LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

Effects of Stress and Working Memory Capacity on Foreign Language Readers' Inferential Processing During Comprehension

Manpreet K. Rai
Kansas State University

Lester C. Loschky
Kansas State University

Richard Jackson Harris
Kansas State University

Nicole R. Peck
Kansas State University

Lindsay G. Cook
Kansas State University

Although stress is frequently claimed to impede foreign language (FL) reading comprehension, it is usually not explained how. We investigated the effects of stress,

---

This study contains some information that has been presented at the Annual Meeting of the Society for Text and Discourse (2008). This work was supported by funds from the Kansas State University Office of Research and Sponsored Programs. The authors wish to acknowledge the work of Bernardo de la Garza, Patricia Barros, Karlie Mann, Becky Steinert, Breeanna McSpadden, Jamie Travis, and Kristen Geri, who helped in discussion and data collection for the experiment. The authors also wish to thank Randall Engle for providing us with the Automated Operation Span task and Scott Smerchek for programming the experiment. Finally, we wish to thank Yasmin Diaz and Bradley Shaw for help in translating and proofreading our Spanish texts, and the Spanish instructors in the Kansas State University Modern Languages department for encouraging their students to participate in our study.

Correspondence concerning this article should be addressed to Manpreet Rai or Lester Loschky, Department of Psychology, Kansas State University, 492 Bluemont Hall, Manhattan, Kansas 66506–5302. Internet: mkrai@ksu.edu; loschky@ksu.edu

Language Learning XX:X, XXXX 2010, pp. 1–32

DOI: 10.1111/j.1467-9922.2010.00592.x

working memory (WM) capacity, and inferential complexity on Spanish FL readers' inferential processing during comprehension. Inferences, although necessary for reading comprehension, vary in inferential complexity and WM demands. We measured 55 intermediate-level Spanish FL learners' reading comprehension, using questions with three levels of inferential complexity: non-inference (factual), bridging inference (pronoun referent), and pragmatic inference. We measured participants' WM capacity and varied their stress level between blocks using a video camera. Results showed that higher WM learners were more accurate overall. Inference construction during comprehension was negatively related to inferential complexity. Stress increased processing time overall, with a trend toward greater effect on response times (RTs) for questions requiring greater inferential complexity. Higher WM learners showed a greater effect of inferential complexity on RTs than lower WM learners. More generally, and consistent with the Eysenck, Santos, Derekschan, and Calvo's (2007) Attentional Control Theory, analyses showed that higher WM learners strategically traded reading speed (processing efficiency) for greater comprehension accuracy (processing effectiveness), whereas lower WM learners only did so under stress and did so less successfully. Thus, stress impedes FL reading comprehension through interactions between WM capacity and inferential complexity, and such effects are moderated by strategy use.

Keywords foreign language; reading comprehension; working memory; inferences; inferential complexity; stress; anxiety; Attentional Control Theory

Stress, Working Memory, and Foreign Language Inference Construction During Reading Comprehension

For foreign language (FL) learners, reading is both a critically important life skill and an important means of acquiring an FL. However, reading can be a complex and difficult process even in one's native language, and it is even more so in an FL (Harrington & Sawyer, 1992; Horwitz, Horwitz, & Cope, 1986; Lally, 1998; Miyake & Friedman, 1998; Saito, Garza, & Horwitz, 1999). In order to understand the complexities involved, it helps to consider reading as a form of discourse processing that involves multiple levels of knowledge representation (Kintsch, 1998). The reader must first decode the letters and words in a text and then parse the string of words to understand their grammatical relationships. This is the FL reader's first major challenge, and the resulting mental framework is referred to as the *surface-level representation* of a text. Next, the reader must interpret and hold in working memory (WM) the meanings of the propositions that have been decoded from the surface level; this is referred to as the *textbase* level of representation. Then the reader must apply his world knowledge to make sense of the individual propositions, creating a global mental representation of the situation they represent, which is known as the *situation model* level of representation. Finally, the reader may make use of their knowledge of social cues and specific contexts to understand subtle contextual meanings, which are referred to as the *pragmatic level* of representation. FL readers often must spend most of their effort on the first two levels of representation, because they have less lexical and grammatical knowledge (Horiba, 1996; Kembo, 2001; Shimizu, 2005), although a lack of culture-specific knowledge can also cause particular difficulties at the last two levels of representation (Brantmeier, 2003; Kispal, 2008).

Importantly, much of the information in any text is not directly stated but must be inferred by the reader. At the textbase level, the reader must make inferences about the missing connections between separate propositions—for example that "he" refers to a previously mentioned entity, "George"—which are known as bridging inferences (Graesser, Singer, & Trabasso, 1994; Kispal, 2008; Singer, Harkness, & Stewart, 1997). In addition, when creating a situation model of the text, readers must use their prior knowledge of the world to infer missing information implied by the text, which are known as pragmatic inferences (Harris & Monaco, 1978; Johnson-Laird, 1993; Kispal, 2008). For example, when reading "The vase fell on the floor. George went to the kitchen to get a broom and dustpan," the reader may infer a meaningful connection between the two sentences—for example, that because the vase dropped on the floor, it must have shattered, thus causing George to get a broom and dustpan. All of these inferential processes require WM. The reader must (a) hold the propositions of the textbase level in WM while (b) retrieving relevant world knowledge from long-term memory into WM (e.g., vases are usually made of breakable material such as glass, which shatters when it falls on the floor), and (c) use that retrieved knowledge to fill in the missing propositions needed for the situation model to make sense (e.g., the fallen vase must have shattered, causing George to get the broom and dustpan *in order to clean up the glass*). As noted above, because FL readers must frequently expend considerable processing resources simply to decode the surface and textbase levels, there are fewer WM resources left to draw pragmatic inferences at the situation model level, thus making FL reading even more challenging (Horiba, 1996; Kembo, 2001; Shimizu, 2005; Walter, 2004).

In addition, FL learners often learn to read in a classroom environment, which for many of them can be a stressful experience, especially if they have test anxiety, are worried about being negatively evaluated, or have difficulty comprehending the reading material (Horwitz & Cope, 1991; Horwitz et al., 1986; Saito et al., 1999). Second, language acquisition researchers have long argued that stress and anxiety inhibit learning an FL (Horwitz, 2000; MacIntyre, 2002; Saito et al., 1999). However, very little research has clearly shown how stress and anxiety directly affect FL processing (although see Blumenthal et al., 2006; Dewaele, 2002; MacIntyre & Gardner, 1994a, 1994b). Interestingly, given that WM limitations can make FL reading comprehension difficult, particularly when one needs to make pragmatic inferences, we will argue that stress and anxiety may inhibit FL readers' ability to draw such inferences, thus impeding comprehension, and that they do so by consuming their WM resources.

Working Memory Model

Before further discussing the roles of WM and stress in drawing inferences while reading, we will discuss the theoretical construct of WM in more detail. Baddeley's model (Baddeley, 1986, 2000, 2003; Baddeley & Hitch, 1974) is perhaps the best known and established conceptualization of WM. In this model, information in WM is actively manipulated, processed, and temporarily stored until it is either forgotten or encoded into long-term memory. This active functioning of WM is a higher level brain function in the prefrontal cortex and can be affected by emotional context, such as that of anxiety (Gray 2001; Gray, Braver, & Raichle, 2002). Importantly, WM is restricted by capacity and time, such that information can only be held in WM for a brief time (e.g., <1 min) without exerting cognitive effort (Baddeley, 1986, 2000; Cowan, 1997). WM has four major components: the central executive and three "slave" systems, known as the phonological loop, the visuospatial sketchpad, and the episodic buffer (Baddeley, 2000). For our purposes, we will limit our discussion primarily to the central executive and the episodic buffer, both of which have been implicated in reading comprehension of extended text (Baddeley, 2000; Daneman & Carpenter, 1980). The central executive actively processes information temporarily held in WM, regardless of whether it is phonological or visual in nature. The episodic buffer is a workspace in which visual, phonological, and semantic information is integrated by the central executive, and it also serves as a buffer, or way station, between long-term memory and the central executive (Baddeley, 2000, 2003). While reading extended discourse, the central executive uses the goal-directed attentional system to retrieve knowledge from long-term memory to the episodic buffer and process it there for meaning. If WM resources become strained, the stimulus-driven (bottom-up) attentional system gains precedence because the central executive is unable to make use of its resource-intensive goal-directed processes (Eysenck et al., 2007; Gray et al., 2002). In other words, once the individual words read have been phonologically decoded, the central executive must actively maintain them online and integrate both their surface-level meaning and any discourse-level inferential meanings they may have in the episodic buffer. If the central executive is unable to process the discourse-level meanings of what is read, the reader will have comprehension difficulties, resulting in a greater reliance on processing of the surface and textbase levels to comprehend the text.

Working memory capacity (sometimes referred to as WM span) is generally defined as the amount of information that can be both actively processed and stored in a limited period of time (Daneman & Carpenter, 1980), and there are large individual differences in such capacity (Daneman & Carpenter, 1980; Unsworth, Heitz, Schrock, & Engle, 2005). One way to understand WM capacity is in terms of effectiveness (accuracy of comprehension) and efficiency (speed of comprehension) (Eysenck et al., 2007) of information processing by the central executive.

WM and Reading Comprehension

Working memory plays a critical role in all language comprehension (Ardila, 2003; Daneman & Carpenter, 1980; Harrington & Sawyer, 1992; Juffs, 2004; Leeser, 2007; Miyake & Friedman, 1998; Walter, 2004). Daneman and Carpenter (1980) conducted an important series of experiments in which they devised a reading span task that predicted reading comprehension better than a simple digit span task did. Reading ability was positively correlated with WM span, with respect to both processing and storage capacity. Of particular interest for our present purposes, this correlation between WM span and reading was shown in readers' ability to answer factual questions (involving no inferences), pronoun reference questions (involving bridging inferences), and topic summary questions (involving pragmatic inferences) after reading short passages.

Working memory capacity is also important in second language (L2) comprehension for the same reasons as those involved in first language (L1) reading comprehension. For example, Harrington and Sawyer (1992) found that FL readers with higher WM spans performed better on both the reading and grammar/vocabulary sections of the Test of English as a Foreign Language (TOEFL) proficiency exam. Hulstijn and Bossers (1992) tested the influences of both L2-nonspecific and L2-specific factors on FL reading comprehension. They found that at early stages of L2 learning, reading comprehension is dependent on L2-specific knowledge, such as vocabulary and grammar rules, whereas at later stages, L2 reading involves more L2-nonspecific factors, such as WM and L1 reading proficiency. Thus, at higher levels of L2 proficiency, WM becomes increasingly important for FL reading comprehension.

Interestingly, research by Van den Noort, Bosch, and Hugdahl (2006) showed that when moving from a given person's L1 to their L2 and L3, his available WM resources diminish from L1 to L2 and from L2 to L3. Because the language proficiency of their participants also declined from L1 to L2 to L3, this suggests that WM interacts with FL proficiency. Importantly, however, the authors also found significant correlations between the reading span scores of their participants in their three languages "(all $r > .75$, $p < .02$)" (p. 294), strongly suggesting a constant underlying WM capacity regardless of the language one is using (see Osaka & Osaka, 1992; Osaka, Osaka, & Groner, 1993, for similar results). Van den Noort et al. (2006) explained the relationship between the language-independent WM capacity and FL proficiency in the following way. When processing of an FL is less fully automatized, it will require more attention, which is a central executive resource (Service, Maury, & Luotoniemi, 2002). Thus, processing an FL in which one has low to moderate proficiency will lead to a reduction in executive WM resources (Service et al., 2002).

Inferences and WM

A handful of studies have investigated the role of WM in drawing inferences during reading comprehension. These studies have shown that readers with lower WM capacity are less likely to draw pragmatic inferences (Calvo, 2001; Linderholm & van den Broek, 2002) or require more processing time to do so (Estevez & Calvo, 2000) than readers with higher WM capacity. Readers with low WM capacity are also less accurate in identifying the referent of a pronoun (a type of bridging inference) or identifying the general theme of a story (a global inference) (Daneman & Carpenter, 1980). In fact, research suggests that when readers with lower WM capacity do draw inferences, they are more likely to be bridging inferences than pragmatic inferences (St. George, Mannes, & Hoffman, 1997). Together, these results suggest that as the degree of inferential processing required for reading increases, WM capacity becomes more important.

Inferences in L2 Reading

Other studies suggest that there is a positive correlation between FL proficiency and the likelihood of drawing higher level inferences during reading (Horiba, 1996; Kembo, 2001; Shimizu, 2005). Specifically, studies have shown that less proficient FL learners are likely to draw either no inferences or only bridging inferences while reading, whereas more proficient FL learners are more likely to draw inferences, including pragmatic inferences, while reading (Horiba, 1996; Kembo, 2001; Shimizu, 2005). In addition, Hammadou (1991) found that more proficient readers made more inferences supported by the text than did the less proficient learners, who tended to make more unsupported inferences. As noted earlier, research suggests that executive WM resources are less taxed when FL proficiency is higher (Harrington & Sawyer, 1992; Service et al., 2002; Van den Noort et al., 2006). Likewise, executive WM resources constrain the level of inferential processing during reading (Calvo, 2001; Daneman & Carpenter, 1980; Estevez & Calvo, 2000; Linderholm & van den Broek, 2002; St. George et al., 1997). Together, the above two general findings explain why learners with greater FL proficiency would be more likely to draw higher level inferences while reading (e.g., pragmatic inferences) that involve the situation model level of representation (Horiba, 1996; Kintsch, 1998; Shimizu, 2005)—namely, as the FL becomes more automatized, it requires fewer central executive resources, thus leaving more WM resources available for processing pragmatic inferences.

Stress, Anxiety, and WM

Stress increases demands on WM resources, which can be explained by the distraction theory of choking under pressure (Baumeister, 1984). Choking under pressure occurs when people perform poorly under high-stress situations (i.e., "pressure"), although they can perform well under low stress.[1] According to distraction theory, performance deteriorates on cognitive tasks, such as mathematics, or we would argue, drawing inferences during comprehension, that put heavy demands on the executive functions of WM when anxious thoughts distract a person's attention from the task, thus depleting the limited executive resources needed to complete it (Ashcraft & Kirk, 2001; Baumeister, 1984; Beilock & Carr, 2005). Therefore, an important part of understanding the relationship between stress and WM is to understand the relationship between stress and anxiety, which Spielberger (1983) defined as "the subjective feeling of tension, apprehension, nervousness, and worry associated with an arousal of the autonomic nervous system" (p. 1). In addition to this general definition, Spielberger (1972, 1983) further made the distinction between trait and state anxiety, with the former referring to a relatively stable personality variable and the latter referring to a transient emotional state where feelings of anxiety can change based on dangerous or threatening situations. Thus, both state and trait anxiety can mediate the effects of stress in depleting executive WM resources. Consistent with the above ideas, people who are higher in trait anxiety tend to have lower WM capacities (Darke, 1988a; MacLeod & Donnellan, 1993; Sorg & Whitney, 1992), but this depends on whether they are exposed to stressful situations (Sorg & Whitney, 1992).

The ideas of distraction theory have been elaborated upon by Attentional Control Theory (Eysenck et al., 2007). According to Attentional Control Theory, anxiety decreases performance by decreasing goal-driven attentional control—specifically the ability to inhibit information, shift attention, and update information in WM—thus forcing a person to rely more on stimulus-driven attention, which may be drawn to task-irrelevant anxiety-provoking stimuli.

In addition, Attentional Control Theory distinguishes between the effects of stress and anxiety on effectiveness (accuracy of performance), and efficiency (speed of performance). According to Eysenck et al. (2007), anxiety often leads to a speed-accuracy trade-off, as decreasing effectiveness can be compensated for by increasing processing time, thus decreasing efficiency. According to this view, anxiety will more commonly lead to decreases in efficiency than in effectiveness, so long as the person has control over how quickly they respond (Eysenck et al., 2007). Such speed-accuracy trade-offs specifically occur when the central executive is under such taxing conditions (e.g., anxiety) that the attentional resources needed to perform a task are strained, thus impeding performance. If, however, a person must make a speeded response, they cannot trade efficiency for effectiveness, and stress and WM can interact to affect performance in surprising ways. For example, Beilock and Carr (2005) instructed participants to respond "as quickly as possible, without sacrificing accuracy" (p. 103) and found that when the task was most demanding of WM resources, stress caused greater impairments in mathematics performance for those having *higher* WM capacities than those with *lower* capacities. The authors argued that this was because higher WM capacity participants had more WM resources to lose under stress than those who already had lower WM capacity. Importantly, all participants followed the speeded response instructions and actually responded faster under stress, which for the higher-WM participants resulted in reduced effectiveness. This raises the question of what would happen to those with higher WM capacity when under stress if speeded responses are *not* required. If people with higher WM capacity have more WM resources to lose when under stress (Beilock & Carr, 2005), but under stress people trade efficiency for effectiveness when allowed to control their response speed (Eysenck et al., 2007), it suggests that when under stress, those with higher WM capacity would be more likely to trade efficiency for effectiveness than those with lower WM capacity. Such trade-offs would likely involve compensatory strategies such as making regressions in the text, engaging in subvocal rehearsal, and generally taking longer to process the text (Eysenck et al., 2007).

Stress, Anxiety, and L2

Horwitz and Cope (1991; as cited in Aida, 1994) argued that FL anxiety is "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning processes" (p. 31), which can manifest in communication apprehension, test anxiety, and fear of negative evaluation. MacIntyre and Gardner (1994a) argued that stress and resultant anxiety affect virtually all stages of FL processing, including "input" (e.g., attention to items, and encoding them in memory), "processing" (the "organization, storage, and assimilation of the material" [p. 286]), and "output" (e.g., written or spoken language production, or test performance) stages, with effects on FL reading potentially occurring at any of these levels (e.g., encoding of text to WM, organizing it into a coherent situation model, or responding to a test question). Nevertheless, there has been much debate about whether stress and resultant anxiety are an impediment to L2 learning and performance (e.g., Dewaele, Petrides, & Furnham, 2008; Horwitz, 2000; MacIntyre & Gardner, 1994a, 1994b; Saito et al., 1999; Sparks, Ganschow, & Javorsky, 2000). For example, Saito et al. (1999) and Horwitz (2000) found that the trait of FL reading anxiety was negatively correlated with grades in FL courses and positively correlated with reading difficulty. On the other hand, Dewaele et al. (2008) have noted that those with higher self-rated proficiency tend to have less FL anxiety, consistent with the arguments of Sparks et al. (2000). The crux of this debate is the direction of causality among stress, resultant anxiety, and FL processing and acquisition—that is, do stress and anxiety produce worse FL performance, or does worse FL performance produce stress and anxiety? Because correlation does not equal causation, what is required are experimental studies that manipulate stress as an independent variable, and measure its effects on FL processing and/or acquisition dependent variables. A rare example of such an experimental study was that of MacIntyre and Gardner (1994b) who used French as their target L2. To induce stress, they videotaped the learners and showed that (a) doing so increased state anxiety and (b) more anxious learners had greater problems in the input, processing, and output stages of learning than did less anxious learners. Nevertheless, MacIntyre and Gardner (1994b) did not measure FL proficiency and thus cannot exclude the possibility that only less proficient learners showed an increase in anxiety and worse performance. Thus, FL processing and learning can be impaired by the anxiety resulting from the stresses placed upon learners, although this may only be the case for lower proficiency learners (and see Bailey, 1983; MacIntyre & Gardner, 1994a, 1994b; Sheen, 2008, for examples of facilitative anxiety in certain situations). Thus, to better understand the relationship among stress, anxiety, and FL reading comprehension, further experimental research that measures learners' FL proficiency, manipulates their stress, and measures its subsequent effects on their FL reading comprehension is needed.

The Current Study

Predictions

Working memory capacity is a strong predictor of reading comprehension in one's native language (Daneman & Carpenter, 1980; Miyake & Friedman, 1998) and even more so in an FL (Harrington & Sawyer, 1992; Hulstijn & Bossers, 1992). Reading comprehension requires drawing inferences, and the more inferential processes required to comprehend a text, the stronger the influence of WM resource limits (Calvo, 2001; Linderholm & van den Broek, 2002; St. George et al., 1997). Likewise, the less proficient an FL learner is, the less inferential processing in which they will engage (Horiba, 1996; Kembo, 2001; Shimizu, 2005). Importantly, anxiety caused by stress disrupts performance on tasks requiring executive WM resources (Baumeister, 1984), including inferences drawn while reading in one's native language (Darke, 1988b). This may be especially true for those with higher executive WM capacity, if they are forced to make speeded responses, because they have more resources to lose (Beilock & Carr, 2005). However, to the degree that readers can trade speed for accuracy (Eysenck et al., 2007), the effects of anxiety are more likely to produce a greater loss of processing efficiency than effectiveness. Furthermore, because the central executive is heavily taxed by anxiety when under stress, the central executive should be less able to process complex inferential information (especially in an L2).

From the above, we would predict that stress will disrupt FL reading comprehension, with stronger effects on higher level inferences involving the situation model level of representation (i.e., pragmatic inferences) than on lower level inferences such as bridging inferences, or non-inferences, such as memory for facts. We would also predict that such stress effects would be greater for those with higher WM resources but primarily affecting their reaction times.

Based on these general predictions, the present study used three different types of inference questions in an attempt to sample from different representational levels (Kintsch, 1998). First, we used factual questions, which can be answered directly from a single sentence in the text without drawing inferences, thus serving as a non-inference control condition. Next, we used pronominal referent questions. Computing pronoun referents is necessary for discourse coherence and is a simple type of bridging inference. Both factual and bridging inference questions draw primarily from Kintsch's (1998) textbase level, where specific propositions are constructed. Finally, we used pragmatic inference questions, which are highly plausible but not required for a statement to be true or to be meaningfully processed; they require the use of knowledge-based representations from long-term memory (Harris & Monaco, 1978). Such questions draw from Kintsch's situation model level and should require the greatest degree of inferential processing and WM resources. Because FL comprehension might be especially impacted by stress and WM capacity limits, we recruited participants who were intermediate FL learners of Spanish near the end of their fourth semester of study. Based on the above general predictions, the following specific hypotheses were formulated:

Hypothesis 1 (WM). Intermediate FL learners with low WM resources will have lower effectiveness (lower accuracy) and lower efficiency (longer reaction times) answering non-inference questions, bridging inferences, and pragmatic inferences than those with high WM resources. This is based on the wealth of literature showing that WM is a predictor of performance in L2 comprehension (Ardila, 2003; Calvo, 2001; Harrington & Sawyer, 1992; Miyake & Friedman, 1998; Walter, 2004).

Hypothesis 2 (Stress). Intermediate FL learners under high stress will have lower processing efficiency (longer reaction times [RTs]) than those under low stress in answering fact (non-inference), bridging inference, and pragmatic inference questions. This is based on findings showing that stress generally reduces processing efficiency (Eysenck et al., 2007) at the input processing stage in an FL (MacIntyre & Gardner, 1994a, 1994b; Ikeda, Iwanaga, & Seiwa, 1996).

Hypothesis 3 (Inference Type). Intermediate FL learners will show the most effectiveness (greatest accuracy) and efficiency (shortest RTs) for the fact (non-inference) questions, followed by bridging inferences, and, finally, the pragmatic inferences. This is based on the assumption that this hierarchy reflects the degree of inferential processing required by the three question types and that greater inferential processing requires greater WM resources (Calvo, 2001; Darke, 1998b; Linderholm, & van den Broek, 2002; St. George et al.,

1997) and greater FL proficiency (Hammadou, 1991; Horiba, 1996; Kembo, 2001; Shimizu, 2005).

Hypothesis 4 (Interaction of Stress and WM). When under stress, those with high WM resources will show a greater decrement in efficiency (longer RTs) than will those with low WM resources. This is based on the assumptions that (a) stress primarily reduces the WM resources of those having greater resources (Beilock & Carr, 2005) and (b) people under stress tend to trade efficiency for effectiveness if they are able to control their response speed (Eysenck et al., 2007).

Hypothesis 5 (Interaction of Inference Type and WM). Because more WM resources are required for greater inferential processing, particularly for pragmatic inferences (Darke, 1988b; St. George et al., 1997; Whitney, Ritchie, & Clark, 1991), those with lower WM capacity will have lower effectiveness (lower accuracy) and lower efficiency (longer RTs) for questions requiring more complex inferential processing than will those with high WM capacity.

Hypothesis 6 (Interaction of Stress and Inference Type). Because the need for WM resources increases with the degree of inferential processing, and stress reduces WM resources, the effects of stress and resultant anxiety should be greater for questions requiring greater inferential processing (Darke, 1988b). This should primarily be reflected in lower efficiency (longer RTs).

Method

Participants

Fifty-nine intermediate level (fourth semester) Spanish FL students from Kansas State University participated in this study for extra course credit. All were English native speakers. Four participants were eliminated for indicating suspiciousness about the stress manipulation on our "suspiciousness questionnaire" (see below), leaving a sample size of 55. All participants self-reported their proficiency levels using the Demographic Questionnaire for Hispanophones (DQH; see below for more details). Participants' self-reported Spanish reading proficiency levels were quite homogenous, with 89% of participants rating themselves good or average on a 5-point reading proficiency scale (poor = 1, fair = 3, average = 26, good = 21, excellent = 3).

Measures

Suspiciousness Questionnaire

This brief questionnaire was administered in order to determine whether participants had any suspicions that the experiment involved manipulation of their stress level. The questionnaire asked whether anyone had told them anything about the study beforehand, and, if so, what they had been told. Finally, it asked if they had any suspicions about the experiment. All questionnaire responses were read independently by four of the authors, and those participants judged by all four to have suspected that their stress was manipulated were excluded from the data analyses.

Demographic Questionnaire for Hispanophones

This scale assesses participants' background in their Spanish usage, including 10 questions with multiple parts and 5-point Likert scale self-assessments of their reading, writing, listening, and speaking proficiency (Cronbach's alpha = .70). Self-reported proficiency scales have been shown to be highly correlated with objective measures of proficiency (Dufour & Kroll, 1995; Kroll, Michael, Tokowitz, & Dufour, 2002; MacIntyre, Noels, & Clément, 1997) and are widely used in studies of bilingualism (Bialystok, Martin, & Viswanathan, 2005; Colzato, Ramos, van den Wildenberg, & Paolieri, 2008; Dewaele et al., 2008; Van den Noort et al., 2006).

Foreign Language Reading Anxiety Scale

Developed by Saito et al. (1999), "FLRAS elicits students' self reports of anxiety over various aspects of reading, their perceptions of reading difficulties in their target language [in this case Spanish], and their perceptions of the relative difficulty of reading as compared to other language skills" (Saito et al., 1999, p. 204). The Foreign Language Reading Anxiety Scale (FLRAS) was used because it specifically taps into FL *reading* anxiety rather than general FL anxiety, which have been shown to be separate psychological constructs (Matsuda & Gobel, 2004; Saito et al., 1999). The FLRAS consisted of items such as "I get upset when I'm not sure I understand what I'm reading in Spanish" and "I would be happy just to learn to speak Spanish rather than having to learn to read as well." The FLRAS, a trait anxiety scale, contains 20 items, each with a 5-point Likert Scale. The range of possible scores was thus 20–100. The FLRAS was found to have a high internal consistency reliability coefficient (Cronbach's alpha = .86) in the current study, consistent with other published norms for the scale (Matsuda & Gobel, 2004; Saito et al., 1999).

STAI-Trait Anxiety Inventory

The STAI-Trait Anxiety Inventory (TRAIT) scale is a 20-question, 5-point Likert scale developed by Spielberger, Gorsuch, and Lushene (1970). The TRAIT scale was found to have high reliability in our study (Cronbach's alpha = .85).

STAI-State Anxiety Inventory

The *STAI-State Anxiety Inventory* (STATE) also consists of 20 questions, each with a 5-point Likert scale (Spielberger et al., 1970). In the current study, this scale was similarly found to have high reliability (Cronbach's alpha = .90).

Automated Operation Span Task

This computerized version of the Operation Span Task (OSpan) task developed by Unsworth et al. (2005) is an established measure of WM, which was quite reliable in the current study (Cronbach's alpha = .78). The OSpan task particularly taps into the central executive component of WM by forcing participants to remember a string of letters while performing simple mathematical operations between the presentation of each letter. Because the OSpan measures the capacity of the central executive in WM, it predicts reading ability without measuring reading ability itself. For example, it is both highly correlated with reading span tasks (.88), which directly involve reading, as well as with mathematical operation span tasks (.77) that do not (Unsworth et al., 2005).[2] In the same way, the OSpan task seems logically independent of FL proficiency. This is consistent with the idea that each person has an underlying WM span regardless of language, as shown by Van den Noort et al.'s (2006) significant correlations between WM spans across learners' L1, L2, and L3. However, we would also predict, consistent with Van den Noort et al.'s arguments, that performing a reading task in one's L2 will add a processing load that taxes one's executive WM resources. In sum, we used the "language-free" OSpan executive WM measure based on the assumption that executive WM capacity and L2 proficiency are separable but that they interact, such that those lower in L2 proficiency will put a greater strain on their existing central executive WM resources.

Reading Comprehension Task

We created a computerized Spanish reading comprehension task (RCT) to measure FL readers' ability to draw inferences during comprehension. This task consisted of reading 12 short Spanish stories, each of similar length in number of sentences ($M = 13.11$, $SD = 1.43$). The three question types for each passage showed no reliable differences between the number of words for non-inference ($M = 12.75$, $SD = 3.62$), bridging inference ($M = 12.42$, $SD = 3.09$), and pragmatic inference ($M = 11.42$, $SD = 2.50$) questions, $F(2.22) = 0.471$, $p = .79$. The level of reading difficulty for the passages was intended to be something that a typical fourth-semester undergraduate Spanish FL student would find neither overly difficult nor overly easy to comprehend, based on the judgments of three faculty members of the Spanish Language Program (two of them native Spanish speakers), who read and offered revisions of the texts.

Two example passages together with their questions are given in the Appendix. Like many other psychometric measures, we used a two-alternative forced choice (2-AFC) procedure, having a range of .5 (chance) to 1 (perfect). After reading each passage, participants answered three 2-AFC questions: a non-inference (factual) question, a bridging inference (pronoun-referent) question, and a pragmatic inference question. For the pronoun questions, the distance between pronouns and their referents varied from zero to two sentences, which is the typical range of distances found in natural texts (Hobbs, 1977). The RCT task was self-paced with one sentence shown per screen and participants pressing the space bar to proceed to the next sentence. Sentences were presented one at a time in order to require participants to use their WM to integrate their textbase representations across sentences. The measure was quite reliable (Cronbach's alpha = .72).

Design and Procedure

Each participant was tested individually. Upon arrival, they first completed the DQH self-proficiency questionnaire. Next, participants completed the OSpan task. Both tasks were administered in the participants' native language, English. Participants then completed the computerized Spanish RCT. The RCT was divided into two blocks of six passages, each block taking approximately 20 minutes to complete. Prior to reading the experimental RCT passages, participants were given a practice passage to familiarize them with the task.

Stress was a within-participant factor, with every participant being in the high-stress condition for one block of six passages and in the low-stress condition for the other block of six passages, with order of stress and passages randomized and counterbalanced across participants. In order to experimentally manipulate stress, we used a video camera, a technique employed successfully by others to induce anxiety through social evaluative stress (Beilock & Carr, 2005; Calvo, Ramos, & Eysenck, 1993; MacIntyre & Gardner, 1994a). We define *social evaluative stress* here as a situation in which an individual will be publicly evaluated on their performance, consistent with others who have used social evaluation as a stressor (Avero & Calvo, 1999; Dandeneau, Baldwin, Baccus, Sakellaropoulo, & Pruessner, 2007; Garner, Mogg, & Bradley, 2006; Mogg, Philippot, & Bradley, 2004; Viglione & Exner, 1983). Thus, because participants were informed that their video-taped performance would be evaluated by a Spanish professor, based on the Attentional Control Theory, the video camera was the external task-irrelevant anxiety-provoking stimulus, which would divert attention from the task at hand (i.e., the reading comprehension task), thereby hampering performance.

In the high-stress condition, the participants (a) read a Spanish tongue twister aloud into the camera and (b) were told that a Spanish teacher would be evaluating their video-taped performance on the RCT task. In the low-stress condition, the above-described stressors were absent. The cover story regarding the video camera varied depending on whether the high-stress condition occurred in the first or second block of RCT passages. If participants had the high-stress condition in the first block, at the beginning of the low-stress condition in the second block, participants were told that no more video was needed but to please continue to do their best, as their responses would continue to be recorded by the computer. If participants had the high-stress condition in the second block, at the beginning of the second block, participants were told that now that they had practiced the task, they would be video-taped for evaluation by a Spanish professor.

After completing all trials in the RCT task, participants immediately completed the STATE, TRAIT, and FLRAS anxiety questionnaires (Saito et al., 1999; Spielberger et al., 1970) and, finally, a suspiciousness questionnaire. After completing the experiment, all participants were debriefed, the deception of the study was revealed, and a piece of candy was given for mood repair.

Results

A first question was whether WM capacity could predict reading comprehension performance (as measured by accuracy and RT measures) independently of L2 reading proficiency or whether variation in performance, which would otherwise be attributed to WM capacity, could in fact be explained more simply in terms of variations in L2 reading proficiency. To answer this question, simple hierarchical regression analyses were conducted on accuracy and RT data from the RCT task using WM span and self-reported L2 reading proficiency as predictors. In the first regression, WM and self-reported reading skills were the predictors, with accuracy (proportion of correct RCT answers) as the criterion. This analysis showed a significant positive relationship between WM span and accuracy ($\beta = .369$, $t = 2.93$, $p = .005$), indicating that those with greater WM span also had greater accuracy. Self-reported reading skills, however, did not reliably predict accuracy ($\beta = .212$, $t = 1.68$, $p = .099$), indicating that self-reported reading proficiency did not predict accuracy above and beyond WM span alone, $R^2 = .191$, R^2 change $= .045$, $F(1, 52)$ change $= 0.099$, n.s. A similar regression analysis was performed on RT data (time to read and answer a question) and showed that neither WM ($\beta = .026$, $t = 0.184$, $p = .854$) nor self-reported reading proficiency ($\beta = .080$, $t = 0.56$, $p = .573$) reliably predicted reaction times. Thus, in the current study, differences in reading comprehension effectiveness can be attributed to differences in WM independently of participants' proficiency. Because self-rated proficiency did not significantly explain differences in either accuracy or RTs, it was not included in further analyses.

We next conducted two separate mixed factorial ANOVAs for the dependent variables of accuracy and RT. According to the Attentional Control Theory, accuracy is a measure of processing effectiveness and RT is a measure of processing efficiency (Eysenck et al., 2007). The design of the ANOVAs was 2 (WM Capacity: high vs. low, based on a median split of OSpan scores[3]) × 2 (Stress: high vs. low) × 3 (Inference Type: non-inference control, bridging inference, or pragmatic inference). WM capacity was a between-subjects variable, and stress condition and inference type were within-subjects variables. Participants' OSpan scores were distributed quite similarly to the published norms for the measure, with both the current study and Unsworth et al. (2005) having a median score of 59 ($M = 55.97$, $SD = 13.55$).

Accuracy

We first analyzed participants' RCT accuracy as a function of WM capacity, Stress, and Inference Type. Consistent with hypothesis 1, Figure 1 shows a

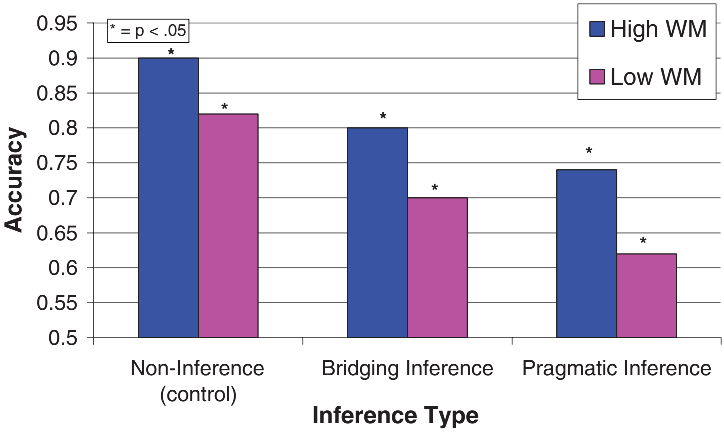

Figure 1 Accuracy as a function of Inference Type for high and low WM.

significant main effect of WM capacity, $F(1, 52) = 9.75, p < .001$, Cohen's $f = .402$, such that participants with higher WM capacity had greater accuracy across all three inference types than those with lower WM. Because Cohen's f values of .10, .25, and .40 are typically considered to be small, medium, and large effect sizes, respectively (Kirk, 1995), these results suggest that WM played a large role in reading accuracy. As can also be seen in Figure 1, consistent with hypothesis 3, participants' accuracy scores showed a significant main effect of inference type, $F(2, 52) = 30.38, p < .001$, Cohen's $f = .426$. Bonferroni multiple comparisons procedures revealed significantly greater accuracy for non-inference control questions than for either bridging inferences or pragmatic inferences (ps $< .01$) and a marginally significant trend for more accuracy for bridging inferences than for pragmatic inferences ($p = .056$), consistent with the idea that greater inferential processing makes comprehension more difficult for intermediate FL learners. Inspection of Figure 1 shows a nonsignificant trend, which suggests that the effect of WM capacity increased across the three inference types as more inferential processing was required, although this interaction was not reliable. Furthermore, there was no significant main effect of stress on accuracy. In fact, there were no statistically significant interactions between inference type and WM (Fs < 1) for accuracy or any interactions of stress with WM or inference type on accuracy (all Fs < 1).

A stepwise regression was also conducted in order to assess the variance in accuracy accounted for by the above factors as well as the self-reported anxiety measures. Consistent with the ANOVA, it was found that WM capacity significantly predicted 16% of the variance in accuracy across all question types ($R^2 = .160, p = .003$), showing that increasing WM capacity led to greater accuracy ($\beta = .401, t = 3.15, p = .003$). However, the STATE ($M = 37.18$, $SD = 8.77$) self-report anxiety questionnaire further predicted an additional 8% of the variance in accuracy above and beyond that of WM ($R^2 = .243$, change in $R^2 = .083, p = .022$). This showed that those with higher STATE anxiety had less accurate performance overall than did those with lower STATE anxiety ($\beta = –.288, t = –2.359, p = .022$). The TRAIT ($M = 37.83, SD = 8.73$) and FLRAS ($M = 54.03, SD = 9.75$) measures did not significantly predict accuracy. t-Test analyses revealed that there was no difference in the STATE scores for those who received stress first ($M = 38.12, SD = 7.99$) and those who received stress last ($M = 36.33, SD = 9.87$), t(46) = 0.383, $p = .493$. Because the questionnaires were administered after the completion of the RCT task, any stress effects may have dissipated before the participants answered the STATE anxiety questionnaire. However, the STATE anxiety measure predicted 8% of the variance in accuracy, regardless of the stress manipulation.

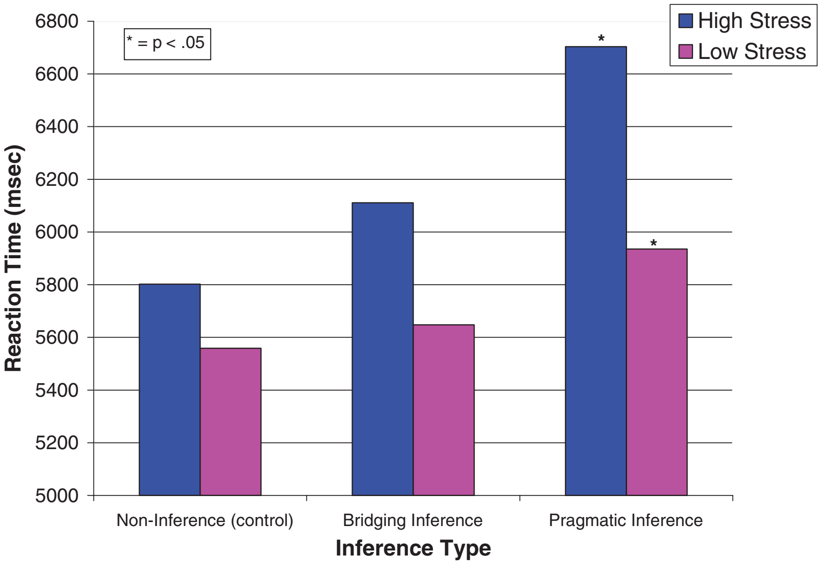

Figure 2 Reaction time as a function of Inference Type for high and low stress.

Reaction Time

We next analyzed RT effects. We first trimmed the RT data, eliminating cases that were 2.05 *SD*s from the mean (the top and bottom 2% of the distribution), conditionalized on participant and stress condition, leaving a total of 1,904 observations.[4] We then carried out the same ANOVAs as described earlier. Importantly, Figure 2 shows that, consistent with hypothesis 2, stress significantly increased RT (i.e., decreased efficiency) across all inference types, $F(1, 52) = 10.43$, $p = .002$, Cohen's $f = .171$. In addition, Figure 2 shows that, consistent with hypothesis 3, there was a main effect of inference type, $F(2, 104) = 15.04$, $p < .001$, Cohen's $f = .294$, such that greater inferential complexity required greater processing time. Specifically, Bonferroni corrected post hoc pairwise comparisons revealed that there were significant differences in RT between all three question types: non-inference questions ($M = 5{,}910$ ms, $SE = 236$) were faster than bridging inference questions ($M = 6{,}329$ ms, $SE = 218$), $p = .044$, or pragmatic inference questions ($M = 6{,}908$ ms, $SE = 278$), $p < .001$; bridging inference questions were faster than pragmatic inference questions, $p = .006$. Contrary to hypothesis 6, there was no interaction between stress and inference type, although inspection of Figure 2 shows a nonsignificant trend in the data in the predicted direction, with stress seeming to have a greater effect on RTs for those questions requiring greater inferential processing.

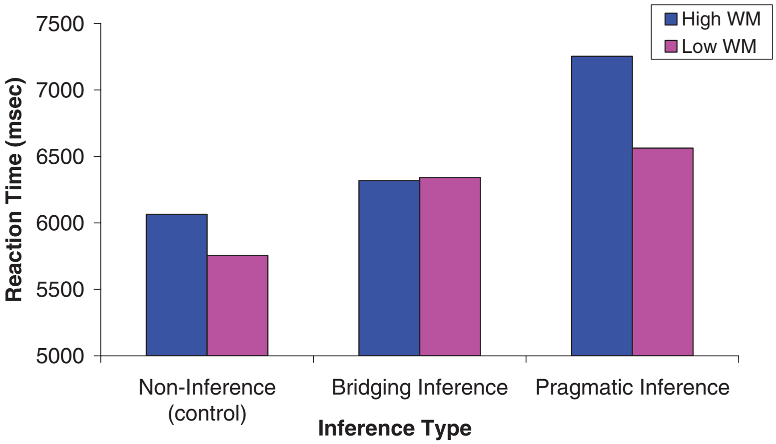

Figure 3 Reaction time as a function of Inference Type for high and low WM.

Working memory capacity was found to have no effect on RT ($F < 1$), nor was there a significant interaction between WM capacity and inference type on RT, $F(2, 104) = 1.90, p = .154$. Nevertheless, Figure 3 shows a nonsignificant trend for participants with higher WM capacities to take longer than those with lower WM capacities for non-inference questions and pragmatic inference questions, which would be in apparent contradiction to hypothesis 5.

We also carried out a stepwise regression analysis to determine the effects of the self-report anxiety measures on RTs. It was found that the FLRAS predicted 11% of the variance in RT ($R^2 = .110, p = .014$). Specifically, those with higher FL reading anxiety took longer to respond across all question types ($\beta = .332$, $t = 2.531$, $p = .014$). The other self-report state and trait anxiety measures did not significantly predict RT. *t*-Test analyses revealed that there was no difference in the FLRAS scores for those who received stress first ($M = 52.75, SD = 10.58$) and those who received stress last ($M = 55.75, SD = 9.33$), $t(46) = -1.04$, $p = .303$. This lack of effect of the stress manipulation is not surprising, as FLRAS is a measure of a trait, which is an enduring disposition. In sum, those with higher FL reading anxiety have lower processing efficiency.

Finally, there was a significant stress by WM interaction for RT, $F(1, 52) = 5.69, p = .021$, Cohen's $f = .120$. Figure 4 shows that stress increased RT, but this increase was more pronounced in those with low WM than in those with high WM. This is reflected in a negative correlation between WM capacity and the effect of stress on RT ($r = -.268$, $p = .025$). Although hypothesis 4 correctly predicted that processing efficiency would differentially decrease (i.e.,

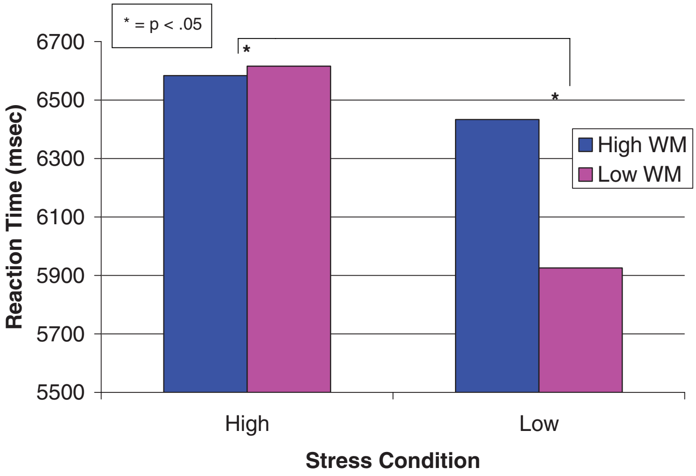

Figure 4 Reaction time as function of stress for high and low WM.

increasing RTs) between the high- and low-WM groups, the trend across WM groups and stress conditions was in the opposite direction to that predicted by hypothesis 4. Specifically, we had predicted that under high stress, those with high WM would show a greater decrease in processing efficiency (i.e., a greater increase in RTs) than those with low WM. However, Figure 4 shows that under *low* stress, RTs were longer for readers with higher WM capacity than for those with lower WM.

Discussion

The current study investigated the roles of stress and WM capacity in FL inferential processing during reading comprehension. Inferences are critical for reading comprehension but are demanding of WM resources and thus pose a particular challenge to FL learners (Ardila, 2003; Daneman & Carpenter, 1980; Hammadou, 1991; Horiba, 1996; Miyake & Friedman, 1998; Walter, 2004). Based on previous research, we predicted that both stress, which situationally reduces WM resources, and readers' inherent WM resource limits would affect FL inferential processing during comprehension, especially because stress reduces the resources available for WM to complete the task. Our results were largely consistent with both predictions. As expected, greater WM resources led to greater FL comprehension for all three levels of inferential complexity. In addition, as predicted, stress exacerbated the difficulty of comprehending text in an FL, by reducing processing efficiency (increasing processing time) for all three levels of inference. Also consistent with predictions, participants took longer to respond to pragmatic inference questions than to bridging inference questions, which, in turn, took longer than non-inference questions. There was also a nonsignificant trend for stress to interact with this ordering of inferential complexity such that the effects of stress on processing efficiency (RTs) appeared to be greatest when the required inferential processes were maximal. This is consistent with previous research showing that stress and resultant anxiety decreases available WM resources (Baumeister, 1984; Beilock & Carr, 2005; Darke, 1988a; Eysenck et al., 2007; Sorg & Whitney, 1992) and that increased inferential processing requires greater WM resources (Darke, 1988b; St. George et al., 1997). Thus, when participants were under stress, either the retrieval of relevant prior knowledge from long-term memory or the inferential gap-filling processes (or both) became more difficult, leading to longer processing times. The fact that stress did not affect processing effectiveness (accuracy) for any of the question types is consistent with the Attentional Control Theory (Eysenck et al., 2007). According to this theory, whenever possible, people will try to compensate for the loss of WM resources caused by stress by engaging in compensatory strategies that require more processing time for the task.

In apparent contradiction to our predictions, there was a nonsignificant trend for those with higher WM resources to take longer on both fact and pragmatic inference questions than those with lower WM resources (Figure 3). It seems puzzling that readers with greater WM resources would show a trend to take longer to answer reading comprehension questions than those with fewer WM resources. However, this result can be better understood in terms of the speed-accuracy trade-offs engaged in by both the higher and lower WM learners. Specifically, the higher WM learners also showed greater accuracy on all question types. Thus, it appears that learners with higher-WM resources had enough executive resources to do well in the RCT task but only if they allocated extra processing time to it. Conversely, learners with lower-WM resources may have felt overwhelmed by many of the RCT questions, particularly those requiring more inferential complexity, and therefore did not bother to allocate more processing time to them.

Consideration of speed-accuracy trade-offs also helps to make sense of the otherwise puzzling interaction between WM capacity and stress, which is shown in Figure 5. Based on the results of Beilock and Carr (2005), together with Eysenck et al.'s (2007) Attentional Control Theory, we had predicted that readers with higher WM capacities would show a greater effect of stress on processing efficiency (i.e., RTs) than readers with low WM capacities.

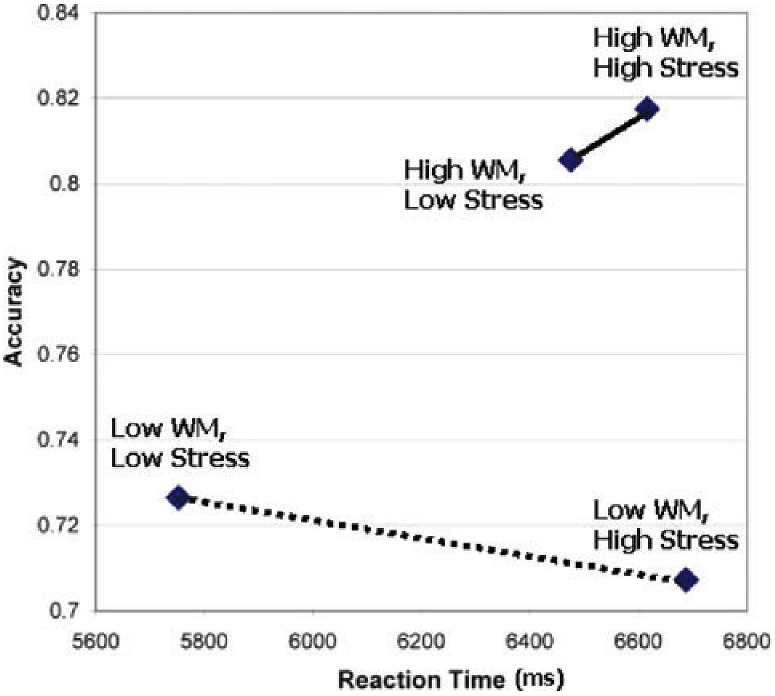

Figure 5 Speed-accuracy trade-off as a function of WM capacity and stress.

However, we found the opposite—namely, a larger effect of stress on those with low WM capacity. What was most surprising was to find that under low stress, the readers with more WM resources took more processing time than those with fewer WM resources. The speed-accuracy trade-offs engaged in by the higher and lower WM learners under low versus high stress helps to understand these results. As shown in Figure 5, for higher WM learners, going from low to high stress produced a small (nonsignificant) increase in RT together with a concomitantly small (nonsignificant) increase in accuracy. Thus, higher WM learners had the WM resources needed to successfully trade processing efficiency for effectiveness when under stress.

This was not the case for the lower WM learners, for whom the effect of going from low to high stress produced a large (and significant) decrease in processing efficiency (i.e., an increase in RT) but also a small (nonsignificant) *decrease* in effectiveness (accuracy), reflecting an overall drop in performance due to stress. Wickens (1984) noted that such negative speed-accuracy trade-offs commonly occur in tasks imposing large WM loads, because longer RTs can result in greater decay from short-term storage. In the current study, increased decay from WM with longer RTs could be particularly problematic for those with lower WM resources because participants could not reread previously presented sentences but instead had to rely on their WM of the text. The above discussion helps shed light on the fact that the higher WM learners took longer to respond when under low stress than the lower WM learners. It appears that the higher WM learners responded to this challenging FL reading task with a consistent strategy of trading efficiency for effectiveness, regardless of the stress level, and that this was a successful strategy for them.[5] Conversely, the lower WM learners only engaged in this strategy when they were under stress, but they lacked the resources to do so effectively, thus showing negative effects of stress. Thus, those teaching reading in an FL should realize that pragmatic inference questions (i.e., those requiring the learner to use their prior knowledge to build situation models of the text) may take more time and effort to process than simpler factual or bridging inference questions and that social evaluative stress in the classroom (e.g., during reading tests) can exacerbate these processing difficulties.

Although numerous studies have shown that self-report and objective measures of proficiency are highly correlated (Dufour & Kroll, 1995; Kroll et al., 2002; MacIntyre et al., 1997), further research may want to include both types of proficiency measures. Furthermore, although the current study only allowed participants to read text line by line, in order to require them to maintain information in WM (e.g., the episodic buffer), reading line by line is admittedly unnatural. Thus, in future studies we might instead use an eyetracker to determine whether WM capacity, stress, and inferential complexity affect reading rates, fixation durations, and the rate of regressive eye movements, all of which are natural ways of trading processing efficiency for effectiveness and, in the case of regressions, are a means of refreshing the storage component of WM (Ehrlich & Rayner, 1983).

Previous studies have suggested that stress negatively affects FL reading comprehension (Horwitz, 2000; Saito et al., 1999). Because reading clearly provides learners with an excellent opportunity to encounter comprehensible input in an FL, the question then becomes *how* stress might reasonably be expected to reduce comprehension, given what we know about factors influencing reading comprehension processes. The current study suggests that individual differences in WM capacity, together with the degree of inferential complexity required to comprehend a piece of text, may play critical roles in the effects of stress on FL comprehension. Having a better understanding of these complex processes should help improve FL reading pedagogy, which is becoming increasingly important in our more globalized community. For example, the current study suggests that social evaluative stress, which commonly occurs in classroom settings, increases the time FL learners need in order to comprehend text, particularly when they must draw high-level inferences. In addition, the current study suggests that readers who have more WM resources may in some cases actually spend more time processing complex inferences in text than readers with fewer WM resources. Although our results are based on a somewhat unnatural reading task, both results may suggest that encouraging students to take time to read carefully when they are under stress and giving them that extra time may aid their comprehension. The current study seems to suggest that such additional processing time might not be particularly helpful for students with lower WM capacity; however, under normal reading conditions, students can use extra processing time to reread the text in order to refresh their WM, thus circumventing the problem faced by the lower WM learners in our study (i.e., the decay of WM with longer processing times because they could not reread previous sentences). This speed-accuracy trade-off would seem to be particularly important during reading tests, which combine social evaluative stress with time limits—and provides a rationale for allowing extra time on such tests. Further research should address the question of whether such differences in comprehension also translate into differences in retention and learning of FL forms (Loschky, 1994).

Revised version accepted 9 September 2009

Notes

1 As used here, the terms *stress* and *pressure* both refer to external factors (such as being video-taped) that can cause *anxiety*, which, in contrast, is an internal affective state.

2 There are various types of WM span tasks, including digit span tasks (e.g., Baddeley & Hitch, 1974), reading span tasks (e.g., Daneman & Carpenter, 1980), and operation span tasks (e.g., Unsworth et al., 2005). The digit span task requires participants to recall numbers while performing a secondary distractor task between the presentation of the digits, and it is generally assumed to primarily tax the storage capacity of the phonological loop (Unsworth et al., 2005). Reading span tasks, in contrast, involve not only the storage function of WM but also the processing function, by requiring participants to recall the last word in each of a series of sentences while also reading for comprehension, as measured by decisions about the truth value of each sentence. Operation span tasks (e.g., Unsworth et al., 2005) are similar in that they require both the storage and processing functions of

WM, by requiring that participants recall a series of letters, presented after each of a series of mathematical operations. Importantly, the operation span task is predictive of reading ability without involving reading itself, by virtue of tapping into the central executive component of WM, which is why we used it rather than a digit span or reading span task.

3 A median split was used, consistent with numerous studies in the WM literature (Beilock & Carr, 2005; Unsworth et al., 2005). Regression analyses using the OSpan as a continuous variable are also reported in the current study.

4 Reaction times were conditionalized on participant and stress condition in order to ensure that we neither trimmed out specific participants' data more than others (i.e., the faster or slower participants) nor did we preferentially trim data from faster or slower stress conditions (e.g., trimming the faster responses primarily from the low-stress condition and trimming slower responses primarily from the high-stress condition) but instead trimmed the excessively slow or fast individual responses of each subject in each stress condition. In this way, in our ANOVAs, each participant's mean RT for each stress condition is based on an equivalent and unbiased trimming procedure.

5 Kroll et al. (2002) noted that FL learners with higher WM capacity spend more time than learners with lower WM capacity attempting to determine whether cognates are real cognates or false cognates. This would certainly use more resources and processing time and thus is a possible explanation for our results.

References

Aida, Y. (1994). An examination of Horwitz, Horwitz and Cope's construct of foreign language anxiety: The case of students of Japanese. *Modern Language Journal, 78*(2), 155–168.

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders, 36*, 233–240.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*(2), 224–237.

Avero, P., & Calvo, M. G. (1999). Emotional reactivity to social-evaluative stress: Gender differences in response systems concordance. *Personality and Individual Differences, 27*(1), 155–170.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews: Neuroscience, 4*, 929–939.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory,* Vol. 111. New York: Academic Press.

Bailey, K. (1983). Competitiveness and anxiety in adult second language learning: Looking at and through the diary studies. In H. W. Seliger & M. H. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 67–102). Rowley, MA: Newbury House.

Baumeister, R. F. (1984) Choking under pressure: Self-consciousness and paradoxical effects on incentives on skillful performance. *Journal of Personality and Social Psychology, 46*(3), 610–620.

Beilock, S. L., & Carr, T. H. (2005). When high powered people fail: Working memory and "choking under pressure" in math. *Psychological Science, 16*(2), 101–105.

Bialystok, E., Martin, M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism, 9*, 103–119.

Blumenthal, P., Britt, T. W., Cohen, J. A., McCubbin, J., Maxfield, N., Michael, E. B., et al. (2006). Stress effects on bilingual language professionals' performance. *International Journal of Bilingualism, 10*(4), 477–495.

Brantmeier, C. (2003). Beyond linguistic knowledge: Individual differences in second language reading. *Foreign Language Annals, 36*(1), 33–44.

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations while reading. *Memory, 9*(6), 365–381.

Calvo, M. G., Ramos, P., & Eysenck, M. W. (1993). Text anxiety and reading: Efficiency vs. effectiveness. *Cognitiva, 5*(1), 77–93.

Colzato, L. S., Baja, M. T., Van Den Wildenberg, W., & Paolieri, D. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*(2), 302–312.

Cowan, N. (1997). The development of working memory. In N. Cowan (Ed.), *The development of memory in childhood* (pp. 163–200). Hove, UK: Psychology Press.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory and Cognition, 23*, 166–180.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Dandeneau, S. D., Baldwin, M. W., Baccus, J. R., Sakellaropoulo, M., & Pruessner, J. (2007). Cutting stress off at the pass: Reducing vigilance and responsiveness to social threat by manipulating attention. *Journal of Personality and Social Psychology, 93*(4), 651–666.

Darke, S. (1988a). Anxiety and working memory capacity. *Cognition & Emotion, 2*(2), 145–154.

Darke, S. (1988b). Effects of anxiety on inferential reasoning task performance. *Journal of Personality and Social Psychology, 55*(3), 499–505.

Dewaele, J.-M. (2002). Individual differences in L2 fluency: The effect of neurobiological correlates. In V. Cook (Ed.), *Portraits of the l2 user* (pp. 219–250). Clevedon, UK: Multilingual Matters.

Dewaele, J. M., Petrides, K. V., & Furnham, A. (2008). Effects of trait emotional intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning, 58*(4), 911–960.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior, 22*(1), 75–87.

Estevez, A., & Calvo, M. G. (2000). Working memory capacity and time course of predictive inferences. *Memory, 8*(1), 51–61.

Eysenck, M. W., Santos, R., Derekshan, N., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional Control Theory. *Emotion, 7*(2), 336–352.

Garner, M., Mogg, K., & Bradley, B. (2006). Orienting and maintenance of gaze to facial expressions in social anxiety. *Journal of Abnormal Psychology, 115*(4), 760–770.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395.

Gray, J. (2001). Emotional modulation of cognitive control: Approach-withdrawal states double-dissociate spatial from verbal two-back task performance. *Journal of Experimental Psychology, 130*(3), 436–452.

Gray, R., Braver, T., & Raichle, M. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences, 99*(6), 4115–4120.

Hammadou, J. (1991). Interrelationships among prior knowledge, inference, and language proficiency in foreign language reading. *Modern Language Journal, 75*(1), 27–38.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25–38.

Harris, R. J., & Monaco, G. E. (1978). Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology: General, 107*, 1–22.

Hobbs, J. R. (1977). Pronoun resolution. *SIGART Newsletter, 61*, 28.

Horiba, Y. (1996). Comprehension processes in L2 reading: Language competence, textual coherence, and inferences. *Studies in Second Language Acquisition, 18*(4), 433–473.

Horwitz, E., Horwitz, M., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70*(2), 125–132.

Horwitz, E. K. (2000). It ain’t over ’til it’s over: On foreign language anxiety, first language deficits, and the confounding of variables. *Modern Language Journal, 84*(2), 256–259.

Horwitz, M. B., & Cope, J. A. (1991). Foreign language classroom anxiety. In E. K. Horwitz & D. L. Young (Eds.), *Language anxiety from theory and research to classroom implications* (pp. 27–39). Englewood Cliffs, NJ: Prentice Hall.

Hulstijn, J. H., & Bossers, B. (1992) Individual differences in L2 proficiency as a function of L1 proficiency. *European Journal of Cognitive Psychology, 4*(4), 341–353.

Ikeda, M., Iwanga, M., & Seiwa, H. (1996). Test anxiety and working memory system. *Perceptual and Motor Skills, 82*(3, Pt. 2), 1223–1231.

Johnson-Laird, P. N. (1993). The mental representation of the meaning of words. In A. I. Goldman (Ed.), *Readings in philosophy and cognitive science* (pp. 561–583). Cambridge, MA: MIT Press.

Juffs, A. (2004). Representation, processing, and working memory in a second language. *Transactions of the Philological Society, 102*(2), 199–225.

Kembo, J. A. (2001). Testing of inferencing behaviour in a second language. *International Journal of Bilingual Education and Bilingualism, 4*(2), 77–96.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Kispal, A. (2008, May). *Effective teaching of inference skills for reading* (Research Report DCSF-RR031). Cheshire, UK: National Foundation for Educational Research, Department of Education (Division of Children, School and Families).

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical influency in a second language. *Second Language Research, 18*, 141–175.

Lally, C. (1998). Friend or foe: The use of direct object pronouns by native English speakers in both English and French. *Reading Improvement, 35*(1), 31–37.

Leeser, M. J. (2007). Learner based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning, 57*(2), 229–270.

Linderholm, T., & Van Den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology, 94*(4), 778–784.

Loschky, L. C. (1994). Comprehensible input and second language acquisition: What is the relationship? *Studies in Second Language Acquisition, 16*, 303–323.

MacIntyre, P. D. (2002). Motivation, anxiety and emotion in second language acquisition. In P. Robinson (Ed.), *Individual differences in second language acquisition* (pp. 45–68). Amsterdam: Benjamins.

MacIntyre, P. D., & Gardner, R. C. (1994a). The effects of induced anxiety on three stages of cognitive processing in second language learning. *Studies in Second Language Acquisition, 16*(1), 1–17.

MacIntyre, P. D., & Gardner, R. C. (1994b). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44*(2), 283–305.

MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning, 47*, 265–287.

MacLeod, C., & Donnellan, A. M. (1993). Individual differences in anxiety and the restriction of working memory capacity. *Personality and Individual Differences, 15*(2), 163–173.

Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System, 32*, 21–36.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.

Mogg, K., Philippot, P., & Bradley, B. (2004). Selective attention to angry faces in clinical social phobia. *Journal of Abnormal Psychology, 113*(1), 160–165.

Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English readings span tests. *Bulletin of the Psychonomic Society, 30*, 287–289.

Osaka, M., Osaka, N., & Groner, R. (1993). Language independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychonomic Society, 31*, 117–118.

Saito, Y., Garza, T. J., & Horwitz, E. K. (1999). Foreign language reading anxiety. *Modern Language Journal, 83*(2), 202–218.

Service, E., Maury, S., & Luotoniemi, E. (2002). Individual differences in phonological learning and verbal STM span. *Memory and Cognition, 35*(5), 1122–1135.

Sheen, Y. (2008). Recasts, language anxiety, modified output, and L2 learning. *Language Learning, 58*(4), 835–874.

Shimizu, M. (2005). Effects of a question on generating bridging inferences: L2 reading proficiency and a qualitative analysis of reading processes. *Bulletin of the Chubu English Language Education Society, 34*, 353–360.

Singer, M., Harkness, D., & Stewart, S. (1997). Constructing inferences in expository text comprehension. *Discourse Processes, 24*, 199–224.

Sorg, B., & Whitney, P. (1992). The effect of trait anxiety and situational stress on working memory capacity. *Journal of Research in Personality, 26*(3), 235–241.

Sparks, R. L., Ganschow, L., & Javorsky, J. (2000). Déjà vu all over again: A response to Saito, Horowitz, and Garza. *Modern Language Journal, 84*(2), 251–255.

Spielberger, C. D. (1972). *Anxiety: Current trends in theory and research: I*. New York: Academic Press.

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (STAI)*. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. C., & Lushene, R. E. (1970). *Manual for the State Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

St. George, M., Mannes, S., & Hoffman, J. (1997). Individual difference in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience, 9*(6), 776–787.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research and Methods, 37*(3), 498–505.

Van Den Noort, M. W., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist, 11*(4), 289–296.

Viglione, D. J., & Exner, J. E., Jr. (1983). The effects of state anxiety and limited social evaluative stress on the Rorschach. *Journal of Personality Assessment, 47*(2), 150–154.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics, 25*(3), 315–339.

Whitney, P., Ritchie, B. G., & Clark, M. B. (1991). Working memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes, 14*, 133–145.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63–101). New York: Academic Press.

Appendix

Below are two example reading comprehension text passages and their accompanying questions. The passages given below are in English but were presented in Spanish to learners in the study. The question-type labels ("Bridging Inference," "Non-inference," and "Pragmatic Inference") were not seen by participants in the experiment.

Example Passage 1 and questions:
Since we are thinking of adopting a new pet, yesterday my family and I visited the animal shelter. It is a sad yet hopeful place. There is a room filled with puppies only. Unfortunately, most people want young dogs, so the old dogs don't get much attention. One room was completely filled with breeds of larger dogs. My father spent most of his time looking at these larger breeds. My mother and I walked around, trying to find where the cats were located. We were surprised to note that the shelter also had rare animals. My mother thought the pig and chicken were disgusting. We saw both a small lizard and a

hen waiting to be adopted! Vicente, my brother, saw a pot-bellied pig. As my mother was explaining how dirty farm animals are, the pig suddenly snorted at him!

Bridging Inference: At whom did the pig snort? a. father *b. Vicente
Non-inference: What type of dogs weren't getting much attention? *a. older dogs b. puppies
Pragmatic Inference: Did the narrator's father want to get a small dog? a. yes *b. no

Example Passage 2 and questions:
There are fifty students in young Santino's economics class. Unfortunately, the four most annoying people in the class happen to sit nearby. The student who sits to his right, Ricardo, props his book up on his desk and sleeps through the class. Sebastian, who sits to Santino's left, fidgets a lot. He spends the class period practicing tricks he can do with his pen. About every half minute, he drops it and has to look for it. Manuel, who sits in front of Santino, reads the newspaper all through class. There just can't be that much interesting news to fill the whole hour! And the red-haired kid who sits behind him constantly kicks his chair. It's as if the red-haired kid and his friends are trying to make Santino fail his class! Every time he kicks Santino's chair, Santino remembers how much he can't stand economics.

Bridging Inference: Who kicks Santino's chair? *a. the red-haired kid b. Manuel
Non-inference: What is Santino studying? a. current events *b. economics
Pragmatic Inference: Does Santino like the people sitting near him? a. yes *b. no