*Reading in a Foreign Language* October 2018, Volume 30, No. 2
ISSN 1539-0578 pp. 189–208

# An examination of self-assessment and interconnected facets of second language reading

Haley Dolosic
Washington University in St. Louis
United States

## Abstract

As second language (L2) reading assessment and proficiency grow in importance globally, many questions remain about the multifaceted nature of L2 reading. This study examined the interconnected relationships of text type, test method, topic familiarity, and self-assessment, as they relate to L2 reading through correlation and regression analyses with Chinese English as a foreign language (EFL) students at a university in China, enrolled in English for Academic Purposes (EAP) courses. Although findings of this study indicate that topic familiarity did not hold its traditionally influential role in L2 reading with these learners, both test method and text type demonstrated statistically significant relationships with L2 reading performance, including a statistically significant interaction between text type and test method. Criterion-referenced self-assessment demonstrated statistically and practically significant associations between learners' self-ratings and L2 reading performance. However, when examined by test and text type, further complexities indicated these associations were tied specifically to narrative texts and multiple-choice test methods.

***Keywords***: English as a Foreign Language (EFL), L2 reading, self-assessment, assessment, test method, text type, topic familiarity

Reading is seen as a crucial skill for the citizens of modern societies to learn information and communicate (Grabe, 2009). Recently, these skills have grown in importance as individuals around the world increasingly reach across not only national but also linguistic borders for their own academic and economic pursuits, as they strive to attain more desirable positions in society both within their homelands and abroad. As English increasingly becomes a medium of instruction and business negotiation globally, access to such success is often driven through successful performance on high stakes tests of English which frequently focus on reading skills (Baker, 2015; Cheng, 2008). In China, the importance of successful English reading is intensified for English as a foreign language (EFL) students, as this growing population is developing its reading skills earlier and with greater focus than ever before (Pan, 2007; Wu, 2016). However, despite studies which have examined L2 reading comprehension across varied dimensions, second language (L2) reading comprehension is not yet fully understood as an integrated process (Bernhardt, 2011; Grabe, 2009; Koda, 2005). Specifically, Bernhardt (2011) highlights the need for comprehensive, empirical investigations of the “array of variables” (p. 136) that shape L2 reading comprehension, building from prior research to examine these interacting variables.

Furthermore, when learners encounter tests of L2 reading, they may also not be aware of their own strengths and weaknesses (Brantmeier, 2006b; Schultz, 2017). Such an awareness, or ability to self-assess, could prove central to learners’ capabilities to become autonomous, life-long L2 readers (Little, 2009). Although prior research examines self-assessment as it relates to general reading comprehension tests (Ross, 1998) and in reference to its ability to accurately place students (Brantmeier, 2005a, 2006b), few studies examine L2 reading self-assessment in relation to the variables that shape L2 reading comprehension. These comparisons could provide crucial insight on understanding L2 self-assessment and reading comprehension, offering implications for supporting learners in their quests for success both on tests of English and throughout their lives as L2 readers. Therefore, this study unites some of these often-separated features of self-assessment and L2 reading, seeking a comprehensive understanding of the interactive relationships among key variables such as text type, test method, topic familiarity, self-assessment, and L2 reading comprehension with Chinese EFL university students.

## Literature Review

### *Second Language Reading*

Although reading comprehension competence can be described in many ways, reading is generally understood to be the process by which one perceives, decodes, and seeks to gain meaning from text on a page or screen by combining the extracted information with the knowledge that the reader already possesses (Koda, 2005). Yet, Grabe (2009) highlights that reading is much more than the process of decoding and incorporating information into pre-existing knowledge systems. Rather, he contends that reading is an efficient, purposeful, interactive, strategic, and fundamentally linguistic process, calling attention to the multifaceted and multipurposed nature of reading (Grabe, 2009; Perfetti, Landi, & Oakhill, 2005). When this process takes place in an L2, it is further complicated by orthography, grammar, text conventions, and other knowledge that is key to processing and understanding text in the other language (Koda, 2005, 2007). Readers must overcome these complexities in order to develop a mental representation of the text, often referred to as a situation model (Kintsch & van Dijk, 1978). Bernhardt (2011) examines this complex process in her synthesis of research on the topic of advanced L2 reading, developing upon her Compensatory Model of Advanced L2 Reading (Bernhardt, 1991, 2011). Unifying research in the field of L2 reading, this Compensatory Model (Bernhardt, 2011) confirms that first language (L1) literacy can account for approximately 20% of variance in performance on reading assessment tasks, whereas L2 linguistic knowledge, in terms of grammar and vocabulary knowledge, explains an additional 30% with advanced L2 readers. This model also accounts for the fact that individuals may compensate for weakness in one category, such as L1 literacy, with strength from another, such as L2 vocabulary knowledge. Yet, there is a remaining 50% of the variance in reading comprehension performance that has not yet been reliably explained (Bernhardt, 2011). Brantmeier (2004) contends that further investigations should examine factors that are found within this remaining variance, seeking factors which may be more influential in yielding better L2 reading comprehension. At present, test method, text features, and topic familiarity are recognized by many researchers for their nuanced effects on L2 reading (Brantmeier, 2005a, 2006a; Hammadou, 1991; Riley & Lee, 1996). However, further research uniting these features is necessary to determine their place within the Compensatory Model (Bernhardt, 2011). A great deal of the remaining variance could be explained with the relationships among these key factors that have been shown to shape L2 reading, expanding current theoretical understandings of L2 reading and providing insights that may be applied in classrooms.

*Testing methods*. Researchers and educators alike use various types of tests and assessments to build an understanding of language learners' reading abilities. Within these tests, students are often asked to read a passage appropriate to their stage of acquisition or course level. Then, they complete a task with the text that they have been provided. These tasks can be quite different, asking the learner to select an answer, match two words or ideas, complete a passage by filling in blanks, or simply recall everything that they possibly can from the passage they have just read (Alderson, 2000). Researchers frequently use three different types of assessment tasks in a single study of reading comprehension: free recall, sentence completion, and multiple choice (Brantmeier, 2005b, 2006b), and research has also examined differences by testing methodology (Brantmeier, 2006b; Shohamy, 1984). In their examinations, Bachman (1990) and Wolf (1993) divided these testing tasks, often referred to as test methods, into two categories where the learner is either selecting or constructing a response. Studies have compared findings and correlations across testing tasks, demonstrating differences among results based on type of test (Alderson, Bachman, Perkins, & Cohen, 1991; Alderson, Clapham, & Wall, 1995; Brantmeier, 2006b; Carrell, 1984a, 1984b, 1985; Riley & Lee, 1996; Shohamy, 1984). Results to date indicate that students consistently achieve different scores depending upon the type of testing task they encounter (Brantmeier, 2006b; Wolf, 1993).

When synthesizing research on the topic of L2 reading assessment, Bernhardt (1991, 2011) supported the use of free recall to measure individuals' abilities to comprehend a passage, suggesting that assessment of learners' comprehension is best demonstrated when the assessor does not interfere with the frame of understanding. Yet, Alderson (2000) supports test methods that respond more specifically to the information that the test constructor is trying to gain. Further, Alderson (2000) recommends that a good assessment incorporates multiple methods of assessment, creating a comprehensive and statistically reliable picture of an individual's comprehension abilities. In addition, both Bernhardt (2011) and Alderson (2000) highlight the variety of factors outside of the type of assessment that could shape an individual's performance on a given method of reading assessment such as the learner's proficiency or topic familiarity. Drawing from Bernhardt's Compensatory Model (2011), it is possible that these factors interact, shaping learners' abilities to successfully comprehend advanced L2 texts.

Lee (1987) suggested that learners' ability to communicate what they understand varies according to the task used to measure comprehension. While Wolf (1993) proposed that difficulty in completing a task relies only on the skills required to complete it, the novelty of the task may also need to be considered. For example, according to the theoretical framework Transfer Appropriate Processing (TAP) (Morris, Bransford, & Franks, 1977), students may be more able to succeed on a task that aligns closely with how they were instructed. Recently, in an

L1 study endeavoring to grow learner autonomy, Thomas and McDaniel (2007) found that where the learning task was consistently closely aligned with the assessment task, students consistently performed better, as TAP would predict. The TAP framework is also consistent with L2 frameworks of vocabulary acquisition such as the Type of Processing—Resource Allocation (TOPRA) model (Barcroft, 2002, 2003, 2004, 2013), which emphasizes that tasks must engage specific types processing in order for students to learn certain aspects of form and meaning mapping for novel L2 words. Therefore, prior results that demonstrate greater success when students select their responses could also be associated with instructional methods, such as test preparation, that emphasize selecting the correct response.

Although these examinations do provide an understanding of L2 reading assessment, there is a need for more empirical work examining varied types of assessment, especially as these test types may interact with other key variables in the reading process (Brantmeier, 2005a). Further, as tests of language proficiency have become embedded in the larger culture of testing and access to the higher status in society within China (Cheng, 2008), it is vital that these test types be understood within the often under-researched context of Chinese EFL learners in China (Wu, 2016). Chinese EFL university students may have different approaches to such text-based tasks, based on their language learning contexts and motivations that differ greatly from many more closely examined contexts of language learning (Li & Cutting, 2011). Therefore, there is a need for a study which examines test methods in concert with other variables of L2 reading, particularly within the context of Chinese EFL learners.

*Text types*. Knowing and being able to predict the structure of a text can also be beneficial on a test of reading comprehension (Alderson, 2000) and has been seen as a possible variable to include as part of the Compensatory Model. Within the broader field of applied linguistics, the structure or organization of a text has been examined largely from three different vantage points: (a) the cultural conventions present or not present in a text, (b) the coherence of the passage, or (c) the narrative and expository spectrum of texts. In examining the role of text structure in terms of cultural coherence among 240 Taiwanese university students majoring in English, Chu, Swaffar, and Charney (2002) asked students to read and recall passages that were tailored either toward traditional Taiwanese writing conventions or more English writing conventions. Findings indicated that both experience with the language and reading a text with English conventions were predictive of performance (Chu et al., 2002). Greater experience with the English language led to better recall whereas a text that was tailored toward English or Western conventions yielded a less complete recall (Chu et al., 2002). When Riley (1993) examined text features in terms of coherence, findings indicated that with university students studying French, coherence breakdowns in a story were most impactful for intermediate level students. Riley (1993) argued that the students of lowest proficiency may not have enough processing capacity to comprehend this change of coherence when working to understand the text in their L2. At the same time, he suggested that advanced students may have enough mental processing available to overcome such problems in coherence even after using their attentional resources on language processing, leaving intermediates as the most sensitive group to such coherence issues (Riley, 1993). Horiba (1996) found similarly that L2 speakers of Japanese and English were sensitive to the coherence of a text only when they were proficient in the L2, indicating that the factors of coherence are tied to key facets of the L2 reading experience such as L2 proficiency.

Yet, in general, when examining these textual factors researchers tend to discuss the differences in comprehension for narrative and expository passages (Koda, 2005). Typically, these are defined such that narratives tell stories with a causal chain of events whereas expository passages report facts and information, often in hierarchies (Koda, 2005). DuBravac and Dalle (2002) took up these differences, examining university students' abilities to comprehend varied texts in their L2, French. Findings indicated that students were better able to comprehend the narrative text, having poorer comprehension when reading an expository text. Donin, Graves, and Goyette (2004) similarly found that military officers studying French recalled more from the narrative than the expository texts. Further, with Japanese EFL university students, Yoshida (2012) found that more was recalled for narratives than expository texts, and that proficiency was likewise a key predictor of students' success. Despite the differing foci of these studies, together their findings indicate that the structure of a text shapes readers' comprehension. However, these differences may be affected by the reader's L2 proficiency or familiarity with the test method. Despite a need to better understand Chinese EFL learners' approaches to such varied texts both within and outside of reading comprehension tests, no prior study has examined the role of text organization with advanced EFL university students in China, particularly as text features relate to other key features of L2 reading.

*Topic*. In addition to test and text type, first and second language research has reliably found that readers' familiarity with the topic of the passage that they are reading impacts their understanding of the text (Koda, 2005). For example, Barry and Lazarte (1998) found with 48 university students studying Spanish that participants who demonstrated "high knowledge" were able to generate more inferences, fewer incorrect inferences, recall more text information, and generate qualitatively better representations of the text than "low knowledge" participants on a three-passage free recall comprehension assessment. Further, Brantmeier (2005a) examined the role of analogies to facilitate comprehension in L2 reading with Costa Rican and American students studying English and Spanish respectively, yet topic familiarity variance came to the forefront as the greater predictor of students' successes, with analogies failing to help students who had little knowledge of the topics of the passage to comprehend the text. Uso-Juan (2006) found similar results with Spanish-speaking EFL students. Dividing students into high and low knowledge groups based upon their majors in the university setting, Uso-Juan (2006) found that background knowledge was integral to students' comprehension. More recently, Horiba and Fukaya (2015) also found that a high knowledge group of Chinese EFL university students, as determined by major, were more successful in comprehending reading passages with qualitatively different recalls being evident between the high and low knowledge groups. Together these findings clearly indicate that there are strong effects associated with the topic of the reading passage, particularly with high and low knowledge groups, and appear to support the claim that topic knowledge is central to comprehension.

However, as Nassaji (2002) cautions, background or topic knowledge is often found to be more complex than it is straightforward. Prior knowledge is not a simple mechanism that clearly and directly predicts success on reading comprehension. Rather, Nassaji (2002) contends, it is "related to far more complex cognitive processes," (p. 93) involving the readers' previous mental representation, the structure and nature of the text, and how the reader is able to incorporate it into their memory. Such dynamic representations have been found in studies which examine proficiency and topic familiarity together. For example, Chen and Donin (1997) examined reading with graduate students who were L2 speakers of English, finding that, for these individuals, language proficiency and background knowledge could compensate for one another in order for learners to successfully comprehend the text. Therefore, these students could use language proficiency to overcome a lack of background knowledge (Chen & Donin, 1997). Further, few studies examine topic knowledge when it is less extreme, as the topics of standardized assessment might be. Research is needed to understand the exact role of topic familiarity in L2 reading, particularly in China where few studies to date have taken up this area of inquiry and many high-stakes English reading tests occur.

### *Second Language Self-Assessment*

Within this study, self-assessment (SA) is defined as one's own evaluation of one's performance or capabilities. SA has been studied in the field of language learning since the late 1970s both to understand the mechanisms and instruments of SA themselves and to assess their impact on students' learning. At the institutional level, Oskarsson (1978) examined correlations between written test scores and students' self-ratings, finding a positive relationship between university students' own views and their actual scores on language assessments. Seeking to better understand these self-evaluative techniques, LeBlanc and Painchaud (1985) found that, with a highly-contextualized SA instrument, students were able to accurately represent their own L2 abilities through SA, meaning that their self-ratings were closely associated with their actual performance. Bachman and Palmer (1989) continued this trend by examining the validity of SA as a reliable construct, finding that it was statistically reliable with their students. From these early stages, authors have argued for the use of SA to develop learner autonomy (Blanche & Merino, 1989; Harris, 1997; McNamara & Deane, 1995).

As Little (2009) argued, students must be involved in "planning, monitoring, and evaluating" (p. 224) their language learning in order to become autonomous, life-long language learners. Specifically, an accurate understanding of one's own strengths and weaknesses could provide language learners with opportunities to set realistic, challenging goals which are central to such life-long language learning (Sweet & Mack, 2017; Ziegler, 2014). Therefore, SA may be an important indicator of students who are developing the skills they need to become these life-long language learners. For example, Sweet and Mack (2017) investigated the effects of reflective, collaborative speaking SA tasks with 367 university students studying Spanish in the United States. For this program, learners were provided with speaking tasks to complete and reflect upon, both individually and with a peer. Findings demonstrated that these learners were able to develop an awareness of their capabilities, the process of learning language, and an ability to monitor their own progress (Sweet & Mack, 2017). SA itself may also become more accurate, with learners being better able to represent their own strengths and weaknesses following other types of training. For example, with 91 Vietnamese university students studying EFL, Nguyen and Gu (2013) provided half of the students with a metacognitive, reflective training for English writing. The students who received the training gained significantly greater accuracy in their SA than students who did not experience the training. Further, these students also produced English writing that was scored more highly (Nguyen & Gu, 2013). Similarly, regular SA training has demonstrated improvement across the semester in terms of linguistic development for beginning and intermediate reading, writing, and listening (Mazloomi & Khabiri, 2016; Shahrakipour, 2012). Such results may indicate that understanding when and where students are aware of their own strengths and weaknesses, both with and without specific SA training, could be vital in developing courses and programs that support life-long language learning.

However, nuances of SA have also been discovered through closer examination of L2 SA as a measurement and placement tool. For example, in his meta-analysis of SA, Ross (1998) found that SA is more accurate when the items used to self-assess are more functional and less abstract. Brantmeier (2005b), in seeking a fine-tuned placement tool, continued this line of inquiry, finding that with adults studying Spanish in the United States, descriptive SA did not correlate with multiple choice assessment, yet it did correlate with free recall measures of reading comprehension. In another study, with university students studying Spanish, Brantmeier (2006b) found that descriptive SA did not correlate with students' performance on a computer-based proficiency exam. Yet, when Brantmeier and Vanderplank (2008) introduced criterion-referenced items that aligned with DIALANG proficiency standards (https://dialangweb.lancaster.ac.uk/) and situated the learner within a specific context of language use, students were able to accurately self-assess their reading comprehension as measured on multiple choice items. These items were statements of actual concrete skills that the learner rated his or her ability to complete on a five-point Likert scale. When these criterion-referenced items were expanded to include other skills, Brantmeier, Vanderplank, and Strube (2012) discovered that advanced language learners were able to represent their abilities by answering these items across listening, reading, and writing. Their self-evaluations were significantly correlated with their performance (Brantmeier, et al., 2012).

Yet, it appears that the format of the questions alone cannot account for students' responses. For example, with EFL learners, Butler and Lee (2006) found that elementary students who completed a series of language tasks before completing a self-assessment demonstrated associations between their self-ratings and performance. However, students who did not have this "on-task" condition did not demonstrate the same associations, meaning that having the recent experience may have led to different results (Butler & Lee, 2006). In another study, with Chinese learners of Japanese at a university, Suzuki (2015) found that experiential factors were heavily associated with students' ability to self-assess. Those with greater experiences in the target language seemed to underrate their skills whereas those with little experience appeared to overrate their skills (Suzuki, 2015). In addition, Dolosic, Brantmeier, Strube, and Hogrebe (2016) found that adolescents studying French in an immersive setting were unable to self-assess their speaking abilities on a criterion-referenced instrument when they arrived, but they were able to accurately self-assess on a criterion-referenced measure after spending four weeks in the immersive environment, demonstrating the power of experience for successful SA. Further, for adult students studying English in China who lacked these immersive experiences, criterion-referenced SA scores did not share a relationship across performance scores (Schultz, 2017). Yet, Ding and Stapleton (2016) also found through qualitative inquiry that Chinese university students studying in an English medium university in Hong Kong were able to develop SA abilities from exposure to an immersive context of learning. These findings taken together indicate that cultural and educational experiences as well as contextualized items are key for successful SA. With the possible positive benefits of SA in terms of learner autonomy, it is vital to better understand SA, particularly with this growing population of EFL students in China.

## This Study

In the modern, globalized world it is vital that researchers continue to examine features and facets that make up L2 reading in order to gain a full understanding. Despite a need to examine the constructs of topic familiarity, text type, test method, SA, and L2 reading performance comprehensively in a single study in order to gain a more complete understanding of the multivariate nature of reading, to date no study examines these constructs in a single investigation, particularly with the growing and developing population of Chinese EFL university students. Motivated by this need, the following study was guided by the following research questions:

1. What is the relationship among text type, test method, and L2 reading performance with Chinese EFL university students studying English in China?
2. With these students, what is the relationship between topic familiarity and L2 reading performance?
3. How do these students self-assess their L2 reading abilities? Are they able to accurately self-assess their L2 reading abilities as measured by an L2 reading task across varied text types and test methods?

### *Context & Procedure*

This study was conducted at a medium-sized university in Northern China specializing in the training of teachers. All students (*N* = 77) were English majors, enrolled in an English for Academic Purposes (EAP) course during the third year of study at the university. Many students planned to become English teachers, continue for advanced degrees, or use their English language skills to succeed in business. Only 64 students out of 77 completed all measures successfully. As is typical for the context, 59 students self-identified as female; five identified as male. These students were brought together in a single lecture hall to complete all instruments at one time. All students came to the lecture hall and completed all measures within one paper packet with the researcher and course instructor present. Students progressed linearly, from start to finish, over the course of a two-hour session.

### *Instruments*

During the two-hour session, students completed multiple measures, including a demographic questionnaire, criterion-referenced self-assessment questionnaire, and a reading comprehension assessment. The criterion-referenced SA was tailored from prior research, having been validated with larger samples (Brantmeier, et al., 2012). Developed from the DIALANG framework, this questionnaire was made up of criterion-referenced items that situated the learner in a specific language use context, asking learners to rate their ability to complete the task on a scale of one to five, with one meaning that the learner “Strongly disagreed” that they would be able to complete the task while five meant that the learner “Strongly agreed” that they would be able to complete the task (Brantmeier et al., 2012). The reading comprehension assessment was carefully constructed according to the guidelines established by Wolf (1993) and Alderson (2000). It included three forms of assessment: Free Recall, Sentence Completion, and Multiple Choice. Four passages were carefully selected to be of equal length and difficulty, differing only in topic and text type (with two fitting the description of "narrative" and two fitting the description of "expository"). The four passages that were tested were qualitatively checked for agreement that they were narrative or expository in nature among two researchers and the author. Instructions for all tasks were presented bilingually (in Mandarin and English). Topic familiarity was measured through a five-point Likert scale question following each passage where the learner indicated their familiarity with the topic of the passage on a scale of one to five. These items were also presented bilingually (in Mandarin and English). Free recall was scored on pausal units. (See Brantmeier, Strube, & Yu, 2014 for full discussion.) Sentence completion and multiple choice were scored as correct or incorrect based on pre-determined answers drawn from the texts by the researcher, with one point given for each correct response. For further information about the number of items on each section and possible number of items for which an individual participant could have been given a higher score, see Table 2.

### *Analysis*

All analyses were conducted with complete observations of all variables in the open-source statistical program, R version 3.3.3 (R Core Team, 2017), using additional packages including psych (Revelle, 2017), psychometric (Fletcher, 2010), and car (Fox & Weisberg, 2011). Figures were generated using GGPlot2 (Wickham, 2009).  Although data were transformed to run analyses that corresponded to the research questions outlined here, final results are presented with un-transformed data because all results and interpretations were equal across Box-Cox transformed data and the original data (Box & Cox, 1964). Data were cleaned by removing outliers and incomplete questionnaires. Removed outliers were determined to be significantly different from the rest of the sample when examined through univariate and multivariate techniques. Analyses comparing different measures were conducted with percentages in order to account for the number of possible points one could obtain on any given measure.

## Results

Descriptive statistics were examined, and correlations, regression, and ANOVA were conducted to answer the research questions. Descriptive statistics are provided in Table 1 to demonstrate the means and variability of the data collected.

### *RQ1: What is the Relationship among Text Type, Test Method, and L2 Reading Performance with Chinese EFL University Students Studying English in China?*

Descriptive statistics indicate a great variability to students reading performance as measured through composite scores of all test types with a standard deviation of 15.97 (Figure 1). Scores were examined together as percent scored correct because percentage allowed for all text and test types to be compared on the same metric. Using the percentage of correct answers as a metric, there are evident differences among these types of texts and tests. Table 2 presents the composite and the sub-sections of reading comprehension scores. The columns are organized to provide a clear understanding of the average raw score, called "Mean Correct" compared to both the total number of points possible on any given section, called "Possible Correct," and the variability of students' performance or "Standard Deviation (SD)." In addition, there is a clear metric to make some comparisons, such as expository compared to narrative texts, through the column "Mean

Percent Correct" where students' average scores are converted to percentages based on the number of possible points to obtain on a given section. Together these columns should provide an understanding of the spread and scoring of the data.

Table 1. *Descriptive statistics of all key variables*

| | Minimum | Maximum | *M* | *SD* |
|---|---|---|---|---|
| Topic Familiarity | 6 | 21 | 10.44 | 3.12 |
| Self-Assessment (SA) | 47 | 73 | 58.45 | 6.05 |
| Free Recall | 4 | 50 | 26.73 | 11.55 |
| Sentence Completion | 9 | 27 | 19.69 | 3.97 |
| Multiple Choice | 8 | 28 | 22.61 | 3.28 |
| Narrative | 12 | 56 | 34.88 | 9.78 |
| Expository | 17 | 47 | 34.16 | 7.23 |
| Composite | 33 | 98 | 69.03 | 15.97 |

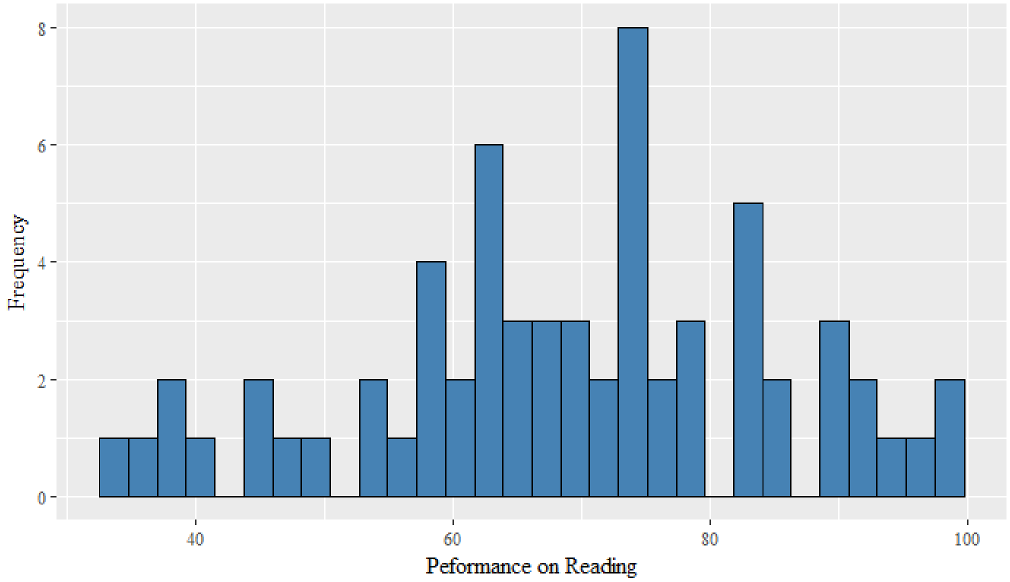

**Figure 1.** A histogram of overall reading comprehension performance

Table 2. *L2 reading performance measures*

| | Mean Correct | Possible Correct | *SD* | Mean Percent Correct |
|---|---|---|---|---|
| Composite | 69.03 | 177 | 15.97 | 39% |
| Narrative Texts | 34.88 | 97 | 9.78 | 36% |
| Expository Texts | 34.16 | 82 | 7.23 | 42% |
| Free Recall | 26.73 | 122 | 11.55 | 22% |
| Sentence Completion | 19.69 | 28 | 3.97 | 70% |
| Multiple Choice | 22.61 | 28 | 3.28 | 81% |

In order to gain further understanding of these variables, a 2 x 3 within-subjects ANOVA was conducted to investigate direct and indirect effects of both text and test type, as an interaction between these effects was possible. Results of this analysis are presented in Table 3.

Table 3. *ANOVA results*

| | *F* | *p* | $\eta^2$ |
|---|---|---|---|
| Test Method | 921.02 | < 0.001 | 0.162 |
| Text Type | 37.68 | < 0.001 | 0.003 |
| Text & Test Type Interaction | 26.82 | < 0.001 | 0.002 |

These results indicate that scores, when separated by test method, had statistically significant differences. This association had an effect size ($\eta^2$) of 0.162, meaning that a small yet significant effect is related to the test type. Follow-up analyses indicated that students performed significantly better on multiple choice items than sentence completion or free recall and were more successful on sentence completion than free recall (Figure 2). All such test method differences were statistically significant ($p < 0.05$), accounting for Bonferroni corrections, and held considerable effect sizes ($d = 0.58$–$6.39$). It is evident that test type had a clear relationship with students' performance as students performed significantly better on tests with discrete question types.

Statistically significant differences were also evident between the two text types (Figure 2); however, such differences in terms of practical significance were much smaller, with an eta squared of only 0.03 ($\eta^2 = 0.03$). In addition, there is a statistically significant interaction between text and test type, indicating that the differences among test scores is inconsistent across text types. However, this effect is small, with an eta squared of only 0.02 ($\eta^2 = 0.02$). This can be seen in Figure 2 by the relative column heights within each grouping that demonstrate slight differences. Follow-up tests of simple main effects using a Bonferroni correction were conducted to examine specific text and test type comparisons. Results indicated that narrative and expository text types demonstrated statistically significant differences for sentence-completion ($d = 0.91$) and multiple-choice test types ($d = 0.52$), meaning that the scores on both sentence completion and multiple-choice sections were different, depending upon the type of text the learner was reading.

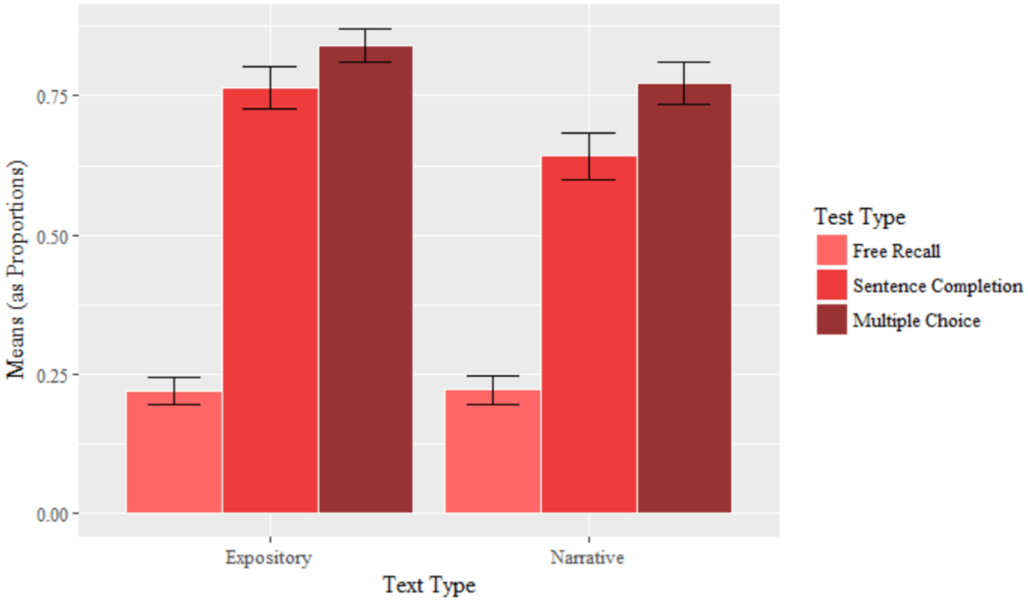

**Figure 2**. Test type & test method mean responses with 95% confidence intervals

*RQ2: With these Students, What is the Relationship between Topic Familiarity and L2 Reading Performance?*

Correlations between topic familiarity and performance are negligible and not statistically significant across all texts ($r < 0.2$; $p > 0.05$). Despite desires to further examine possible relationships with this construct, due to the lack of a significant correlation among topic familiarity and composite reading comprehension, other statistical tests were deemed no longer appropriate.

*RQ3: How do Chinese EFL University Students Self-assess their L2 Reading Abilities? Are Students Able to Accurately Self-assess their L2 Reading Abilities as Measured by an L2 Reading Task across Varied Text Types and Test Methods?*

Much like reading performance, there is a wide, near-normal distribution in SA scores ($M$ = 58.45, $SD$ = 6.05). Thus, some students assessed themselves as excellent readers while others assessed themselves as poor readers (Figure 3).

As is illustrated in Figure 4, students' SA ratings did correlate with performance ($r = 0.26$, $p < 0.05$). However, this overall effect may have been driven by specific relationships within the overall composite score. When considered by test method and text type (Tables 4 and 5), it is evident that SA correlated with multiple choice tasks and narrative text types, reaching statistical significance with both of these sub-categories. However, SA did not significantly correlate with reading comprehension assessment scores for expository texts, free recall, or sentence completion. When further broken down by text and test type in a single statistical test, only multiple-choice assessment of the narrative passages held a statistically significant relationship with criterion-referenced SA ($r = 0.21$, $p < 0.05$). Therefore, although SA appears to be significantly related to performance when these measures are combined and when, therefore, reliability is at its highest, there is evidence here that such findings may be driven by specific sub-categories of tests and texts.

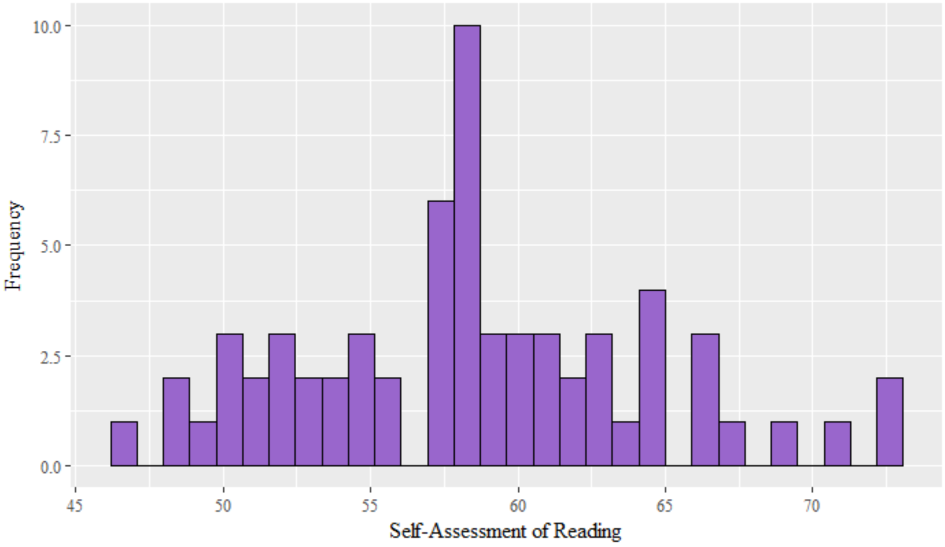

**Figure 3**. Histogram of overall self-assessment (SA)

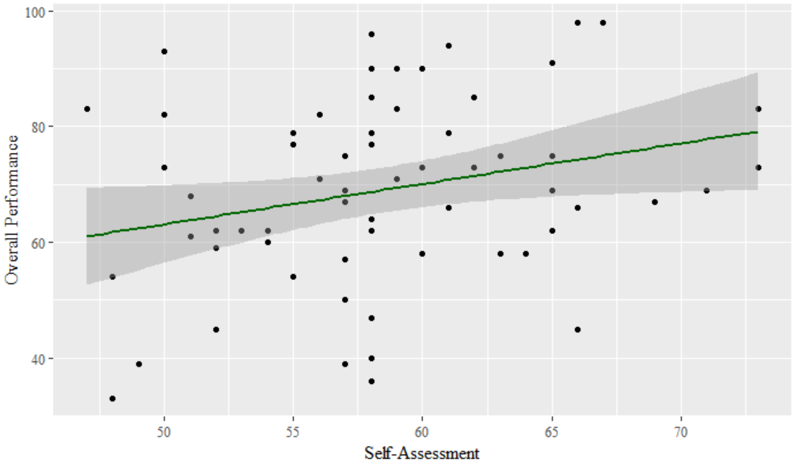

**Figure 4.** Scatterplot of association between self-assessment (SA) and overall performance with 95% confidence band

Table 4. *Correlation of self-assessment and reading performance across text types*

| | Self-Assessment | Narrative Text | Expository Text |
|---|---|---|---|
| Self-Assessment | 1.00 | | |
| Narrative Text | 0.28* | 1.00 | |
| Expository Text | 0.21 | 0.76** | 1.00 |

*Note.* * = *p* < 0.05; ** = *p* < 0.01

Table 5. *Correlation of self-assessment and reading performance across test types*

| | Self-Assessment | Free Recall | Sentence Complet. | Mult. Choice |
|---|---|---|---|---|
| Self-Assessment | 1.00 | | | |
| Free Recall | 0.19 | 1.00 | | |
| Sentence Completion | 0.23 | 0.54** | 1.00 | |
| Multiple Choice | 0.35* | 0.45** | 0.42** | 1.00 |

*Note.* * = $p$ < 0.05; ** = $p$ < 0.01

## Discussion

These findings, which indicate that test type or assessment task did have an impact on participants' reading comprehension performance, mirror prior research, with students scoring more correct answers when they were able to select rather than construct responses (Alderson, 2000; Wolf, 1993). While the increased cues present in the selection-based responses could explain some of this test method difference, as TAP might suggest (Morris et al., 1977), such an effect could also be borne out of students' preparation and training for reading exams, which are largely multiple choice in the English language classroom in China (Cheng, 2008). These effects could be further heightened by the context of these Chinese university students' language learning experiences, which culminate in high-stakes English language exams including their National College Entrance Exam (高考 /gāo kǎo) and two levels of the College English Test (Cheng, 2008). Thus, such exams may result in classroom instruction and student preparation to match these test formats, leading students to develop strong abilities with specific testing tasks. Further, the statistically significant small effects of text type could also be tied to learning experiences and fit within the paradigm of TAP, as these students are often asked to read from a variety of genres and answer questions in order to prepare them for possible short passages that exist on these high-stakes texts (Cheng, 2008; Morris et al., 1977). The importance and centrality of these assessments within Chinese culture and societal structure require that research continue to explore within China to fully capture the realities of students' capabilities and expectations within the local context.

Results of this study also suggest that topic did not play a key role for these learners in comprehending these texts. These findings contradict much of existing scholarship on topic familiarity, which indicate that topic can have a central role in understanding a text (Barry & Lazarte, 1998; Brantmeier, 2005a; Uso-Juan, 2006; Horiba & Fukaya, 2015). However, these prior studies have often looked at the extremes of high-knowledge and low-knowledge readers, without giving weight to the variability of possible topic familiarity on a more typical reading task, such as a standardized assessment. Much like what Bugel and Buunk (1996) found with a more neutral passage on a standardized examination, this study suggests that when topic familiarity is not at the extremes of high or low knowledge but rather lies across a continuous spectrum, it may impact reading comprehension less completely. Further, the importance of topic familiarity may be clouded by the proficiency of these students, which was at an advanced level (Bernhardt, 2011; Chen & Donin, 1997). At this level, it could be possible that students have developed their strategies and skills of reading to compensate for a lack of prior knowledge about a given topic much like what Chen and Donin (1997) found with graduate students studying in English (Bernhardt, 2011). Yet, with similarly advanced students, Brantmeier (2005a) found that topic knowledge was associated with comprehension outcomes. In addition, this relationship was similar for intermediate and advanced students (Brantmeier, 2005a). As findings of both this study and prior research do not coincide perfectly with our understanding, it is evident that this complexity of topic familiarity must be addressed further in future studies. While this study sought to understand topic familiarity with the context of the many variables that shape successful L2 reading (Bernhardt, 2011; Grabe, 2009), such as text type and test method, the resulting data prevented statistical tests which would allow for a multivariate examination of text, test, and topic together. These results harken to Nassaji (2002), who recommended that the relationship between topic familiarity and L2 reading be examined with caution, seeking out the nuance of this multifaceted intersection. As a result, it is vital that researchers continue to examine topic familiarity throughout the world, especially in nations such as China that rely heavily on L2 reading assessment as a pathway to success. In these high-stakes circumstances, test constructors must be certain that they are measuring reading capabilities rather than topic knowledge. In such future examinations across language learning contexts, researchers should study topic familiarity with more nuanced passages along a range of familiarity using instruments attuned to measuring topic familiarity alongside measures of the key variables of L2 reading comprehension so as to better understand this relationship among topic familiarity and L2 reading comprehension while honoring its compensatory nature (Bernhardt, 2011; Nassaji, 2002).

In addition, findings of the present study suggest that these students' self-ratings do relate to their reading comprehension performance. Such findings substantiate prior findings that support the use of contextualized items for successful SA (Brantmeier, 2005b; Brantmeier & Vanderplank, 2008; Brantmeier et al., 2012; Butler & Lee, 2006; Dolosic et al., 2016; LeBlanc & Painchaud, 1985; Ross, 1998). In addition, these findings may provide insight into learners' experiences that may have allowed them to successfully self-assess, as experience has been found to be a key factor in students' ability to self-assess (Suzuki, 2015). However, these findings differ somewhat from prior research on SA in China (Schultz, 2017), which indicates that adult Chinese EFL students are unable to self-assess. In addition, these new findings indicate that there is a test method effect present in SA as well as performance within reading comprehension. Such findings, when put in conversation with prior findings and the context of language learning, appear to indicate that SA, like performance assessment, may be affected by TAP, wherein learners are better able to perform on tasks with which they already have extensive experience.

## Conclusion

Despite limitations such as the lack of an L1 reading test and exact L2 proficiency measure, these results respond to a need to better understand Chinese EFL readers and L2 reading comprehension as they increasingly encounter high-stakes uses of their L2 reading skills. Findings of this study demonstrate that the role of topic in L2 reading could be more nuanced than prior research has indicated, particularly with advanced Chinese students of EFL. Further, SA, text types, and assessment tasks need to be considered in concert, as findings here demonstrate that they are related and interacting variables that may provide an opportunity to unravel the underlying, multifaceted nature of L2 reading. In addition, these findings suggest that educators and researchers alike should carefully select text and test types for local contexts and the purposes of the assessment.

With consistent changes and developments to high-stakes testing of English both in China and around the world, it is vital that findings such as these are considered when constructing tests. The reality that preparation and assessment must align for students both to have a good understanding of their abilities and to succeed fully on the task are supported both by this study and prior research. These results might also indicate that instructors should prepare students to succeed by providing them with experiences that are similar to the tests their students will face. Such techniques may bolster both students' abilities and their awareness of their strengths and weaknesses in completing the L2 reading assessments.

In the future, researchers should undertake similar studies across varied contexts, examining these intersecting factors alongside variables that have been established in the field to impact L2 reading such as L2 language proficiency, L1 reading abilities, and topic interest. Together, these factors should be examined to expand directly upon the Compensatory Model of Advanced L2 Reading and wider understandings of L2 reading. Furthermore, as suggested by Brantmeier et al. (2012), future research should examine the value of incorporating SA as part of course requirements where students could begin to self-diagnose their reading abilities and may become better equipped to be lifelong L2 readers. Through such investigations, it may be possible to gain a more complete understanding of the complex and multifaceted experience of L2 reading comprehension.

## References

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

Alderson, J. C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge, UK: Cambridge University Press.

Alderson, N. J., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing*, *8*, 41–66.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, *6*, 14–29. doi: 10.1177/026553228900600104

Baker, W. (2015). Culture and complexity through English as a lingua franca: Rethinking competences and pedagogy in ELT. *Journal of English as a Lingua Franca*, *4*, 9–30. doi: 10.1515/jelf-2015-0005

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52*, 323–363. doi:10.1111/0023-8333.00186

Barcroft, J. (2003). Effects of questions about word meaning during L2 Spanish lexical learning. *The Modern Language Journal*, *87*, 546–561. doi: 10.1111/1540-4781.00207

Barcroft, J. (2004). Second language vocabulary acquisition: A lexical input processing approach. *Foreign Language Annals*, *37*, 200–208. doi: 10.1111/j.1944-9720.2004.tb02193.x

Barcroft, J. (2013). Input-based incremental vocabulary instruction for the L2 classroom. In J. W. Schwieter (Ed.) *Innovative research and practices in second language acquisition and bilingualism* (pp. 107–138). Amsterdam, Netherlands: John Benjamins.

Barry, S., & Lazarte, A. A. (1998). Evidence for mental models: How do prior knowledge, syntactic complexity, and reading topic affect inference generation in a recall task for nonnative readers of Spanish? *The Modern Language Journal*, *82,* 176–193. doi: 10.1111/j.1540-4781.1998.tb01191.x

Bernhardt, E. (1991). *Reading development in a second language: Theoretical, empirical, & classroom perspectives*. Norwood, NJ: Ablex Publishing.

Bernhardt, E. (2011). *Understanding advanced second-language reading*. New York, NY: Routledge.

Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, *39*, 313–338. doi: 10.1111/j.1467-1770.1989.tb00595.x

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological), 26*, 211–252.

Brantmeier, C. & Vanderplank, R. (2008). Descriptive and criterion-referenced self-assessment with L2 readers. *System*, *36*, 456–477. doi:10.1016/j.system.2008.03.001

Brantmeier, C. (2004). Building a comprehensive theory of adult foreign language reading: A variety of variables and research methods. *The Southern Journal of Linguistics*, *27*(1), 1–7.

Brantmeier, C. (2005a). Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension. *The Modern Language Journal*, *89*, 37–53. doi: 10.1111/j.0026-7902.2005.00264.x

Brantmeier, C. (2005b). Nonlinguistic variables in advanced L2 reading: Learner's self-assessment and enjoyment. *Foreign Language Annals*, *38*, 493–503. doi: 10.1111/j.1944-9720.2005.tb02516.x

Brantmeier, C. (2006a). Toward a multicomponent model of interest and second language reading: Sources of interest, perceived situational interest, and comprehension. *Reading in a Foreign Language*, *18,* 89–115.

Brantmeier, C. (2006b). Advanced L2 learners and reading placement: Self-assessment, computer-based testing, and subsequent performance. *System*, *34*(1), 15–35. doi: 10.1016/j.system.2005.08.004

Brantmeier, C., Strube, M., & Yu, X. (2014). Scoring recalls for L2 readers of English in China: Pausal or idea units. *Reading in a Foreign Language*, *26*, 114–130.

Brantmeier, C., Vanderplank, R., & Strube, M. (2012). What about me? Individual self-assessment by skill and level of language instruction. *System*, *40*, 144–160. doi: 10.1016/j.system.2012.01.003

Bugel, K., & Buunk, B. P. (1996). Sex differences in foreign language text comprehension: The role of interests and prior knowledge. *The Modern Language Journal*, *80,* 15–31. doi: 10.1111/j.1540-4781.1996.tb01133.x

Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessment among Korean elementary school students studying English. *The Modern Language Journal*, *90*, 506–518. doi: 10.1177/0265532209346370

Carrell, P. L. (1984a). Evidence of a formal schema in second language comprehension. *Language Learning*, *34*(2), 87–108. doi: 10.1111/j.1467-1770.1984.tb01005.x

Carrell, P. L. (1984b). The effects of rhetorical organization on ESL readers. *TESOL Quarterly*, *18*(3), 441–469. doi: 10.2307/3586714

Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, *19*, 727–752. doi: 10.2307/3586673

Chen, Q., & Donin, J. (1997). Discourse processing of first and second language biology texts: Effects of language proficiency and domain-specific knowledge. *The Modern Language Journal*, *81*, 209–227. doi: 10.1111/j.1540-4781.1997.tb01176.x

Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*, *25*, 15− 37. doi: 10.1177/0265532207083743

Chu, H. C. J., Swaffar, J., & Charney, D. H. (2002). Cultural representations of rhetorical conventions: The effects on reading recall. *TESOL Quarterly*, *36*, 511–541. doi: 10.2307/3588239

Ding, F., & Stapleton, P. (2016). Walking like a toddler: Students' autonomy development in English during cross-border transitions. *System*, *59*, 12–28. doi: 10.1016/j.system.2016.04.003

Dolosic, H. N., Brantmeier, C., Strube, M., & Hogrebe, M. (2016). Living Language: Self-Assessment, Oral Production, and Domestic Immersion. *Foreign Language Annals*, *49*, 302–316. doi: 10.1111/flan.12191

Donin, J., Graves, B., & Goyette, E. (2004). Second language text comprehension: Processing within a multilayered system. *Canadian Modern Language Review*, *61*, 53–77. doi: 10.3138/cmlr.61.1.53

DuBravac, S., & Dalle, M. (2002). Reader question formation as a tool for measuring comprehension: Narrative and expository textual inferences in a second language. *Journal of Research in Reading*, *25*, 217–231. doi: 10.1111/1467-9817.00170

Fletcher, T. (2010). Psychometric: Applied psychometric theory (R package version 2.2) [Computer Software].

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression.* Thousand Oaks, CA: Sage Publications.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge, UK: Cambridge University Press.

Hammadou, J. (1991). Interrelationships among prior knowledge, inference, and language proficiency in foreign language reading. *The Modern Language Journal*, *75*, 27–38. doi: 10.1111/j.1540-4781.1991.tb01080.x

Harris, M. (1997). Self-assessment of language learning in formal settings. *ELT Journal*, *51*, 12–20. doi: 10.1093/elt/51.1.12

Horiba, Y. (1996). Comprehension processes in L2 reading: Language competence, textual coherence, and inferences. *Studies in Second Language Acquisition*, *18*, 433–473. doi: 10.1017/S0272263100015370

Horiba, Y., & Fukaya, K. (2015). Reading and learning from L2 text: Effects of reading goal, topic familiarity, and language proficiency. *Reading in a Foreign Language*, *27*, 22–46.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394. doi: 10.1037/0033-295X.85.5.363

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge, UK: Cambridge University Press.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning, 57*(1), 1–44. doi: 10.1111/0023-8333.101997010-i1

LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly, 19*, 673–687. doi: 10.2307/3586670

Lee, J. F. (1987). Comprehending the Spanish subjunctive: An information processing perspective. *The Modern Language Journal*, *71*, 50–57. doi: 10.1111/j.1540-4781.1987.tb01055.x

Li, X., & Cutting, J. (2011). Rote learning in Chinese culture: Reflecting active confucian-based memory strategies. In L. Jin & M. Cortazzi (Eds.), *Researching Chinese learners: Skills, perceptions and intercultural adaptations* (pp. 21–42). Basingstoke, UK: Palgrave MacMillan.

Little, D. (2009). Language learner autonomy and the European language portfolio: Two L2 English examples. *Language Teaching*, *42*, 222–233. doi: 10.1017/S0261444808005636

Mazloomi, S., & Khabiri, M. (2016). Diagnostic assessment of writing through dynamic self-assessment. *International Journal of English Linguistics*, *6*(6), 19–31.

McNamara, M. J., & Deane, D. (1995). Self-assessment activities: Toward language autonomy in language learning. *TESOL Journal*, *5*, 17–21.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533. doi: 10.1016/S0022-5371(77)80016-9

Nassaji, H. (2002). Schema theory and knowledge-based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning*, *52*, 439–481. doi: 10.1111/0023-8333.00189

Nguyen, L. T. C., & Gu, Y. (2013). Strategy-based instruction: A learner-focused approach to developing learner autonomy. *Language Teaching Research*, *17*, 9− 30. doi: 10.1177/1362168812457528

Oskarsson, M. (1978). *Approaches to self-assessment in foreign language learning.* Strasbourg, France: Council for Cultural Cooperation.

Pan, J. (2007). Facts and considerations about bilingual education in Chinese universities. In A. Feng (Ed.), *Bilingual education in China: Practices, policies and concepts* (pp. 200–218). Clevedon, UK: Multilingual Matters.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Malden, MA: Blackwell Publishing.

R Core Team (2017). R: A language and environment for statistical computing [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://cran.r-project.org

Revelle, W. R. (2017). psych: Procedures for personality and psychological research [Computer Software]. Retrieved from https://cran.r-project.org

Riley, G. L. (1993). A story structure approach to narrative text comprehension. *The Modern Language Journal*, *77*, 417–432. doi: 10.1111/j.1540-4781.1993.tb01989.x

Riley, G. L., & Lee, J. F. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. *Language Testing*, *13*, 173–189. doi: 10.1177/026553229601300203

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, *15*, 1–20. doi: 10.1177/026553229801500101

Schultz, L. M. (2017). Affect with Chinese learners of English: Enjoyment, self-perception, self-assessment, and abilities across levels of language learning. *Quarterly Journal of Chinese Studies*, *5*(2), 65–81. ISSN: 2224-2716

Shahrakipour, H. (2012). On the impact of self-assessment on EFL learners' receptive skills performance. *International Research Journal of Arts and Humanities*, *40*(40), 1–13.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing, 1*, 147–170. doi: 10.1177/026553228400100203

Suzuki, Y. (2015). Self-assessment of Japanese as a second language: The role of experiences in the naturalistic acquisition. *Language Testing*, *32*, 63–81. doi: 10.1177/0265532214541885

Sweet, G., & Mack, S. (2017, March). *Self-assessment and learner agency: A new approach.* Paper presented at the meeting of American Association for Applied Linguistics Conference, Portland, OR.

Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition*, *35*, 668–678.

Uso-Juan, E. (2006). The compensatory nature of discipline-related knowledge and English-language proficiency in reading English for academic purposes. *The Modern Language Journal*, *90*, 210–227. doi: 10.1111/j.1540-4781.2006.00393.x

Wickham, H. (2009). ggplot2: Elegant graphics for data analysis [Computer Software]. New York, NY: Springer-Verlag. Retrieved from https://cran.r-project.org

Wolf, D. F. (1993). A comparison of assessment tasks used to measure FL reading comprehension. *The Modern Language Journal*, *77*, 473–489. doi: 10.1111/j.1540-4781.1993.tb01995.x

Wu, S. (2016). *The use of L1 cognitive resources in L2 reading by Chinese EFL learners*. London, UK: Routledge.

Yoshida, M. (2012). The interplay of processing task, text type, and proficiency in L2 reading. *Reading in a Foreign Language*, *24*, 1–29.

Ziegler, N. A. (2014). Fostering self-regulated learning through the European language portfolio: An embedded mixed methods study. *The Modern Language Journal*, *98*, 921–936. doi: 10.1111/modl.12147

## About the Author

Haley Dolosic is a Doctoral Candidate studying Applied Linguistics in Education in the Education Department at Washington University in St. Louis. Her current research interests include self-assessment across diverse linguistic backgrounds and advanced research methodology in applied linguistics. E-mail: dolosichn@wustl.edu.