MEASURING INTEGRATED ENGLISH PROFICIENCY IN JAPANESE HIGH SCHOOLS: EVIDENCE FROM THE C-TEST

MAI ARISUDA
Saishigakusha Junior High School, Japan
16ab084.seinan@gmail.com

AKIHIRO ITO
Seinan Gakuin University, Japan
ito@seinan-gakuin.jp

ABSTRACT

This study examines the applicability of the C-Test as a reliable and valid assessment tool for measuring English proficiency among Japanese senior high school students. The C-Test, developed by Klein-Braley and Raatz, is internationally recognized for its brevity and high internal consistency. Despite these advantages, its use in Japanese secondary education remains limited. In this study, a five-passage C-Test was administered to 59 students at a vocational high school, and its internal consistency, construct validity (via factor analysis), and criterion-related validity (via correlations with GTEC, a widely used English proficiency test in Japan) were examined. A post-test survey captured students' subjective evaluations. The results showed high reliability (Cronbach's α = 0.93), unidimensionality (76.4% variance explained), and moderate-to-high correlations with General Tests of English Communication (GTEC) scores (r = 0.66 overall). Additionally, 94.9% of students reported that the C-Test was less stressful than the GTEC. These findings suggest that the C-Test is a low-stress, time-efficient assessment tool that aligns well with existing proficiency measures, especially in resource-limited contexts. While pedagogically promising, generalizability is limited by the localized sample. Future studies should clarify the construct domain of the C-Test within the Japanese curriculum, align scores with performance benchmarks, and explore broader applications across diverse school types. A limitation of this study is that participants were drawn from a single vocational high school, which constrains the generalizability of the findings.

KEYWORDS

C-Test, high school, English proficiency, reliability, validity

INTRODUCTION

In Japanese high school English education, there is a growing need for methods that can efficiently and objectively evaluate students' overall English proficiency in a comprehensive manner. One potential method that could meet such needs is the use of the C-Test.

The C-Test was developed with the aim of efficiently measuring learners' overall language proficiency within a short time frame, and its effectiveness has been demonstrated in numerous empirical studies on its reliability and validity (e.g., Eckes & Grotjahn, 2006; Raatz & Klein-Braley, 2002). This test consists of several short textual passages in which, with the exception of the first sentence, the second half of every second word is deleted (e.g., inform___, which should be restored to information). Test takers are required to complete the missing portions using contextual clues. In other words, the C-Test is a testing format designed to intentionally reduce redundancy within a text so as to measure the learner's integrated language proficiency (Grotjahn, 1996; Sigott, 2004).

Recent research has indicated that answering C-Test items requires not only grammatical and lexical knowledge but also higher-order language abilities such as contextual understanding and semantic inference. Consequently, the scores are considered valid indicators of learners' overall language proficiency. Indeed, C-Tests frequently report high reliability coefficients above 0.90, suggesting strong score stability (Eckes, 2011). Moreover, C-Test scores have been found to correlate moderately to highly with scores on other standardized English tests, providing empirical evidence for their validity.

In Japan, Mochizuki (1994) found a statistically significant positive correlation between C-Test scores and Grade Pre-2 results on the EIKEN Test in Practical English Proficiency, suggesting that the C-Test can function as a substitute for standard English tests. High correlations have also been observed between C-Test and TOEIC scores, and its validity is widely recognized internationally (Dörnyei & Katona, 1992). These findings support the view that the C-Test is effective in measuring English proficiency with both reliability and validity.

The C-Test is relatively easy to create, administer, and score, imposing minimal burden on teachers and researchers. Compared with administering separate skills-based tests, the C-Test offers greater practicality and cost-effectiveness. Therefore, in terms of ease of implementation, cost, and measurement validity, the C-Test can be considered to have high educational value.

In recent years, English education in Japanese high schools has adopted a policy of fostering the four skills—listening, speaking, reading, and writing—in an integrated manner, and developing corresponding measurement and evaluation methods has become a challenge. Persistent criticism

2

remains that even after six years of English study in secondary education, students' practical language skills are not sufficiently developed. As part of addressing this, reforms to the university entrance examination system have been implemented, with attempts to introduce external certification tests capable of measuring all four skills.

While the four-skills approach is pedagogically meaningful in that it seeks to evaluate comprehensive English proficiency by assessing each skill separately, external tests are expensive and require considerable time and effort to administer, making it difficult to conduct them continuously at the school level. Therefore, there is a need in the classroom for more affordable, easier-to-administer tests that nonetheless possess both reliability and validity. The C-Test should be considered as one such viable option.

Although research on the C-Test has accumulated to some extent, its application to Japanese high school English education has not been sufficiently examined. In particular, empirical studies verifying its validity by focusing on correlations with widely used standard English tests such as the GTEC remain scarce.

RESEARCH QUESTIONS

In response to these issues, the present study administered a C-Test created and implemented in an actual high school setting, with the aim of examining its effectiveness through the following four research questions:

- To examine the internal consistency of the C-Test and verify its reliability.
- To determine the degree of unidimensionality in C-Test scores and clarify its characteristics as a test.
- To calculate correlation coefficients between C-Test scores and those of a standard English test, thereby verifying its validity.
- To conduct a survey of high school students who took the C-Test and explore its role and potential in English education.

METHODOLOGY

Purpose

The purpose of this study is to examine the potential of the C-Test as an assessment tool in the context of Japanese high school English education, by analyzing both the score data obtained from its administration and the results of a post-test student survey. The investigation focuses on two key psychometric qualities—reliability and validity—while also considering the practical applicability of the C-Test in real educational settings.

Participants

The study was conducted with a total of 59 first- and second-year students enrolled in a vocational high school located in the Kyushu region of Japan. Prior to the administration of the test, participants were informed that the results would have no bearing on their academic grades, and written informed consent was obtained from all students. All participants had received at least three years of formal English instruction in junior high school and more than one year in high school, ensuring that they possessed a basic foundation in English learning.

Materials

The C-Test used in this study was developed in accordance with the construction criteria proposed by Klein-Braley and Raatz (1985), and consisted of five English texts.

The number of items per text was as follows:
- Text A: 20 items
- Text B: 22 items
- Text C: 28 items
- Text D: 30 items
- Text E: 30 items

The total number of items across all texts was 130. The source texts were drawn from commercially published textbooks not used in the participating school's classes, in order to avoid any familiarity effects. For test construction, the texts were arranged in order of ascending difficulty, starting with the easiest in terms of vocabulary and syntax.

In line with the C-Test format, the first sentence of each text was presented intact, while in the remainder of the text, the second half of every second word was deleted. Although the original guidelines recommend adjusting the number of deleted letters according to word length (i.e., deleting exactly half of the letters for even-length words and rounding up for odd-length words), this study employed a simplified approach: for five-letter words, the final two letters were deleted, and similar fixed deletions were applied to other lengths. This modification was intended to improve readability and reduce cognitive load for test takers. All deletions were visually highlighted to make the missing segments easily identifiable.

Procedure

The C-Test booklet (A3 paper folded into A4 size) was distributed to participants. The front cover contained a brief explanation of the C-Test format and one practice item with a five-minute time limit. Given that most participants were encountering the C-Test for the first time, they were guided through the practice item by the instructor to ensure familiarity with the format.

4

The main test was then administered under a 30-minute time limit. Upon completion, an answer sheet containing the correct responses was distributed, and participants self-scored their responses using the exact-word method, in which only answers perfectly matching the original word were counted as correct.

Ethical Considerations

Written informed consent was obtained from all participants prior to data collection. As this was a non-invasive educational study involving no sensitive personal information, formal approval from the institutional ethics committee was not required.

RESULTS AND DISCUSSION

Descriptive Statistics

Table 1 presents the descriptive statistics for the five subtests (Texts A–E) that comprise the C-Test, as well as the total score.

Table 1. Basic Statistics for the C-Test ($N = 59$)

Subtest	Max Score	Mean	SD
A	20	15.58	2.59
B	22	14.97	3.17
C	28	20.36	3.29
D	30	18.14	4.38
E	30	18.59	4.82
Total	130	87.63	16.03

Intercorrelations Among Subtests

Table 2 shows the Pearson product–moment correlation coefficients among the five subtests.

Table 2. Pearson Correlations Among C-Test Subtests ($N = 59$)

Subtest	A	B	C	D	E
A	1.00				
B	0.65	1.00			
C	0.66	0.72	1.00		
D	0.71	0.79	0.68	1.00	
E	0.70	0.72	0.70	0.72	1.00

All correlations were statistically significant at the P < .01 level and exceeded the benchmark value of 0.30, indicating moderate to strong relationships among the subtests.

Factor Analysis

To examine the underlying structure of the test, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was calculated and found to exceed the 0.60 criterion (Kaiser, 1974), indicating suitability for factor analysis. An exploratory factor analysis using the principal factor method with varimax rotation was conducted on the subtest scores. The first factor yielded an eigenvalue of 3.82, accounting for 76.43% of the variance. All subsequent factors had eigenvalues below 1.0 (Factor 2 = 0.37, Factor 3 = 0.34, Factor 4 = 0.28, Factor 5 = 0.19) and were not considered meaningful. This supports a unidimensional model, suggesting that the C-Test primarily measures a single latent construct. The confirmation of unidimensionality strengthens the theoretical basis for interpreting the test scores as a coherent measure of general English proficiency. This unidimensional structure also provides an empirical foundation for discussing the construct measured by the C-Test, an issue that is revisited in the following section on assessment perspectives

Reliability

Cronbach's alpha was computed to assess the internal consistency of the overall C-Test. The coefficient was $\alpha = 0.93$, indicating a high degree of reliability. Such strong reliability suggests that the C-Test can be repeatedly administered in educational contexts with consistent results, reinforcing its practicality as a classroom assessment tool.

Relationship with GTEC Scores

Table 3 shows descriptive statistics for the GTEC scores alongside C-Test scores, while Table 4 presents their intercorrelations.

Table 3. Descriptive Statistics for GTEC and C-Test Scores ($N = 59$)

Test	Max Score	Mean	*SD*
GTEC Reading	250	115.95	32.86
GTEC Listening	250	118.53	31.60
GTEC Writing	160	176.59	31.12
GTEC Total	660	411.07	76.94
C-Test Total	150	87.63	16.03

6

IJLLALW

Table 4. Pearson Correlations Between GTEC and C-Test Scores (*N* = 59)

	GTEC Reading	GTEC Listening	GTEC Writing	GTEC Total	C-Test Total
GTEC Reading	1.00				
GTEC Listening	0.59	1.00			
GTEC Writing	0.42	0.40	1.00		
GTEC Total	0.84	0.82	0.75	1.00	
C-Test Total	0.61	0.56	0.42	0.66	1.00

The results revealed moderate-to-high positive correlations between the C-Test and GTEC scores. The correlation between the C-Test total score and the GTEC total score was r = 0.66, suggesting that the C-Test scores are closely associated with performance on a widely recognized standardized English proficiency test. This provides statistical evidence of the C-Test's criterion-related validity.

Student Perceptions of the C-Test
In the post-test survey, 56 students (94.9%) reported that the C-Test was less stressful than the GTEC, while only 2 students (3.4%) preferred the GTEC in terms of reduced stress; 1 student (1.7%) did not respond. Qualitative comments indicated that students appreciated the shorter testing time and the absence of lengthy reading passages, both of which made it easier to maintain concentration.

These findings suggest that the C-Test imposes relatively low psychological demands on test takers, making it a practical option in classroom contexts where reducing test anxiety is a priority. Taken together, these findings demonstrate that the C-Test is internally consistent, unidimensional, and moderately correlated with an established standardized test, while also being perceived positively by students. Building on these empirical results, the following section discusses broader interpretive and practical implications of the C-Test in Japanese high school education.

Taken together, the results indicate that the C-Test possesses strong internal consistency, a unidimensional factor structure, and moderate-to-high correlations with an established standardized test. Moreover, it was perceived by students as a low-stress format. Overall, the empirical evidence points to both strengths and challenges of the C-Test. These can be broadly categorized into theoretical issues, concerning construct clarity and benchmarking, and practical

7

issues, concerning classroom use and future applications. These attributes position the C-Test as a viable alternative or supplementary tool for assessing English proficiency in Japanese high schools, particularly in resource-limited settings. Building on these empirical results, the following section discusses broader interpretive and practical implications of the C-Test in Japanese high school education.

The discussion is organized into two parts: theoretical challenges related to construct clarity and proficiency benchmarks, and practical implications and directions for future research.

Ambiguity in the Assessment Perspective

One of the major concerns is that the scope of language abilities measured by the C-Test has not been clearly defined. Within the framework of the new national curriculum guidelines, English is evaluated across three perspectives: **(1) knowledge and skills**, **(2) thinking, judgment, and expression**, and **(3) proactive attitude toward learning**. Determining which of these the C-Test most appropriately addresses is not straightforward. While it is reasonable to assume that the C-Test primarily corresponds to the "knowledge and skills" domain, empirical research explicitly supporting this classification remains insufficient.

Clarifying the construct measured by the C-Test—that is, identifying precisely which aspects of language proficiency are being assessed—is a critical step. Without a clearly articulated theoretical definition, there is a risk of misinterpretation or misapplication in educational contexts. Establishing this construct clarity will enable educators to better understand what the C-Test measures and how its results can be meaningfully integrated into broader assessment frameworks.

Difficulty in Linking Scores to Proficiency Levels

Another obstacle to the widespread adoption of the C-Test is the absence of clear benchmarks for linking C-Test scores to proficiency bands or grades (e.g., A/B/C ratings). For example, in the Ministry of Education, Culture, Sports, Science and Technology (MEXT) "Action Plan for the Improvement of Students' English Proficiency," the target for high school graduates is set at the level equivalent to EIKEN Grade Pre-2 to Grade 2 (CEFR A2 to B1). However, no established equivalency currently exists between C-Test scores and such external standards.

Moreover, the C-Test primarily measures receptive skills—particularly reading, vocabulary, grammar knowledge, and contextual understanding—and does not directly assess productive skills such as listening and writing. This limitation constrains its utility as a comprehensive four-skills assessment tool. Consequently, further empirical research is necessary to verify its construct validity and its correlations with other measures, especially those assessing productive abilities.

8

Practical Advantages Despite Limitations

Despite these challenges, the C-Test offers notable practical advantages. It is relatively easy to create and score, allowing for efficient measurement of English proficiency even under conditions of limited time and resources. In particular, for vocational high schools or other institutions where opportunities to take external proficiency tests are limited, the C-Test could serve as a means to track students' progress and provide continuous feedback on learning outcomes.

From a practical standpoint, the C-Test's objective scoring method—based on exact matches between responses and target words—minimizes discrepancies among raters. This objectivity stands in contrast to many in-house school tests that inevitably involve subjective judgment. As such, the C-Test can complement existing assessment tools while contributing to more consistent grading practices. Additionally, the efficiency of scoring can reduce teacher workload, a benefit that is especially significant in the context of the heavy demands placed on high school teachers in Japan.

Directions for Further Validation

To further support the use of the C-Test, it will be important to examine its correlations with various external proficiency measures, such as GTEC, EIKEN, TOEFL Junior, and TEAP. This will help determine the degree to which the C-Test aligns with a wide range of assessment indicators. Establishing such connections will provide a stronger theoretical and empirical foundation for the test's reliability and versatility, thereby broadening its potential applications in educational practice.

Directions for Future Research

This study examined whether the C-Test can serve as a reliable and valid method of measuring and evaluating English proficiency within Japanese high school education. By developing, administering, and analyzing a C-Test in an actual educational setting, the research demonstrated its potential utility.

From the perspective of internal consistency, the C-Test showed high reliability (Cronbach's α = .93). In terms of validity, the test displayed moderate-to-high correlations with scores from a widely recognized standardized proficiency test (GTEC), and approximately 95% of participants indicated in post-test surveys that they found the C-Test less stressful than the GTEC. Taken together, these results suggest that the C-Test is a feasible and effective measurement tool in Japanese high school English classrooms.

One notable advantage of the C-Test is its ability to assess multiple aspects of language proficiency—vocabulary knowledge, grammatical competence, and contextual understanding—within a single format. This efficiency enables flexible use across a range of proficiency levels and learning contexts. In school environments where opportunities to take external tests are limited, the C-Test can function as both a diagnostic instrument and a motivational tool, helping to make students' language development more visible.

Nevertheless, the present study is limited by its focus on a single school type (vocational high school) in a specific region. This raises questions about the generalizability of the findings. Future research should involve cross-sectional studies in diverse regions and school contexts, as well as investigations into how the C-Test aligns with other skill areas, particularly listening and writing, to evaluate its applicability as a complementary component of a four-skills assessment framework.

Moreover, it is essential to explore institutional models for the systematic integration of the C-Test into high school curricula. This includes examining its compatibility with external proficiency assessments, its alignment with the national curriculum guidelines, and the standardization of test construction and scoring procedures. Teacher training to ensure effective administration and interpretation of results will also be a crucial element in sustainable implementation.

Clarifying the construct targeted by the C-Test and strengthening its construct validity should be prioritized in future work. Specifically, research should investigate which of the three curriculum evaluation perspectives—knowledge and skills, thinking, judgment, and expression, or proactive attitude toward learning—the C-Test most directly addresses, through both theoretical analysis and empirical evidence. Furthermore, studies should explore the possibility of linking C-Test scores to performance grades (e.g., A/B/C) and external standards such as CEFR levels or EIKEN grades.

Given that the C-Test primarily measures receptive skills, further investigation into its relationship with performance-based assessments of productive skills will help clarify its complementary role within a balanced evaluation system. To ensure sustainable use in schools, practical issues such as standardized item development, efficient scoring systems, and the provision of professional development opportunities for teachers must be addressed.

Through continued theoretical research and practical validation, the C-Test has the potential to be firmly established as a highly effective assessment tool in Japanese high school English education.

CONCLUSION

This study examined the potential of the C-Test as a reliable and valid method of measuring English proficiency within Japanese high school education. The findings demonstrate that the C-Test is both feasible and effective, and its practicality and positive reception among students make it a promising option for classroom implementation. Despite these promising findings, the study is limited by its focus on a single school type (vocational high school) in a specific region, which constrains generalizability. Further research should involve diverse school types and investigate correlations with additional proficiency measures to broaden applicability.

Limitations of the study

This study has several limitations that should be acknowledged. First, it was conducted at a single vocational high school in the Kyushu region, which restricts the representativeness of the sample and limits the generalizability of the findings. Second, the focus of the assessment was primarily on receptive skills such as vocabulary, grammar, and contextual understanding, while productive skills like speaking and writing were not directly measured. Finally, because no established benchmarks currently link C-Test scores to external standards such as CEFR levels or EIKEN grades, the interpretability of the results in broader educational contexts remains limited.

Acknowledgements

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

Dörnyei, Z., & Katona, L. (1992). Validation of the C-Test amongst Hungarian EFL learners. *Language Testing, 9*(2), 187–206. https://doi.org/10.1177/026553229200900206

Eckes, T. (2011). *Introduction to many-facet Rasch measurement: Analyzing and evaluating raters and rating scales.* Peter Lang.

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-Tests. *Language Testing, 23*(3), 290–325. https://doi.org/10.1191/0265532206lt330oa

Grotjahn, R. (1996). Theoretical foundations of the C-Test. In R. Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen* (pp. 9–30). Brockmeyer.

Ito, A. (1996). Testing English tests: A language proficiency perspective. *JALT Journal, 18*(2),

11

183–197.

Kiya, M., & Ito, A. (2017). Validation of the C-Test among Japanese learners of English as a foreign language. *The European Journal of Applied Linguistics and TEFL, 6*(1), 91–110.

Klein-Braley, C., & Raatz, U. (1985). The C-Test: A valid operationalization of the reduced redundancy principle in language testing. *Language Testing, 2*(1), 41–54. https://doi.org/10.1177/026553228500200104

Mochizuki, M. (1994). Assessing English language proficiency using the C-Test: An analysis of Japanese university students' performance. *JALT Journal, 16*(1), 19–37.

Raatz, U., & Klein-Braley, C. (2002). Introduction to the C-Test. In J. Coleman, R. Grotjahn, & U. Raatz (Eds.), *Practice and problems in language testing 8* (pp. 79–108). Narr.

Sigott, G. (2004). *Towards identifying the C-Test construct.* Peter Lang.