

IMPLICATING THE INTEGRATED FORMAT ON READING TEST ASSESSMENT: AN EVALUATION OF RELEVANT FACTORS

Trinh Ngoc Thanh*

HCMC University of Technology and Education

Received: 27/02/2020; Revised: 18/03/2020; Accepted: 28/04/2020

Abstract: The present study in general evaluates relevant factors involved in assessing reading performance in the context of English classroom teaching. Unlike the traditional way of placing the reading text and test items separately in the split format, the test items were placed in accordance with relevant parts of the reading text in the integrated format and this study compares reading performance on the ground of the following variables: test formats, study subject, learning task, and pre-proficiency level. Findings from the study firstly indicated the influence of test formats on reading performance from the evidence that participants in the integrated format performed better than those in the split format. Second, findings also showed significant effect of the interaction between test formats and learning task as well as the interaction between pre-proficiency level and study subject in reading test performance. Third, in combination with test formats, task design also had an influence on reading test performance.

Keywords: Integrated format, reading assessment, reading performance, split-attention

1. Introduction

The present study reconsiders the extent to which test formats contribute to the reduction of cognitive loads in language learning. From this consideration, the purpose of the study is to evaluate relevant factors in reading test assessment. More specifically, this study examines the scope of factors including study subject, learning task, and pre-proficiency level under the manipulation of the integrated test format. Under the setting of these factors, three research questions for the present study are as follows:

1. What are the combined effects of test formats, study subject, and learning task on test score performance?
2. What are the combined effects of pre-proficiency level, test formats, study subjects, and learning task on test score performance?
3. What is the single effect of learning task and of test formats on task performance?

2. Theoretical framework

This section reviews major points on cognitive load theory and its application in learning. Depending on the amount of cognitive demands for the learning task, the component of elements interactivity can be classified into the low element and high element interactivity where the informational complexity determines how much effort should be gained from the learners in the learning of the materials (Paas, Rekl, & Sweller, 2004). In a further elaboration

* Email: thanhtn@hcmute.edu.vn

from Sweller (1994), while low element interactivity may work best in the simultaneous learning model where the reference of other elements is not necessary in the structure of the learning task, the high element interactivity occurs from the connectedness of simultaneous learning and thus requires a higher order of organizing abstract elements in the promotion for successful learning task.

A crucial factor in explaining the constitution of low-and high-element interactivity in the success of the learning task is the concept of cognitive load. Schnotz and Kürschne (2007) mentioned a general division of cognitive load into intrinsic and extraneous cognitive load. The intrinsic cognitive load refers to the type of cognitive load which is inherent from the learning material itself and thus imposes on the capacity of working memory in the acquisition of knowledge from different resources of information. Meanwhile, the type of instruction given to the learning task also plays a role in the knowledge acquisition; and thus, when the information is excessively delivered in the learning task, the excessive information in turn results becomes the by-product of extraneous or ineffective cognitive load (Paas et al., 2004).

Therefore, to eliminate the split-attention effect in the processing of learning material, the usage of worked examples is more often hypothesized as a means to initiate the mental integration of relevant referents only when the meaningful learning information is embedded in the representation of knowledge (Chandler & Sweller, 1991). Another hypothesis related to the split-attention effect is whether the physical integration in the integrated format can prevent the split-attention effect from emerging in the interaction with the learning material; and as being noted in the integrated format the material designers should closely align the combination of two different sources of information with each other (Sweller, Ayres & Kalyuga, 2011).

There are major arguments regarding cognitive load theory. First, according to Sweller (1994), the major concern in cognitive load theory is given for the reduction of extraneous cognitive load in the design for the learning tasks. Second, the reduction of extraneous cognitive in return goes along the concern that learners have certain constraints in their cognitive system and thus the instruction should appropriately adapted to amend the gap between the cognitive levels and the instructional design (Schnotz & Kürschne, 2007). Third, an implication of cognitive load theory in teaching and learning is reserved for resolving complex learning tasks via means of simplifying information elements so that the cognitive load from long-term memory is sufficient for information retrieval and knowledge storage.

3. Methods

3.1. Reading task design

The present study follows the reading task design in Huynh's (2015) study in which the reading assessment was taken in classroom context and the test formats were divided into the split and integrated form. While the split format contained the separation between the reading text and the set of reading questions, the reading questions were inserted in accordance with the relevant text in the integrated format. In the selection of reading text, an online free-access article entitled "Robin Hood: Fact or Fiction" from Linguapress publisher was chosen as the main text for the present study. Within the text length of approximately 530 words, the text

selection design was suitable for a classroom reading assessment mini-test in 20 minutes. Details of the three reading tasks of the mini-test were as in Table 1.

3.2. Participants

The present study collected data from 84 English-major students at tertiary level across four study subjects: Reading 1 (n=17), Reading 2 (n=22), Reading 4 (n=21), and British-American Civilization (n=24). While the Reading courses had the focus on developing reading skills for students from Year 1 to Year 2, the British-American Civilization was a specialized course in historical-political-cultural knowledge about the United Kingdom and the United States. According to scores of pre-tests and previous courses, the English proficiency level of participants was ranged from pre-intermediate to upper-intermediate.

It is noted that the pre-test score (from the pilot test) was only applicable for participants in Reading 1 because there was no previous marking record of reading courses for this group of participants. Furthermore, among the four groups, Reading 1 and British-American Civilization were experimental group in which they were assigned a recording exercise with a set of comprehension questions regarding the content of the reading text. Due to the course arrangement over two semesters, the time span of collected data was from November 2015 to June 2016.

Table 1. Criterion-referenced design of the classroom reading mini-test

Task	PET reading test structure	Classroom reading mini-test (IELTS test format)	Description
Task 1	Multiple choice questions (part 4)	Multiple choice questions	This task involves the selection of the correct answer among four choices (A, B, C, or D). The classroom reading assessment test similarly adapts test item format in PET and IELTS reading tests.
Task 2	True-False (part 3)	True-False-Not Given	This task involves the main activity of validating the accuracy of the given statements. PET reading test only requires the performance of reading the given statement and defining True if the statement agrees with the information and False if the statement contradicts the information. Adapting IELTS test format in the classroom reading assessment test, the value Not Given is defined when there is no information derived from the text on this statement.
Task 3	Multiple choice cloze test (part 5)	Fill in the blank with no more than two words	This task requires test-takers to fill in the gap with the appropriate word or group of words. The multiple choice cloze test in PET requires test takers to choose the right answer among four choices (A, B, C, or D) on vocabulary and grammar items. In the adaptation of IELTS test format, the answer is limited within the length of two words and the gap-filled words should be appropriate and grammatical.

3.3. Description of variables

The raw data was recorded in Excel 2016 and converted into a csv file before it was processed by R software (version 3.6.1). Table 2 below presents the description of variables in the csv file:

Table 2. Description of variables in the study

Variables	Description
class	the study subjects: Reading 1(read1), Reading 2 (read2), Reading 4 (read4), British-American Civilization (civil)
class_numeric	the study subjects transferred into numeric value: 1=read 1; 2=read 2; 3=read 4; and 4=civil
id	the last one or two digits of student identification number
ver	test formats (0=split; 1=integrated)
score	the total score of the mini-test (from 0 to 12)
pre-score	the score (converted to the 10-point scale) indicating the English proficiency level of participants from PET mock test (for participants in Reading 1) and final grade of previous courses of reading skill (the remaining study subjects)
condition	the unavailability (0=control) and availability (1=experimental) of the recording exercise before participants took part in the class mini-test in split and integrated format.

3.4. Quantitative analysis

With regards to the setting of the variables, the present study firstly analyzed the combined effects of relevant factors on test score performance using ANOVA analysis. More specifically, three-way ANOVA analyses were applied to analyze (1) the combined effects of test formats, study subject, and learning task on test score performance and (2) the combined effects of pre-proficiency level, test formats, study subjects, and learning task on test score performance in two models. Second, Welch two-sample t-test was applied to compare the mean scores of test performance under the division of factors with significant effects from ANOVA analyses.

4. Results

4.1. Descriptive statistics of test score performance

The following descriptive statistics firstly summarizes the number of participants allocated into test formats and study subjects, to be followed by the mean scores and its corresponding SD. The mean scores between two test formats overall indicate that participants in the integrated format scored higher than those in the split format:

Table 3. Descriptive statistics of test score performance

Test format x Study subjects

	civil	read1	read2	read4
0(split)	13	9	13	10
1(integrated)	11	8	9	11

Mean and SD

	Test format	Subject	Mean	SD
1	0	Civil	5.53	1.19
2	1	Civil	7.63	1.28
3	0	read1	5.88	1.36
4	1	read1	7.12	1.35
5	0	read2	6.15	1.57
6	1	read2	6.66	1.11
7	0	read4	5.90	1.79
8	1	read4	6.09	2.02

4.2. Analyzing the combined effects test formats, study subject, and learning task on test score performance

A three-way between subjects 2x4x2 ANOVA analysis was conducted to measure the effects of test formats (integrated and split formats; var. name: ver), study subjects (civil, read 1, read 2, read 4; var. name: class_numeric), and learning task (with and without recording exercise; var. name: condition) on test score performance. There was a significant effect of test formats on test score performance at the 0.01 level ($F(1, 76)=9.37, p=0.003$). There was also an interaction between test formats and learning task on test score performance and this interaction is statistically significant at the 0.05 level ($F(1, 76)=4.29, p=0.04$).

Table 4. Analysis of Variance for test formats, study subject, learning task, and test score performance

<i>Response: mydata\$test_score</i>						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
mydata\$ver	1	21.108	21.1077	9.37890	0.0030	**
mydata\$class_numeric	1	0.047	0.0473	0.0210	0.885074	
mydata\$condition	1	2.037	2.0373	0.9052	0.344399	
mydata\$ver:mydata\$class_numeric	1	2.203	2.2034	0.9790	0.325574	
mydata\$ver:mydata\$condition	1	9.671	9.6708	4.29710	0.041568*	
mydata\$class_numeric:mydata\$condition	1	1.683	1.6831	0.7479	0.389875	
mydata\$ver:mydata\$class_numeric:mydata\$condition	1	0.876	0.8756	0.3891	0.534665	
Residuals	76	171.042	2.2505			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.3. Analyzing the combined effects of pre-proficiency level, test formats, study subjects, and learning task on test score performance

The follow-up analysis on the combined effects of four factors on test score performance (var. name: test_score) of three groups of participants from Reading 2 (read 2), Reading 4 (read4), and British-American Civilization (civil) and was divided into two models. The first model involves the combined effects of pre-proficiency level (in the form of pre-test scores; var. name: pre_score), learning task (with and without recording exercise; var. name: condition), and test formats (integrated and split formats; var. name: ver).

A three-way ANOVA analysis was conducted to measure the effects of three factors on test score performance in the first model. Results of the test indicated no significant effects of pre-proficiency level ($F(1,59)=0.73, p>0.05$) and no combined effects in the interaction of pre-proficiency level with learning task ($F(1,59)=0.34, p>0.05$) as well as with test formats ($F(1,$

59)=0.02, $p>0.05$) at the 0.05 level. Apart from the significant effect of test formats on test score performance, the interaction between learning task and test formats was also significant at the 0.05 level ($F(1,59)=4.94, p=0.02$).

Table 5. Analysis of Variance Table of Model 1

<i>Response: mydata\$test_score</i>						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
mydata\$pre_score	1	1.806	1.8060	0.7397	0.39322	
mydata\$condition	1	1.003	1.0033	0.4110	0.52396	
mydata\$ver	1	13.935	13.9348	5.7078	0.02011	*
mydata\$pre_score:mydata\$condition	1	0.854	0.8538	0.3497	0.55654	
mydata\$pre_score:mydata\$ver	1	0.050	0.0504	0.0206	0.88627	
mydata\$condition:mydata\$ver	1	12.082	12.0820	4.9488	0.02995	*
mydata\$pre_score:mydata\$condition:mydata\$ver	1	0.258	0.2582	0.1058	0.74617	
Residuals	59	144.041	2.4414			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The second model involves the combined effects of pre-proficiency level (in the form of pre-test scores; var. name: pre_score), learning task (with and without recording exercise; var_name: condition), and study subjects (civil, read 2, read 4; var. name: class_numeric). The second three-way ANOVA indicated that in the second model the effect of pre-proficiency level on test score performances was not significant ($F(1,61)=0.69, p>0.05$) at the 0.05 level. Meanwhile, the interaction of pre-proficiency level with study subjects was significant at 0.05 level ($F(1,61)=4.29, p=0.04$).

Table 6. Analysis of Variance Table of Model 2

<i>Response: mydata\$test_score</i>						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
mydata\$pre_score	1	1.806	1.8060	0.6984	0.40658	
mydata\$condition	1	1.003	1.0033	0.3880	0.53568	
mydata\$class_numeric	1	1.172	1.1717	0.4531	0.50341	
mydata\$pre_score:mydata\$condition	1	1.204	1.2041	0.4656	0.49759	
mydata\$pre_score:mydata\$class_numeric	1	11.105	11.1047	4.2943	0.04247	*
Residuals	61	157.740	2.5859			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

4.4. Analyzing the single effect of learning task and of test formats on task performance

Previous findings showed significant effect in the interaction between pre-condition and test formats on test score performances. First, the analysis of task performance based on the single effect of learning task was conducted by means of comparing the differences in score performance of task between two conditions: with recording exercise (experimental; $n=41$; class_numeric: read 1 and civil) and without recording (control; $n=43$; class_numeric: read 2 and read 4). The task performance was analyzed using Welch two-sample t-test in two modes: single and dual task performance. Results of Welch two-sample t-test in the first analysis indicated no significant effect of learning task on task performance at the 0.05 level even though participants in the experimental group scored higher than those in the control group (except for the performance of task 1 and in the pair of task 1 3).

Table 7. The single effect of learning task on task performance

	Mcontrol	M experimental	t	df	p-value
Task 1	2.13	2.09	0.206	79.113	0.83
Task 2	1.697	1.878	-0.7675	81.871	0.445
Task 3	2.348	2.512	-0.85	81.04	0.397
Task 1 2	3.837	3.975	-0.436	81.49	0.663
Task 1 3	4.48	4.39	0.38	81.7	0.70
Task 2 3	4.04	4.60	-1.91	81.802	0.059

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Second, the analysis of task performance based on the single effect of test formats was conducted by means of comparing the differences in score performance of task between two formats: split (n=45) and integrated (n=39). On similar analysis using Welch two-sample t-test, participants in the integrated format scored higher than those in the split format. However, significant effect of test formats on task performance was found in the performance of task 1 ($t=-2.51$, $df=77.9$, $p=0.01$), the performance of dual task 1 2 ($t=-2.94$, $df=76.99$, $p=0.004$), and the performance of dual task 1 3 ($t=-2.63$, $df=75.89$, $p=0.01$) at the 0.05 level.

Table 8. The single effect of test formats on task performance

	M Split	M integrated	t	df	p-value
Task 1	1.88	2.38	-2.51	77.9	0.01*
Task 2	1.6	2.0	-1.72	80.77	0.08
Task 3	2.37	2.48	-0.559	76.597	0.577
Task 1 2	3.48	4.38	-2.94	76.99	0.004*
Task 1 3	4.13	4.79	-2.638	75.89	0.01*
Task 2 3	4.11	4.56	-1.52	81.072	0.13

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5. Discussion and implications

The purpose of the present study is to investigate the effects of test formats, study subjects, learning task, and pre-proficiency level on test score performance in the evaluation of combined effects and single effect. Findings of the study firstly revealed that test formats have a significant effect on test score performances, considering that participants in the integrated format performed better than those in the split format. Despite being different in research scope, while investigating the influence of material design on geometry test performance, Tindall-Ford et al. (2015) similarly found that participants in the integrated format earned a higher mean score than those in the split format. In a similar concern with their study, it could be the employment of self-management strategy by learners as participants while making their efforts in reducing the split-attention effect; in addition, participants in the present study may also understand how to respond the test items by integrating the questions with the relevant text.

Furthermore, the present study also showed that there was a significant effect in the interaction between test formats and learning task, and this finding can be linked with the role of worked examples in lowering cognitive load for the purpose of relating isolated information from the reading material. In another study on the scope of material designing for an introductory accounting course, Sithole (2017) also indicated participants showed a better

performance in the integrated format for the recall and transfer tests and suggested the usage of integrated textbook as an effective worked example for learning.

Regarding the usage of recording task prior to the participation of the reading performance test for the experimental group, the present study refers to Pouw et al.'s (2019) study on the cognitive basis for split-attention effects to suggest the implementation of recording exercise as the means to lower the extraneous cognitive load for the purpose of maintaining and searching information in working memory. In other words, the recording exercise as the learning task in this study may reduce the amount of time for memory traces while learners perform the reading task in the assessment phase in classroom.

In the present study, the evaluation of single effect for learning task in the treatment of recording exercise presented that there was no significant difference in the mean score between split and integrated formats. In correspondence with task requirements from the design of MCQs in split and integrated formats, this finding further raises practical concerns about how individual differences differentiate their reading performance in MCQs, considering the impact of MCQs task in magnifying reading self-efficacy (Solheim, 2011).

6. Conclusion

The following study is conducted to analyze relevant factors which are likely to affect reading test performance in the context of language classroom. Under the main involvement of the split and integrated test formats, findings from the study generally support the reduction of split-attention effect from the implication of integrating relevant texts with their relevant questions in the reading task. The study also sheds lights on contributors to reading test performance, considering that these contributors are likely to promote and demote the process of reading comprehension.

There are some suggestions for future research investigating similar scope. In manipulating the condition which can reduce the split-attention effects in reading assessment in the context of language classroom, future studies can refer to the integration of hypertext glosses as the technological aids of explaining word meanings in reading assessment (Chen, 2014). The assessment of reading comprehension in different formats can also be adapted to longitudinal design of extensive reading in the evaluation of its effectiveness in the development of reading proficiency and related variables (Jeon & Day, 2016)

References

- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293-332.
- Chen, I.J. (2014). Hypertext glosses for foreign language reading comprehension and vocabulary acquisition: Effects of assessment methods. *Computer Assisted Language Learning*, 29(2), 413-426.
- Huynh, C.M.H. (2015). Split-attention in reading comprehension: A case of English as a foreign/second language. *6th International Conference on TESOL* (pp. 1-12). SEAMEO. Ho Chi Minh City, Vietnam.
- Jeon, E.Y., & Day, R.R. (2016). The effectiveness of ER on reading proficiency: A meta-analysis. *Reading in a Foreign Language*, 28(2), 246-265.
- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, 32(1), 1-8.

Pouw, W., Rop, G., De Koning, B., & Paas, F. (2019). The cognitive basis for the split-attention effect. *Journal of Experimental Psychology*, 148(11), 2058-2075.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, 19(4), 469-508.

Sithole, S.T. (2017). Enhancing students understanding of introductory accounting by integrating split-attention instructional material. *Accounting Research Journal*, 30(3), 283-300.

Solheim, O.J. (2011). The impact of reading self-efficacy and task value on reading comprehension scores in different item formats. *Reading Psychology*, 32(1), 1-27.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295-312.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.

Tindall-Ford, S., Agostinho, S., Bokosmaty, S., Paas, F., & Chandler, P. (2015). Computer-based learning of geometry from integrated and split-attention worked examples: The power of self-management. *Journal of Educational Technology & Society*, 18(4), 89-99.

NGHIÊN CỨU CÁC YẾU TỐ TRONG THỰC NGHIỆM DẠNG ĐỀ TÍCH HỢP VÀO ĐÁNH GIÁ ĐỌC HIỂU TIẾNG ANH

Tóm tắt: Nghiên cứu thực nghiệm này được thực hiện nhằm định trị các nhân tố trong việc đánh giá kỹ năng đọc hiểu tiếng Anh ở môi trường lớp học. Dựa trên sự so sánh kết quả đánh giá đọc hiểu giữa hai dạng đề: dạng đề không phân vùng (split format) và dạng đề tích hợp (integrated format), nghiên cứu phân tích các nhân tố bao gồm dạng đề, môn học liên quan, bài tập tìm hiểu, và trình độ đọc hiểu. Kết quả thực nghiệm cho thấy người tham gia làm dạng đề tích hợp có kết quả cao hơn dạng đề không phân vùng. Bên cạnh đó, có mối liên hệ trong sự tương tác giữa dạng đề và bài tập tìm hiểu cũng như là giữa trình độ đọc hiểu và môn học liên quan. Nghiên cứu cũng chỉ ra cùng với dạng đề, việc thiết kế dạng bài đánh giá cũng có ảnh hưởng tới kết quả đánh giá.

Từ khóa: Split-attention, integrated format, reading assessment, reading performance