Classification of multiple text reading test formats by analyzing English language proficiency tests

Yosuke Mikami[1*]

*Correspondence:
mikami.yosuke@surugadai.ac.jp

[1] Global Education Center, Surugadai University, 698 Asu, Hannno-City, Saitama 357-8555, Japan

Abstract

This study proposes a classification method for multiple text reading test formats in English language proficiency tests. A preliminary study involving 11 proficiency tests revealed two tests that fit the scope of the main study. Results show that multiple text reading test formats use complementary texts rather than conflicting texts. As for questions in a set of test forms in multiple text reading test formats, cognitive processing on integrating the contents of texts was different in proficiency tests. Moreover, the type of connection formation required by the questions differed among the proficiency tests. Implications for pedagogy are presented.

Keywords: Multiple texts, English language proficiency tests, Integration, Cognitive framework

Introduction

While the reading formats of traditional English language proficiency tests (EPTs) have focused on measuring the reading comprehension of single texts, the trend in recent years has been toward multiple text reading test formats (MTRFs), which are a set of multiple text reading forms to assess learners' understanding of multiple texts. One rationale for incorporating MTRFs into proficiency tests is authenticity. With the proliferation of the Internet, it has become the norm, not the exception, for learners to encounter multiple texts in real-life situations (Alexander & The Disciplined Reading and Learning Research Laboratory, 2012). Furthermore, MTRFs allow learners to demonstrate various aspects of language proficiency. For example, the processing of integration between texts, which occurs only in multiple text reading, is the core of text comprehension and is influenced by proficiency (Karimi & Shabani, 2013). By focusing on integrating multiple texts, learners' proficiency can be measured from different angles. In addition, when comparing the content of multiple texts, learners must possess critical thinking skills to evaluate the information sources and judge the authenticity of the information.

 Despite a general trend toward integrating these MTRFs into proficiency tests, specifically some standardized tests having incorporated this format (e.g., Test of English

for International Communication; TOEIC), questions remain about the validity of this approach. According to the validity framework proposed by Messick (1995), the content aspect (referring to whether test contents suitably evaluate all aspects of the topic they are designed to measure) is essential to ensure the validity of tests. Specifically, an important issue for the content aspect of construct validity is deciding the knowledge, skills, attitudes, motivations, and other attributes that the assessment task uncovers (Messick, 1995). While many studies on single text reading tests have examined the aspects measured, few studies have compared MTRFs across proficiency tests to identify the test takers' skills that MTRFs try to capture and the assessment tasks used. The current situation of MTRF implementation in proficiency tests remains to be understood, primarily when focusing on testing formats, such as the types of texts employed and questions readers must answer. This study proposes a new classification for MTRFs and analyzes the types of texts and questions adopted in proficiency tests.

Literature review

Understanding multiple texts

Reading in real-life situations aims to understand texts from different sources, such as various authors and publishers (Karimi, 2015). Thus, understanding multiple texts is not an exception but is typical behavior in modern society. However, reading multiple texts is more complicated than reading individual ones. First, the structures of different texts deviate from the principles of consistency and cohesion. Unlike single texts, multiple texts do not explicitly state how individual texts should be associated; therefore, readers must consciously choose to integrate and interpret the information they contain (Anmarkrud et al., 2014). Second, integrating multiple texts requires the reader to infer connections between texts. To understand them in an integrated manner, readers must form connections by making inferences about individual texts that are free of consistency and cohesion (Schedl et al., 2021). Thus, understanding multiple texts is challenging, even for first language (L1) readers (Bråten & Braasch, 2018).

Furthermore, since second language (L2) learners encounter various texts in both academic scenarios and everyday discourse, their multiple text comprehension skills affect their academic success and value formation in everyday life (Karimi & Richter, 2021). Therefore, acquiring these comprehension skills is critical for L2 learners.

Multiple text integration

Multiple text integration is a unique reading process that takes place only when reading multiple texts (List, 2020); it involves the process of forming connections between intertextual information, which is crucial for meaning construction and the core of multiple text comprehension (Primor et al., 2021). If connections are not formed, mental representations of individual texts remain disparate, failing to lead to integrated comprehension. Therefore, multiple text integration—the process of identifying and integrating connections between texts—is necessary for successful comprehension (Bråten et al., 2011).

Three types of multiple text integration occur when reading multiple texts (List, 2020). The first is the combination of existing knowledge with text-based information (i.e., *elaboration*). The second is combined with other information presented in the

same text (i.e., *intratextual integration*). The third is integrating information presented in unrelated texts (i.e., *intertextual integration*). Among these the third type of multiple text integration is the most challenging reading process for L1 readers (List, 2020). The most straightforward action for a reader when encountering multiple texts is to leave the individual texts as they are, as "islands," without forming textual connections (Wineburg, 1991).

Barzilai et al. (2018) defined the third type of integration as "connecting, combining, or organizing information from different texts to achieve diverse aims, such as meaning-making, problem solving, or creating new texts." In this study, we adopted this description as the definition of multiple text integration. Within this definition, the word *connecting*, or *connection forming*, is considered a subdivided concept of multiple text integration. Connection formation is a reading process that causes commonalities and conflicts between texts (List et al., 2021a).

If multiple text integration is often tricky for L1 readers, it is even more challenging for L2 learners. Compared with L1 readers, who can allocate cognitive resources to higher-level processing—such as inference and comprehension monitoring—L2 learners expend cognitive resources in lower-level processing, such as lexical access, acronym processing, and proposition formation (Morishima, 2013). Moreover, less proficient learners face even greater difficulties when integrating multiple texts. When L2 learners were divided by proficiency levels and asked to read multiple texts, the low-proficiency group stuck to decoding vocabulary and did not use global strategies to grasp the overall message of the texts (Plakans, 2009). In general, L2 learners tend to ignore information discrepancies when reading conflicting texts (Braasch & Kessler, 2021). Thus, successful multiple text integration is a significant cognitive achievement for L2 learners.

Text relationship

In comprehending multiple texts, readers must integrate information from diverse viewpoints and scrutinize semantic correlations across a spectrum of texts (Karimi, 2015; List et al., 2021a; Rouet et al., 2021). The challenges of interpreting and comprehending multiple texts are significantly influenced by the relationships between them, encompassing instances of texts containing both complementary and conflicting ideas (Britt et al., 1999; Perfetti et al., 1999). For instance, when readers delve into two complementary texts, they formulate mental representations from a singular standpoint. After reading the first text and perusing the second text, the reader revises the constructed mental representation, aligning information from the second text with the initial one (Perfetti et al., 1999). Conversely, navigating conflicting texts entails a more intricate process (Kobayashi, 2014; Kurby et al., 2005; Rouet et al., 2021).

The discrepancy-induced source comprehension (D-ISC) model by Braasch and Bråten (2017) posits that when confronted with conflicting texts, readers experience cognitive disequilibrium or imbalance due to incongruities in the presented information. Grappling with these conflicting texts challenges readers to cultivate a more profound and adaptable understanding of specific issues or subjects by considering diverse information sources (van den Broek & Kendeou, 2015). Nevertheless, discerning connections across texts and evaluating information from conflicting texts entails cognitively taxing and intricate processes, even for readers proficient in their first language (Anmarkrud

et al., 2014). Without explicit guidance, they frequently overlook conflicting information and abstain from constructing coherent mental representations of the connections between the texts (Stadtler et al., 2020). However, limited test analysis research has focused on the connections readers must form when engaging multiple texts in proficiency tests.
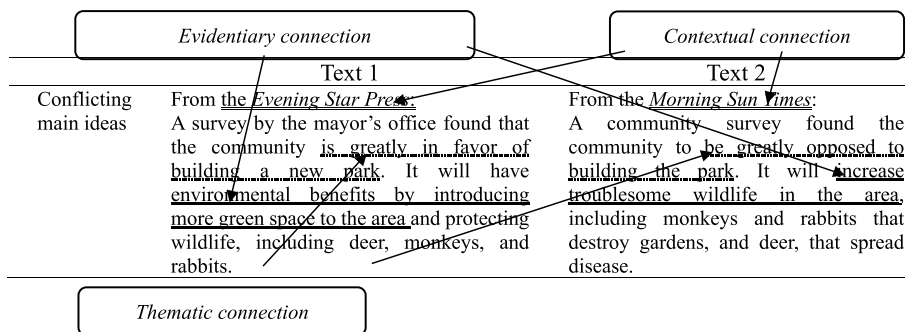
Specificity of connections

The connections formed when reading multiple texts include specificity (Fig. 1). List et al. (2019) found that L1 readers form intertextual connections with at least three different types of specificity. *Evidentiary connections* are formed when readers confirm specific information or data and link them to the text. *Thematic connections* occur when readers connect main ideas across texts. *Contextual connections* are combined when readers compare or link metatextual features, including information about an author's trustworthiness or disciplinary perspectives. According to studies involving L1 readers, evidentiary and thematic connections are more likely to form, while contextual connections are less likely (List et al., 2019, 2020).

Additional research found that the text relationships between multiple texts affect specificity. List et al. (2021a) examined the specificity of connections by reading different multiple texts for L1 readers through online processing. They found that reading multiple conflicting texts promoted the formation of contextual connections while reading complementary texts promoted evidentiary and thematic connections (List et al., 2021a). Thereby, they showed that multiple texts with different relationships affected the formation of connections across texts. However, there are fewer empirical studies involving L2 learners than those involving L1 readers. Research on multiple texts reading has been conducted primarily outside the L2 reading context (Karimi, 2015). Whether different relationships affect the specificity of connections in L2 empirical studies, including proficiency tests, remains untested.

Previous empirical and EPT studies on multiple texts

Barzilai et al.'s (2018) synthesis reviewed empirical studies on multiple text comprehension in terms of text relationships and tasks for assessment. Regarding text relationships, conflicting texts were used in 44.3% of the empirical studies, whereas complementary

Note. This figure was created by the author, based on List et al. (2021a).

Fig. 1 Specificity of connections on expository texts

texts were adopted in only 4.9%. Further, the number of texts used in the studies ranged from 2 to 12 (*Mean* = 4.85, *SD* = 2.96).

Regarding tasks, 96.7% of the studies used writing tasks with the following factors: argument tasks (50.8%), synthesis tasks (19.7%), inquiry tasks (19.7%), and summary tasks (16.4%). In one study that compared different types of tasks, a positive effect on integration was found in writing tasks that required learners to proactively integrate and relate opinions and arguments between texts. Writing was the most common method used to assess learners' integration (63.9%), while multiple-choice questions were used in 4.9% of the cases.

Although writing is widely used to evaluate the integration of multiple texts, this assessment type has limitations. Specifically, the results of the comparison between argument and summary writing tasks in terms of the quality of integration were inconsistent (Barzilai et al., 2018). This is because the results were susceptible to differences in instructions, methods of validating tasks, and individual factors (e.g., the presence or absence of background knowledge and epistemic beliefs; Barzilai et al., 2018).

Compared with the empirical studies on multiple text comprehension, few studies have focused on MTRFs in proficiency tests. To the best of our knowledge, Shimizu (2010) performed the only existing study, focusing on questions assessing multiple text integration in proficiency tests. Three of the six proficiency tests included in the study had questions that assessed multiple text integration. Among them was TOEIC, of which approximately 70% of the questions could be answered by reading a single text, while only 21% of the questions required the integration of texts. However, this study excluded other proficiency tests used worldwide (e.g., Cambridge English Qualifications).

In the past 10 years, TOEIC and Cambridge English Qualifications have undergone revisions in 2016 and 2015, respectively. Therefore, the results of Shimizu's analysis might not be in line with those of the current examinations. Furthermore, Shimizu (2010) did not examine cognitive processing in MTRFs across proficiency tests. Therefore, knowledge of the type of cognitive processing required by the examinees in MTRFs is lacking.

Limitations of previous studies and research questions

The preceding studies were subject to certain limitations. First, empirical studies on multiple text reading have found greater use of conflicting than complementary texts. In addition, as the relationship between texts is reported to affect reading comprehension processing, reading processing differs between complementary and conflicting texts (Braasch & Bråten, 2017; List et al., 2021a). In general, L2 learners tend to ignore information discrepancies when reading conflicting texts (Braasch & Kessler, 2021). Consequently, text relationships among multiple texts may affect test takers' performance on MTRFs. However, research focusing on text relationships in multiple text reading forms of the proficiency tests has been limited.

Second, although there are few examples, Shimizu (2010) analyzed the proportions of integration assessment questions on a set of reading formats in TOEIC. However, integration studies for other proficiency tests have been limited to multiple text reading comprehension questions.

Third, categorization within multiple text reading comprehension questions is limited. The level of integration formation and the type of specificity affect reading processing (e.g., List et al., 2019, 2021a, 2021b). Therefore, reading comprehension questions requiring connection formation across texts might impact the test takers' performance. However, the research focusing on the level of integration formation and specificity for test sets of MTRFs are limited. To address these research gaps, the following research questions were formulated in this study:

> RQ1: What kind of text relationship is used in MTRFs across proficiency tests? (text relationship).
> RQ2: What level of integration is required for questions in MTRFs across proficiency tests? (integration level).
> RQ3: What specificity is required for the formation of integration across proficiency tests? (specificity of connections).

Method

Definition of multiple texts

To analyze MTRFs, terms related to multiple texts must first be defined. Based on previous studies (Britt & Rouet, 2020; Rouet et al., 2021), the following definition of multiple texts was adopted in this study: "a set of generally consistent textual discourses, with or without source information." For example, if a reading format in proficiency tests contains two or more independent texts, it is regarded as an MTRF. Based on the definition, reading formats that include written text with graphs or figures were excluded from this study.

Test forms refer to complete content from a single form of a test. When there are multiple items or questions associated with a reading passage (or group of passages), all associated material is referred to as a "set;" (e.g., the set includes 2–3 text passages and 2–4 questions). The analysis in this study was conducted at the set level, as only texts and items within a set potentially focus on connection. Questions refer to multiple items or questions associated with a reading passage. Integration assessment questions are defined as questions that specifically require test takers to connect and integrate the contents among multiple texts.

Cognitive processing of multiple text integration in MTRFs

Cognitive processing of multiple text integration is gradual and progressive (List & Alexander, 2019; Primor et al., 2021). Primor et al. (2021) divided the integration process of multiple text readings into two levels: "*information selection*," the process of identifying relevant information from each text according to the purpose of the reading and the rest of the text (Level 1); and "*intertextual links*," the various links formed by the reader between texts, including temporal, causal, contradictory, similar, and generalized connections (Level 2).

List and Alexander (2019) classified the four levels of integration when connections are formed between texts in more detail: Level 1 involves awareness or attention to potential overlaps or connections between texts (*relational identification*), while Level 2 involves individual understanding of the potential relationships of information

presented within individual texts (*separate representation*). Level 3 is a single combinatory statement using conjunctions, such as linking texts with predicate verbs that are not explicitly presented between the texts (*simultaneous relation*). Finally, Level 4 elaborates on the relationships identified in the previous level (*relational elaboration*). According to previous studies (List & Alexander, 2019; Primor et al., 2021), multiple text integration is characterized by progressive formation, which refers to connections across texts and is assumed to be formed progressively.

List and Alexander's (2019) perspective on integration can be applied to the analysis of the cognitive processing required for readers in MTRFs, through the traditional study of reading a single text. Alderson's (2000) reading comprehension domain evaluation specification for a single text can be used as a reference. This specification categorizes the cognitive processing of reading formats into three domains: (a) comprehension; (b) comprehension plus transfer or knowledge construction/connection formation; and (c) comprehension plus reasoning/inference. Based on Alderson's (2000) research, this study aggregated the classifications of Primor et al. (2021) mentioned previously and List et al. (2021a) to create a *reading format cognitive framework* for MTRFs (Table 1). We use this framework to analyze MTRFs in proficiency tests.

In the reading format cognitive framework, Multiple Text Question 1 (MT1) is a question that can be answered by simply understanding a single text among texts; MT1 does not require text integration. Multiple Text Question 2 (MT2) requires identifying the connections between multiple texts and answering questions. These questions can be answered by identifying the superficial connections between texts. For example, in *Cambridge English: Advanced* (CAE) Part 6, where MTRF contains a set of four texts (A, B, C, D), the question "Which expert has the same view as B about whether ecosystems will adjust to the consequences of climate change?" is classified as MT2 (superficial intertextual integration). As it is explicitly stated that the target of integration is Text B, the reader must retrieve information superficially related to Text B.

Finally, Multiple Text Question 3 (MT3) requires the formation of connections between texts, similar to MT2; however, it can only be answered when an integrated understanding of multiple texts has been achieved. In other words, this question requires a more elaborate understanding of the texts. For example, the question in CAE Part 6 "Which museum director has a different opinion from the others about how well museums cater to all levels of society?" is classified as MT3 (integration with elaboration). This question does not specify the information to be integrated into the texts; therefore, the reader must read each of the four texts to organize and understand the individual arguments in an integrated manner.

Table 1. Reading Format Cognitive Framework of Multiple Text Reading Test Formats

	Low	→	High
List et al. (2021b)	Separate representation	Simultaneous relation	Relational elaboration
Primor et al. (2021)	N/A	Information selection	Intertextual integration
	MT1	**MT2**	**MT3**
Classification of cognitive processing	*Single text comprehension*	*Intertextual integration (superficial connections)*	*Intertextual integration (elaborate connections)*

Scope of assessment tasks and preliminary study

Although the main methods for assessing readers' multiple text comprehension include writing or multiple-choice questions, this study included only multiple-choice questions on MTRFs. The reasons for excluding writing assessments are as follows: First, based on this study's definition of multiple texts, integration assessment questions involving writing were not found among the targeted proficiency tests. Second, the results of empirical studies on multiple texts are inconsistent because they are influenced by task settings and instructions (Barzilai et al., 2018).

The purpose of the preliminary study was to clarify which proficiency tests fit the scope of this study. This preliminary study consisted of two steps. In Step 1, proficiency tests with and without MTRFs in the reading section were identified. In Step 2, among proficiency tests with MTRF, the presence or absence of integration assessment questions was distinguished. Finally, proficiency tests with integration assessment questions were identified as the scope of this study.

In Step1, the preliminary study included assessing the Reading sections of the following recently published official textbooks: three TOEIC Listening and Reading books, five Internet-based Test of English as a Foreign Language (TOEFL iBT) books, five International English Language Testing System (IELTS) books, five Cambridge English: Key (KET) books, five Cambridge English: First (FCE) books, five CAE books, three Cambridge English: Proficiency (CPE) books (Reading and Use of English sections), five Global Test of English Communication (GTEC) books, and four Test of English for Academic Purposes (TEAP) books. Regarding the number of texts, different tendencies were found among proficiency tests. The average number of texts included in MTRFs was 2.5 in TOEIC, 4.1 in CAE Part 6, 4.5 in CAE Part 8, 5 in CPE Part 7, and 4.0–4.3 in FCE Part 7. After Step 1 of the preliminary study, TOEIC Part 7 (multiple text passage), FCE Part 7, CAE Parts 6 and 8, and CPE Part 7—which included MTRF with multiple-choice questions—were found to potentially adhere to the criteria of this study (Table 2).

On TOEIC, Part 7 is the third part of the TOEIC Reading section and focuses on reading comprehension passages. It includes Single Passage (10 texts, 29 total questions) and Multiple Passage (5 sets of double or triple passages, 25 total questions) item types (Ashmore et al., 2018). The scope of the analysis of TOEIC Part 7 excludes Single Passage item types (Table 3).

Among the official EPT texts that contained 417 of all reading formats in 55 tests, there were 65 MTRFs. That is, the percentage of MTRFs in all proficiency tests was 11.9%. Of these, TOEIC Part 7 had the highest number (23), followed by CAE Part 6 (17). Regarding the number of texts per set (Primor & Katzir, 2018), TOEIC had the lowest average number of texts (2.5), whereas the others ranged from four to five or

Table 2 English Proficiency Tests Covered in This Study

	MTRF & multiple-choice	Single text reading format	MTRF & writing (integrated tasks)
Text only	TOEIC / FCE / CAE / CPE	IELTS / GTEC / TEAP / KET	TOEFL iBT
Text & graph/figure	TOEIC	N/A	TEAP

Table 3 Summary of Preliminary Study

|	CAE		TOEIC	CPE	FCE	Total
Test forms	**20**		**6**	**9**	**20**	**55**
All reading formats	**129**		**114**	**54**	**120**	**417**
	Part 6	**Part 8**	**Part 7**	**Part 7**	**Part 7**	**Total**
Average number of texts per set in MTRFs	4.1	4.4	2.5	5.0	4.3	4.0
MTRF (% in all reading formats)	17 (13.2)	7 (5.4)	23 (20.2)	5 (9.3)	13 (10.8)	65 (11.9)
Complementary texts	8	6	23	5	12	54
Conflicting texts	9	1	0	0	1	11
Average number of texts	4.1	4.4	2.5	5.0	4.3	4.0
Integration assessment questions	**Yes**	No	**Yes**	No	No	-
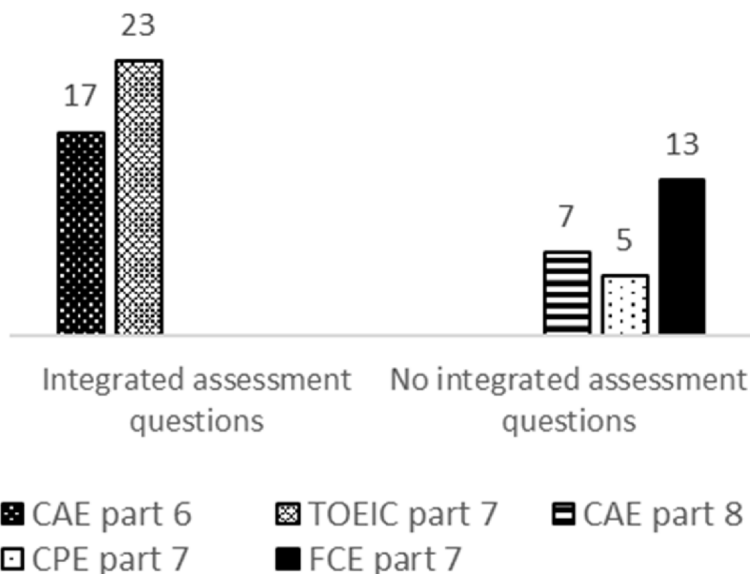
Fig. 2 Frequencies of MTRFs with and without integration assessment questions

more. In the four proficiency tests, text relations in the MTRF were 54 (83.1%) for complementary texts and 11 (16.9%) for conflicting texts.

In Step2 of the preliminary study, the percentage of integration assessment questions in sets of test forms among the proficiency tests was calculated according to Shimizu (2010). The results revealed that almost half of the proficiency tests included integration assessment questions. The proficiency tests comprised two types of MTRFs: those with and without integration questions (Fig. 2). Of the five test sections covered, CAE Part 6 and TOEIC Part 7 contained integration assessment questions. By contrast, CAE Part 8, FCE Part 7, and CPE Part 7 had no integration assessment questions. Specifically, CPE Part 7 did not include any integration assessment questions despite having the highest proficiency test. In sum, among proficiency tests with MTRFs, CAE Part 6 and TOEIC Part 7 included integration assessment questions. Based on Step1 and Step2 of the preliminary study, CAE Part6 and TOEIC Part7 were thus within the scope of this study.

Perspectives of analysis

The author categorized MTRFs in TOEIC and CAE according to the following perspectives (Table 4). (1) The definition of text relationship was adopted from List et al (2021a). Explanatory texts with overlapping claims and rationales were considered multiple texts with a complementary relationship. Explanatory texts in which the claim and the rationale were in conflict were classified as multiple texts with a conflicting relationship.

Questions were analyzed based on the following criteria (Table 1): (2) the cognitive processing required by the questions: (a) single text comprehension (MT1); (b) text integration (MT2; superficial connection formation); (c) text integration (MT3; elaborate connection formation).

(3) Specificity of connection formation refers to the specificity (evidentiary, thematic, contextual) of the connections that the reader is asked to form by the integration assessment questions (List et al., 2021a). Among MT2 and MT3, one question may be intended to form multiple specificity connections, with the number of specificity connections (118) compared to the number of questions in MT2 and MT3 (107). For example, TOEIC Part 7 indicates that test takers must understand a business situation by reading a set of texts such as letters and invoices. To answer the question "Who is most likely Ms. Burstein?", the name of the sender of the letter and the invoice must be compared in the context in which the text was created. In this case, the specificity required by this question was determined to be a contextual connection. To answer the question "How much did Ms. Burstein probably charge Azin Shinwai?", specific evidential information from the two texts must be integrated and understood. In this case, the specificity of the question was determined to be an evidentiary connection.

Following the criteria described above, two Japanese English language teachers coded the MTRFs and the questions in proficiency tests. The agreement rate was 86.8% (Cohen's $\kappa = 0.774$, $p < 0.001$). Any discrepancy in their assessments was resolved through discussion. After all texts were classified using the procedure described above, the frequency of each item was calculated.

Statistical analysis

Descriptive statistics were used to examine the characteristics of the texts and questions used in the MTRF for the two proficiency tests (CAE Part 6 and TOEIC Part 7). This approach was employed for several reasons. First, since previous study has also interpreted trends and characteristics using frequency of values rather than estimated statistics (i.e., Shimizu, 2010), this study adoped the same procedure. Second, there are

Table 4 Analytical Criteria of This Study

Criterion		Item	Existing study
1	**Text relationship**	**Texts (complementary / conflicting)**	**List et al. (2021a)**
2	Cognitive processing on questions in MTRF	Single text comprehension (MT1), intertextual integration (superficial connections; MT2), intertextual integration (elaborate connections; MT3)	Alderson (2000), Primor et al. (2021), List et al. (2021b)
3	Specificity of connection formation	Connection (evidentiary / thematic / contextual)	List and Alexander (2019)

limitations in comparing different tests with estimated statistics. For example, the chi-square test is one of the analytical methods that compares frequencies for each test item. However, the result of chi-square test might remain unstable if the frequency of the items contains zero values (Bolboacă et al., 2011). Therefore, descriptive statistics were considered the better alternative to examine the characteristics and trends among the proficiency tests in this study.

Results

Text relationship

Table 5 summarizes the results of this study. Regarding text relationships, complementary texts accounted for 82.5% (54 of 65) of the total MTRFs, with 17.5% (11 of 65) of the total number comprising conflicting texts. Focusing on individual proficiency tests, CAE Part 6 contained mainly complementary or conflicting texts, with the former accounting for 47.1% (8 of 17) and the latter for 52.9% (9 of 17). TOEIC Part 7, however, was complementary text-centered, consisting solely of complementary texts.

Analysis of text relationships in terms of average word count and text count data showed different trends for the two proficiency tests. CAE Part 6 contained four texts with around 140 words on average in a set of MTRF. There were eight complementary texts and nine conflicting texts in sets of test sets, in roughly equal proportions. A set on MTRF in TOEIC Part 7 consisted of two to three texts with an average of over 150 words. While no conflicting texts were found, complementary texts were mainly presented in the TOEIC Part 7. The number of texts and words on a set in TOEIC Part 7 was lower than in CAE Part 6. Thus, comparing the MTRFs of the two proficiency tests, texts in CAE Part 6 had fewer words per set than in TOEIC Part 7, while the number of texts on a set in CAE Part 6 was higher and the text relationship more complicated than in TOEIC Part 7.

Table 5 Summary of Multiple Text Reading Formats in English Proficiency Tests

	CAE Part 6	TOEIC Part 7	Total
Test forms	20	6	26
All reading formats	129	114	243
Sets of multiple texts (% in all reading formats)	17 (13.2)	23 (20.2)	40 (16.5)
Complementary texts	8	23	37
Conflicting texts	9	0	9
Average number of words in all texts	582.9	405.3	-
Average number of words in each text	141.6	154.7	-
Average number of texts in MTRF	4.1	2.5	-
Questions on sets	64	115	179
MT1 (% on all questions)	0 (0.0)	72 (62.6)	72
MT2	48 (75.0)	43 (37.4)	91
MT3	16 (25.0)	0 (0.0)	16
Evidentiary connection (% on connections)	51 (77.3)	41 (78.8)	91
Thematic connection	15 (22.7)	1 (1.9)	16
Contextual connection	0 (0.0)	10 (19.3)	10

Note. "Test forms" refer to complete content from a single form/version of a test. "All reading formats" includes all individual passages and all sets that include multiple passages

Cognitive processing of questions

One hundred seventy-nine questions in the proficiency tests involved MTRF (64 in CAE Part 6 and 115 in TOEIC Part 7; Fig. 3). CAE Part 6 required integrated processing for all questions (75.0% in MT2 and 25.0% in MT3). In contrast, 62.6% (72 of 115) of the TOEIC Part 7 questions required single text comprehension (MT1), 37.4% (43 of 115) required text integration (MT2), and 0% required text integration (MT3). In other words, more than 60% of the TOEIC Part 7 questions could be answered by understanding a single text without integration tasks.

Although CAE Part 6 and TOEIC Part 7 are considered a type of MTRF, the framework of MTRF presented in this study revealed that the tendency across proficiency tests was different. Questions on set in TOEIC Part 7 adopted mainly MT1 (single text comprehension) and MT2 (superficial connections). Meanwhile, CAE Part 6 used MT2 and MT3 (elaborate connections) requiring elaborate connections.

Connection forming

Regarding the connections required by questions on TOEIC Part 7, Fig. 4 and Table 5 illustrate that evidentiary connections accounted for 78.8% (41 of 52), thematic connections for 1.9% (1 of 52), and contextual connections for 19.3% (10 of 52). Regarding CAE Part 6, evidentiary connections accounted for 77.3% (51 of 66) and thematic connections for 22.7% (15 of 66); there were no contextual connections. Thus, from the perspective of connection forming, the adoption of questions to form contextual connections was different across TOEIC Part 7 and CAE Part 6. Considering that TOEIC Part 7 mainly adopted complementary texts, and only CAE Part 6 used conflicting texts, a combination of contextual connections with conflicting texts was not found in any MTRFs.
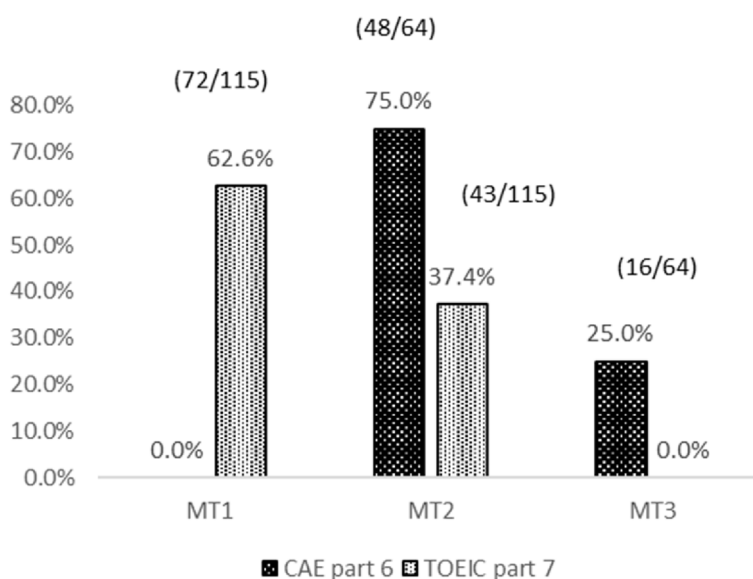

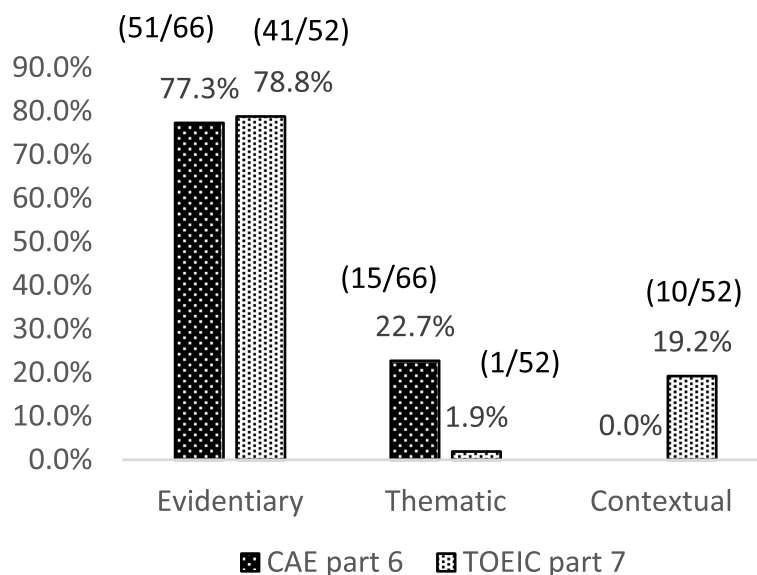**Fig. 3** Proportion of questions based on cognitive processing on multiple text reading test formats

Fig. 4 Specificity of connections formed on multiple text reading test formats

Discussion

Text relationships

Regarding text relationship across proficiency tests (RQ1), CAE Part 6 and TOEIC Part 7 showed different trends in the text relationships used. One reason for the difference in text relationships used in the two tests is the construct of the proficiency test (Table 6). For CAE Part 6, both complementary and conflicting texts were used. The subconstruct of multiple text comprehension on CAE defines it as follows:" able to understand opinion and attitude; comparing and contrasting of opinion and attitude" (Cambridge Language Assessment, 2015). This suggests that CAE Part 6 uses conflicting texts because the focus is on measuring the ability to contrast and understand opinions.

In contrast, TOEIC Part 7 used only complementary texts. The subconstruct of reading multiple texts on TOEIC defines it as "can infer gist by connecting information across two texts" (Schedl, 2010, p. 7). In other words, rather than measuring the ability to process contrasting information, TOEIC Part 7 focuses on measuring the ability to connect information between texts and grasp the main points. Therefore, complementary texts, the structure of which is simpler than that of conflicting, may be mainly used. This difference in subconstructs for multiple text comprehension may have affected the text relationships used in CAE Part 6 and TOEIC Part 7.

The second reason is that CAE and TOEIC target different ranges of test takers. At the Common European Framework of Reference for Languages (CEFR) levels, CAE is intended for test takers at advanced level C1. The results of this study show that text sets for CAE Part 6 had approximately equal numbers of complementary and conflicting texts. On conflicting texts, L2 learners find it difficult to process conflicting information and tend to ignore it at times (Braasch & Kessler, 2021). Therefore, CAE, which targets advanced test takers, may include in its MTRFs not only

Table 6 Comparison of Construct, Genre, and Test Format on Reading Section Between CAE and TOEIC

	CAE	TOEIC
CEFR levels	C1	A1 to C1
Construct	Facilitate processing, which approximates more closely to real-life processing in various reading types (Weir et al., 2013)	Measure the everyday English skills of people working in an international environment by a wide variety of genres regarding real-life work settings (Educational Testing Service, 2017b)
Subconstructs—understanding multiple texts	Able to understand opinion and attitude; comparing and contrasting of opinion and attitude (Cambridge Language Assessment, 2015)	Can infer gist by connecting information across two texts (Schedl, 2010)
Discourse mode genre	Newspapers, magazines and journals, fiction and non-fiction books (extracts), promotional and informational sources (e.g., guides, manuals; Khalifa & Weir, 2009)	- Written texts concerning businesses, hotels, hospitals, restaurants, international meetings, conventions, and sporting events (Educational Testing Service, 2017b) - Text messages, online chat dialogues (Im & Cheng, 2019)
Rhetorical task	Descriptive, narrative, expository, argumentative, instructive (Khalifa & Weir, 2009)	A wide variety of genre regarding real-life work settings (Educational Testing Service, 2017b)
Number of sets in overall reading section	34 sets (per test format)	13 sets (per test format)
Number of questions in overall reading section	55 questions (per test format)	100 questions (per test format)
Number of MTRF set in overall reading section	A set with 4 questions each (Part 6)	4 sets with 5 questions each (Part 7)
Number of questions in MTRF	4 questions per test format (Part 6)	20 questions per test format (Part 7)

complementary texts that are easier for test takers to deal with, but also conflicting texts that require more complex processing.

In addition, the difference of target test takers may affect the number of texts per set in proficiency tests. For example, the average number of texts per set in CAE Part 6 was 4.1, compared to 2.5 in TOEIC Part 7. By including more texts in the MTRF, CAE Part 6 is intended to require a more complex reading process for test takers. The construct of CAE is described as follows: "… facilitate processing which approximates more closely to real-life processing in various reading types …" (Weir et al., 2013, p. 162). In other words, the Cambridge English Qualifications may focus on assessing reading comprehension skills in real-life situations by using more texts, specifically conflicting texts.

On the other hand, this study indicates TOEIC Part 7 had only complementary texts. The reason could be attributed to the wider range of targeted test takers. TOEIC is intended from entry level A1 to the advanced level C1. Therefore, TOEIC, which targets test takers ranging from high to low-proficiency test takers, may use complementary texts while refraining from adopting conflicting texts. Specifically, the construct of TOEIC is described as follows: "measure the everyday English skills of people working in an international environment by a wide variety of genres regarding real-life work settings" (Educational Testing Service, 2017b). TOEIC thus may aim

to test learners' ability to comprehend everyday discourses from a limited number of texts per set by using complementary texts, which have simpler structures of text relationships than conflicting texts. Although the proficiency tests may have different purposes and targets for test takers, the feature of text relationship explored in this study could provide for the understanding of MTRFs in proficiency tests.

The study also reveals that the text relationships used tend to differ between proficiency tests and empirical studies. For example, the results showed that TOEIC had a high use of complementary texts. This finding was similar to the MTRFs in other proficiency tests. As for text relationships, complementary texts accounted for 82.5% of the total MTRFs in proficiency tests, while conflicting texts comprised 17.5%. In contrast, Barzilai et al. (2018) showed in their review that in empirical studies on multiple text comprehension, complementary texts accounted for 4.9% of the total, while conflicting texts accounted for 44.3%, indicating a significant dissociation of text relationships between proficiency tests and empirical studies. Two possible reasons exist for mainly using complementary texts in MTRFs. The first is the intention to manage the difficulty level as the cognitive process of reading conflicting texts is more complex than that of reading complementary texts (Braasch & Kessler, 2021). Therefore, complementary texts in proficiency tests may be used to avoid unduly increasing the difficulty of the assessments. The second reason is that it reflects real-life situations. In the real world, people encounter complementary texts more often than conflicting ones (Hess, 2008; McCrudden et al., 2021). The test developers' intention to reflect real-life situations may therefore result in the selection of natural complementary texts.

Required cognitive processing

Regarding the cognitive processing of integration (RQ2), the results suggest that the required cognitive processing in MTRFs differs in CAE Part 6 and TOEIC Part 7. CAE Part 6 had many questions regarding intertextual integration (superficial connections; MT2; 75%) and multiple text comprehension (elaborative connections; MT3; 25%). The construct of multiple texts might explain why CAE Part 6 focuses on MT2 and MT3. The subconstruct of MTRF in CAE is as follows: "understand opinion and attitude; *comparing and contrasting* of opinions and attitudes" (Cambridge Language Assessment, 2015, p. 10). This construct reveals that using integration assessment questions, such as MT2 and MT3, in set of MTRFs is intended to align with the construct. Adopting questions that require elaborative integration, as measured in MT3, in CAE might reflect the aim of CAE. This test is designed to undergo development to assess its appropriateness for evaluating suitability for academic purposes at the undergraduate level, according to the academic descriptors specified in the CEFR C1 framework (Cambridge Language Assessment, 2015). Thus, requiring test takers to engage in complex reading activities, such as MT3, might be appropriate in CAE.

TOEIC Part 7 contains, however, many questions that require (1) individual text comprehension (MT1; 62.6%) and (2) intertextual integration (superficial connections; MT2; 37.4%). The construct on MTRF for TOEIC Part 7 states the following: "understand explicitly stated information in order to infer non-explicit information" (Schedl, 2010, p. 7). This statement suggests that the questions on integration in TOEIC Part 7 are tests of superficial connections, although the constructs do not describe the level of integration.

By focusing on the integration assessment questions, we might be able to predict future trends in MTRFs. For example, in the TOEIC Part 7 assessed in this study, 62.6% of questions on MTRFs were single text reading comprehension questions, and 37.4% were intertext integration questions. Conversely, in Shimizu's (2010) study, single text reading comprehension questions accounted for approximately 78% of the questions, while intertext integration questions accounted for approximately 22%. In other words, after the revision of TOEIC test formats in 2016, the number of questions assessing intertext integration increased by approximately 15%. However, research on MTFRs for TOEIC is limited, warranting further investigation to accurately understand trends and determine the future direction of MTRFs for TOEIC.

Regarding instruction, teachers should pay close attention to the level of cognitive processing required by the questions in the MTRF. When an MTRF does not include integration assessment questions, what it asks examinees to do is rather simple. According to Khalifa and Weir's (2009) reading comprehension model, after the mental representation of a single text is constructed, an intertextual mental representation is naturally formed, which enables the examinee to read the single text well. In this case, it can be said that no special instruction from the teacher is required when teaching reading comprehension.

Conversely, when integration questions are included in the MTRF, different cognitive processing is required of examinees (e.g., List & Alexander, 2019). Multiple text integration is a reading process that occurs only in multiple text reading. It includes the formation of connections and the integration of intertextual information, which is essential for meaning construction, and the core of multiple text comprehension (Primor et al., 2021). In other words, in the case of MTRFs, which include integration assessment questions, the reader must infer and integrate the connections between texts. If the MTRF is viewed as an extension of single text reading, some learners may have difficulty integrating and understanding multiple texts. Thus, even for the same MTRF, the load placed on the examinee varies greatly, depending on whether an intertext integration question is included.

Specificities of connections

Regarding the cognitive processing of connections (RQ3), how test takers are asked to form connections also differs among proficiency tests. Compared with TOEIC Part 7, CAE Part 6 asks readers to form thematic more connections because the text formatting in the exam is of the persuasive type with claims and evidence. Meanwhile, for contextual connections, TOEIC Part 7 requires learners to pay attention to source information, such as the writing and organization of the text. However, CAE Part 6 contains many questions that ignore the source information. Thus, none of the questions in CAE Part 6 require the formation of contextual connections in MTRFs.

The adaptation of source information may differ between CAE Part 6 and TOEIC Part 7. Although source information is a text condition that reflects the real world, the reader's existing knowledge may affect its understanding and use. By allowing readers to judge the trustworthiness and knowledge level of source information, the task facilitated readers' integrated understanding of text information (Rouet et al., 2021). This judgment of the source information may be influenced by the reader's pre-existing knowledge and

use of strategy, which includes the potential to interfere with the measurement of learners' skills as intended by the proficiency tests. Therefore, contextual connections in proficiency tests should be used with caution. Specifically, care should be taken to ensure that test developers consider whether the reader's existing knowledge and strategic skills affect their answers.

Conclusion

Key findings

The findings of this study are as follows. First, a prior review of empirical research and the present study show differences in the text relationship of MTRFs. In terms of text relationships, MTRFs in proficiency tests tend to use complementary texts. However, most empirical studies on multiple text comprehension have employed conflicting texts and rarely used complementary texts (Barzilai et al., 2018). Considering these empirical studies, future research on multiple text comprehension with complementary texts is necessary, whereas for proficiency tests, the use of conflicting texts should be considered more carefully.

Second, regarding questions, the reading format cognitive framework clarifies the different tendencies of cognitive processing on questions in proficiency tests. CAE Part 6 required integrated processing for all questions. Meanwhile, more than 60% of questions on TOEIC Part 7 could be answered by understanding a single text without integration tasks. Specifically, CAE Part 6 was more advanced in terms of cognitive processing than TOEIC Part 7. In addition, based on the comparison between the previous study (i.e., Shimizu, 2010) and the present study, focusing on TOEIC Part 7, the percentage of integration assessment questions increased by approximately 15% over the last 10 years.

Third, the specificity of the required connection formation also differed among the proficiency tests: CAE Part 6 required the formation of thematic connections in many questions, whereas only TOEIC Part 7 required the formation of contextual connections. In summary, the results revealed that the texts and questions used in MTRFs showed differences among proficiency tests.

Limitations and pedagogical implications

A few limitations exist in the present study. First, the classification of text relationships used in this study might not have fit perfectly with some proficiency tests. List et al. (2021a) classified texts in terms of complementary or conflicting relationships, based on typical expository texts, such as newspaper articles and persuasive texts. However, because the text in TOEIC Part 7 includes a wide range of text forms, such as e-mails and advertisements, classifying them comprehensively using this definition of text relations might be challenging.

Second, there is the difficulty of comparing proficiency tests. As constructs differ among proficiency tests, direct comparison may not be possible in a strict sense. For example, as the context and purpose of use vary among proficiency tests, text and questions in a set of multiple texts in test forms may appear to be different. However, from a broader perspective, proficiency tests have one thing in common: they are used in research to measure L2 learners' proficiency in English. Moreover, while the purpose of using MTRFs may be different, the purpose of comprehending multiple texts is at least

congruent. According to the new criteria used in this study, classifying sets in MTRFs sheds light on the current status of multiple text reading comprehension in proficiency tests. It may also increase the possibility of developing new MTRFs in the future.

Third, in this study, the analysis of MTRFs was limited to multiple-choice questions, which have the following advantages (Hughes & Hughes, 2020): First, the reliability of the scores tends to be high; and second, it can measure receptive skills without forcing test takers to perform writing or speaking tasks. In addition, multiple-choice questions provide a detailed measure of test takers' understanding. However, there are some drawbacks to multiple-choice questions. As most of the choices in these questions are used as distractors, an increased number of incorrect answers is likely. Furthermore, multiple-choice questions may reduce the positive test effect because the items permit many possible answers. Consequently, multiple-choice questions run the risk of unintentionally creating false knowledge for test takers (Roediger & Marsh, 2005). In short, multiple-choice questions may not always be adequate for measuring the comprehension of multiple texts in MTRFs. Therefore, in the future, MTRFs with different tasks should also be considered.

Given the limitations of multiple-choice questions, integrated writing tasks (e.g., TOEFL iBT writing tasks) are a promising means to assess multiple text comprehension. For example, in empirical research on L1 multiple text comprehension, writing tasks, such as summary and argumentative writing, were employed in more than 90% of studies (Barzilai et al., 2018). Although the number of MTRFs with writing tasks in proficiency tests is still limited, this testing method reflects a real-life situation. With an expected increase in MTRFs with writing tasks, the need for further analysis may increase in parallel.

It is important to consider that the characteristics of writing tasks may be a barrier to the further widespread use of integrated writing tasks in MTRFs in proficiency tests. When multiple text comprehension is assessed with writing products, the quality of integrated writing may be influenced by differences in instructions, ways of examining task effects, and individual differences such as background knowledge and epistemic beliefs (Barzilai et al., 2018). Therefore, MTRFs with writing should be analyzed using perspectives adopted in this study. This may allow for a detailed examination of factors influencing the integrated writing task.

Further, appropriate MTRFs may vary across contexts and text topics. For example, each MTRF contains diverse text relationship and questions types for various proficiency tests. This characteristic implies that a common understanding—that is, what abilities should be measured by having examinees read multiple texts and what abilities are ideal— has not been clarified among test developers. Since research on multiple text comprehension has flourished in L1 studies and is limited in L2 studies, future empirical research on L2 multiple text comprehension should be conducted to develop an optimal MTRF.

Another limitation of this study is the limited amount of information provided by the comparison of frequencies in test formats. A narrower perspective from the present study is its implications for the characteristics of test formats on proficiency tests in terms of text relations, questions on level of integration, and types of specificities on connection forming across texts. The results of this analysis can be used to identify the different characteristics and trends of MTRFs on TOEIC and CAE. A more detailed

analysis, for example by interviewing test takers or by conducting the think-aloud method while reading MTRFs, would provide a more multifaceted picture of the characteristics of MTRFs.

Based on the study results, the following important pedagogical implications are presented. From an educational perspective, first, teachers must explicitly instruct students on the structure of MTRFs. In the reading and processing of multiple texts, which is at the core of multiple text comprehension, integration among texts distinguishes reading a single text from reading multiple texts; therefore, teaching MTRFs, including integration assessment questions, is not simply an extension of traditional reading instruction. Thus, teaching MTRFs from the perspectives of the criteria used in the present study would be effective. For example, by analyzing MTRFs based on the reading format cognitive framework (see Table 1) proposed in this study, it is possible to explicitly indicate the tasks required by the test.

Using the reading format cognitive framework proposed in this study to categorize questions on MTRF, first, we found that each proficiency test tends to have different integration questions and requires different levels of integration. For example, the characteristics of multiple text reading, such as text relationships, level of integration and the type of connections formed, have specific tendencies across proficiency tests. Thus, teachers need to explicitly explain the characteristics of MTRFs to learners.

Second, teachers should actively introduce task interventions to promote learners' understanding of multiple texts. For example, text highlighting may be a possible task intervention for L2 learners. According to Leroy et al. (2021), tasks requiring readers to highlight connections between texts while reading multiple texts improved their comprehension of multiple texts. Leroy et al. (2021) state that this highlighting task led to the inclusion of more connections in their writing products, thus facilitating readers' integrated understanding of texts. When teaching L2 students how to address MTRFs, task interventions such as highlighting might support a good understanding of multiple texts, especially for low-proficiency L2 learners.

Third, the results of this study show that some proficiency tests had a frequency of zero in the text relations, integration level, and connection formation criteria. This may indicate that there are constructs that are not measured by proficiency tests. For example, no MTRFs for conflicting texts were found in TOEIC Part 7 and there were no questions asking for MT3, the highest integration level. However, CAE Part 6 had no integration questions requiring the formation of contextual connections. Future tests may need to consider developing questions that incorporate a perspectives of text relations and cognitive processing level.

Fourth, test developers can produce new assessment tasks by focusing on connection formation. Higher-proficiency learners form a higher total number of connections across texts while reading multiple texts (Mikami, 2024). Therefore, by adjusting the number of connections formed and the depth of processing in MTRFs, test developers may be able to manipulate the difficulty of MTRFs. Additionally, because few MTRFs currently adopt the question of integrating contextual information across conflicting texts, this combination of text and task factors could potentially create new reading comprehension assessment tasks.

Abbreviations

CAE	Cambridge English: Advanced
CEFR	Common European Framework of Reference for Languages
CPE	Cambridge English: Proficiency
D-ISC	Discrepancy-Induced Source Comprehension
EPT	English language proficiency tests
FCE	Cambridge English: First
GTEC	Global Test of English Communication
IELTS	International English Language Testing System
KET	Cambridge English: Key
L1	First language
L2	Second Language
MTRF	Multiple Text Reading Test Format
MT1	Multiple Text Question 1
MT2	Multiple Text Question 2
MT3	Multiple Text Question 3
RQ1	Research Question 1
RQ2	Research Question 2
RQ3	Research Question 3
TEAP	Teaching English for Academic Purposes
TOEFL iBT	Internet-based Test of English as a Foreign Language
TOEIC	Test of English for International Communication

Availability of data and material
The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests
 The authors declare that they have no competing interests.

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.

Alexander, P. A., & The Disciplined Reading and Learning Research Laboratory. (2012). Reading into the future: Competencies for the 21st century. *Educational Psychologist, 47*(4), 259–280. https://doi.org/10.1080/00461520.2012.722511

Anmarkrud, Ø., Bråten, I., & Strømsø, H. I. (2014). Multi-document literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences, 30*, 64–76. https://doi.org/10.1016/j.lindif.2013.01.007

Ashmore, E., Duke, T., & Sakano, J. (2018). Background and goals of the TOEIC Listening and Reading update project. In D. Powers and J. Schmidgall (Eds.), *The research foundation for the TOEIC tests* (Vol. 3, pp. 31–38). ETS. https://www.ets.org/s/toeic/pdf/research-compendium.pdf

ALC Press. (2019a). *GTEC kako mondai shu advanced* [GTEC official tests for advanced learners]. Authors.

ALC Press. (2019b). *GTEC kako mondai shu basic* [GTEC official tests for beginners]. Authors.

Barzilai, S., Zohar, A. R., & Mor-Hagani, S. (2018). Promoting the integration of multiple texts: A review of instructional approaches and practices. *Educational Psychology Review, 30*(3), 973–999. https://doi.org/10.1007/s10648-018-9436-8

Bolboacă, S. D., Jäntschi, L., Sestraş, A. F., Sestraş, R. E., & Pamfil, D. C. (2011). Pearson-Fisher Chi-Square Statistic Revisited. *Information, 2*(3), 528–545. https://doi.org/10.3390/info2030528

Braasch, J. L. G., & Bråten, I. (2017). The Discrepancy-Induced Source Comprehension (D-ISC) model: Basic assumptions and preliminary evidence. *Educational Psychologist, 52*(3), 167–181. https://doi.org/10.1080/00461520.2017.1323219

Braasch, J. L. G., & Kessler, E. D. (2021). Working toward a theoretical model for source comprehension in everyday discourse. *Discourse Processes, 58*(5), 449–467. https://doi.org/10.1080/0163853X.2021.1905393

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). Role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist, 46*(1), 48–70. https://doi.org/10.1080/00461520.2011.538647

Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J. F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative, comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 209–233). Erlbaum.

Bråten, I., & Braasch, J. L. G. (2018). The role of conflict in multiple-source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (1st ed., pp. 184–201). Routledge. https://doi.org/10.4324/9781315627 496-11

Britt, M. A., & Rouet, J.-F. (2020). Multiple document comprehension. *Oxford Research Encyclopaedia of Education*. https://doi.org/10.1093/acrefore/9780190264093.013.867

Bennesse Corporation. (2018). *GTEC CBT koushiki mondai shu* [GTEC CBT official tests]. Authors.

Bennesse Corporation. (2019a). *GTEC CBT koushiki mondai shu speaking* [GTEC CBT official tests for speaking]. Authors.

Bennesse Corporation. (2019b). *GTEC CBT koushiki mondai shu writing* [GTEC CBT official tests for writing]. Authors.

Cambridge University Press. (2019a). *A2 KEY for schools 1 with answers*. Authors.

Cambridge University Press. (2019b). *A2 KEY for schools trainer 1 with answers*. Authors.

Cambridge University Press. (2022c). *A2 KEY for schools 2 with answers*. Authors.

Cambridge University Press. (2020a). *B2 first 4 with answers*. Author.

Cambridge University Press. (2020c). *C1 advanced trainer: Six practice tests with answers*. Author.

Cambridge University Press. (2012a). *Cambridge English advanced 5 with answers*. Author.

Cambridge University Press. (2013). *Cambridge English objective proficiency student's book with answers* (2nd ed.). Author.

Cambridge University Press. (2014c). *Cambridge English KEY 4 with answers*. Authors.

Cambridge University Press. (2014a). *Cambridge English advanced 1 for revised exam from 2015 with answers*. Author.

Cambridge University Press. (2014b). *Cambridge English first 1 for revised exam from 2015 with answers*. Author.

Cambridge University Press. (2014d). *Cambridge English KEY 7 with answers*. Authors.

Cambridge University Press. (2015a). *Advanced trainer: Six practice tests with answers* (2nd ed.). Author.

Cambridge University Press. (2015c). *First trainer: Six practice tests with answers* (2nd ed.). Author.

Cambridge University Press. (2015b). *Cambridge English proficiency 2 with answers*. Author.

Cambridge University Press. (2016b). *Cambridge English first 2 with answers*. Author.

Cambridge University Press. (2016a). *Cambridge English advanced 2 with answers*. Author.

Cambridge University Press. (2018b). *Cambridge English first 3 with answers*. Author.

Cambridge University Press. (2018a). *Cambridge English advanced 3 with answers*. Author.

Cambridge University Press. (2019c). *Cambridge IELTS 14 academic with answers*. Authors.

Cambridge University Press. (2020b). *Cambridge IELTS 15 academic with answers*. Authors.

Cambridge University Press. (2021b). *Cambridge IELTS 16 academic with answers*. Authors.

Cambridge University Press. (2021a). *Cambridge English advanced 4 with answers*. Author.

Cambridge University Press. (2022d). *Cambridge IELTS 17 academic with answers*. Authors.

Cambridge University Press. (2023). *Cambridge IELTS 18 academic with answers*. Authors.

Cambridge University Press. (2012b). *Cambridge English proficiency 1 for updated exam with answers*. Author.

Cambridge Language Assessment. (2015). *Research Notes - A quarterly publication reporting on learning, teaching and assessment. 62,* 1–35. https://www.cambridgeenglish.org/images/268809-research-notes-62.pdf

Educational Testing Service. (2016a). *TOEIC test koushiki mondaishu* [TOEIC test official tests]. Author.

Educational Testing Service. (2016b). *Koushiki TOEIC listening & reading mondaisyu 1*. [Official TOEIC Listening & Reading practice sets]. Author.

Educational Testing Service. (2017a). *Koushiki TOEIC listening & reading mondaisyu 2*. [Official TOEIC Listening & Reading practice sets]. Author.

Educational Testing Service. (2017b). *TOEIC Examinee Handbook Listening & Reading Test*. https://funge.uva.es/wp-content/uploads/2014/11/Handbook-TOEIC-LR-y-SW.pdf

Educational Testing Service. (2018b). *Official TOEFL iBT tests volume 2*. Authors.

Educational Testing Service. (2018a). *Official TOEFL iBT tests volume 1*. Authors.

Educational Testing Service. (2020a). *Official TOEFL iBT tests volume 1*. Authors.

Educational Testing Service. (2020b). *Official TOEFL iBT tests volume 2*. Authors.

Educational Testing Service. (2020c). *The official guide to the TOEFL iBT test*. Authors.

Hess, D. (2008). Controversial issues and democratic discourse. In L. S. Levstik & C. A. Tyson (Eds.), *Handbook of research in social studies education* (pp. 124–136). Routledge.https://doi-org.ezproxy.tulips.tsukuba.ac.jp/https://doi.org/10.1177/0265532219828252

Hughes, A. (2020). *Testing for language teachers*. (3rd ed.). Cambridge University Press.

Im, G.-H., & Cheng, L. (2019). The Test of English for International Communication (TOEIC). *Language Testing, 36*(2), 315–324.

Karimi, M. N., & Shabani, M. B. (2013). Comparing the strategic behavior of more successful vs. Less successful readers of multiple technical reading texts. *Innovation in Language Learning and Teaching, 7*(2), 125–138. https://doi.org/10.1080/17501229.2012.726223

Karimi, M. N. (2015). L2 Multiple-document comprehension: Exploring the contributions of L1 reading ability and strategic processing. *System, 52,* 14–25. https://doi.org/10.1016/j.system.2015.04.019

Karimi, M. N., & Richter, T. (2021). Text-belief consistency effects in L2 readers. *Discourse Processes, 58*(8), 726–742. https://doi.org/10.1080/0163853X.2021.1913935

Khalifa, H., & Weir, C. (2009). *Examining reading: Research and practice in assessing second language reading.* Cambridge University Press.

Kobayashi, K. (2014). Students' consideration of source information during the reading of multiple texts and its effect on intertextual conflict resolution. *Instructional Science, 42*(2), 183–205. https://doi.org/10.1007/s11251-013-9276-3

Kurby, C. A., Britt, M. A., & Magliano, J. P. (2005). The role of top-down and bottom-up processes in between-text integration. *Reading Psychology, 26*(4–5), 335–362. https://doi.org/10.1080/02702710500285870

Leroy, C., Gerjets, P., Oestermeier, U., & Kammerer, Y. (2021). Simultaneous presentation of multiple documents and text highlighting: Online integrative processes and offline integrated understanding. *SCientific Studies of Reading, 25*(2), 179–192. https://doi.org/10.1080/10888438.2020.1784903

List, A. (2020). Investigating the cognitive-affective engagement model of learning from multiple texts: A structural equation modeling approach. *Reading Research Quarterly, 56*(4), 781–817. https://doi.org/10.1002/rrq.361

List, A., & Alexander, P. A. (2019). Toward an integrated framework for multiple-text use. *Educational Psychologist, 54*(1), 20–39. https://doi.org/10.1080/00461520.2018.1505514

List, A., Du, H., & Lee, H. Y. (2021a). Examining relationship formation across consistent and conflicting texts. *Discourse Processes, 58*(2), 134–154. https://doi.org/10.1080/0163853X.2020.1834328

List, A., Du, H., & Lee, H. Y. (2021b). How do students integrate multiple texts? Investigation of the top-down processing. *European Journal of Psychology of Education, 36*(3), 599–626. https://doi.org/10.1007/s10212-020-00497-y

List, A., Du, H., Wang, Y., & Lee, H. Y. (2019). Toward a typology of integration: Examining the document model framework. *Contemporary Educational Psychology, 58*, 228–242. https://doi.org/10.1016/j.cedpsych.2019.03.003

McCrudden, M. T., Huynh, L., Lyu, B., & Kulikowich, J. M. (2021). Bridging inferences and learning from multiple complementary Texts. *Discourse Processes, 20*(5/6), 529–548. https://doi.org/10.1080/0163853X.2021.1924586

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Mikami, Y. (2024). Fukusu tekusuto dokkai deno tsunagari keisei ga nihonjin EFL gakushusha no tougouteki rikai ni ataeru eikyou no kennshou [Investigation of the influence of connection formation on the integrative understanding of multiple texts for Japanese EFL learners]. *STEP Bulletin, 35*. 74–101. https://www.eiken.or.jp/center_for_research/pdf/bulletin/vol35/vol_35_p74-p101.pdf

Morishima, Y. (2013). Allocation of limited cognitive resources during text comprehension in a second language. *Discourse Processes, 50*(8), 577–597. https://doi.org/10.1080/0163853X.2013.846964

Obunsha. (2015). *TEAP jissen mondai shu.* [TEAP practical tests]. Authors.

Obunsha. (2016a). *TEAP gino betsu mondai shu reading/ listening.* [TEAP practical tests for reading and listening]. Authors.

Obunsha. (2016b). *TEAP gino betsu mondai shu writing/speaking.* [TEAP practical tests for writing and speaking]. Authors.

Perfetti, C. A., Rouet, J. F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp, & S. R. Goldman (Eds.), *The construction of mental representation during reading* (pp. 99–122). Lawrence Erlbaum Associates Publishers。

Plakans, L. (2009). Discourse synthesis in the integrated second-language writing assessment. *Language Testing, 26*(4), 561–587. https://doi.org/10.1177/0265532209340192

Primor, L., & Katzir, T. (2018). Measuring multiple-text integration: A review. *Frontiers in Psychology, 9*(2294), 1–16. https://doi.org/10.3389/fpsyg.2018.02294

Primor, L., Yeari, M., & Katzir, T. (2021). Choosing the right question: Effect of different question types on multiple-text integration. *Reading and Writing, 34*(6), 1539–1567. https://doi.org/10.1007/s11145-021-10127-8

Roediger, H. L., III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning Memory and Cognition, 31*(5), 1155-1159. https://doi.org/10.1037/0278-7393.31.5.1155

Rosvold, K. & Seki, M., (2015). *TEAP kouryaku mondai shu* [Preparation book for TEAP]. Kyogakusya.

Rouet, J. F., Saux, G., Ros, C., Stadtler, M., Vibert, N., & Britt, M. A. (2021). Inside document models: The role of source attributes in readers' integration of multiple text content. *Discourse Processes, 58*(1), 60–79. https://doi.org/10.1080/0163853X.2020.1750246

Schedl, M. (2010). *Background and goals of the TOEIC listening and reading test redesign project*. TOEIC Compendium. https://www.ets.org/Media/Research/pdf/TC-10-02.pdf

Schedl, M., O'Reilly, T., Grabe, W., & Schoonen, R. (2021). Assessment of academic reading. In X. Xi & J. M. Norris (Eds.), *Assessing academic English for higher education admissions* (pp. 22–60). Routledge. https://doi.org/10.4324/9781351142403-2

Sharpe, D. (2015). Chi-square test is statistically significant: Now what? *Practical Assessment, Research, and Evaluation, 20*(1), 1–10. https://doi.org/10.7275/tbfa-x148

Shimizu, H. (2010). Eibun dokkai ni okeru tekusuto kan jouhou tougou nouryoku no kennsho [Verification of the ability to integrate intertextual information into English reading comprehension]. *STEP Bulletin*, *22*, 47–62. https://www.eiken.or.jp/center_for_research/pdf/bulletin/vol22/vol_22_p47-p62.pdf

Stadtler, M., Scharrer, L., & Bromme, R. (2020). How relevance affects understanding of conflicts between multiple documents: An eye-tracking study. *Reading Research Quarterly, 55*(4), 625–641. https://doi.org/10.1002/rrq.282

van den Broek, P., & Kendeou, P. (2015). Building coherence in web-based and other non-traditional reading environments: Cognitive opportunities and challenges. In R. J. Spiro, M. DeSchryver, M. S. Hagerman, P. M. Morsink, & P. Thompson (Eds.), *Reading at a crossroads? Disjunctures and continuities in current conceptions and practices* (pp. 104–114). Routledge.

Weir, C. J., Vidaković, I., & Galaczi, E. D. (2013). *Measured constructs: A history of Cambridge English examinations, 1913–2012*. Cambridge University Press.

Wineburg, S. S. (1991). Historical problem solving: A study of cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*(1), 73–87. https://doi.org/10.1037/0022-0663.83.1.73

Publisher's Note