

Article

The Impact of Scenario-based Settings on Cognitive Reading-to-write Processes

Yumei Zhang

Beijing International Studies University, China

Shaoqian Luo*

Beijing Normal University, China

Abstract

In Scenario-based assessment (SBA), a real-life scenario is created for students to do reading and writing tasks in a computer-mediated environment. It has been recognized that task authenticity will help strengthen assessment validity and task settings and characteristics will influence cognitive efforts. This study attempts to explore the underlying reading-to-write processes in an SBA setting by contrast with those in a non-SBA setting, and thereby investigate the impact of task settings on cognitive processes.

Four participants were invited to complete three forms of integrated reading-writing tasks in different settings and then recall their thinking processes stimulated by screen capture recordings and their writing outcomes. The data were analyzed through parsing, encoding, and classifying.

The results showed that there were differences between the processes elicited from the SBA setting and the non-SBA setting in the saliency and frequencies of some process categories, and in the ways the source texts were used. Comparatively, the processes summarized in the SBA setting were more congruent with those summarized in theoretical frameworks of integrated reading-writing processes and findings from previous studies in other contexts. The SBA setting not only encouraged a more comprehensive understanding of the multiple texts provided but also found the participants more active in making connections between new and prior ideas rather than scanning for supporting piecemeal information and patch-making. The findings affirmed the effectiveness of the authentic scenario created by computer technology, which provides insights into the constructs and validity of the assessment.

Keywords

Scenario-based assessment, computer-mediated language tasks, task setting, reading-to-write processes

***Corresponding Author**

Address: No. 19 Xin Jie Kou Wai Street, Haidian District, Beijing, 100875, China.

Email: sqluasheila@bnu.edu.cn

1 Introduction

The development of information and communication technology as well as digital devices has offered us a new channel to obtain information and to communicate. Correspondingly, reading and writing in the cyber community have become inevitable for students in their academic and future vocational life.

For education, it is necessary to reflect the real-world requirements and prepare students for future challenges. In terms of educational assessment, however, it has been critically reviewed that contemporary assessments are not keeping pace with the changing contextual needs and the nature of digital literacy (van den Broek, 2012). Usually, reading comprehension tests are prevalently comprised of isolated passages followed by several questions on paper-based forms (Rupp, Ferne, & Choi, 2006; van den Broek, 2012), and writing performance is traditionally elicited by a prompt (Plakans, 2007). Furthermore, it is noted that there is a lack of attention to cross-text integration (Britt & Rouet, 2011; Strømsø, Bråten & Britt, 2010) and that the item questions are relatively independent and do not map onto a real-world purpose (Rupp, Ferne, & Choi, 2006). Therefore, a new generation of literacy assessments is called for (Partnership for 21st Century Skills, 2008) and is expected to reflect the characteristics of new literacy scenarios.

Recognizing the contextual change and the opportunity brought by technology development, researchers at Educational Testing Service (hereafter ETS) proposed an innovative approach to literacy assessment – scenario-based assessment (hereafter SBA) (Bennett, 2010; O'Reilly & Sabatini, 2013; Sabatini et al., 2014a; 2014b). SBA was lately adopted by the Organization for Economic Co-operation and Development (hereafter OECD) in PISA reading assessment in 2018 (OECD, 2019).

A growing body of research has been conducted to investigate the plausibility of SBAs used in literacy assessments in different contexts (e.g., Deane et al., 2018; Deane & Song, 2014; O'Reilly et al., 2014; Sabatini et al., 2019; van Rijn, Graf, & Deane, 2014; Zhang et al. 2019). Generally, assertive claims were made about the reliability and other psychometric properties of the scenario-based reading or reading-writing assessments. However, most of the studies were conducted by researchers at ETS in the high school context in the United States. Moreover, test-takers' cognitive processes have not been sufficiently understood while completing the literacy tasks with the scenario-based features. As it has been acknowledged, task features will influence how test-takers approach the tasks and the difficulty of the tasks (Bachman & Palmer, 1996; 2010). More importantly, exploration on the underlying processes can aid in evaluating the cognitive validity (Field, 2013) and the demonstration of cognitive validity implies that the task could assess on the same range of cognitive operations or skills as those required in the target domain (Bax, 2013). Similarly, it is indicated in Kane's (1992; 2006; 2013) argument-based framework when the extrapolation inference is examined.

This study is part of a larger research project that aims to validate the computer-mediated SBA developed to assess Chinese EFL learners' reading and writing competence. More specifically, this present study attempts to explore the impact of the task settings (SBA vs. non-SBA setting) on the cognitive reading-to-write processes. Through the comparison, it could shed light on the validity of the design of the SBA setting under Kane's (1992; 2006; 2013) argument-based validation framework. In this framework, the investigation into cognitive processes is to examine the extrapolation inference.

2 Literature Review

2.1 Scenario-based assessment

Generally, SBA has been defined as a methodology or an approach to designing and delivering more authentic assessment tasks with the assistance of computer technology (Bennett, Deane, & van Rijn, 2016; OECD, 2019; Sabatini et al., 2019). Sabatini et al. (2019) from ETS define SBA as a cluster of

techniques or assessment design methodology that creates a real-life scenario under which a series of thematically related sources and assessment tasks and items are sequenced and organized. Purpura (2016) explains that it is not a construct *per se* but a way of delivering assessment tasks in a scenario narrative. Similarly, in PISA 2018 framework for assessing reading literacy, OECD (2019) introduced SBA as an assessment approach to provide an overarching purpose that leads students to read a collection of related texts and to complete the task.

In terms of writing from sources, the scenario-based design presents students with a real-life topical scenario or context along with a communication goal of writing. Then they will be encouraged to search for information online and read some articles so that their ideas put into writing will be more informed and better-structured. Usually, there are virtual characters (e.g., a tutor or peers) communicating with the test-takers through an online platform, including introducing the task, asking lead-in questions, offering information and guidance throughout the process (Deane et al., 2018; Guo et al., 2020).

SBA tasks designed for different contexts may vary in the specific constructs, item formats, etc., but usually carry certain features. The typical features of a computer-based SBA task mainly include (Boveri, 2017; O'Reilly & Sabatini, 2013; Sabatini et al., 2019):

- (1) design of a real-life scenario that introduces an overarching goal;
- (2) a sequence of materials and tasks that lead up to an end goal;
- (3) social interactions or collaborations with virtual characters;
- (4) a narrative storyline that guides, motivates, and engages students in the process.

These features represent largely the differences between an SBA design and a non-SBA design. The birth of SBA is a reaction to the changes of the literacy practices in the digital environment and enabled by the advancement of computer technology. It aims to deliver an assessment that is more relevant to real-world challenges and it is designed in light of cognitive theories of discourse process (Magliano, et al., 2018; Sabatini et al., 2019), learning theories (Ercikan & Pellegrino, 2017), and measurement theories to ensure its validity and usefulness (Sabatini et al., 2019).

This paper focuses on the cognitive aspect and attempts to gain insight into what processing activities are activated inside the black box (the test-takers' mind) through stimulated recalls. Cognitive processes in the scenario-based reading-to-write setting and a non-SBA setting are compared, as it is very likely that the setting or features of the language use tasks will exert an impact on students' cognitive activities (Bachman & Palmer, 2010; Lei, 2008; Weigle, 2002). The comparison could also help identify which setting could activate cognitive processes that are more congruent to theoretical models or the intention of the assessment tasks. In this way, evidence for the validity of the SBA tasks could be collected.

2.2 Context and cognition

Writing has been frequently analyzed through a cognitive perspective (Weigle, 2002) and there are multiple models portraying writers' cognitive efforts during the writing processes. One of the most cited models includes planning, translating, and reviewing (Hayes & Flower, 1980). This process is regarded non-linear but generative, during which the writers make online planning, reformatting, and revising back and forth (Zamel, 1983; Lei, 2008).

Similarly, integrated reading and writing are also depicted as a constructive process (e.g., Flower, 1987; Spivey, 1984; 1990; 1997; Spivey & King, 1989). Reading is usually conceptualized as support to writing and as resource input for writers (Hayes, 2012; Grabe & Zhang, 2016). However, the importance of reading has become prominent, as reading comprehension can be as problematic as writing for EFL learners (Grabe & Zhang, 2016). The extent to which they have understood the reading texts is related to how they would make effective use of the input in their own writing, further influencing their writing performance. Therefore, apart from the writing product, research efforts have been made on the cognitive processes of the reading-to-write activity (e.g., Delaney, 2008; Plakans, 2008).

The scrutinization of cognitive processes is essential to understand the task and students' skills or abilities (Bachman, 2004) in reading and writing. But it should be noted that reading and writing are social activities in the real world (Hayes, 1996; Weigle, 2002), which entails that one's thinking activities will be very likely shaped by the social and cultural context (Weigle, 2002). Writers write for a certain purpose in a social setting to a certain group of the target audience. Therefore, it is necessary to analyze cognition in reference to the goals by which the participants are motivated and the structure or context in which the task is embedded (Cumming, Busch, & Zhou, 2002; Lei, 2008).

Some influencing factors have been encompassed in some cognitive frameworks, such as socio-cultural conventions, task purpose, and conditions (e.g., Cumming, Busch, & Zhou, 2002; Hayes & Flower, 1980; Hayes, 1996; 2012). In addition, guidelines for developing language assessment also suggest considerations be made on the impact of task context and features on students' performance (e.g., Bachman & Palmer, 1996; 2010). For instance, the contextual factors identified include task characteristics and situation or setting (Bachman & Palmer, 2010).

With the appliance of technology in language assessment, the features and context of language assessments are changing. The initiation of the SBA was to take advantage of computer technology to situate students in a task setting with some features of real-life communication in the cyber community. It has been attracting discussions on how the use of computer technology could impact the construct of language assessment and on students' performance (Chapelle & Voss, 2017).

2.3 The use of computer technology in language assessment

Technology has offered many opportunities for language assessment, and it is promising to allow for more possibilities in the future. Meanwhile, there are critical voices that encourage further research on the impact of technology use in language assessment. To date, research has been focusing on the efficiency and authenticity of language assessment through computer technology (Chapelle & Douglas, 2006).

The primary impetus for utilizing technology in assessment is to improve the efficiency of assessment practice (Chapelle & Voss, 2017). It could make it more efficient to collect response data, analyze the data, and provide impromptu feedback to test-takers.

Second, technology could enhance the authenticity and contextualization of language assessment (Alderson, 2000; Chapelle & Douglas, 2006). It is assumed that the authentic context for language use will help collect students' performance data that are more reflective of their performance in the target domain, thus strengthening the validity of the assessment tasks (Bachman & Palmer, 1996; Purpura, 2016). On this basis, another question being asked is that whether communicative task-based language testing is more operationalizable by computer (Corbel, 1993).

There are many other instances to support that computer technology could make language assessment more efficient and authentic. But no matter what, one fundamental issue concerning the utilization of technology is what impacts it may have on students' performance. Researchers have been making efforts in this regard, but most studies focus on the comparability of students' performance in computer-based and paper-based settings (e.g., Choi, Kim, & Boo, 2003; Schaeffer, et al., 1993; Taylor, et al., 1999). One concern underlying this line of research is that students' familiarity with doing a test on a computer might be irrelevant to the construct of a language assessment. But these studies found very few and slight differences that could warrant the claim that the settings of the tasks would affect students' performance. However, the impact of task setting is not only about the physical settings of test delivery, but also concerns contextual features or characteristics of the tasks and how the tasks are presented to test-takers (Chapelle & Douglas, 2006).

Moreover, when computer technology is used for language assessment, its function is not merely to transform the tasks from a paper to a computer. As Chapelle and Douglas (2006) commented, our goal of

using technology in assessment probably should not be limited to the pursuit of efficiency when it could be helpful to create better assessments. For instance, technology makes it possible to design more real-life and interactive scenarios and thereby facilitate understandings of students' performance in the target domain. Therefore, the current study is to investigate students' cognitive processes of language tasks in two computer-mediated settings, one with the real-life features of online communication afforded by computer technology and the other one without the features. The study could advance our understanding of the impact of technology on language assessment and students' performance.

3 Research Design

3.1 Research questions

As the research purpose purports, this study mainly attempts to answer the following question: To what extent do the task settings (SBA and non-SBA setting) impact the test-takers' cognitive reading-to-write processes?

3.2 Participants

There were four participants. Two of them were 11th graders at a senior high school in Beijing, and the other two were English majors in a key university in Beijing. To note, as the data were collected at the first term of the academic year in early November, the two groups of students had just embarked on their new period of academic life. The two newly enrolled college students were involved as approximate 12th graders in high school, given that it was infeasible to recruit the 12th graders who were engaged in preparing for the college entrance examination.

The two high school students (pseudonyms: Leo and Judy) were rated at the upper-intermediate level of English proficiency by their English teacher. Their English scores at the last final examination were 124 and 118 (out of 150) respectively. The two college students (pseudonym: Linda and John) received high English scores in the college entrance examination. Linda got 145 and John got 139 out of 150. In the study, one 11th grader (Judy) and one college freshman (John) were introduced to the SBA setting, while the other two (Leo from 11th grade and Linda from college) were involved in the non-SBA setting.

The four participants said that they used computers frequently for surfing online, searching for information and writing their homework, etc. They thought the internet and computers were indispensable to their social and academic lives. All of them had taken online tests once and claimed they had no problem using a computer to read or write.

3.3 The tasks

There are three paralleled forms of tasks for both settings and the difference among the different forms is mainly the topical issue being discussed. The topics are (1) should students be allowed to take phones to schools (Phones to Schools), (2) should plastic bags be banned (Plastic Ban), and (3) should we keep being ourselves or change to fit in (Be Self or Fit in). There are three passages in each form and the writing task is to write an opinion essay about the given topic.

The difference between the tasks used in the SBA setting and the non-SBA setting mainly lies in whether they carry the SBA features or not. For the non-SBA setting, the three passages in each form and the writing tasks are equivalent to those in the SBA setting. But the passages and writing tasks are presented to the participants one by one without any connections in between or further instructions besides "write an opinion essay about the following topic and there are three passages you can read before you write". In contrast, in the SBA setting, there is an overarching goal which is that a magazine is

calling for opinions about a controversial issue. A virtual character (the tutor from an online reading-and-writing club) posts the call-for-opinion poster to the club members and encourages them to write one. The tutor and another senior student (Alice) provide instructions, narrative guidelines, and scaffoldings throughout the task. Figure 1 presents the first webpage of the SBA task. The prompts given by the virtual tutor and senior student are summarized and presented in Table 1. They are numbered according to the order of occurrence.

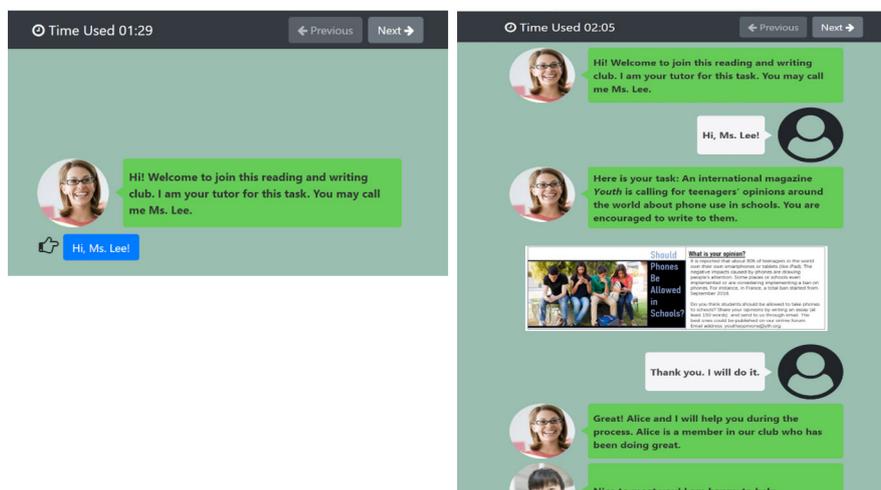


Figure 1. The first webpage of SBA

Table 1

Prompts from the Virtual Tutor and Senior Student through Social Media

Tutor (Ms. Lee):	Senior student (Alice):
1. welcome	4. saying hi
2. posting and introducing the task (call for essays)	9. a brief comment on the articles presented
3. introducing the senior student, Alice	10. reminding students that “every coin has two sides” and recommending another article
5. encouraging students to make a list of the pros and cons of the given issue	11. suggesting students sort out the advantages and disadvantages they have learned in the articles
6. suggesting searching for more information online about the given issue	
7. recommending one article	
8. inviting them to think about the comprehension questions which are key to understand the article	
12. offering tips for writing an opinion essay	

3.4 Stimulated recalls

Stimulated recalls and think-alouds are introspective methods frequently used to elicit one’s thought processes that take place during an event (Gass & Mackey, 2016). In this study, stimulated recall is adopted on the grounds of practicality and research needs. Firstly, it is suggested that a think-aloud session should not last too long (2 hours maximum) to avoid any negative effect on the validity caused by participants’ fatigue (Guo, 2015). The SBA is challenged by this time restriction as it is composed of three passages and a writing task. Secondly, it has been questioned about the extent to which verbalizing

one's internal thoughts can truly represent the natural processes, no matter which method is used (Plakans, 2007; Stratman & Hamp-Lyons, 1994). More importantly, it is more of a concern with concurrent thinking aloud than with stimulated recalls, as recalls afterwards would not affect the ongoing cognitive processes (Bowles, 2010; Gass & Mackey, 2016). Third, though it is possible that the participant cannot remember all their thoughts in the retrospective recall, it is noted that the likelihood of a recall being accurate is greater if the time between the performance of the task and the recall is short enough for the participants to be able to retrieve their thought processes in their memory (Ericsson & Simon, 1996; Gass & Mackey, 2016).

Therefore, the consecutive recalls with no or little time interval between the event and the recall (Ericsson & Simon, 1996) were conducted in the present study. Moreover, the four participants' thoughts were accessed at three phases (1) after they first read the prompt and did some planning, (2) after they finished reading the three articles, and (3) after they completed the writing task. This was based on two considerations. One was that the participants might not be able to recall clearly what they were thinking at the previous moments as the task might take over one hour. The other consideration was that the researcher aimed to capture students' thoughts at each phase and to reveal the process of mental model construction as they proceeded in different settings.

In terms of the stimuli, a video recording is thought to be more effective than an audiotape while a video of eye-tracking could be too much for participants (Gass & Mackey, 2016). Besides, screen capture and mouse tracking device can also be used to stimulate recalls and triangulate verbal reports about mental processes (Castek & Coiro, 2015; Gass & Mackey, 2016; Ziegler & Mackey, 2014). In the current study, a screen capture software (ApowerREC) was used to assist students' recalls. It could capture students' operations on the screen, such as tracking their reading path (the participants are told to use a mouse to indicate roughly the area they are reading), the choice making on the multiple-choice questions, the page-turning, and typing process while answering the open-ended questions and writing. Compared to the eye-tracking technique which is often used to gain insights into test-takers' cognitive processes (e.g., Bax, 2013), this screen capture software is limited in its capability to present accurate data of eye movement and fixation duration, but it is able to keep the tracking going on across multiple screen displays.

About the language used in the stimulated recall protocol, as it is argued that the limitation of L2 proficiency could inhibit them expressing themselves, which confronts the stimulated recall data with more validity threats (Gass & Mackey, 2016), the participants in this study could use Chinese or switch to English anytime as long as they felt comfortable.

When it comes to the implementation of the stimulated recall, it is advised to have a detailed protocol as a checklist while carrying out the procedure (Gass & Mackey, 2016). With reference to the protocols used by Mackey et al. (2000) and Leeman (1999), a protocol was developed for the current study. The recall was conducted in the participants' L1 (Chinese) and the protocol was later translated into English and displayed in Appendix 1. The protocol also drew some insights from previous literature. For instance, as guided by Gass and Mackey (2016) and Lei (2008), interest was shown to the participants' thinking during pauses and revisions. In this reading-to-write study, special attention was also paid to their behaviors of referring to the reading session.

3.5 Data analysis: Coding and categorizing

The analysis of stimulated recall data was carried out from transcription, coding, and description to the analysis *per se* (Gass & Mackey, 2016). To begin with, the recorded recalls from the four participants were transcribed and coded manually by the researcher and double-checked by a fellow researcher.

The transcribed data were analyzed through parsing, encoding, and classifying as suggested by previous researchers working on verbal reports (e.g., Guo, 2015; Plakans, 2007). Parsing was to segment

the transcribed texts into chunks or idea units. Then the encoding and classification were carried out. First, the participants' reading-writing processes were coded and classified with reference to the processes specified by the previous theoretical frameworks or studies which were summarized by Spivey (organizing, selecting, and connecting) and Stein (monitoring, elaborating, structuring, and planning) (Spivey, 1990; 1997; Stein, 1990); Any other featured processes that could not be clearly categorized were also coded for further discussions. Second, during coding and categorizing, particular attention was paid to the participants' use of the source texts and their underlying thinking processes.

The goal of the assessment task was to write an opinion essay and reading played a supportive role. Therefore, the stimulated recall did not investigate the reading processes in so much detail as to the word recognition level. The primary interest was to explore how the participants read when they had a writing goal in mind. As a result, more comprehensive insights could be achieved about the writing processes, for the participants were asked to recall their thinking during the noticeable pauses (longer than 5 seconds) of writing.

4 Results

The processes which emerged at each phase of the task were categorized and summarized. Table 2 presents the frequencies of each category that occurred in each participant's recalls and his/her recorded behaviors on the three forms.

Firstly, the participants involved in the SBA and non-SBA settings all began with reading the prompts. They all reported that they had understood the setting and were aware that the goal of the task was to write an opinion essay and they could take a position upon a quick evaluation of the issue based on their existing knowledge. Their existing knowledge included general world knowledge, personal experiences, etc.

For instance, John supported banning plastic bags as he drew on his knowledge that plastics would take a long time to break down. Judy wanted to write against allowing students to take phones to schools because her experience with phones in the class was not beneficial or necessary. Besides, at this phase, John made an outline after reading the prompts that included his main argument and some keywords to indicate supporting arguments, while the other participants did not make an outline or write down anything at this phase.

At the second phase from where the two pairs of participants entered SBA and non-SBA settings respectively, more differences were pronounced. As it is shown, John and Judy in the SBA setting were more frequently engaged in reading the whole texts and connecting with their existing ideas while reading, as Judy reflected (originally in Chinese):

I read every passage. After reading every passage, I answered the comprehension questions. Sometimes I had to go back to the passage to find the answers or to think more about the question. After reading all the passages, I found that there was one passage mainly arguing for and one against allowing phones to schools. And I was thinking that I could use some ideas to support my own opinion. For example, too much screen time would affect children's cognitive development (Judy-Phones to Schools-Phase 2).

She even modified her perspective after reading the three texts in the third form of the SBA. She originally held the idea that people should change to fit in the society and then thought that "fitting in is not belonging" after reading the last passage, "it won't make you happy if you can't embrace the true self. The meaning of life is to find your own place in the world and make a difference" (Judy-Be Self or Fit in-Phase 2).

Similarly, John read all the texts and answered the comprehension questions. And he emphasized that:

while reading those different opinions, I realized that my original opinion was a little stubborn because what I knew was limited. It is not that simple to just find something to replace plastic bags, because I learned that other types of materials are not necessarily better (John-Plastic Ban-Phase 2).

Table 2
Summary of the Process Categories and Their Frequencies

Forms	John			Judy			Leo			Linda		
	1	2	3	1	2	3	1	2	3	1	2	3
Phase 1: Pre-reading and writing												
(a) read the prompt to understand task goal	1	1	1	1	1	1	1	1	1	1	1	1
(b) activate existing knowledge	1	1	1	1	1	1	1	1	1	1	1	1
(c) take a position	1	1	1	1	1	1	1	1	1	1	1	1
(d) make an outline	1	1	1									
Phase 2: Reading the source texts												
(a) read whole texts	1	1	1	1	1	1	1			1	1	
/skim for arguments								1	1			1
(b) connect with own ideas	3	2	2	2	2	1	2			2	2	
/seek supporting evidence								2	3			2
(c) evaluate ideas	1	1	1		1					1		
Phase 3: Writing (recall on the pauses)												
a. reread the task	1	1	1	1	1	1	1		1	1	1	1
b. make an outline	1	1	1	1	1					1	1	
c. organize local ideas	2	3	3	6	2	5	7	8	6	6	5	9
d. translate ideas into words	9	9	12	5	7	8	9	9	11	8	10	11
e. go back and reread certain parts of the source texts	3	4	2	4	4	3	5	6	5	6	8	5
f. Reread and evaluate what has been written to...												
f-1. develop/modify ideas	3	3	2	2	2	3	4	3	3	2	2	2
f-2. edit language	5	7	8	2	2	3	4	5	6	5	4	5
f-3. adjust cohesion	1	1	3		2	1	1				1	
No. of pauses during writing	23	27	31	19	21	23	30	31	31	27	30	32

In contrast, Leo and Linda in the non-SBA setting were more likely to skim for each author's positions and focus more on the articles that were congruent with their positions to seek more supporting ideas. But it was found that they also read the texts one by one when they were doing the first form of the assessment, and when they got familiar with the whole task procedure, they tended to skim and seek corresponding evidence for later use in writing. This tendency could be identified through the recordings and the participants' recall. While doing the third form of the assessment, Linda read the prompts and the recording showed that she read the passages quickly and spend more time reading the second passage. According to her recall, she thought,

People are social animals and they need to change to fit in society so that they can have a good environment for living. I knew that there were usually three passages following, and I can find some ideas from the passages. So, when I was reading, I was clear about what I wanted for writing..... Of course, I read about the different ideas as well. But quickly. Because my original opinion was really what I believed. What I needed was to find more evidence to support my idea (Linda-Be Self or Fit in-Phase 2).

It shows that the SBA participants were more engaged in reading the texts closely, probably because they had to answer the comprehension questions next to each text, while the non-SBA participants had the tendency to skim for the main arguments and scan for congruent ideas. Moreover, in the SBA setting, the participants who read more carefully seemed to be more aware of different perspectives, which motivated them to connect and evaluate on their own opinions, rather than directly seeking supporting ideas.

For the third phase, most of the participants read the prompts again but not all of them wrote down outlines or planned the content. It appeared in Table 1 that the SBA participants were more active in preparing and planning and it was found that they incorporated some ideas from source texts in their writing plans. For John who made an outline when he first read the first outline, he added some points taken from the source texts to his original plan and reorganized the structure. Comparatively, the participants in the non-SBA setting were less frequently found making plans before writing. Leo made no claim about making any outlines nor was he observed doing so, while Linda noted down her general plan at the first and second time and reported that she did not do so for the third form as she thought the topic was quite subjective and vague.

According to the participants' recalls on their thinking processes during composing especially during the pauses of typing, there were four main process categories summarized in Table 1, including generating and planning local ideas, translating ideas into words, rereading, selecting, and using information from source texts and rereading and evaluating their own writing. For the last one, the writers read what they had written to develop ideas further, to modify the language, and to check cohesiveness.

Firstly, no matter how much effort they had made for planning before writing, they would make online planning of ideas and translate them into words. However, the readiness of the ideas and language could be reflected by the fluency of typing. As it is indicated by the total frequency of pauses during writing, the two participants in the SBA setting paused less frequently than the other two participants during writing and produced relatively longer texts.

In terms of what they were processing during the pauses, all the four participants paused most frequently for translating their ideas into proper words. In their verbal reports, they recalled that:

I was thinking about how to express my opinion in one sentence (John-Phones to Schools-Phase 3);

I was thinking about using a more sophisticated syntax to say ... (Judy-Be Self or Fit In-Phase 3);

I was mainly considering which word to use, drawbacks, or weaknesses (Linda-Plastic Ban-Phase 3).

With respect to the source use, it was interesting to discover that Leo and Linda, went back to the source texts more often than John and Judy. Moreover, it was found that Linda tended to copy the original sentences and paste them into her own writing. She explained in the interview that she thought those sentences copied from the text were useful and fitted with what she wanted to say. Therefore, she used them in her writing and made some minor changes so that they could cohere with the preceding words grammatically. This behavior of patch-making in writing could also be seen in Leo's essays, but he used clauses or parts of the original sentences instead of the whole sentences. On the contrary, John and Judy seemed more able to integrate the information learned in the texts into their writing. This could be reflected in the fact that they went back to the texts less often but made more effective use of the information. Sometimes, they did not turn to the texts even when they were writing about the points

in the texts, which might indicate that the information had already been constructed into their existing mental frameworks of knowledge and their writing plans. The difference might also be attributable to the tip given by the virtual tutor in the SBA setting which warned students not to copy the authors' original words. This was warranted by the interviews with John and Judy.

Finally, the four participants frequently reread what they had written previously while writing. They read the paragraph just written to check the completeness of meaning-making and to develop further ideas. Similarly, they read the previous sentence or segment to make further development. More frequently, they read the previous parts and corrected language mistakes or modified the wording. Besides, there were occasions when they reorganized the structure after reading what they had written to make it more cohesive. For instance, John first planned to respond to the opposite perspective at the end, but he gave it a second thought before proceeding to the last paragraph. He moved to the beginning and added his response to the opposite opinion in the opening paragraph before declaring his main argument.

To sum up, the SBA and non-SBA were both reading-to-write tasks in nature and the results show that all the students went through three main phases: pre-reading/writing, reading, and writing. In terms of the sub-processes at each phase, there were similarities and differences between the participants involved in the two settings. The major difference lay in how they processed reading and source use while writing.

Additionally, through the recording and students' recalls, it was found that the participants in the SBA setting could learn more about the topical information, some vocabulary, and expressions. More importantly, when they had understood the texts comprehensively, they could shape more informed perspectives, modifying or even altering positions.

5 Discussion

The exploration of the impact of the SBA and non-SBA setting on the reading-to-write processes showed that the two groups were involved in similar processes at the pre-reading/writing phase, including understanding the task, activating topical knowledge, and taking a position. Differences were mainly detected while the participants were reading. The SBA tasks engaged the participants in more careful reading, probably because they were required to answer comprehension questions. Besides, the SBA participants were found more active in connecting the ideas learned through reading with their own ideas and evaluating the ideas. On the other hand, in the non-SBA setting, the participants were more likely to skim the texts and scan for supporting evidence while reading. The differences in the processes in reading may contribute to a conclusion that the SBA setting encouraged a closer and more comprehensive understanding of the multiple texts provided. This promoted the participants to make connections among new and prior ideas rather than simply selecting discrete information. Connecting was a process depicted in the theoretical framework proposed by Spivey (1990; 1997) and in Yang and Plakans' (2012) study.

During writing, the process categories summarized from both settings were largely in sync, including making outlines, planning local ideas, translating ideas into language, going back to the source texts, rereading, and evaluating their own writing, and making modifications in ideas, language, and organization. These processes largely corresponded with those portrayed in writing models (e.g., Hayes, 2012). But the two groups differed in frequencies of some processes. For instance, the non-SBA participants went back to the source texts and made local planning more frequently while writing. This might imply that the SBA could enhance the efficiency of reading and writing. This could also be indicated by fewer typing pauses and less frequent returning to the source texts during writing.

As stated earlier, the exploration of the underlying processes students perform during the tasks could offer important insights into the validity of the assessment, and it is crucial to examine the plausibility of the extrapolation inference in Kane's validation framework (Kane, 1992; 2006; 2013). The extrapolation inference suggests how the tasks could be used to elicit students' performance in the real-

life target domain. The results from the stimulated recalls strengthened the claim for the extrapolation inference as the processes involved in the SBA tasks were largely congruent with those summarized in theoretical frameworks of integrated reading-writing processes and the findings from previous studies in other contexts (Flower, 1987; Hayes, 1996; 2012; Plakans, 2007; Spivey, 1990; 1997; Stein, 1990). Like Plakans' (2008) findings, the writers in the SBA setting went through the processes that included understanding the task, taking a position, planning content, writing, and reading for evaluation. It is discovered that reading-to-write tasks involve multifaceted KSAs, including knowledge, cognitive and metacognitive skills (Yang & Plakans, 2012), which reflects the demands of real-world tasks.

Furthermore, it has been acknowledged in the literature that reading and writing are constructive processes (Flower, 1987; Spivey, 1990; 1997; Yang & Plakans, 2012). Through observing the participants' behaviors during the task and their recalls, this study found that the four participants all made attempts to use the information in the source texts to construct their writing. But those in the SBA setting were more likely to make effective use of the source texts for meaning construction. Their opinions became clearer and more informed after reading.

In terms of technology in language assessment, it was introduced earlier that technology could enhance the authenticity and contextualization of language assessment (Alderson, 2000; Chapelle & Douglas, 2006). SBA was initiated with the acknowledgment that the information and communication technology has been changing the context of reading and writing (OECD, 2019; Sabatini, et al., 2014a). In return, it makes use of computer technology to create such scenarios for students to perform writing tasks to exchange opinions in the cyber community where they have access to reading materials and external guidance or interactions with others. The SBA setting with these authentic features is assumed to have some impact on students' reading-to-write performance. And it is hypothesized that students' performance in a more authentic assessment will better reflect their performance in the real world and thus enhance the validity of the assessment (Bachman & Palmer, 1996; Purpura, 2016). This study compares students' reading-to-write processes in the computer-mediated SBA and non-SBA settings. It is to evaluate the extrapolation inference in Kane's (1992; 2006; 2013) argument-based framework of validation. The findings provide evidence that the scenario features of the SBA are effective to enhance the validity in the cognitive perspective. This further supports that task contexts or settings, including the features and ways we present the tasks, are playing a role in language assessment, and technology is potential to make language assessment tasks more authentic and valid.

6 Conclusions

Information and communication technologies provide us a digital community for reading and writing, and computer technology has made it possible to create virtual scenarios for learning and assessment. The computer-mediated SBA is an assessment innovation that responds to the contextual change, attempting to assess students' literacy skills in real-life scenarios of online communication.

The SBA used in the present study was used to assess Chinese students' English writing skills with the support of multiple reading texts. With the task authenticity strengthened by computer technology, it is assumed that the SBA could be exploited to collect data that could better represent students' performance in similar real-world tasks.

The purpose of this present study was to examine whether the setting of the SBA tasks would bring up differences in cognitive processes, compared with a non-SBA setting. The investigation into the cognitive processes helps us understand the skills or the ability activated by the tasks, therefore shedding light on the validity of the computer-mediated SBA. In Kane's (1992; 2006; 2013) argument-based framework, it concerns the extrapolation inference which links the performance on the assessment to that in the target domain.

Based on the four participants' stimulated recalls, it was found that the processes elicited in the SBA

setting were more congruent to those summarized in theoretical frameworks. Moreover, the participants in the SBA setting could make more effective use of the source texts in their writing and learn more actively about the topical issues.

The finding implies that the computer-mediated SBA tasks developed in this context are valid from a cognitive perspective, offering support for the use of the tasks to infer EFL learners' reading-to-write skills in similar contexts. Furthermore, it contributes insights into the relationship between cognition and task settings or conditions. The implication is that one's cognitive activity is influenced by task settings. Therefore, it is necessary to take into account the task settings and features while evaluating students' cognitive traits. And the task settings should reflect the real-world scenarios so that more valid and reliable claims could be made about students' skills or abilities. It also shed light on the potential role of computer technology in increasing task authenticity and assessment validity.

The limitations of this study should be noted along with suggestions for future research. First, methodologically, the study on the cognitive processes was limited in the use of stimulated recalls and it might be helpful to adopt other methods for triangulation. For instance, follow-up interviews for participants' explanations would promote further understanding of their behaviors.

Second, the four participants were all at intermediate English proficiency level, according to their scores on a standardized English test in senior high school or the college entrance exam. In the future, participants with lower proficiency could be involved in both settings to examine whether any variance of the study finding would be yielded.

Third, the study did not separate the effect of the additional reading comprehension tasks in the SBA settings from that of the computer-mediated scenario features. The reason was that the reading comprehension tasks were considered an integral part of the SBA design as scaffolding from the virtual tutor. Nonetheless, future research could evaluate the impact of these comprehension items so that more specific insights could be gained on the impact of the scenario features afforded by computer technology.

Finally, future studies could also compare interactions from a real tutor or peer with the pre-prepared interactions delivered automatically through specific cues or at specific times. In this way, the effectiveness of the interactions assisted by computer technology could be better understood.

Acknowledgments

This study is funded by the following research grants: (1) 2021 Qihang Research "A Longitudinal Study on College Students' English Competence Development in New Liberal Arts Construction" funded by Beijing International Studies University (No. 21110012042), and (2) "A Study on the Impact of The National Curriculum on the Development of EFL Learners' Core Competencies" funded by Beijing Educational Science Projects for the 13th Five-Year Plan (No. CADA17076).

References

- Alderson, J. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Bachman, L. (2004). *Statistical analysis for language assessment*. Cambridge, UK: Cambridge University Press.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessment and justifying their use in the real world*. Oxford: Oxford University Press.
- Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking.

Language Testing, 30(4), 441-465.

- Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning: A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives*, 8, 70-91.
- Bennett, R., Deane, P., & van Rijn, P. (2016). From-domain theory to assessment practice. *Educational Psychologist*, 51(1), 82-107.
- Boveri, D. (2017). *The influence of scenario-based assessment on reading comprehension*. [Unpublished Doctoral Dissertation], Northern Illinois University.
- Bowles, M. (2010). *The think-aloud controversy in second language research*. New York: Routledge.
- Britt, M. A., & Rouet, J. F. (2011). Research challenges in the use of multiple documents. *Information Design Journal*, 19(1), 62-67.
- Castek, J., & Coiro, J. (2015). Understanding what students know: Evaluating their online research and reading comprehension skills. *Journal of Adolescent and Adult Literacy*, 58(7), 546-549.
- Chapelle, C., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge: Cambridge University Press.
- Chapelle, C. A. & Voss, E. (2017). Utilizing technology in language assessment. In E. Shohamy & I. Or (Eds.). *The Encyclopedia of Language and Education. Language testing and assessment* (Vol. 7). Springer Press.
- Choi, I.-C., Kim, K. S. & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20(3), 295-320
- Corbel, C. (1993). Computer-enhanced language assessment. In G. Brindley (Series Ed.) Research Report Series 2. *National Centre for English Language Teaching and Research*, Macquarie University, Sydney, Australia.
- Cumming, A., Busch, M., & Zhou, A. (2002). Investigating learners' goals in the context of adult second-language writing. In S. Ransdell & M. Barbier (Eds.), *New Directions for Research in L2 Writing* (pp.189-208). Dordrecht, Netherlands: Kluwer Academic Publisher.
- Deane, P., & Song, Y. (2014). A case study in principled assessment design: Designing assessments to measure and support the development of argumentative reading and writing skills. *Educativa Psicologia*, 20(2), 99-08.
- Deane, P., Song, Y., van Rijn, P., O'Reilly, Fowles, M., Bennett, R., Sabatini, J., & Zhang, M. (2018). The case for scenario-based assessment of written argumentation. *Reading and Writing*, 32(6), 1575-606.
- Delaney, A. Y. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes*, 7, 140-50.
- Ercikan, K., & Pellegrino, J. W. (2017). *Validation of score meaning for the next generation of assessments: The use of response processes*. New York, NY: Taylor & Francis.
- Ericsson, K., & Simon, H. (1996). *Protocol analysis: Verbal reports as data* (3rd ed.). Cambridge, MA: MIT Press.
- Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening* (pp. 77-151). Cambridge: Cambridge University Press.
- Flower, L. (1987). Interpretive Acts: Cognitive and the construction of discourse. *Poetics*, 16, 109-130.
- Gass, S., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research*. New York: Routledge.
- Grabe, W., & Zhang, C. (2016). Reading-writing relationships in first and second language academic literacy development. *Language Teaching*, 49(3), 339-355.

- Guo, C. (2015). *Think-aloud protocols in foreign language teaching research*. Beijing: Foreign Language Teaching and Research Press.
- Guo, H., Zhang, M., Deane, P., Bennett, R. (2020). Effects of scenario-based assessment on students' writing processes. *Journal of Educational Data Mining*, 12(1), 19-45.
- Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In C. Levy & S. Ransdell (Eds.), *The Science of Writing* (pp. 1-27). Mahwah, NJ: L Erlbaum.
- Hayes, J. (2012). Modeling and remodeling writing. *Written Communication*, 29(3), 369-388.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach*. Hillsdale, NJ: Lawrence Erlbaum Associates: 3-30.
- Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112, 527-535.
- Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education and Praeger.
- Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, 42(2), 448-457.
- Leeman, J. (1999). *Recasts in Spanish as a second language: An empirical study of negative evidence and enhanced saliency*. Presented at Second Language Research Forum, Minneapolis, MI, September.
- Lei, X. (2008). Exploring a sociocultural approach to writing strategy research: Mediated actions in writing activities. *Journal of Second Language Writing*, 17, 217-236.
- Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive implicit negative feedback? *Studies in Second Language Acquisition*, 22, 471-497.
- Magliano, J. P., McCrudden, M. T., Rouet, J. F., & Sabatini, J. (2018). The modern reader: Should changes to how we read affect research and theory? In M. F. Schober, M. A. Britt, & D. N. Rapp (Eds.), *A handbook of discourse processes* (2nd ed.) (pp. 342-361). New York, NY: Routledge.
- O'Reilly, T., & Sabatini, J. P. (2013). *Reading for understanding: How performance moderators and scenarios impact assessment design* (Research Rep. No. RR-13-31). Princeton, NJ: Educational Testing Service.
- O'Reilly, T., Weeks, J., Sabatini, J., Halderman, L., & Steinberg, J. (2014). Designing reading comprehension assessments for reading interventions: How a theoretically motivated assessment can serve as an outcome measure. *Educ Psychol Rev*, 26, 403-424.
- OECD. (2019). *PISA 2018 assessment and analytical framework*. PISA, OECD Publishing. Paris. Retrieved on Jan. 19th, 2020 from <https://doi.org/10.1787/b25efab8-en>.
- Partnership for 21st Century Skills. (2008). *21st century skills map*. Washington, DC: Author. Retrieved July 3rd, 2019 from http://www.p21.org/storage/documents/21st_century_skills_english_map.pdf
- Plakans, L. (2007). *Second language writing and reading-to-write assessment tasks: A process study*. [Unpublished Doctoral Dissertation], University of Iowa.
- Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13, 111-129.
- Purpura, J. E. (2016). Second and foreign language assessment. *Modern Language Journal*, 100(Suppl), 190-208.
- Rupp, A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective. *Language Testing*, 23(4), 441-474.
- Sabatini, J. P., O'Reilly, T., Halderman, L., & Bruce, K. (2014a). Broadening the scope of reading comprehension using scenario-based assessments: Preliminary findings and challenges. *L'Année psychologique*, 114, 693-723.

- Sabatini, J. P., O'Reilly, T., Halderman, L., & Bruce, K. (2014b). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disabilities Research and Practice, 29*(1), 36-43.
- Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2019). Engineering a twenty-first century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing*, online, DOI: 10.1080/15305058.2018.1551224
- Schaeffer, G. A., Reese, C. M., Steffen, M., McKinley, R. L. & Mills, C. N. (1993). *Field test of a computer-based GRE general test, ETS RR 93-07*. Princeton, NJ: Educational Testing Service.
- Spivey, N. (1984). *Discourse synthesis: Constructing texts in reading and writing*. Newark, DE: International Reading Association.
- Spivey, N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication, 7*, 256-287. doi:10.1177/0741088390007002004
- Spivey, N. (1997). *The constructivist metaphor: Reading, writing, and the making of meaning*. New York, NY: Academic Press.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly, 24*, 7-26.
- Stein, V. (1990). Exploring the cognition of reading-to-write. In L. Flower, V. Stein, Ackerman, J., and M. Kantz (Eds.), *Reading to write: Exploring a cognitive and social process (35-75)*. Oxford: Oxford University Press.
- Stratman, J. F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 89-111). Thousand Oaks, CA: Sage.
- Strømsø, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction, 20*, 192-204.
- Taylor, C., Kirsch, I., Eignor, D. & Jamieson, J. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning, 49*(2), 219-74.
- van den Broek, P. (2012). Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. *School Psychology Review, 41*, 315-325.
- van Rijn, P., Graf, E., & Deane, P. (2014). Empirical recovery of argumentation learning progressions in scenario-based assessments of English language arts. *Psicología Educativa, 20*, 109-115.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge: University Press.
- Yang, H-C., & Plakans, L. (2012). Second language writers' strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly, 46*(1), 80-103.
- Zhang, M., van Rijn, P., Deane, P., & Bennett, R. (2019). Scenario-based assessments in writing: an experimental study. *Educational Assessment, 24*(2), 73-90. DOI: 10.1080/10627197.2018.1557515
- Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly, 17*, 165-187.
- Ziegler, N., & Mackey, A. (2014). *Pre-task planning, performance, and perceptions: A study of L2 text-chat*. Paper presented at Second Language Research Forum, Columbia, SC.

Yumei Zhang, Ph.D. in applied linguistics. Lecturer at Beijing International Studies University. Her research interest includes second language acquisition, language teaching, and assessment. She has

published several articles in applied linguistics and co-authored a book on language testing with Prof. Shaoqian Luo.

Shaoqian Luo, Ph.D. in applied linguistics. Professor at Beijing Normal University. Her research interest includes second language acquisition, task-based learning, and language assessment. Her work has been published extensively in English and Chinese, including journal articles, book chapters, and books.