Effects of Working Memory Capacity and Content Familiarity on Literal and Inferential Comprehension in L2 Reading

CEM ALPTEKIN
Bogazici University
Istanbul, Turkey

GÜLCAN ERCETIN
Bogazici University
Istanbul, Turkey

This study examines the effects of working memory capacity and content familiarity on literal and inferential comprehension in second language (L2) reading. Participants were 62 Turkish university students with an advanced English proficiency level. Working memory capacity was measured through a computerized version of a reading span test, whereas content familiarity was achieved through nativization of a narrative, that is, textual and contextual modification to reflect the reader’s own culture. After completing the reading span test, the participants were randomly divided into two groups, one being exposed to the original text and the other to the nativized version. They then answered multiple-choice comprehension questions aiming to check literal and inferential comprehension. The results revealed independent and additive effects of working memory capacity and content familiarity on inferential comprehension. No effects were observed on literal understanding. These findings have implications for the design of assessment instruments in L2 reading comprehension.

doi: 10.5054/tq.2011.247705

Significant positive correlations between working memory (WM) capacity and language comprehension have been found in numerous first language (L1) studies (see Daneman & Merikle, 1996, for a review). In particular, WM, with its restricted functions of processing and storage, is said to play an important role in distinguishing efficient and inefficient readers, as indicated by Swanson and his colleagues (Swanson, 1999; Swanson & Alexander,
Among others. Research to date, albeit scant, suggests that the relationship between WM capacity and reading comprehension also seems to apply to second language (L2) reading (Geva & Ryan, 1993; Harrington & Sawyer, 1992; Leeser, 2007; Miyake & Friedman, 1998; Walter, 2004). Given that the cognitive resources underlying reading as a whole can be associated with the processing and storage functions of WM capacity, it is important to probe what role WM plays in reading comprehension taken in terms of its multilevel representational architecture, particularly with respect to its specific dimensions of literal and inferential reading. This is a sensitive issue, in that WM capacity may be differentially affected, depending on whether reading comprehension is of a literal or inferential nature, because there could be qualitative differences in the complexity level of the reading tasks involved in each case (Sasaki, 2000) and the degree of activated and (re)constructed schematic information stored in long-term memory (LTM). Specifically, it is said that the more complex a task, the more it implicates the contribution of LTM-based knowledge to WM processing (Kintsch, Healy, Hegarty, & Pennington, 1999). A difficult task such as inferential elaboration, for example, cannot normally be tackled adequately without the efficient use of LTM-based knowledge structures (Calvo, 2001; Singer & Ritchot, 1996).

In the case of L2 performance, WM operations are affected by what is commonly referred to as a state-level cognitive deficit, which involves processing limitations by the WM in an L2, irrespective of the individuals’ trait-level cognitive abilities (Ardila, 2003; Cook, 1997; Proverbio, Roberta, & Alberto, 2007). For example, when learners complete standardized tests in their L2, their performance tends to be lower, thus underestimating their true ability (e.g., Lee, 1986; Mestre, 1986). Other research findings involving L2 performance point to a reduced span in processing the L2 input due to the lack of sufficient LTM contributions to WM capacity (Brown & Hulme, 1992; Service & Craik, 1993). With regard to reading in particular, L2 readers, compared to efficient L1 readers, tend to become more involved with processing the text literally, such that they fail to call on higher-level conceptual processes of reading. This propensity for text-based processing (Alptekin, 2006; Horiba, 1996, 2000; Jonz, 1989; Taillefer, 1996), stemming from inadequate language proficiency (e.g., Clarke, 1980) rather than WM limitations as a trait, leads to excessive focus on surface- and propositional-level features (e.g., lexical decoding, syntactic parsing, coreferencing), leaving few cognitive resources available

---

1 The construct of WM has evolved from the traditional concept of short-term memory, which refers to a rather passive temporary memory store, whose function is to retain transient information over short intervals. WM goes beyond the mere retention of information, in that it is also involved in the concurrent processing of mental operations in a variety of cognitive tasks.
for allocation to LTM-based data, which would normally contribute to the
generation of a meaningful text representation (Alptekin & Ercetin, 2009; Pulido, 2009). In addition, L2 readers’ inadequacy in the kind of “socioculturally appropriate background knowledge shared between L1 writers and readers” compels them to rely more on both the textual linguistic data and their L2 proficiency to extract meaning from L2 texts, compared to L1 readers (Nassaji, 2002, p. 643). Whatever the reason, linguistic and/or cultural, text-boundedness normally impedes reading comprehension, in that the rapid retrieval from LTM of domain-specific schemas and their contribution to information processing in the WM become quite difficult, thereby paving the way to a shallow textual representation.

Another important factor that affects reading comprehension is the degree of interaction between the reader’s domain knowledge and textual content, as has been illustrated amply in L2 schema-theoretic research (e.g., Carrell, 1987; Carrell & Eisterhold, 1983; Lee, 2007) as well as in recent construction-integration models of comprehension focusing on L1 (e.g., Kintsch, 1998) and on L2 (e.g., Nassaji, 2002). When text content and domain knowledge are congruent, L2 readers perform more like efficient L1 readers, making adequate use of both their higher- and lower-order cognitive operations for comprehension. It follows that L2 readers’ familiarity with textual content tends to improve their comprehension, in particular, their inferential understanding, which results from knowledge-driven processes (Fincher-Kiefer, 1992). However, given L2 readers’ heavy reliance on the text, content familiarity does not necessarily lead to improved comprehension accuracy in literal reading, because this involves both processing the surface-level features of the text and constructing the text-based propositional meaning—rather than forming a mental model of the situation it depicts.

Furthermore, an investigation of the combined effects of WM capacity and content familiarity on the two dimensions of reading is essential, because this is an area that still remains largely unexplored, at least to our knowledge. Whether reading performance is affected by WM capacity limitations with or without the role played by domain knowledge has important implications. Several models have been proposed regarding the combined effects of WM capacity and domain knowledge on cognitive performance in general (Hambrick & Engle, 2002). The compensation model suggests that domain knowledge attenuates the effects of WM capacity, in that high levels of domain knowledge compensate for low levels of WM capacity. As such, no difference would be observed between high- and low-WM capacity individuals when sufficient domain knowledge is available (Ackerman & Kyllonen, 1991; Ericsson & Kintsch, 1995). The rich-get-richer model predicts that WM capacity strengthens the
effect of domain knowledge, because individuals with high levels of WM capacity benefit from prior domain knowledge more extensively, compared to those with lower levels of WM capacity (Just & Carpenter, 1992; Leeser, 2007). Finally, the independent influences model posits that WM capacity and domain knowledge have additive and independent effects on individuals’ cognitive performance (Hambrick & Engle, 2002; Hambrick & Oswald, 2005; Payne, Kalibatseva, & Jungers, 2009).

To elucidate the effects of WM capacity and content familiarity on the literal and inferential aspects of L2 reading comprehension, we deem it necessary to first discuss how we perceive the WM as a limited-capacity processing and storage system in relation to literal and inferential types of reading performance, how content familiarity can be enhanced through text genre and content modification, and whether WM capacity and content familiarity mediate each other’s effects on literal and inferential understanding.

OPERATIONALIZING WM

Despite a number of controversies surrounding the conceptualization and operationalization of WM capacity (Bunting, Conway, & Heitz, 2004; Duff & Logie, 2001; Towse & Hitch, 1995; Whitney, Arnett, Driver, & Budd, 2001), there is general consensus among cognitive psychologists that WM is a limited-capacity information processing system that allows for the active maintenance of information in the face of concurrent distraction while tackling a variety of cognitive tasks. With regard to language acquisition in particular, WM plays an important role in both L1 and L2 learning (Miyake & Shah, 1999), not to mention its significant relationship with reading comprehension, as mentioned earlier.

WM’s most detailed architectural description, offered by Baddeley (1986, 2000), entails a multicomponent model with a supervisory attentional system, the central executive, which is responsible for complex processing operations. The central executive is supported by three domain-specific slave systems (phonological loop, visuospatial sketch pad, episodic buffer), each of which has a specific function to cater to the efficiency of its performance. The phonological loop is the verbal component of WM, involved in temporarily storing phonological or auditory information. The visuospatial sketchpad is responsible for generating and temporarily storing images. The episodic buffer combines information from different sources (including the LTM) and modalities into a single, multifaceted code or episode. Certain researchers, including Baddeley (2003), view span tasks (e.g., reading span, counting span, operation span), designed to assess the processing and storage functions of WM, as common measures of the central
executive capacity (Baddeley, 2003; Baddeley & Hitch, 1994; Conway & Engle, 1994; Engle, Kane, & Tuholski, 1999; Engle & Oransky, 1999; Just & Carpenter, 1992; Turner & Engle, 1989).

Span tasks aim at measuring the active maintenance of information in the face of concurrent processing and/or distraction. Even though they differ in surface level features, these tasks are structurally similar, in that they consist of a dual-task format combining a storage measure (primary task), such as recalling the target words presented, with a processing measure, (secondary task), such as reading sentences for accuracy, solving mathematical operations, and counting shapes (see Conway et al., 2005, for a comprehensive review of span tasks).

One commonly used instrument of WM capacity assessment is the reading span task (RST), designed by Daneman and Carpenter (1980). Its popularity seems to come from its construct validity in a wide array of complex cognitive behaviors for which the ability to control attention and thought is crucial (Conway et al., 2005), along with the ability to overcome interference—both of which it is said to do reliably (Whitney et al., 2001). Consequently, the RST is widely used in investigating the relationships between WM capacity and reading comprehension, often with variations made in the types of tasks tapping processing and storage (e.g., Chun & Payne, 2004; Georgiou, Das, & Hayward, 2008; Harrington & Sawyer, 1992; Leeser, 2007; Walter, 2004). For example, although some studies follow Daneman and Carpenter’s original model and use oral reading as an intrusion into one’s storing performance (e.g., Georgiou et al., 2008; Harrington & Sawyer, 1992), others deploy sentence-level acceptability judgments based on such intruders as syntactic accuracy, semantic plausibility, or sentence veracity (e.g., Chun & Payne, 2004; Leeser, 2007).

An important issue concerning the RST is that its combination of processing and storage is considered too similar to a measure of reading comprehension itself (Kintsch, 1998). True, the RST is, in the final analysis, a test that requires reading. Test takers have to understand sentences while trying to remember each sentence-final word. As such, the test is deemed problematic by some critics: “Performance may be partly, or even largely, dependent on general reading ability, which is correlated with reading comprehension skill, but which deploys many processes other than working memory. So, although the test has a working memory component, this component may not be responsible for the correlation between test scores and comprehension skill” (Seigneuric, Ehrlich, Oakhill, & Yuill, 2000, p. 82).

Nevertheless, Daneman and Hannon (2007), based on evidence coming from one of their previous studies (Hannon & Daneman, 2001), refute this claim in unequivocal terms. They refer to their research that aimed at comparing the relative powers of the RST and a standardized
test of reading comprehension (the Nelson-Denny) at predicting test-takers’ performance on another standardized test of reading comprehension (the Verbal Scholastic Assessment Test, or VSAT), with and without VSAT text availability. Under text-available conditions, reading span was found to be a good predictor of VSAT performance, yet the Nelson-Denny was a better predictor. Under text-unavailable conditions, even though reading span remained a good predictor of VSAT performance, the predictive power of the Nelson-Denny was considerably reduced. The reversal of the predictive powers, they maintain, is an indication that reading span, far from being simply another measure of reading comprehension skill, is actually “a measure of dynamic working memory system that processes and temporarily stores information in the service of complex cognitive tasks such as reading comprehension . . . and verbal reasoning” (Daneman & Hannon, 2007, p. 40).

Another important issue in examining the relationship between WM capacity and L2 comprehension is whether WM capacity should be measured in the L1 or L2. At first, this may appear to be a trivial issue because WM effects are reported to be independent of any specific L1 or L2 (Osaka & Osaka, 1992; Osaka, Osaka, & Groner, 1993) and performance of WM tasks in the L1 and L2 show strong correlations (Alptekin & Erçetin, 2010; Miyake & Friedman, 1998; Van den Noort, Bosch, & Hugdahl, 2006). Nonetheless, L2 reading comprehension seems to be more closely related to L2 WM than to L1 WM. For instance, Geva and Ryan (1993) noted that L2 WM contributes to predicting performance in L2 reading, whereas L1 WM does not. Thus they maintained that “seemingly parallel L1 and L2 memory measures are not completely interchangeable” (Geva & Ryan, 1993, p. 30). Havik (2005) found that, although WM capacities in L1 and L2 are highly correlated, it is only L2 WM that correlates significantly with L2 reading ability. Harrington and Sawyer (1992) found significant correlations between L2 WM and L2 reading proficiency. Similarly, Walter (2004) found a significant correlation between L2 WM and L2 reading comprehension for both upper- and lower-intermediate L2 learners. Going beyond correlations, Miyake and Friedman (1998) examined the relationship between L1 and L2 listening span and L2 syntactic comprehension through a path analysis. They showed that L2 syntactic comprehension was directly linked to L2 WM capacity. Although there was a moderate relationship between WM capacity for L1 and that for L2, L1 WM capacity was found to be a mediator variable. That is, it was directly related to L2 WM capacity but indirectly related to L2 syntactic comprehension. In sum, research investigating the relationship between WM capacity and comprehension in the L2 provides support for the view that L2 WM capacity is directly related to L2 comprehension, with L1
WM capacity being chiefly a mediator, as maintained by Miyake and Friedman (1998). It is perhaps safe to suggest that the more advanced the individual’s L2 reading skill, the more the processing of L2 is likely to share the same pool of WM resources as the processing of L1 (Carpenter, Miyake, & Just, 1994).

LITERAL AND INFERENTIAL DIMENSIONS OF READING COMPREHENSION

Current theories of L1 text comprehension view reading as an interaction between the reader’s text-based and knowledge-based processes, both of which involve multilevel representations of a text and its content (Kintsch, 1998). In terms of levels of interactive processing, the reader engages first in the linguistic processing of surface-level textual features. This process gradually paves the way to the construction of a text microstructure, which further includes relating propositions that are in close proximity in the text so as to form a coherent semantic whole. When the reader combines the locally built semantic wholes, a textbase is constructed in the form of a macrostructure. The textbase, which captures the text-internal meaning of the passage, contains the propositions embedded in the sentences and their interrelationships. In addition to text-based procedures involving the surface code (e.g., lexical decoding, word-to-text interpretation, syntactic parsing), the extraction of meanings from sentences, and the gradual accumulation of meanings as a result of processing successive sentences, textbase construction further involves the generation of inferences that are necessary for discourse coherence. These are essentially text-connecting inferences (Graesser, Singer, & Trabasso, 1994), technically referred to as “bridging inferences” (Singer, 1994, p. 500), whose function is to relate new to previous information in two distinct ways. First, there are those that are driven by explicit textual features, such as anaphoric references, connectives, signaling devices, transitional phrases, and rhetorical predicates, which normally bind intrasentential and intersentential text constituents (particularly adjacent sentences in the latter case). Kintsch (1998) argued that the term “inference” is a misnomer for these text-connecting devices, because they do not generate new information based on text content through strategic, controlled, and resource-demanding processes of deduction (pp. 189–190). Elsewhere, these devices serve to trigger text-based “local bridging” through routinely generated automatic processes, as opposed to proper text-based inferences, which lead to “global bridging” and require effortful cognitive processes (Ozuru, Dempsey, & McNamara, 2009), because of their implicating the integration of information
located across larger distances and relying on controlled operations that
tap logical and pragmatic resources. It is this latter type of bridging
inferences that actually “fill in the gaps” (p. 231) in what has been
explicitly stated, thereby leading to newly constructed information.
Evidently, the more conceptual gaps there are in text-based information,
the more global bridging inferences need to be made in order to
connect the various ideas in the text so as to generate a coherent whole.

It is safe to posit, at this stage, that the literal level of reading
comprehension, which is generally defined as the reader’s ability to
“gain meaning directly from the print” (Walker, Munro, & Rickards,
1998, p. 88), essentially captures surface code features and textbase
meanings explicitly stated in the text as well as the connecting devices
that bind these text constituents locally. As such, it simply represents the
author’s propositional message, falling short of generating new
information that would extend and refine the textbase on its way to
becoming integrated with a situational representation of what the text is
truly about. That is, it fails to point to what authors mean, even though it
is able to reflect what they say. In this sense, literal reading has been
perceived as failing to provide a deep understanding of text content
(King, 2007) and has been associated with the performance of unskilled
readers, who are thought to be unable to go beyond the information
contained in a text (Walker et al., 1998).

Perfetti (1989) maintained that inferences constitute the vital
distinction between text meaning, as determined by the textbase
representation, and text interpretation, as determined by the situation
model. Although strategically formulated bridging inferences of a global
nature help set up textbase coherence, elaborative inferences (Singer,
1994), which are technically “extratextual inferences” (Graesser et al.,
1994, p. 376), expand upon and embellish textual content to form a
coherent mental representation of the text. In elaborative inferencing,
the inference is derived from readers’ knowledge structures that are
relevant to textual content, requiring them to reason beyond the text in
order to generate new information.

**WM SUBSERVING LITERAL AND INFERENTIAL READING**

It is clear that both the textbase and the situation model are
underpinned by WM resources. The construction of a coherent textbase
requires incoming information from the text to be connected with
information currently active in WM, so as to enable the reader to
integrate successive propositions in a text. Similarly, the construction of
a situation model relies on the maintaining by WM of currently
processed textual information and relevant information retrieved from
LTM with a view toward integration (Andreassen & Bråten, 2010). However, the demands on WM of each layer of mental representation for the text are different. The textbase, which is associated with literal understanding, implicates cognitive operations that do not necessarily carry serious *intrinsic cognitive load* for WM capacity (Sweller, 1994) in relation to inferential comprehension. It is unlikely for cognitive processes involving the handling of explicit textual data (e.g., lexical decoding, syntactic parsing, etc.) or the automatic activation of connecting devices (e.g., anaphoric resolution) to overload WM resources unless the reader’s L2 proficiency level is quite low.

Generating strategic bridging inferences, on the other hand, is relatively more cumbersome, because it requires the replacement of implicit propositional meanings with explicit ones through logical and pragmatic means across sentences and, at times, paragraphs. Implicitly expressed meanings (e.g., syntactically or lexically ambiguous sentences) and text distance place considerable demands on the limited capacity of WM resources.

Nevertheless, it is inferencing of an elaborative nature that constitutes the most resource-demanding process in reading, because, as mentioned above, it implicates the reader’s reasoning beyond the text with a view to creating new information. This, in fact, places more and heavier demands on WM capacity to the extent that, if the requirements exceed the upper bound of capacity limitations, there may be serious deterioration of comprehension processes, with fewer and less accurate inferences generated (Graesser et al., 1994). Consequently, readers may find themselves incapable of constructing the conceptual links between the textbase and the situation model.

It follows that an L2 reader’s WM capacity, if overloaded with low-level cognitive operations, is unable to tackle adequately a complex process like inferential comprehension. First, the trade-off between maintaining local coherence and global coherence suffers seriously, because the reader’s chief concern is with the step-by-step efficiency of the lower-level processes of generating meaning out of sense, reference, and syntax. The focus on lower-level linguistic processes leaves fewer or no available resources to engage in higher-level comprehension processes. Second, if readers cannot construct a proper textbase, it is unlikely for them to generate relevant elaborative inferences, which would prevent a deeper comprehension of the text. More specifically, without a proper textbase, the propositions that are being processed by WM may not trigger the relevant mental representations stored in LTM, impeding the retrieval process and thereby leaving the reader at loose ends. As such, interactions between the text and the reader’s domain knowledge may not materialize, rendering the formation of a coherent mental model of the situation unlikely.
ENHANCING FAMILIARITY THROUGH GENRE CHOICE AND CONTENT MODIFICATION

In addition to the mismatch between textual content and readers’ domain knowledge, comprehension impairment may result from the role text genre plays in the activation and (re)construction of schemas. It is often said that expository texts, for example, demand more attentional resources than narrative texts, because their internal structure is more densely content-laden (Budd, Whitney, & Turley, 1995). Furthermore, their content is not only decontextualized but also written to inform readers about new concepts, generic truths, and technical data for which they may not have extensive background knowledge (Graesser et al., 1994). Rhetorically speaking, expository texts are thought to be less cohesively organized by temporal and causal relationships than narratives (Budd et al., 1995) and to induce the reader to focus more on the propositional textbase (Zwaan, 1994). By contrast, narratives are easier to process because they have a close correspondence to daily events in contextually specific situations that are deeply embedded in one’s perceptual and social experience (Graesser et al., 1994, p. 372). Thus, narratives are more conducive to the reader’s inference generation and mental model construction than, for instance, expository texts.

Inferential comprehension can be enhanced particularly well with a literary text that contains a coherent narrative with a well-developed storyline. As indicated by Kintsch and Rawson (2007), the situation model for a literary text may require construction at more than one level, in that thorough comprehension of a story would require the reader not only to infer the protagonists’ motivations but also to interpret their arguments in light of their conceptual components (p. 221). Therefore, the author’s choice of words, sentence formation, and semantic relationships in the textbase create particular effects that play an important role in the reader’s integration process. Moreover, in processing literary narratives, readers have the added advantage of identifying and associating themselves with the characters, events, and places in the story, as if these were part of their own everyday experiences.

Nevertheless, despite writers’ intentions to be understood by their readers, it is possible for comprehension failures or misinterpretations to occur, even in their own cultural context. Kintsch (1988), for instance, argued that writers’ assumptions about their readers’ knowledge base could actually be wrong. A reader may not have the relevant background knowledge and, therefore, withdraw from engagement with the text. Clearly, the matter is more serious in the case of L2 reading, because almost all narrative texts take for granted the underlying cultural knowledge of native speakers of that language. Hence, it is often quite a challenge for L2 readers to identify and associate themselves with the
characters, events, and places from the target language culture. What is needed for genuine comprehension to take place is some sort of cultural membership which, as Fish (1980) pointed out, leads to the development of “interpretive communities,” through which readers interpret the meaning of a text by virtually “rewriting” it in their minds (p. 182). Needless to say, the rewriting process is based not only on shared values, customs, and assumptions but also on shared rules of textual interpretation, all of which can be considered as “shared knowledge” (Sinclair, 2004, p. 85). It is therefore imperative for any type of reading research in the L2 to take into account the factor of culture-specific interpretive community to address concerns about explanatory adequacy. Otherwise, as shown by Murata (2007), based on readers’ answers given to inferential questions, the same text could be interpreted quite differently by readers from different cultural groups, even resulting in contradictory answers at times.

One way of compensating for the lack of a proper interpretive community in L2 reading research has been the application of linguistic and rhetorical criteria to text modification, with a view to making passages more accessible to L2 readers. These criteria involve vocabulary range, structure control, sentence length, and plot complexity. For example, in most graded readers in English, the degree of structural simplicity determines the selection of syntactic forms, whereas lexical frequency and relevance are instrumental in word choice, as indicated by Hill (2008).

Research findings on text modification, whether in the form of simplification or elaboration, point to their influencing the literal and inferential aspects of reading in diametrically opposed ways. As a case in point, Yano, Long, and Ross (1994), with the use of thematically different texts that were simplified, elaborated, or left unmodified, showed that simplification improves literal understanding, yet it does not enhance inferential comprehension because it removes “the richness in detail and connections that help a reader perceive implicational links” (p. 214). They thus advocated the use of elaboration which, in their view, improves readers’ ability to generate inferences. However, they cautioned, elaboration may hamper readers’ processing of surface-level features, due to the additional information introduced into the text. Similarly, based on their research findings that involved 105 passages from nine textbooks (some authentic and others simplified), Crossley, McCarthy, Louwerse, and McNamara (2007) criticized simplified texts on account of their failure to demonstrate cause-and-effect relationships and to develop plots and ideas adequately. Elsewhere, elaboration in the form of explanatory notes is shown to help reading comprehension only in the L1, reducing comprehension altogether in the event the reading task is in the L2 (Yeung, Jin, & Sweller, 1998).
A further argument against the use of simplification or elaboration is that where these operations have been applied to the text, the readers’ processing load cannot be kept stable. Where there is the need to maintain this stability, neither simplification, which alleviates the intrinsic cognitive load of the reading task, nor elaboration, which increases extraneous load (Sweller, 1994), is appropriate. Likewise, the use of multiple equivalent texts based on readability formulas raises a number of problems in terms of construct validity. Readability is a multifaceted construct involving both text-specific and reader-specific factors (Castello, 2008). Without due consideration given to both, readability formulas can at best yield crude measures of text difficulty and, as such, should not be relied on uncritically. Last but not least, these formulas have been criticized for failing to account for deeper levels of text processing, textual cohesion, syntactic complexity, rhetorical organization, and propositional density (Crossley, Greenfield, & McNamara, 2008).

It follows that, in L2 reading comprehension research, the various issues involved in genre choice and text equivalency could somewhat be alleviated through the use of a single narrative text (preferably one that is a full-scale literary text), so as to be conducive to inference generation and referential situation models. The authenticity of the text could further do away with the problems stemming from simplification or elaboration, used as text modification procedures. In this respect, a possible way out could be the “nativization” (Adaskou, Britten, & Fahsi, 1990, p. 9; Alptekin, 2006) of an authentic text, which involves the sociological, semantic, and pragmatic adaptation of the textual and contextual cues of the text into the reader’s own culture-specific mental framework, while keeping its linguistic and rhetorical content essentially intact.

INTERACTION BETWEEN WM CAPACITY AND CONTENT FAMILIARITY

To our knowledge, with the exception of the two studies done by Hambrick and colleagues involving listening comprehension (Hambrick & Engle, 2002; Hambrick & Oswald, 2005), there are no studies that have investigated the combined effects of both WM capacity and domain knowledge on reading comprehension in the L1. Focusing on listening, Hambrick and colleagues demonstrated that the effects of WM capacity and domain knowledge on memory-based tests of comprehension were independent and additive. In a large-sample study conducted with participants with differing degrees of baseball knowledge, Hambrick and Engle (2002) asked the participants to listen to simulated radio broadcasts of baseball games and then to perform memory tests that required answering multiple-choice and open-ended questions on the
listening passages. WM capacity was measured through operation span and counting span tasks, whereas domain knowledge was determined based on tests of baseball rules, regulations, and terminology, as well as participants’ self-ratings of baseball knowledge. A hierarchical regression analysis revealed that domain knowledge was the strongest predictor of overall memory performance, accounting for almost 55% of variance, followed by WM capacity accounting for almost 5% variability. The results also revealed independent contributions of these factors to test performance, because the interaction between WM capacity and baseball knowledge was not significant. In other words, the relationship between working memory and test performance was found to be similar at low and high levels of domain knowledge. Even when a significant interaction was detected on another performance measure, the combined effects of these variables explained only 1% variability. Elsewhere, Hambrick and Oswald (2005) replicated Hambrick and Engle’s (2002) study by adding a domain-irrelevant task in order to have a more direct observation of the interaction between WM capacity and domain knowledge. The results indicated that the relationship between WM capacity and test performance did not change between domain-relevant and domain-irrelevant tasks, providing further evidence for the independent effects of WM capacity and domain knowledge on test performance.

Research on the topic is also scant in L2 reading comprehension studies. Two recent studies come to mind in this connection. The first, using an RST as a measure of WM capacity, tested the interaction between WM capacity and topic familiarity in text recall of L2 learners of Spanish (Leeser, 2007). It was found that both WM capacity and topic familiarity associated with domain knowledge significantly affected learners’ text recall. Although the interaction between these factors was not statistically significant \(p = 0.058\), post-hoc comparisons showed that learners benefited from higher WM capacity only if they were familiar with the topic. This was interpreted as support for the rich-get-richer hypothesis. However, these findings should be viewed with caution, because the overall \(F\) test was not statistically significant, casting a shadow on the researcher’s overall conclusion. The second study (Payne et al., 2009), done with adult native speakers of English learning Spanish, provides further evidence in support of the independent influences model, in agreement with the findings of Hambrick and his colleagues. Measuring WM capacity through a counting span task and operationalizing domain knowledge as domain experience (the number of Spanish courses taken and the years spent actively learning Spanish), the researchers found that these two factors make significant yet independent contributions to L2 reading comprehension. In sum, research on the combined effects of WM capacity and domain knowledge on L2 reading comprehension is scarce, and findings are inconclusive at best.
THE PRESENT STUDY

Based on the above considerations about treating literal and inferential comprehension separately in L2 reading, assessing L2 WM capacity in literal and inferential L2 reading, enhancing topical familiarity for L2 readers through genre choice and content modification, and exploring the nature of the interaction between WM capacity and content familiarity in relation to the two dimensions of reading, we sought to investigate the following research questions:

1. Does L2 WM capacity affect comprehension accuracy in literal versus inferential reading in the L2?
2. Does content familiarity affect comprehension accuracy in literal versus inferential reading in the L2?
3. Is there an interaction between WM capacity and content familiarity in terms of their effects on literal versus inferential reading in the L2?

It was hypothesized that, in view of the close ties between the storage function of WM and complex cognitive processes (Miyake & Shah, 1999, p. 445), there would be a positive relationship between reading span and inferential comprehension accuracy but that no relationship would be found between reading span and literal understanding because of L2 readers’ propensity for text-biased processing (Hypothesis 1). The next hypothesis, which took into account domain-specific familiarity brought about by textual nativization effects, predicted that nativization would allow for better inferential comprehension, yet not necessarily text-bound literal understanding (Hypothesis 2). Given the few available research findings suggesting independent effects of WM capacity and domain familiarity on comprehension, the third hypothesis predicted no interaction between WM capacity and content familiarity (Hypothesis 3).

METHODOLOGY

Participants

The participants in the present study were Turkish undergraduate students enrolled in an English-medium university in Turkey. They had been successful on the university’s English proficiency test, the minimum pass mark of which is accepted as the equivalent of 550 on the paper-based version of the TOEFL. The students had also obtained high scores on the verbal sections of the national university entrance examination (ÖSS), which is administered in Turkish and is similar to the critical reading section of the SAT Reasoning Test. It was thought that, with a minimum competence level in the range of 550 on the TOEFL and high reading proficiency in Turkish, along with a sufficient
world knowledge base, the students would make use of higher-order reading comprehension processes. Student ages ranged from 20 to 23 years, with an average of 21.24 years. Of the 62 students who participated in the study, 54 were female and 8 were male. They formed a rather homogeneous group in terms of their educational background, in that they had all completed study at a teacher-training high school and were enrolled in university-level English language teaching courses in order to become teachers of English.  

**Materials and Procedures**

Materials for the study consisted of an RST (Daneman & Carpenter, 1980), a short story presented in its original and nativized versions in English, and a reading comprehension test with multiple-choice items based on the two versions of the narrative.

**RST**

Given the direct and language-independent relationship between L2 WM capacity and L2 reading (see earlier under Operationalizing Working Memory), it was deemed appropriate to use an RST measure in English rather than in Turkish, in which no pretested and reliable version exists.

This test consisted of 70 unrelated simple sentences in the active voice, each 11–13 words in length. Each sentence ended with a different word. The test comprised four levels, starting at Level 2 and extending up to Level 5, with each level containing five trials. A grammaticality judgment task was incorporated into the RST to ensure that participants processed every sentence syntactically and did not simply focus on the final words. There were 35 grammatical (e.g., *He looked across the room and saw a person holding a gun*) and 35 ungrammatical sentences (e.g., *The girl picked up her bag and down to went the gym*), arranged randomly. Each sentence was presented only once for participants to judge its grammaticality and to memorize the sentence-final word. The total number of words recalled across all trials was recorded as the storage measure of the participant's reading span. The Cronbach’s alpha for the storage task was found to be 0.872. On the other hand, participants’ judgments concerning sentence grammaticality represented the processing measure of their reading span.

The test, administered in a computer lab, was delivered online by displaying one sentence after another in 7-s intervals until all the

---

2 A teacher-training high school is a secondary school where students, in addition to the regular curriculum, take vocational courses in education to increase their chances of being accepted by colleges of education at tertiary level.
sentences in a set were viewed. While processing the sentences, the participants pressed one of two computer keys to indicate whether a given sentence was grammatical or ungrammatical. After all the sentences in a set had been viewed, a field box appeared on the screen for the participants to enter the sentence-final words that they were able to recall.

Scoring the test involved obtaining composite scores by converting word recall and sentence judgment scores to z-scores and taking their average, in light of Waters and Caplan’s (1996) criticism concerning early RST evaluations prioritizing recall at the expense of processing by focusing solely on storage scores. The participants were then divided into high- and low-WM capacity groups based on their composite scores. In order to maximize the differences between the two groups, the high-WM capacity group had composite scores that were at least a standard deviation above the mean. An independent samples t-test conducted on the composite scores indicated that the group means were significantly different, $t_{60} = 8.89$, $p < 0.001$.

Reading Text and its Nativized Version

The narrative text used for reading comprehension was an American short story by Delmore Schwartz (1978). The story, “In Dreams Begin Responsibilities,” is autobiographical in nature and takes place in New York City in the early 1900s, when immigrants were struggling to find their way in the New World. The two conflicting themes in the narrative are success in business and worldly accomplishment on one hand, and social problems caused by quick financial gains in a new culture on the other.

Through a process of nativization, the story was adapted to the Turkish readers’ own social setting, using a conservative number of textual (semantic) and contextual (conceptual) cues. The textual cues that were nativized in the narrative involved changing data that had to do with settings and locations (e.g., New York City>İstanbul; Brooklyn>Taksim; church>mosque; ocean>sea) as well as with characters and occupations (e.g., motorman>ticket collector; organist>piano player; President Taft>Prime Minister İnönü). Contextual cues that underwent nativization involved culture-specific customs, rituals, beliefs, values, and structures (e.g., relevant changes in holidays, cuisine, clothing, currency, manners). For example, the traditional American Sunday dinner was replaced by a traditional religious holiday meal. Likewise, catering to the more conservative Turkish customs of the time, the characters who actually dated in the original story became an engaged couple in the nativized version. Finally, whereas the protagonist’s actions in the original

---

3 The two groups were compared, using the scores on the university’s English proficiency test and the national university entrance examination (ÖSS). Independent samples t-tests revealed nonsignificant results in terms of both proficiency in the L2 ($t_{60} = 0.213$, $p > 0.05$) and verbal ability in the L1 ($t_{60} = 0.012$, $p > 0.05$).
story typified his sense of American individualism, the Turkish protagonist’s deeds in the nativized version were indicative of the value he placed on family opinion and group solidarity. Such minor modifications allowed the participants to process an unabridged authentic text (the original story) or its close equivalent (the nativized version) as the source of linguistically and rhetorically “authentic” L2 input (see Appendix A for a sample text).

Following the adaptation process, the two texts differed by only 58 words, the original comprising 2,270 words (total of 17 paragraphs) and the nativized version 2,328 words (total of 17 paragraphs). A Latent Semantic Analysis, conducted on the two versions of the text to ensure that their linguistic features were comparable, revealed almost identical numerical values for all selected features. The participants were randomly assigned to two groups, one being exposed to the original text and the other to its nativized version.

**Reading Comprehension Test**

Multiple-choice items were used to measure the participants’ comprehension of the text while reading, instead of a free recall procedure after reading. Despite the controversies surrounding the validity of multiple-choice tests for assessing reading comprehension (e.g., Rupp, Ferne, & Choi, 2006), a multiple-choice format was preferred over free recall, not simply because of the former’s great popularity and provision of scorer reliability but also because of the unsuitability of using free recall questions with the reading text made available. First and foremost, given that the current research design was based on Kintsch’s (1988) construction-integration perspective of reading, in which the construction and integration of both the textbase and the situation model draw on WM resources during reading, it was necessary to make the text available for such reading processes as rereading and searching for relevant information. In addition, WM is shown to consistently predict unique variance on multiple-choice comprehension tests in text available conditions and when the questions tap inferential comprehension (Andreassen & Bråten, 2010). Second, with its strong reliance on memory, free recall makes it difficult to distinguish recalled elements from the text from those retrieved from knowledge bases (Koda, 2005, p. 237). Besides, it primarily relies on stored knowledge from explicit memory rather than delving into information stored in implicit memory. That is, it implicates the conscious

---

4Latent Semantic Analysis was conducted through Coh-Metrix, a computational tool measuring text difficulty and cohesion at various levels of language, discourse, and conceptual analysis (Crossley et al., 2007). The five Coh-Metrix measures that were selected were causal cohesion, connectives, logical operators, coreference, and syntactic complexity.
recall of salient textual information from explicit memory far greater than
the less obvious data, which are more dependent on implicit memory.

Adapted from Pearson and Johnson’s (1978) taxonomy of reading
questions, the test contained a total of 20 items, each with four options,
designed according to criteria for constructing literal and inferential
comprehension questions (Day & Park, 2005; Goldman & Durán, 1988).
Half of the questions were textually explicit, while the other half consisted
of textually or scriptally implicit questions. The former measured readers’
literal understanding, with answers that could be directly derived from the
explicitly stated propositional meanings of the text. The latter, on the
other hand, measured readers’ inferential comprehension, with answers
resting either on the ability to generate bridging inferences to fill in the
conceptual gaps arising from the lack of fully explicit data (textually
implicit), or on the ability to generate elaborative inferences by moving
beyond the text to construct a mental representation of the situation
model to which it referred (scriptally implicit). The questions were
identical for both versions of the text, and they followed the order of
events in the story, irrespective of their being explicit or implicit.
Examples of test items for each category appear in Appendix B.

The participants were instructed to read the text on the computer screen
and to answer the questions, which appeared one by one next to the text.
The text could be scrolled independent of the questions. The participants
were told that their comprehension accuracy scores would be based on the
number of correct answers out of the total number of questions. They had
continuous access to the text throughout the administration of the test.
During the 50 min they were given for the test as a whole, they were free in
terms of the processing time for individual questions.

The test was piloted 4 weeks before the main study for validation
purposes. First, two experts, who were experienced in the teaching and
testing of English as a foreign language, were asked to work independently
and to classify each question as either explicit or implicit. The ratio of
consistent classifications to the total number of classifications revealed
93% agreement between the readers. Questions on which no consensus
was reached were revised. Second, the test was administered to a small but
similar sample of Turkish students in terms of age, academic background,
and L2 proficiency level. Using Cronbach’s alpha, the internal consistency
of the test was found to be 0.71. Very easy and very difficult items were
revised for the subsequent version of the test.

RESULTS

An examination of the descriptive statistics for literal and inferential
comprehension (Table 1) indicated that the variability was similar for
the two types of tasks and the distributions were normal. In general, the participants’ performance was significantly better on literal understanding than on inferential comprehension ($t_{61} = 11.27, p < 0.001$).

Figure 1 provides a graphical display of performance differences on the two types of comprehension.

Both the descriptive statistics and box-and-whisker plots point to the difficulty of inferential comprehension. Moreover, the Pearson product-moment correlation between literal and inferential comprehension was found to be low ($r = 0.20, p > 0.05$), which suggests that they are independent components of reading ability.

To explore whether readers’ WM capacity and the type of content to which they were exposed were related to these two types of comprehension,

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literal comprehension</td>
<td>7.77</td>
<td>1.37</td>
<td>8</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Inferential comprehension</td>
<td>4.77</td>
<td>1.34</td>
<td>5</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

FIGURE 1. Box-and-whisker plots for literal and inferential comprehension.
The marginal means indicate that the low-span ($M = 7.63$) and high-span ($M = 7.96$) participants performed similarly on literal comprehension, yet the mean difference between the two groups was noticeable in terms of inferential comprehension. High-span participants had a higher mean ($M = 5.18$) than the low-span participants ($M = 4.45$).

### TABLE 2
Descriptive Statistics for Literal and Inferential Comprehension in Relation to Working Memory Capacity and Text Version

<table>
<thead>
<tr>
<th>WMC</th>
<th>Text version</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Literal comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Original</td>
<td>7.74</td>
<td>1.24</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Nativized</td>
<td>7.50</td>
<td>1.46</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>7.63</td>
<td>1.33</td>
<td>35</td>
</tr>
<tr>
<td>High</td>
<td>Original</td>
<td>7.58</td>
<td>1.50</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Nativized</td>
<td>8.27</td>
<td>1.33</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>7.96</td>
<td>1.43</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>Original</td>
<td>7.67</td>
<td>1.33</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Nativized</td>
<td>7.87</td>
<td>1.43</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>7.77</td>
<td>1.37</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Inferential comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Original</td>
<td>4.05</td>
<td>1.47</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Nativized</td>
<td>4.93</td>
<td>1.00</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4.45</td>
<td>1.34</td>
<td>35</td>
</tr>
<tr>
<td>High</td>
<td>Original</td>
<td>4.75</td>
<td>1.29</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Nativized</td>
<td>5.53</td>
<td>1.13</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>5.18</td>
<td>1.24</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>Original</td>
<td>4.32</td>
<td>1.42</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Nativized</td>
<td>5.23</td>
<td>1.09</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4.77</td>
<td>1.34</td>
<td>62</td>
</tr>
</tbody>
</table>

Note. WMC = working memory capacity.

descriptive statistics for literal and inferential reading across the levels of the WM capacity and text version were obtained (Table 2).

### TABLE 3
Tests of Between-Subjects Effects for Literal and Inferential Comprehension

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent variable</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Partial η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMC</td>
<td>Literal comprehension</td>
<td>1.418</td>
<td>1</td>
<td>1.418</td>
<td>0.75</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Inferential comprehension</td>
<td>6.308</td>
<td>1</td>
<td>6.308</td>
<td>4.07*</td>
<td>0.07</td>
</tr>
<tr>
<td>Text version</td>
<td>Literal comprehension</td>
<td>0.752</td>
<td>1</td>
<td>0.752</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inferential comprehension</td>
<td>10.496</td>
<td>1</td>
<td>10.496</td>
<td>6.77*</td>
<td>0.11</td>
</tr>
<tr>
<td>WMC× text version</td>
<td>Literal comprehension</td>
<td>3.194</td>
<td>1</td>
<td>3.194</td>
<td>1.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inferential comprehension</td>
<td>0.039</td>
<td>1</td>
<td>0.039</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>Literal comprehension</td>
<td>109.534</td>
<td>58</td>
<td>1.889</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inferential comprehension</td>
<td>89.868</td>
<td>58</td>
<td>1.549</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < 0.05.
on inferential comprehension. A similar pattern was observed between the text versions. Although the means for the original ($M = 7.67$) and nativized ($M = 7.87$) texts were similar on literal comprehension, the mean for the nativized text ($M = 5.23$) was higher than that for the original text ($M = 4.32$) on inferential comprehension.

To determine whether the mean differences were statistically significant and whether there was an interaction between WM capacity and text version in terms of their effects on literal and inferential comprehension, a $2 \times 2$ between-subjects multivariate analysis of variance was performed. The assumptions of normality and homogeneity of variance-covariance matrices were satisfactory.

While the effect of WM capacity on the combined dependent variables approached significance, Wilks’ $\Lambda = 0.901$, $F(2, 57) = 3.10$, $p = 0.053$, the effect of text version was highly significant, Wilks’ $\Lambda = 0.869$, $F(2, 57) = 4.30$, $p = 0.018$. However, no interaction effect on the combined dependent variables was observed.

Univariate between-subjects tests showed that WM capacity and text version were significantly related to inferential comprehension, but not to literal understanding (see Table 3). Thus, high-span readers’ performance on inferential comprehension was considerably higher than that of low-span readers, regardless of the text version to which they were exposed. Moreover, content nativization enhanced inferential comprehension for both high- and low-span readers.

These findings confirm our first hypothesis, which predicted a significant relationship between L2 WM capacity and inferential comprehension but not literal comprehension in L2 reading. Our second hypothesis, which predicted that content nativization would lead to better inferential but not literal comprehension in L2 reading, was also confirmed. Finally, as predicted by our third hypothesis, there was no interaction effect between WM capacity and content familiarity on either literal or inferential comprehension.

**DISCUSSION**

One of the objectives of the present study was to determine whether there was a meaningful relationship between L2 WM capacity and the inferential and literal dimensions of L2 reading comprehension. The first hypothesis predicted a significant relationship between reading span and inferential comprehension accuracy but not literal understanding. The hypothesis was supported, in that while both high- and low-span readers’ performance in literal understanding was similar irrespective of the content type used, as shown by the insignificant difference between their means, high-span readers outperformed low-span readers in inferential comprehension.
comprehension. In tune with L2 readers’ propensity for text-biased processing, both groups seem to have relied heavily on surface-level features, as expected. One possible explanation for both groups’ similar performance on literal understanding may have to do with literal reading ability being chiefly dependent on language ability. Because the participants in the present study showed homogeneity in terms of their L2 proficiency, dealing with explicit textual features did not produce an inordinate amount of cognitive load for either group’s WM, which in fact is said to be an important determinant of syntactic comprehension (Miyake & Friedman, 1998, p. 346). On the other hand, given the vital role played by WM in tackling complex cognitive operations such as inference generation (Singer & Ritchot, 1996), it was the high-span readers who obviously moved beyond the confines of sentence comprehension, with a view to integrating information across sentence boundaries and drawing inferences. This is in line with Miyake and Friedman’s (1998, p. 345) observation that the effects of WM constraints become more manifest between high- and low-span individuals when they perform complex tasks that impose heavy demands on WM.

Moreover, based on Koda’s (2005, pp. 199–200) notion that it is possible for tasks used to measure WM capacity to also measure similar or even identical abilities in reading, one can argue that both inferential reading and WM processing exhibit similar higher-order cognitive operations. In fact, one observes a positive relationship between WM capacity and inferential comprehension, which is not the case for literal understanding. In view of this shared variance or perhaps overlap, it is a foregone conclusion that, if a correlation is to be expected between WM capacity and one of the two dimensions underlying reading, this dimension is likely to be inferential comprehension.

The second objective of the study was to probe the relationship between content familiarity and reading comprehension in terms of its literal and inferential dimensions, with each of these treated separately. In this context, the results also support this second hypothesis, that is, that possessing relevant domain knowledge with regard to textual content facilitates inferential comprehension but does not affect literal understanding in L2 reading. Readers’ performance in coping with inferential comprehension improves, possibly due to familiar(ized) subject matter allowing for more LTM contributions that help them generate more and better inferences. On the other hand, literal understanding remains unaffected irrespective of content familiarity, possibly indicative of L2 readers’ propensity for text-biased processing.

An alternative explanation for this interesting finding could be that, as evidenced in studies on child L1 readers (e.g., Cain & Oakhill, 1999; Oakhill, 1984), it is possible that no direct link exists between a given inference and the text content supporting that inference, despite
prevailing theoretical assumptions that a separation of inferential comprehension from the literal understanding of a text is “difficult” (Perfetti, Landi, & Oakhill, 2007, p. 234). In fact, Perfetti et al. point to certain L1 studies that suggest that literal understanding may not be available to readers when they are about to generate an inference. This is particularly true in cases where generated inferences are based not on strongly related concepts in the text but on less strongly related sets of concepts broadly associated with textual content (Sundermeier, Virtue, Marsolek, & Van den Broek, 2005). Consequently, the LTM contributions that allow for content familiarity to promote inferential comprehension in the L2 do not necessarily cater to the improvement of literal understanding, thereby suggesting that higher-order and lower-order reading operations reflect independent cognitive systems at work.

Finally, the third objective of the study was to investigate the combined effects of WM capacity and content familiarity on literal and inferential comprehension. Such an effect was not observed on either literal understanding or inferential comprehension. These results are consistent with the findings of Hambrick and Engle (2002) and Hambrick and Oswald (2005) conducted in L1 settings. Yet they raise questions about Leeser’s (2007) findings pointing to the rich-get-richer hypothesis. It should be noted that the dependent measures in all three of these studies were measured through memory-based tests, all of which required recall of information after listening or reading. Given the memory focus of these measures, it is possible that they hinder an individual’s ability to fully demonstrate comprehension (Chang, 2006). Although the dependent measure used in the current study is different, the conclusion regarding the lack of interaction between WM capacity and content familiarity is consistent with the findings of Hambrick and colleagues. It is also in tune with evidence for the independent influences of domain knowledge and WM capacity on L2 reading comprehension (Payne et al., 2009), which was based on the dependent variable being measured through reader responses to multiple-choice questions (as in the present study). In sum, it is safe to suggest at this point that WM capacity and content familiarity operate independently, and their effects on L2 reading comprehension are additive rather than interactive.

CONCLUSIONS AND IMPLICATIONS FOR ASSESSMENT AND RESEARCH

The present study underlines the complex nature of WM capacity and reading comprehension in the L2. Assessing WM involves the representation of both its storage and processing functions and how each relates to reading tasks that require effortful controlled processes.
Similarly, operationalizing reading comprehension requires the representation of its multicomponential nature, two dimensions of which are treated in this study: literal understanding and inferential comprehension. The findings suggest that content familiarity, working independently from WM, improves inferential comprehension by providing more opportunities for higher-level operations to cater to the situation model of interpretation, yet content familiarity does not seem to affect lower-level operations characteristic of literal understanding. As such, the findings confirm the generally held view that content familiarity has a positive effect on readers’ performance but the findings delimit this effect to inferential comprehension only.

In view of the role content familiarity plays in inferential reading, testers should be sensitive to the type of domain-specific interference that may originate from the text. Focusing on the representations of the surface code and the explicit features of the textbase may camouflage the interference in question, or what Bachman (1990) called the “passage effect” (p. 138), whose manifestation is likely to emerge in cases eliciting textually and scriptally implicit questions that call for inferencing. In testing, then, reading topics should be selected with a view to minimizing interference from the text, so as to have a fair evaluation of test takers’ inferential reading performance.

Next, inferential comprehension questions should be carefully designed so as not to put any extraneous load on WM capacity, because inferential bridging and elaboration, on their own, place heavier demands on WM as a result of the intrinsic complexity of the tasks they involve. As a case in point, measuring inferential comprehension through free recall tasks is likely to exert an extraneous load on WM operations, because the dependence of such tasks on memory alone can require a significant degree of effortful controlled processing to handle both intrinsic and extraneous loads. Recall of knowledge components, added to those embedded in the comprehension task, may cause cognitive overload. Memory as such can be a construct-irrelevant factor that would “bias our understanding of what readers actually comprehend because teachers and researchers are unable to discern readers’ comprehension of those unrecovered units” (Chang, 2006, p. 537).

From a research viewpoint, the fact that neither WM capacity nor content familiarity seems to have a positive relationship with literal comprehension is a vexing issue. Future research might probe this issue with different learner profiles. For example, it would be interesting to examine the potential for interaction between WM capacity and content familiarity in relation to the L2 proficiency level of the readers involved. Research could also investigate the degree of relationships between WM capacity and the type of inference drawn (elaborative versus bridging).
Other areas of inquiry could encompass an examination of the process of content familiarization (nativization versus simplification versus elaboration) on WM’s frequency and richness of inference generation or of the relationships between WM capacity and the genre of discourse (e.g., narrative versus expository versus descriptive), because findings based on one type of text content and genre may not necessarily be applicable to other types of texts (Alderson, 2000, p. 62). Grabe (2009), for instance, maintained that certain textual genres lead readers to emphasize either a text model or a situation model, relating text-model construction to descriptive texts and situation-model construction to narrative texts (p. 48). In the same vein, Tarchi (2010) pointed to the presence of different inference types being generated in relation to expository texts, depending on whether the text mostly requires activation of domain knowledge of facts (e.g., history) or of meanings (e.g., science).

Last but not least, the hypotheses of the present study could be tested in a further study in which the relationship between L1 and L2 WM capacities is investigated in connection with both L1 and L2 reading comprehension, (in)validating in due course the claim that WM for reading is independent of language (Osaka & Osaka, 1992; Osaka et al., 1993).

ACKNOWLEDGMENTS

This research was supported by a grant (07D602) from the Bogaçi University Research Fund, Istanbul, Turkey. We thank the anonymous reviewers of TESOL Quarterly for their helpful comments.

THE AUTHORS

Cem Alptekin is a professor of applied linguistics in the College of Education at Bogaçi University, Istanbul, Turkey. His current research interests are cognitive and linguistic factors in second language reading, English as a lingua franca, and relationships among culture, cognition, and L2 acquisition.

Gülcan Erçetin is an assistant professor in the College of Education at Bogaçi University, Istanbul, Turkey. Her current research interests are cognitive processes in second language reading and second language learning in multimedia/hypermedia environments.

REFERENCES


**APPENDIX A**

**Sample Section From the Original Text**

My father arrives at my mother's house. He has come too early and is suddenly embarrassed. My aunt, my mother's younger sister, answers the loud bell with her napkin in her hand, for the family is still at dinner. As my father enters, my grandfather rises from the table and shakes hands with him. My mother has run upstairs to tidy herself. My grandmother tells my father that my mother will be down soon. My grandfather opens the conversation by remarking about the mild June weather. My father sits uncomfortably near the table, holding his hat in his hand. My grandmother tells my aunt to take my father's hat. My uncle, twelve years old, runs into the house, his hair tousled. He shouts a greeting to my father, who has often given him nickels, and then runs upstairs, as my grandmother shouts after him. It is evident that the respect in which my father is held in this house is tempered by a good deal of mirth. He is impressive, but also very awkward.

**Nativized Version of the Same Section***
My father arrives at my mother’s house. He has come too early and so is suddenly embarrassed. My aunt, my mother’s younger sister, answers the loud bell with her napkin in her hand, for the family has just had their Bayram meal. As my father enters, my grandfather rises from the table and my father kisses his hand. My mother has run upstairs to tidy herself. My grandmother tells my father that my mother will be down soon. My father conveys her his best Bayram wishes. My grandfather opens the conversation by asking how he is. My father sits uncomfortably near the table, holding his hat in his hand. My grandmother tells my aunt to take my father’s hat and bring him a cup of coffee. My uncle, twelve years old, runs into the house, his hair tousled. He comes to my father, who has often given him liras, kisses his hand, and then runs upstairs, as my grandmother shouts after him. It is evident that the respect in which my father is held in this house is tempered by a good deal of mirth. He is impressive, but also very awkward.

*Nativized parts are underlined.

APPENDIX B

Examples of Test Items

Sample Textually Explicit Literal Question

How does the narrator feel about movies?

a. Movies influence people like drugs do.
b. Movies cause people to forget their problems.
c. Movies turn people into anonymous figures.
d. Movies make people imagine wild things.

Sample Textually Implicit Inferential Question

What is the couple’s lack of interest in the “terrible sun” a sign of?

a. The couple’s insensitivity to heat
b. The couple’s awareness of each other
c. The couple’s indifference to a terrifying future
d. The couple’s inability to see what is in sight

Sample Scriptally Implicit Inferential Question

Which of the following best describes the narrator’s state of mind in the story?

a. Indifferent
b. Obsessive
c. Escapist
d. Assertive