by examining how different instructional approaches influence learners' ability to actively use grammatical structures in meaningful output.

3. METHODOLOGY

3.1. The Background of the Quasi-experiment Design

Curriculum design is an ongoing process, driven by the continuous need to align instructional content, teaching methods, and pedagogical strategies with current research insights and the evolving needs of learners. At our department, we have consistently engaged in this process with the aim of improving the overall quality of language education. Our focus has been not only on providing students with theoretical knowledge of the language and its morphosyntactic structure, but also on developing their ability to produce language appropriately and effectively in real-life communicative situations.

In recent years, we had reached a point of relative stability, where students demonstrated consistent linguistic competence across various skills. However, several challenges persisted. These included reduced engagement during practical language classes, slow reaction times, limited working memory in language tasks, and self-reported apprehension toward oral output. These concerns prompted us not to revise the curriculum itself, but to reconsider our instructional approach. In particular, we modified the teaching materials, where we supplemented our primary textbook, *Minna no Nihongo* ([9]), with *Marugoto* ([10]), a resource more oriented toward communicative competence and functional usage, based on You can do framework ([11]). The original instructional structure included guided grammar instruction using explicit methods focused on specific morphosyntactic structures and their forms and their paradigms (we further refer to this structural accuracy centrality as FoFS) once per week, grammar exercises based on FoFS once per week, and grammar-based conversation using FoFS once per week. This was supplemented by grammar practice combining situational accuracy and meaning (we further refer to this as FoF) and output practice (we further refer to this as OTPT) once per week, and conversation activities integrating FoF and OTPT once per week. The revised instructional structure retained guided grammar instruction using FoFS once per week, but grammar exercises were redesigned to combine FoFS and FoF approaches. Grammar-based conversation was adjusted to integrate FoF and OTPT, while grammar practice and conversation were increasingly oriented toward OTPT. Each of these activities was conducted once per week. As a result of these changes, the frequency of FoFS sessions was reduced from three times to two times per week. In contrast, FoF activities were increased from two to three times per week, and output-focused were also increased from two to three times per week.

When describing and explaining the nature of the instructions or the quality of the performance, we draw upon Swain ([12]), Doughty ([13]), Gass ([14]) and Shirahata ([6]).

3.2. Methodology

In order to examine the impact of the revised instructional design on the teaching process and the overall language performance of students, we conducted a quasi-experimental study. The study was carried out over the course of one academic year. The population consisted of first-year undergraduate students majoring in Japanese language and Culture Studies who began their formal instruction in Japanese language at the A1 proficiency level. The population was divided into two groups, Group A and Group B. Group A followed the original teaching structure, while Group B followed the revised instructional model. Group A consisted of 16 students and Group B of 18. The difference in group sizes is due to the fact that only students who were engaged throughout the entire period, attended instructions and practice consistently, and progressed through the course—i.e., those who neither discontinued their studies nor changed their major—were included in the analysis. The number of students with said parameters varied between the two groups. Students were not excluded from the population based on their academic performance. Therefore, even those with lower scores who remained in the course were included in the sample.

Given the nature of a quasi-experimental design, which does not allow for full control over the learning environment or external variables, we did not want to focus solely on comparing overall performance scores between the two groups, in other words, to determine which group was “better” or “more capable” based on their final scores. Instead, in order to minimize the influence of uncontrolled factors that could affect language development (such as motivation, effort and commitment to study, personal circumstances, personality traits, and similar variables), our analysis focused on three key aspects of group performance.

First, we examined the overall distribution of individual language skills within each group. Specifically, we focused on three key areas of performance: (1) form accuracy (FA), referring to students’ ability to correctly produce linguistic forms in a controlled environment, along with their declarative knowledge of these forms; (2) contextual accuracy (CA), which reflects the ability to select and apply the appropriate grammatical and lexical forms depending on the context and communicative setting; and (3) Output (OP), understood as the ability to produce language, such as writing a coherent text, participating in conversation, or responding appropriately to questions in context.

Formal accuracy was measured through tests that measured the knowledge and skills that students had acquired during the whole course, but with emphasis on the FoFS instruction parts of the course. The tests were by nature consistent with the instruction and focused on form accuracy and morphosyntactic knowledge. Contextual accuracy was measured through tests that measured the knowledge and skills that students had acquired during the whole course, but with the emphasis on the FoF instruction parts of the course. The tests were by nature consistent with the instruction and focused on contextual accuracy and grammar and pragmatic competence. Output was measured through tests that measured the knowledge and skills that students had acquired during the whole course, but with the emphasis on the OTPT instruction parts of the course. The tests were by nature consistent with the instruction and focused on language production.

We formulated the following research questions:

- Q1: What is the performance profile of each group in the targeted skill domains at the end of the instructional period?
- Q2: What differences exist between the two groups?

Based on our expectations, we formulated two hypotheses:

- H1: Group A will demonstrate higher performance in the FA domain than Group B.
- H2: Among the three skill areas, Group B will achieve its highest scores in Output.

Second, we analyzed the internal cohesion of progress within each group. Specifically, we were interested in how each group responded to the instructional approach as a collective. In our educational setting, it is essential that the group progresses together at a relatively consistent pace, without becoming

fragmented or overly polarized in terms of individual development. Therefore, we sought to determine the extent to which learning outcomes were evenly distributed within each group.

We posed the following research question:

- Q3: Is there a difference in the internal cohesion of progress between the two groups?

Third, we evaluated the strength of correlations between individual language skills. Existing research suggests a relationship between morphosyntactic competence and productive language ability. However, it is also well established that morphosyntactic competence alone is not sufficient to ensure strong output performance. Therefore, we were particularly interested in how the correlation patterns between specific skills might shift in an instructional setting that places greater emphasis on output practice.

We formulated the following research question:

- Q4: Will the correlation between FA and OP differ significantly between Groups A and B?

In this quasi-experimental study, we were able to control several key variables. The instructors teaching both groups were the same, and the instructional content, teaching methods, and materials were consistent across the groups where the nature of the experimental design allowed for it.

Another important aspect of the study that we could control was the data collection. All data were obtained through formal testing and standardized classroom assessments focused on language performance tasks. In order to ensure the reliability and validity of the results, all student scores were based on tests that were reviewed and administered by at least two different instructors. We took care in order to maintain consistency in test format, scope, and the representation of various task types across all assessments. We focused on gathering data in three targeted areas FA, CA, and OP. Additionally, we also measured overall language proficiency using tasks modeled after the Japanese Language Proficiency Test (JLPT), which serves as a standardized assessment of Japanese as a second language. The tasks were designed to align with the expected proficiency levels of the participants.

This data was analyzed using quantitative methods. Quantitative analysis was carried out through statistical evaluation of the test results, in order to identify patterns in performance and group differences. To enhance our interpretation of the findings, we supplemented the statistical analysis with qualitative data, specifically the insights from teacher journals. The study thus employed a mixed-methods approach, integrating both quantitative and qualitative data to provide a more comprehensive understanding of the instructional impact.

4. RESULTS

4.1. Descriptive Statistics

4.1.1. Descriptive Analysis – JLPT Based Results

Results are summarized in the following Table 1 and visualized in the following Figure 1.

Table 1

	count	mean	std. dev.	min	25% (Q1)	50% (Q2)	75% (Q3)	max.
Group A	16	79.69	14.11	47.00	76.75	83.00	89.50	95.00
Group B	18	67.72	21.83	14.00	64.00	70.00	78.00	97.00

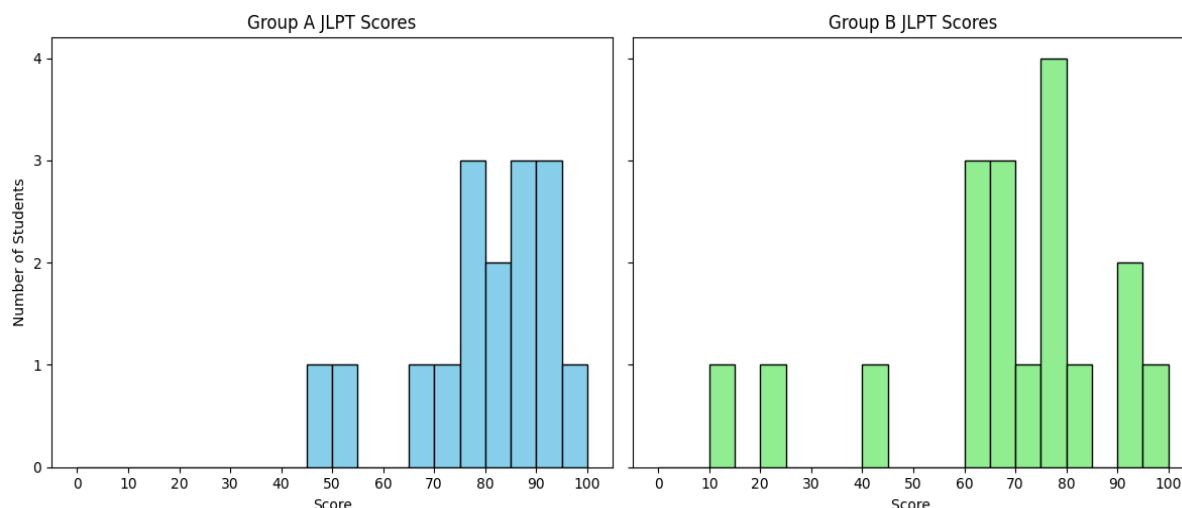


Figure 1

Group A

The mean score of 79.69 suggests that, on average, students achieved a relatively high level of general language performance. The median score of 83.00 is slightly higher than the mean score of 79.69. Furthermore, the upper quartile value (75th percentile) is 89.5, which may be interpreted that number of students were performing at a high level. The interquartile range (from the 25th percentile of 76.75 to the 75th percentile of 89.5) suggests that 50% of the students were within a high-performance scores, further supporting the impression of overall good group competence.

However, there is a moderate degree of variability in the group. The standard deviation is 14.11, minimum score is 47 and a maximum score is 95. The low minimum score suggests that some students struggled with the performance, although the overall trend remains positive.

Group B

The group B's mean score of 67.72 suggests that students achieved a moderate average general language performance across the group. The median score of 70.00 is close to the mean score, which suggests a relatively symmetrical distribution of scores in the central range of the group. However, the minimum score of 14.00 signals that some students had difficulties with the material and/or instruction. In contrast, the maximum score of 97.00 signals that some students in the group achieved very high level of language proficiency.

The interquartile range (64.00 for the 25th percentile and 78 for the 75th percentile) shows that the central 50% of students scored within a 14-point span. 14 points can be considered a wider score gap, though the interquartile range itself indicates that half of the students performed around the median of the group. Still, the standard deviation of 21.83 is relatively high, pointing to substantial variability within the group. This wider score gaps suggest inconsistencies in the rate and quality of language acquisition across learners in this group under the alternated instructions.

Group A vs Group B

Results are visualized in the following Figure 2.

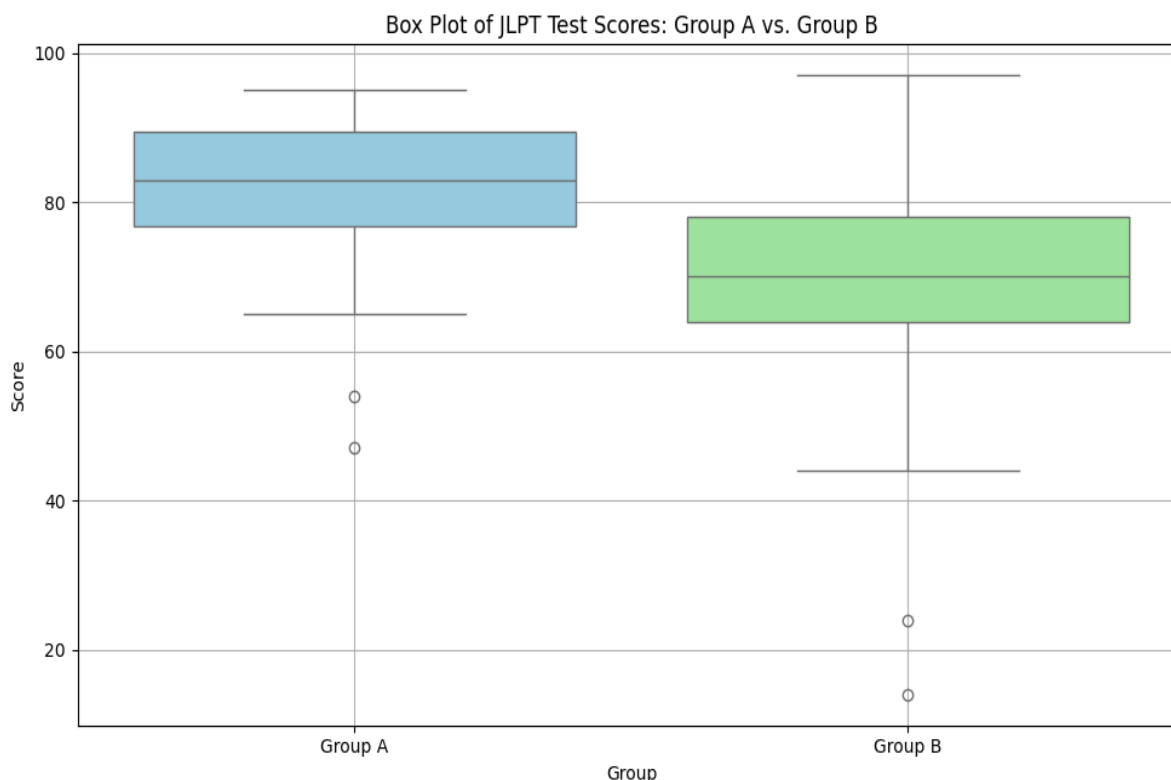


Figure 2

Group A showed a higher mean score of 79.69 compared to Group B's mean score of 67.72, indicating stronger general language proficiency in Group A. This is further supported by the median score, which is 83.00 in Group A, while 70.00 in Group B, suggesting lower central performance in this group.

In terms of variability, Group A's moderate standard deviation of 14.11 indicated relatively consistent students' outcomes within the group. Group B's standard deviation of 21.83 is much higher, which indicates wider performance differences within the group. This may suggest that the modified teaching style may have resulted in more varied learner progress.

The interquartile ranges also differ. Group A's scores were concentrated between 76.75 and 89.50, whereas Group B's scores spread between 64.00 and 78.00, indicating somewhat greater variability among the central performers in Group B.

Interestingly, Group B's maximum score of 97 exceeded Group A's maximum of 95, indicating that some students might have benefited significantly from the modified approach, despite the overall lower average.

4.1.2. Descriptive Analysis – FA Results

Results are summarized in the following Table 2 and visualized in the following Figure 3.

Table 2

	count	mean	std. dev.	min	25% (Q1)	50% (Q2)	75% (Q3)	max.
Group A	16	78.69	13.10	51.00	70.00	84.50	86.50	96.00
Group B	18	68.67	21.26	22.00	59.00	74.00	79.50	96.00

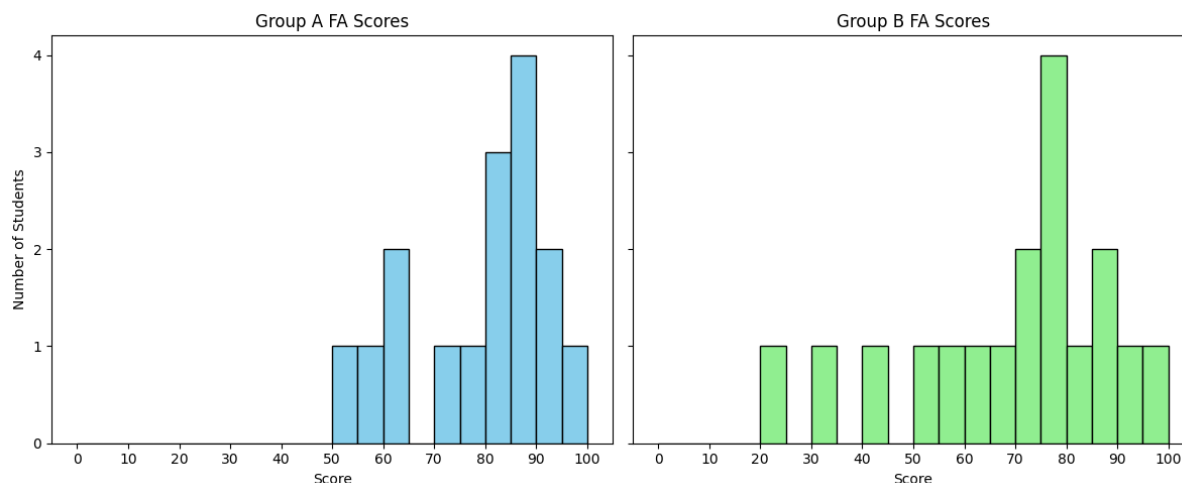


Figure 3

Group A

Results indicate a generally strong performance at the target grammatical structures by the group. The mean score of 78.69 suggests a solid competence within the expected proficiency level. The median score of 84.50, which is higher than the mean, suggests that more than half of the students performed above the average score. The interquartile range (70.00 at the 25th percentile and 86.50 at the 75th percentile) indicates that the middle 50% of students scored relatively high scores, indicating a consistent performance among most students.

The standard deviation of 13.10 indicates a moderate variability in scores. This variability is further indicated by the minimum score of 51, suggesting that a few students faced challenges with the material. The maximum score of 96, on the other hand, suggests that some students demonstrating high mastery of targeted grammatical forms.

Group B

The mean score of 68.67 suggests that Group B demonstrated a moderate level of proficiency in the targeted grammatical areas. The median score of 74.00, which is higher than the mean score, indicates that half of the students scored above the average. The interquartile range (from 59.00 at the 25th percentile to 79.50 at the 75th percentile) shows that the middle 50% of students scored in a moderate range. The standard deviation of 21.26 points indicates high variability within the group, pointing to considerable differences in individual student performances. This is further supported by the minimum score of 22, which indicates that some learners struggled significantly with the material. However, the maximum score of 96 indicates that some students in Group B achieved very high levels of proficiency.

Group A vs Group B

Results are visualized in the following Figure 4.

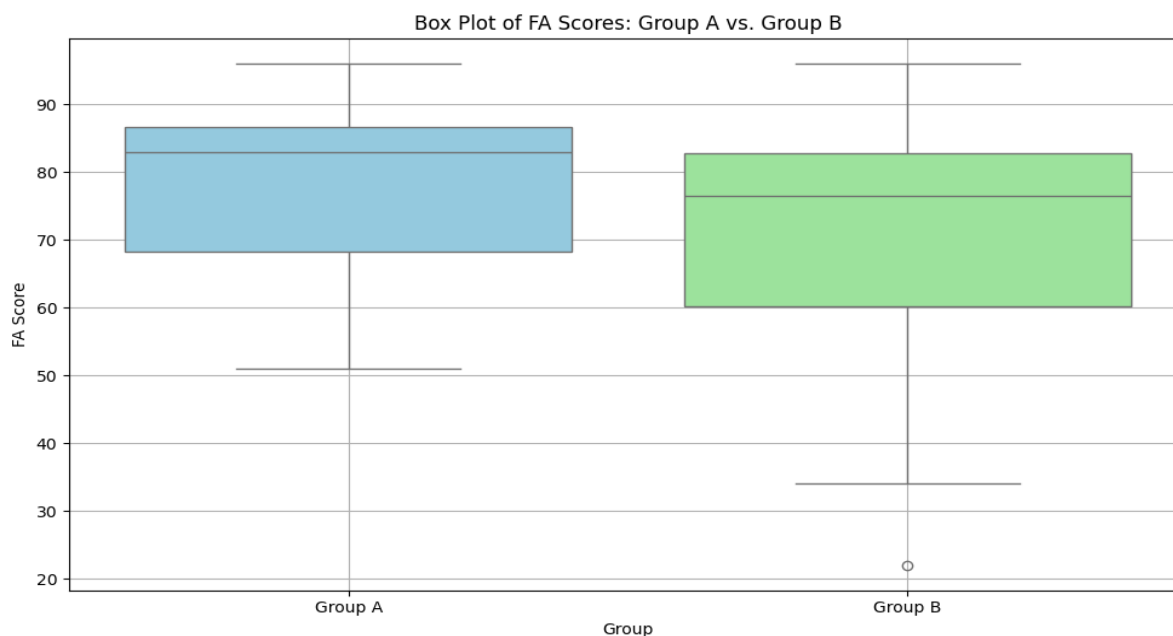


Figure 4

Group A's mean FA score of 78.69 outperforms Group B's mean FA score of 68.67 by a difference of approximately 10 points. This suggests that Group A, who received more traditional form centered and focused instruction, generally gained stronger mastery of grammatical forms by the end of the course. The median score in Group A was 84.50, compared to 74.00 in Group B. This further indicates a higher central tendency in the performance of Group A.

In terms of score distribution, Group A's standard deviation is 13.10, while Group B's standard deviation is 21.26. This difference suggests more consistent outcomes in Group A. The interquartile range in Group A (70.00–86.50) was both narrower and located at a higher performance level than that of Group B (59.00–79.50), which suggests that Group B's results show wider individual variety in students' outcomes. Both groups reached the same maximum score of 96. However, Group B's minimum score of 22, which is much lower than Group A's minimum of 51, indicates wider gap in student outcomes under the alternative teaching method.

Overall, both Group A's interquartile range and lower standard deviation suggest more consistent results and less variability in student outcomes compared to Group B. This may reflect the more structured, traditional teaching approach centered explicitly around the morphosyntactic forms. Group B's more varied results could be attributed to the modified instructional style, which may allow for greater individual learner autonomy but also results in less uniform outcomes. These results thus might suggest that the traditional instructional approach in Group A may foster steadier proficiency of morphosyntactic forms within the group. Meanwhile, the modified teaching style used with Group B might benefit certain learners, especially those who achieve high scores, but may also be challenging for other students.

4.1.3. Descriptive Analysis – CA Results

Results are summarized in the following Table 3 and visualized in the following Figure 5

Table 3

	count	mean	std. dev.	min	25% (Q1)	50% (Q2)	75% (Q3)	max.
Group A	16	89.56	10.53	55	86.75	93	96	99
Group B	18	64.00	20.53	40	46.25	61.5	77.75	100

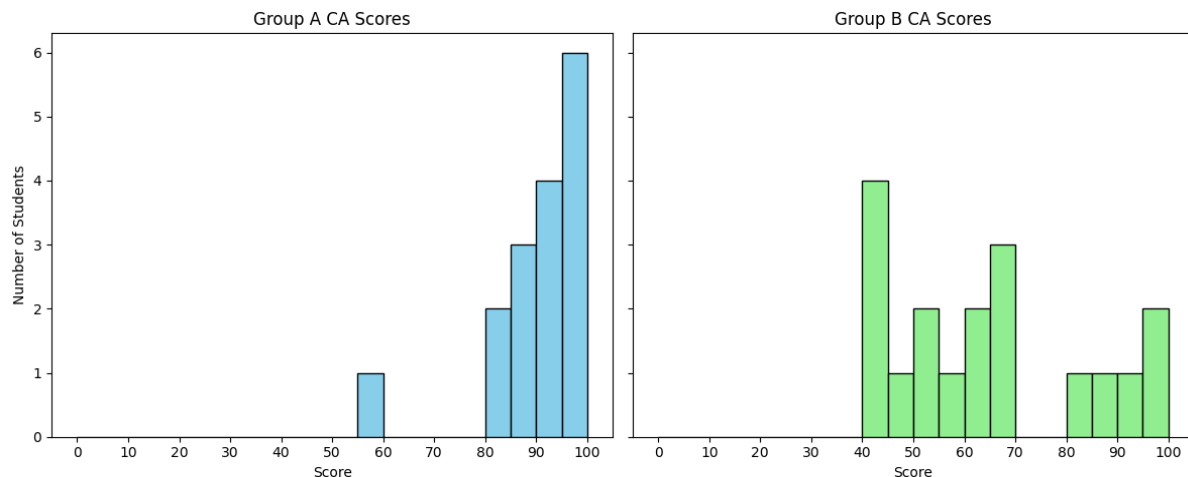


Figure 5

Group A

The results for Group A indicate a generally high level of proficiency in the targeted grammatical structures in terms of their distribution and proper usage. The mean score of 89.56 reflects strong overall competence. The median score of 93.00 is notably higher than the mean, which suggests that over half of the students performed above the average.

The results for the interquartile range (86.75 for the 25th percentile and 96 for the 75th percentile) indicate that the central 50% of students achieved scores within the high-performance range. This may be interpreted as consistent and solid understanding among most learners. The standard deviation is 10.53, which implies that while most students performed strongly, there were some differences in individual achievements. The minimum score of 55 indicates that one student encountered difficulties with the material while the maximum score of 99 indicates exceptional performance by some learners.

Group B

The results for Group B show varied level of proficiency in the targeted language skills. The mean score of 64.00 suggests a moderate overall performance. The median score of 61.50 is slightly below the mean score, which indicates that considerable number of students scored at or below the average.

The results for the interquartile range (46.25 for the 25th percentile and 77.75 for the 75th percentile) is relatively wide, which indicates noticeable variability in students' performance. The standard deviation of 20.53 further supports this. Furthermore, the minimum score of 40 suggest that several students struggled with the material, while the maximum score of 100 suggests, that two students covered the material in its entirety.

Group A vs Group B

Results are visualized in the following Figure 6.

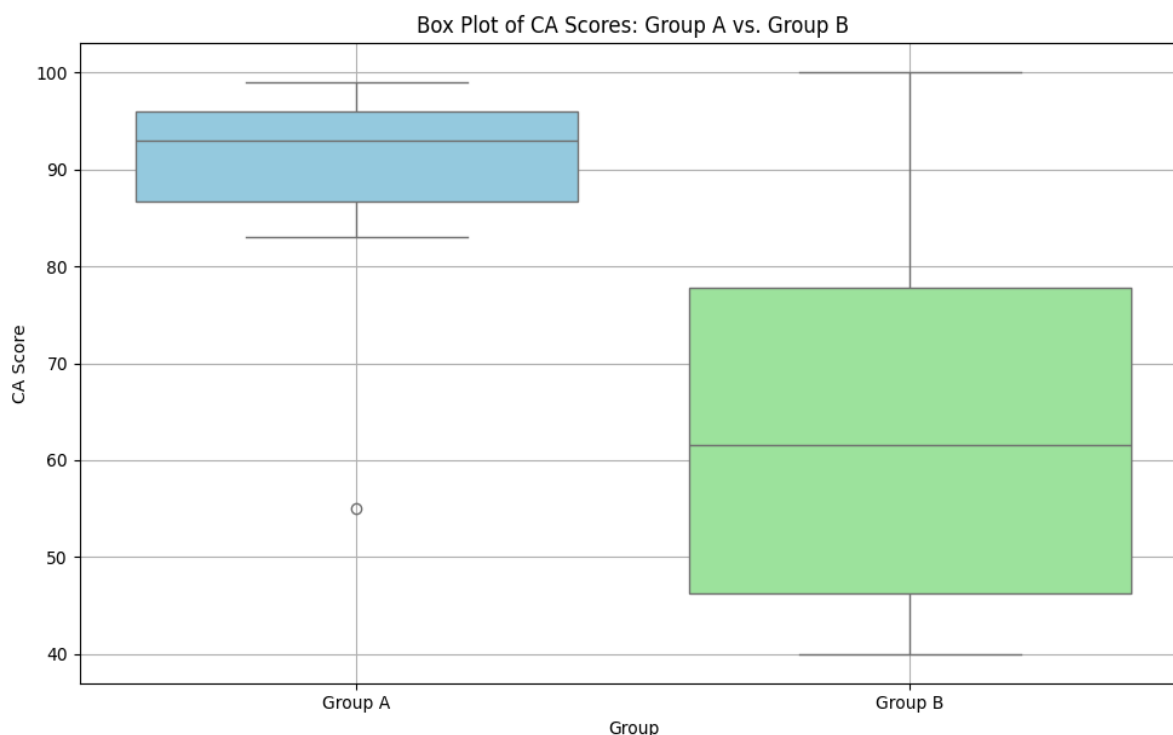


Figure 6

In terms of performance, Group A outperforms Group B. Group A's mean score of 89.56 is 25 points higher than Group B's mean score of 64.00. The median score of Group A is 93.00, while only 61.50 for Group B. This indicates not that Group A has attained higher level of proficiency, but also, that Group A's performance was more homogeneous.

In terms of variability, Group A's standard deviation of 10.53 suggests higher consistency in the distribution of the scores compared to Group B's standard deviation of 20.53. This is further supported by the interquartile ranges. Group A's interquartile range of 86.75-96.00 is much narrow than Group B's interquartile range of 46.25-77.75. This may be interpreted that while Group A likely progress steadily as a group, Group B's progress was more varied and nuanced. This is noticeable on the maximum score, which is very similar for the groups. Group A maximum score is 99 and Group B maxim score is 100. The minimum score though differs by 16 points between the groups at 40 in Group B and 55 in group A. This larger gap between the maximum and the minimum score in Group B points to greater variability and possible inconsistencies in the formal language competence of Group B students.

4.1.4. Descriptive Analysis – OP Results

Results are summarized in the following Table 4 and visualized in the following Figure 7.

Table 4

	count	mean	std. dev.	min	25% (Q1)	50% (Q2)	75% (Q3)	max.
Group A	16	86.88	11.21	58	86.5	88.5	92.75	100
Group B	18	75.44	13.15	34	73.25	78.0	85.00	88

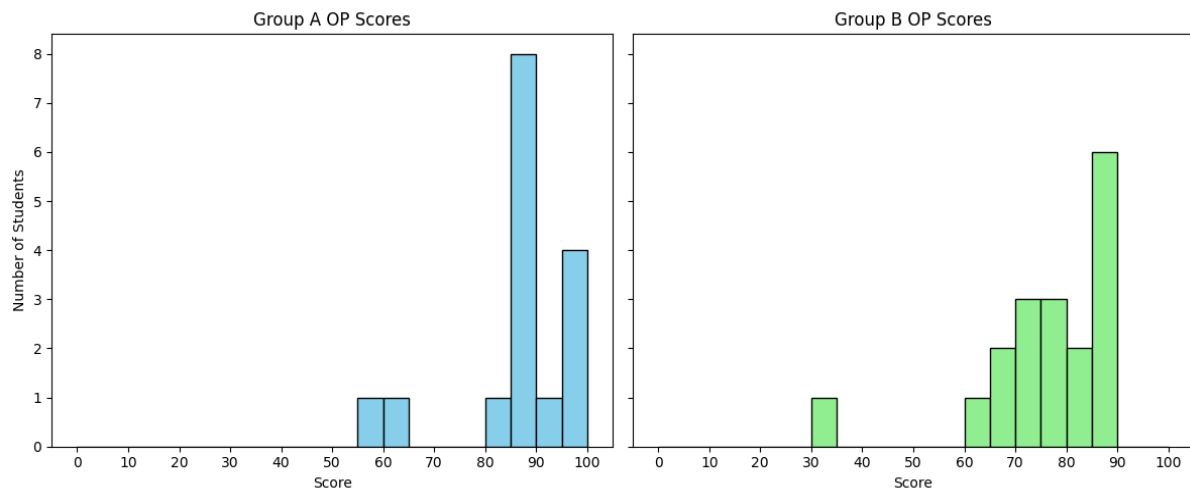


Figure 7

Group A

Results for Group A's OP scores show strong overall output performance among students of the group. The mean score of 86.88 indicates solid level of performance. Moreover, the median score of 88.50 is slightly higher than the mean score, which indicates that more than a half of the students in the group performed above the average of the group. The interquartile range (86.50 at the 25th percentile and 92.75 at the 75th percentile) shows that the middle 50% of the students scored relatively high and relatively comparable.

However, the standard deviation of 11.21 indicates a moderate variability, with some scores notably below the group mean. This is further supported by the minimum score of 58 compared to maximum score of 100.

Group B

The results for Group B suggest a reasonably strong output performance of the group at the end of the course, though there are some inconsistencies in score distribution. The mean score of 75.44 suggests a moderate to high level of output performance. This is further supported by the median score of 78.00, which is higher than the mean, thus indicating that more than half of the students of the group performed above the average. The first quartile (73.25 at the 25th percentile) and third quartile (85.00 at the 75th percentile) show that the middle 50% of the scores are relatively concentrated, suggesting a fairly consistent level of performance among most students. The interquartile range (IQR = 12) indicates moderate dispersion in the central scores.

The standard deviation of the Group for this skill is 13.15, which suggests moderate variability, which would mean that though many students performed close to the average, there were some exceptions. This is further supported by the minimum score of 34, which suggest that at least one student struggled with the material and/or instruction. On the other hand, the maximum score of 88 suggest mastery of the skill adequate to the current study level of the group by several students.

Group A vs Group B

Results are visualized in the following Figure 8.

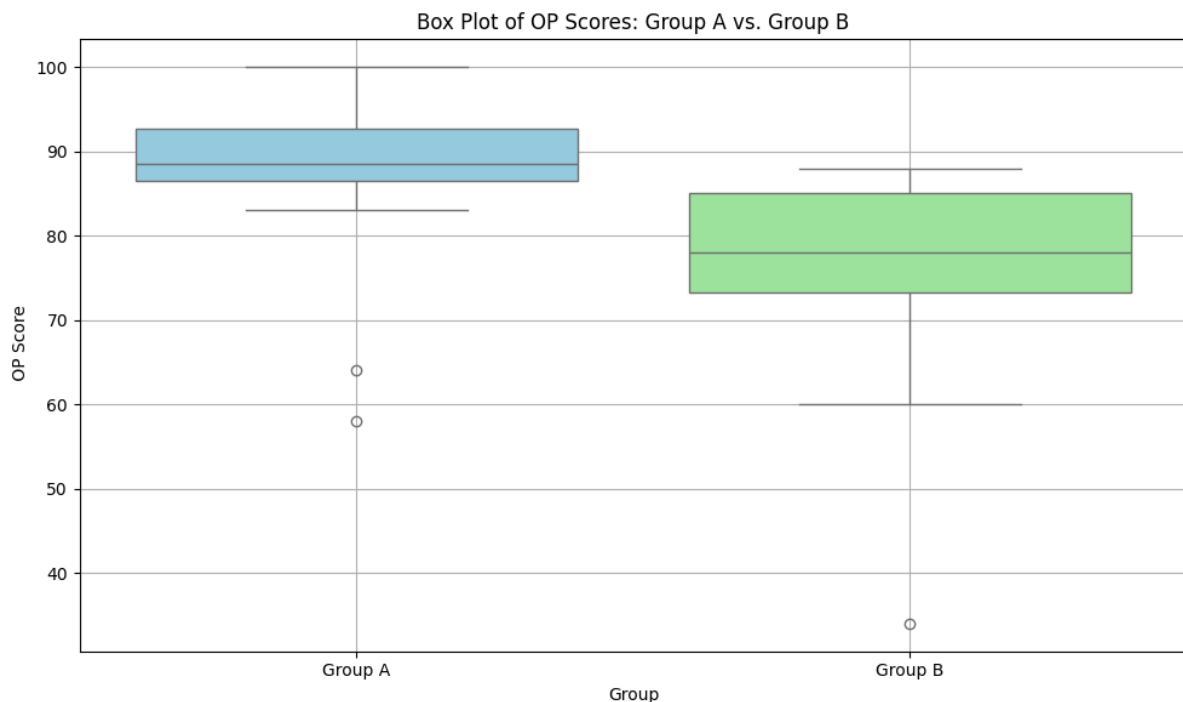


Figure 8

Group A's mean score of 86.88 is higher by 11 points than the mean score of 75.44 of Group B. This indicates strong output performance in Group A, which is further supported by the median score. Group A's median score is 88.50, while median score of Group B stands at 78.00. This suggests that typical student performance was higher in Group A.

In terms of variability, both groups display relatively moderate variability, though difference between the groups could be observed. Group A's standard deviation was 11.21 and group B's standard deviation was 13.15, which indicates that Group A scores were slightly more clustered around the mean. This is further supported by the interquartile range. Group A interquartile range is 86.50 - 92.75, while Group B's interquartile range was 73.25 - 85.00, which means it is more dispersed, which suggests more spread among the middle 50%. This is further supported by Group B's minimum score of 34 in relation to group's maximum score of 88. In contrast, Group A's minimum score was 58 and maximum score was 100, which makes Group B's range 12 points wider, indicating greater dispersion in Group B.

4.2. Correlation of specific skills within groups

Group A

Table 5

correlation pair	correlation (r)	p-value
GLP vs FA	0.8221	0.00009
GLP vs CA	0.7925	0.00025
GLP vs OP	0.8508	0.00003
FA vs CA	0.7211	0.00162
FA vs OP	0.7541	0.00074
CA vs OP	0.7287	0.00136

As is evident from the Table 5, the correlation analysis of Group A indicates strong, statistically significant relationships among the measured variables; moreover, all correlations are statistically significant ($p < 0.01$), with several reaching the $p < 0.001$ level.

General language performance (GLP) scores correlated strongly with all three observed domains. The strongest correlation is observed between general language performance and output performance, which also shows the highest statistical significance, closely followed by form accuracy performance. All correlations were positive and statistically significant, indicating that higher general language performance is associated with better performance across all measured domains.

As for the specific domains, we have observed strongest correlation between form accuracy and output, supporting the argument that grammar skills and output are closely aligned ([15]). Correlation between form accuracy and contextual accuracy was strong, though among the observed domain variables the weakest.

Group B

Table 6

correlation pair	correlation (r)	p-value
GLP vs FA	0.9622	$p < 0.001$
GLP vs CA	0.4062	0.0939
GLP vs OP	0.8994	$p < 0.001$
FA vs CA	0.4710	0.0476
FA vs OP	0.9517	$p < 0.001$
CA vs OP	0.5512	0.0199

As is evident from the Table 6, the correlation analysis of Group B indicates varied results in terms of strength and significance of relationship between observed domain variables.

In Group B, general language performance correlates very strongly ($r=0.9622$) and statistically significantly ($p < 0.001$) with their form accuracy, which might suggest the centrality of form accuracy performance for the group. Similarly, correlation between general language performance and output is both strong (0.8994) and statistically significant ($p < 0.001$). On the other hand, results show that correlation between general language performance and contextual accuracy is much weaker ($r=0.4062$) and moreover statistically not significant ($p=0.0939$). This weaker correlation may reflect the variability in the contextual accuracy performance of the group.

Group B shows particularly strong correlation ($r=0.9517$) between form accuracy and output, which is also statistically significant ($p < 0.001$). This aligns with the general language performance and output correlation results further supporting the prominence of this correlation for the group's language performance.

On the other hand, the contextual accuracy correlations are both moderate and weaker in statistical significance, with CA vs OP at $p=0.0199$, which is higher than the rest of the results, and CA vs FA at $p=0.0476$, almost by the margin of $p=0.05$. These results suggest that even though some relationship between the contextual accuracy performance and knowledge does exist, compared to other domains, it is less consistent, strong and clear. It can also indicate that contextual accuracy is less integrated into the other sub-skills performances in Group B, potentially aligning with the more variable outcomes observed in descriptive statistics.

Group A vs Group B

Both groups exhibit strong, statistically significant correlation between general language performance and output and general language performance and form. This is further supported by the strong and statistically significant relationship between FA and OP. These results suggest that for both groups, output performance was linked to overall performance and form skills. This aligns with the argument, that morphosyntactic skill is important for language output ([1]), and also with the claim, that language production practice and opportunity lead to enhanced grammatical accuracy ([12]).

This is further indicated by the fact, that Group B, which showed lower average morphosyntactic performance in the descriptive results, had stronger correlation between these two domains. Form accuracy and output correlations were stronger for Group B ($r=0.9517$) compared to Group A (0.7541), thus further supporting the input-focused awareness and output-based application. However, it might also reflect the narrower skills set of Group B compared to Group A.

The most notable contrast between the groups is the strength and statistical significance of the contextual accuracy correlations. While Group A exhibits strong and statistically significant correlations between CA and other sub-domains, Group B exhibits only moderate correlations with weaker statistical significance, which suggest that CA contributed less to the overall language performance of students in Group B compared to Group A. These weaker relations may reflect a less cohesive integration of structural grammar knowledge in Group B, which may be the result of the less explicit instruction methods.

5. DISCUSSION

Interpreting the results of pedagogical quasi-experiment is a complex task. It is essential to consider not only the hard results drawn from the hard data, but also the individual variables of students and circumstances of specific instruction settings.

Results from the collected quantitative data indicate that Group A generally outperformed Group B. The comparably higher Group A's performance on the JLPT and FA measures was expected, since this group received more grammar-focused explicit instruction. Additionally, this group was therefore also more used to exercises and tasks specifically designed to test grammar and form skills, which may have influenced the test results. This aspect ought to be considered when interpreting this particular set of data. Furthermore, data from teacher journal entries suggest that the lower performance of Group B, particularly in overall language proficiency and comparably lower morphosyntactic skills, could be attributed to the overall slower pace of instruction during the course, as lot of the instruction was not explicit, thus more time-demanding. In this context, the weaker results of Group B may not necessarily merely indicate only lower overall competence, but rather reflect their relatively lesser knowledge and progress, since they were not able to cover as much material on morphosyntactic forms during the course as Group A.

The results also align with the argument that, especially at the beginner level, explicit focus on morphosyntactic forms leads to stable and rapid results, thus preparing students for higher levels of grammar both cognitively and skill-wise ([16]). However, it is also important to consider these results within the context of different dominant instructions, where implicit instructions generally require more time for students to master a specific form ([17]). Therefore, it is not unreasonable to expect Group B to ultimately catch up with Group A in the future. This expectation is further supported by the fact that, while less cohesive than Group A, Group B still progressed together as a unit, with majority of the students scoring within the moderate performance scores. Additionally, correlations between FA and grammar were stronger in this group than in Group A.

The hypothesis that among the three domains, Group B will achieve its highest scores in output cannot be rejected, since the mean and the median scores of Group B were highest in this particular domain. However, Group B did not outperform Group A, nor in terms of scores, not in terms of number of students with higher performance scores. However, for context, it is worth pointing out, that FA directly correlates with output, and in Group B, this correlation was stronger than in Group A. Since Group B

had less knowledge than Group A, it is not reasonable to expect this group to outperform Group A. Nevertheless, the output results of Group B were consistent, and with the exception of one student, all students scored above the 50% score mark.

For better understanding, it is also important to supplement these data with the data from teacher journals. This data suggests greater willingness in Group B students to participate in output-centered lessons and stronger eagerness for output practice, even on topics, where students had limited or insufficient grammatical and lexical knowledge. In contrast, Group A students, despite the fact that they had better measurable knowledge and broader skill set, showed more prominent reluctance to engage in output-centered lessons, and overall, to spontaneously engage in language production and output. They demonstrated higher levels of apprehension when creating texts, and in general, had longer reaction times in interactions, along with poorer working memory for unfamiliar vocabulary or morphosyntactic structures. In the context of practical language use, these observations are not insignificant. They support the need to create a learning environment that actively addresses such issues and, more broadly, seeks to prevent them.

The most prominent findings were related to the results from contextual accuracy, which may suggest that while students in Group A have developed an internal mental language system that helps them to select appropriate forms and lexicon in context, or retrospectively identify and correct errors in their own language production ([6]), Group B has potentially not developed such a system as yet. This may be supported by Group A's performance scores in this domain, but especially by the correlation between CA and other examined domains. Group B's performance scores were the lowest in this domain and also most varied. Furthermore, correlation between CA and other examined domains were group's weakest.

The context for interpreting these findings is similar to the context of FA findings in that, that it can be presumed that due to more dominant implicit instructions during the course, the development of the language system might simply take longer time ([6], [17]). However, it is worth pointing out, that differences between the two groups were notably significant in this domain. In the context of tertiary education, it is particularly crucial for future graduates not only to use language spontaneously, but also to have a certain level of metalinguistic awareness, which will enable them to correct errors and mistakes, and choose situationally appropriate language forms and strategies. Therefore, based on these findings, it is likely that an intervention is necessary to bridge the gap in this specific domain.

6. CONCLUSION

This study has provided insights into the impact of different instructional approaches on language learners' performance and their ability to produce language and apply grammatical rules. Group A demonstrated stronger overall performance and more cohesive progress. Group B showed weaker and slower development, which may suggest that implicit instruction may take longer to bring measurable results, though it may be linked to other positives relevant to language production and usage.

The findings highlight the importance of balancing explicit and implicit teaching methods, particularly in fostering both grammatical knowledge and active language use. In the light of current knowledge, and as supported by the results of this quasi-experiment, it is essential that curriculum design and instruction methods take into account the needs of the learners. The predominance of either explicit or implicit teaching and instruction methods can contribute to varying rates of group progression and cohesion over specific time period. Therefore, these differences must be taken into account, as motivation and objectives of learners may vary significantly – ranging from hobby-based interest, to professional development, or short-term preparations for work in a different linguistic and cultural environment. The course design should take into account the primary objective of the students – i.e., ease of language production, high language competence, or other specific learning goals - and not only the course content but also the mode of instruction should be adapted accordingly.

REFERENCES

1. Ellis, R 2006, 'Current issues in the teaching of grammar: An SLA perspective', *TESOL Quarterly*, 40(1), 83–107.
2. Fridrikh, A, Bihunova, S, Bihunov, D 2022, 'Focus on teaching grammar in context', *Інноватика у вихованні*, 15, 236–244.
3. Mart, Ç. T 2013, 'Teaching grammar in context: Why and how?', *Theory and Practice in Language Studies*, 3(1), 124–129.
4. Azar, B 2008, 'Grammar Teaching and Communicative Teaching: A Hybrid That Works'. Retrieved from [http://209.166.84.64/assets/authorsCorner/Transcript-TESOL2008 HybridThatWorks.pdf](http://209.166.84.64/assets/authorsCorner/Transcript-TESOL2008%20HybridThatWorks.pdf)
5. Dibekulu, D 2021, 'Teaching grammar: In the view of conventional vs contemporary approach', *NEW ACADEMIA: An International Journal of English Language, Literature and Literary Theory Online*, X(II), April 2021. ISSN 2347-2073.
6. Shirahata, T, Wakabayashi, S, Muranoi, H 2011, *Shousetsu Dai ni gengo shuutoku kenkyuu: Riron kara kenkyuu made*, Kenkyusha, Tokyo.
7. Woymo, T. H., Bachore, M. M. & Jobo, M. M 2024, 'The effects of form-focused communicative grammar instruction on students' pronunciation and grammar in speaking', *Journal of Curriculum and Teaching*, 13(1), 61.
8. Muranoi, H 2007, 'Output practice in the L2 classroom', in R. DeKeyser (ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology*, Cambridge University Press, Cambridge, pp. 51–84.
9. Zanet n.d., *Minna No Nihongo*, 2nd edn, 3acorporation, Tokyo.
10. The Japan Foundation n.d., *Marugoto: Japanese Language and Culture*, Sanshusha, Tokyo.
11. The Japan Foundation n.d., *You-Can-Do*, Sanshusha, Tokyo.
12. Swain, M 2005, 'The output hypothesis: Theory and research', in E. Hinkel (ed.), *Handbook of Research in Second Language Teaching and Learning*, Lawrence Erlbaum, NJ, pp. 471–481.
13. Doughty, C 2003, 'Instructed SLA: Constraints, compensations, and enhancement', in C. Doughty & M. Long (eds.), *The Handbook of Second Language Acquisition*, Blackwell, Oxford, pp. 256–310
14. Gass, S 2003, 'Input and interaction', in C. Doughty & M. Long (eds.), *The Handbook of Second Language Acquisition*, Blackwell, Oxford, pp. 256–310.
15. Izumi, S 2002, 'Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization', in *Studies in Second Language Acquisition*, 24(4), 541–577.
16. Spada, N 2007, 'Communicative language teaching: Current trends and future directions', in N. Long & D. Richards (eds.), *The Handbook of Language Teaching*, Malden, Wiley-Blackwell, MA, pp. 271–287.
17. Muranoi, H 2001, 'Focus on form: Implicit and explicit form-focused instruction incorporated into a communicative task', in Robinson, P, sawyer, M, Ross, S (eds), *Second Language Acquisition in Japan*, 50, 617-673.