Applied Psycholinguistics **22** (2001), 113–128
Printed in the United States of America

Bilingual long-term working memory: The effects of working memory loads on writing quality and fluency

SARAH RANSDELL and M. ROSARIO ARECCO
Florida Atlantic University

C. MICHAEL LEVY
University of Florida

ADDRESS FOR CORRESPONDENCE
Sarah Ransdell, Department of Psychology, 2912 College Avenue, Florida Atlantic University, Ft. Lauderdale, FL 33314. E-mail: ransdell@fau.edu

ABSTRACT
In Experiment 1, 42 multilinguals were able to maintain native language writing quality and fluency in the presence of unattended irrelevant speech while maintaining a concurrent 6-digit memory load. In Experiment 2, 80 bilinguals reduced fluency during writing with the 6-digit load only. In previous research, over 100 monolinguals of comparable verbal and nonverbal skills in three experiments reduced quality and fluency under both secondary tasks (Ransdell, Levy, & Kellogg, 1996). The results are interpreted in terms of a bilingual skill advantage in suppressing irrelevant information. Possessing fluency in another language may confer long-term working memory benefits during dual-task language conditions for bilinguals and even more so for multilinguals.

Essay writing requires multiple cognitive processes including, at the very least, the processes of planning what to write, generating sentences, and reviewing and revising what has been written (Hayes, 1996; Kellogg, 1996; Levy & Ransdell, 1995). Add to that translating from one's native language (L1) to a second language (L2) and the task should be quite demanding, especially when a secondary task is added to the primary task of writing. But there are some reasons to believe it may not be so difficult, especially for fluent multilinguals (Cook, 1997). The purpose of the present research is to investigate the coordination of long-term working memory (LT-WM) resources while writing in L1 and L2. LT-WM has been proposed to explain unusually efficient retrieval and use of domain-specific knowledge (Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995). Over a lifetime, adult multilingual writers develop LT-WM processes that aid the activation and suppression of language of input. The present experiments reveal how this knowledge may be used to advantage in dual-task situations.

Previous research has often failed to find interference during dual-task expert performance (Ericsson & Kintsch, 1995). The present research seeks to add

proficient use of more than one language as a type of expert performance. We first investigate multilingual college writers in a traditional transient-storage approach (Ericsson & Delaney, 1999). According to multicomponent models of WM, such as those originally proposed by Baddeley and Hitch (1974) and later refined by Baddeley (1986), tasks that demand domain-general resources should disrupt primary task performance more than tasks that require only domain-specific resources. Domain-general secondary tasks, such as a concurrent 6-digit memory load, have been shown to require central executive processing because of the necessity to coordinate processing of the main task with near-capacity storage in the second. Domain-specific secondary tasks, such as digit loads of less than 6 and unattended irrelevant speech, have been attributed to phonological loop processes and disrupt writing much less than do central secondary tasks (Ransdell et al., 1996). Multilingual writers may have LT-WM knowledge with which to exceed normal dual-task demands. The research strategy employed here is to pair writing in L1 and L2 with secondary tasks that have been shown to require capacity-limiting peripheral or central WM processing in comparable monolingual writers.

WM loads and L1 writing

In three monolingual experiments, Ransdell et al. (1996) found that unattended and attended irrelevant speech and a concurrent 6-digit load all caused writers to decrease their writing fluency. Only the 6-digit load reduced writing quality in addition to fluency. Fluency refers to the number of words word-processed per minute, controlling for typing speed and including those words deleted before the final draft. Fluency is important because it appears to be a necessary, but insufficient, factor in successful writing quality (Levy & Ransdell, 1995). High-quality writing must be fluent writing, but of course writing well also requires knowledge about grammatical conventions and rhetorical devices, among other things (Hayes, 1996).

In addition to fluency and quality, Ransdell et al. (1996) found that total number of pauses, mean pause duration, and proportion of pauses located at grammatical boundaries were all disrupted by domain-general processing loads in monolingual subjects. However, only writing fluency and total number of pauses were affected by domain-specific processing loads. These results add written language production to the long list of cognitive processes shown to be disrupted by domain-specific and domain-general secondary tasks (see Gathercole & Baddeley, 1993). Writing fluency requires mainly domain-specific, phonological processing in WM, whereas writing quality requires domain-general, central processing. But what if writers have exceptional LT-WM knowledge about suppressing "secondary" tasks such as these or suppressing one language while processing another?

WM loads and L2 writing

It is possible that, because multilinguals must often suppress one or more of their languages at any given time, they may have better access to LT-WM infor-

mation involving such suppression. If so, then multilinguals should also be able to employ strategies that take advantage of suppression skill. Nayak, Hansen, Krueger, and McLaughlin (1990) found that bilinguals exhibited greater flexibility in their use of strategies to accommodate a given situation. Bilinguals used more mnemonic devices when asked to do so and more searching strategies when told to look for formulas; monolinguals were unable to adjust their strategies for each task. McLaughlin (1990) suggested that multilinguals have a greater ability to "learn to learn" than do monolinguals. Vaid and Lambert (1979) found that early bilinguals were significantly faster in detecting embedded figures than either monolinguals or late bilinguals. Early bilinguals (i.e., those who acquire an L2 before about age 11) are thus more "field-independent" than monolinguals. Early bilinguals have become more sensitive to a variety of input cues relative to late bilinguals who learn an L2 as older children or adults, typically in a formal setting. There is even some evidence to suggest that field independence is correlated with successful L2 learning (Naiman, Frohlich, Stern, & Todesco, 1977). It is possible, then, that multilinguals, who are more likely to be early bilinguals in at least one of their L2s, may have some general processing advantages relative to bilinguals and even more so to monolinguals.

Bialystok and Majumder (1998) found that bilingual children performed better than monolingual children on a grammatical judgment task that required selectively attending to the structure of a sentence while ignoring its more salient meaning. Furthermore, fluent bilinguals performed better than either less fluent bilinguals or monolinguals on two tasks that required focusing attention on the relevant verbal aspect of the task while ignoring more salient, but abstract, irrelevant features. The more balanced the bilingual child's languages were (i.e., the more fluent the child's nonnative language was), the higher the score was on the two attentional control tasks. In a review of the "additive" effects of bilingualism, Cook (1997) concluded that L2 expertise yields increased metalinguistic awareness of phonology, syntax, and the arbitrary nature of meaning, as well as gains in cognitive flexibility. Fabbro, Gran, and Gran (1991) found that professional simultaneous translators recognized significantly more sentences containing semantic errors in a dichotic listening task than did interpreting students. Expertise in translating clearly improves simultaneous listening skill. In sum, better selective attention, metalinguistic awareness, cognitive flexibility, and dichotic listening performance are all associated with expertise in L2. It is possible that all of these diverse advantages may form an LT-WM that can help bilinguals perform in dual-task situations.

The strength of one's ability in L1 and L2 should moderate the relationship between LT-WM and dual-task performance. Writers with high self-reported L1 skill and especially high L2 skill should be able to inhibit irrelevant stimuli during writing better than those with low skill, even when controlling for differences in general L1 verbal and nonverbal ability. Several potential covariates are included, such as self-ratings of L1 and L2 language fluency based on the TOEFL exam (Freedle & Kostin, 1999), psychometric measures of nonverbal skill (Cattell & Cattell, 1963), and reading comprehension ability in L1 (Brown, Fishco, & Hanna, 1993).

EXPERIMENT 1

Experiment 1 tested the generality of irrelevant speech and concurrent digit load effects found with monolingual writers using a sample of fluent multilinguals whose L1 is Estonian and who also speak and write English fluently. If fluent multilinguals do have particularly efficient access to LT-WM knowledge, then neither irrelevant speech nor a 6-digit load should disrupt fluency and quality, contrary to such effects found with monolingual English speakers (Ransdell et al., 1996). The extent of self-reported L2 skill should be inversely related to the amount of disruption, if any, caused by irrelevant speech and digit load.

Five essays were written. The first two essays were single tasks, one in Estonian (L1) and one in English (L2). Essay quality and fluency were expected to be less successful in L2 than L1 and to be predicted by composite TOEFL self-ratings of ability in L2. The second two essays were written with a secondary task of Estonian irrelevant speech, once in L1 and once in L2. It was predicted that the effects of irrelevant speech might not occur because of the potential for greater LT-WM knowledge in multilinguals. The final essay was written in L1 with a secondary task of a concurrent 6-digit load in Estonian. As with the monolinguals, the 6-digit load was expected to be more disruptive than the irrelevant speech.

METHOD

Participants

A total of 41 psychology students (39 women, 2 men) who were native speakers of Estonian and who also spoke and wrote English fluently participated for extra credit. Participants were 21.2 years of age on average. The mean L2 composite fluency rating was 2.54/4.0, or 63.5% ($SD = .71$). Participants spoke on average four languages, with a range of three to eight languages reported. All participants were volunteers from psychology classes at Tallinn Pedagogical University in Estonia, a former Soviet bloc country near Finland. Note that an Estonian monolingual group would have been a compelling control group; however, such a group simply does not exist. Nearly all Estonians speak at least one other language unless they are very young or very old.

Reading comprehension in L1 (a translation of the Nelson–Denny reading comprehension subtest), nonverbal skill (Cattell's Culture-Fair test), and self-ratings of speaking, listening, reading, and writing skill in L1 and L2 based on the TOEFL examinations were collected. The average nonverbal skill score was 34% ($SD = 4.2$), compared to 49% in the monolingual sample from Ransdell et al. (1996), $t(1, 124) = 1.66$, $p < .08$. The mean Nelson–Denny reading comprehension score in L1 was 64% ($SD = 5.2$), significantly lower than the 74% in the monolingual sample, $t(1, 124) = 2.31$, $p < .02$.

Design

A repeated-measures design was employed, with essay context as a main factor with five levels. Five essays were written. The first two were single tasks (one

in L1 and one in L2). The second two were written with a secondary task of L1 irrelevant speech, once in L1 and once in L2. The final essay was written in L1 with a secondary task of a concurrent 6-digit load. All instructions were read in L1 by native speakers. All analyses of essay quality were conducted on essays that had been translated into English by a professional translator. Fluency analyses were the same, regardless of original language, as they were based on a word count and time analysis.

Note that a completely factorial design would have included five additional conditions: L1 and L2 essays with L2 irrelevant speech, L2 essays with L1 digit load, and L1 and L2 essays with L2 digit load. Writing five essays is already a very demanding task in terms of time and effort, even across two sessions. Perhaps more serious is the increased potential of carry-over and fatigue effects with 10 conditions. Moreover, a repeated-measures design was necessary, given the important subject variables collected. We therefore chose the present set of five conditions based on the predictions outlined earlier.

The experiment consisted of two sessions. During the first session, participants completed the reading and nonverbal tests. After a short break, participants wrote the two baseline essays in L1 and L2. In the second session, the participants wrote the three remaining essays according to the counterbalanced design. Dependent variables included words created per minute, including those typed and then deleted before the final essay; pauses per minute (of all pauses greater than 5 seconds in length); average pause length (greater than 5 seconds each); sentence length; percentage of pauses located at clause, sentence, or paragraph boundaries; and holistic writing quality. The quality score was derived from the average of two peer raters evaluating each essay in English or an English translation. Raters were blind as to the experimental condition. Scores were totaled from 13 dimensions composed of 6 subcategories: word choice, technical quality, content, purpose, organization, and style. The reliability coefficient for the two independent peer raters was $r = .78$, averaged across all subcategories.

Materials

The Cattell Culture-Fair test consisted of 50 four-choice questions, each of which included a series of simple line drawings. The participants completed the sequence begun by the series of drawings. An Estonian translation of the Nelson–Denny reading comprehension subtest (Brown et al., 1993) was created by a professional translator. Writing fluency, pause duration, frequency, and location were measured by a program called FauxWord (Levy & Ransdell, 1995). This program captured keystrokes as the participant typed and replayed them in real or fast time. The program also displayed the mean pause duration for each essay for pauses over a selected length (in this case, 5 seconds). The essays were replayed and inspected to determine the proportion of pauses located at clause or sentence boundaries. Words per minute (wpm) was determined by recording complete words created during writing but later deleted and by adding this to the final word count. Typing to dictation was used as a measure of simple wpm and was used to qualify the obtained writing fluency measure.

Procedure

The experiment took place in a microcomputer laboratory where the participants read and signed consent forms in L1 and then completed the Cattell Culture-Fair test for 20 minutes and the Nelson–Denny reading comprehension subtest in L1 for 20 minutes. After a 5-minute break, two baseline essays were written, the first in L1 and the second in L2, for 10 minutes each. The first session took about 90 minutes. During the second session, participants were asked to write three 10-minute essays while listening to a tape over headphones. While writing the first two essays, they listened to an L1 recording of one of two children's fairy tales presented in counterbalanced order. Participants did not have to respond to the tape and were told to focus on writing the essay. The first essay was to be written in L1 and the second in L2. During the third essay, the participants listened to a series of 6 random digits in L1 presented about every 30 seconds. When the experimenter said "recall," participants had to say out loud as many of the digits as they could remember, continue writing, and then wait for the next set of digits. The main task was still the essay writing, but the participants had to try to remember as many digits as they could.

The participants were asked to write the best 10-minute essay they could, as if they were writing for a grade. A topic card was placed in front of the participant on top of the personal computer and read aloud. One such topic was the following: "Imagine the best possible college class in an imaginary university with unlimited funds." Presented in counterbalanced order, the other topics took a similar form but were about a college professor, a holiday vacation, a career after graduating from university, and a boyfriend/girlfriend. In all cases, writers were asked to compare their view to other possible opinions and were told when 5 minutes and then 1 minute remained to conclude their essays.

RESULTS

MANOVAs were conducted on each dependent variable with a composite score (speaking, comprehending, reading, and writing) self-report of fluency in L2 as the covariate. L2 composite had two levels divided by a median split. No other subject variables were found to be reliable covariates, so only L2 composite is reported. A priori pairwise comparisons were conducted between baseline L1 and L2 conditions, baseline L1 with irrelevant speech and with a 6-digit load, and baseline L2 with irrelevant speech. Table 1 shows the means and standard errors for all significantly affected dependent variables in each essay context.

Writing fluency

An omnibus MANOVA with an L2 fluency composite score as a covariate revealed a main effect of essay context, $F(4, 37) = 3.80$, $p < .01$, but only a marginal interaction between essay context and L2 composite ($p < .09$). A marginal effect of L2 composite showed that those above the median in self-rated

Table 1. *Effects of essay context on writing performance in Experiment 1 (multilinguals)*

	Essay context condition									
Measure	L1		L2		L1 w/ speech		L2 w/ speech		L1 w/ 6-digit	
Fluency	11.4	(.70)	*10.3*	(.63)	11.9	(.76)	11.5	(.84)	11.0	(.65)
Quality (%)	95.5	(1.8)	*84.3*	(1.7)	94.5	(1.9)	*84.0*	(2.1)	91.6	(2.0)
TP	12.9	(.83)	*15.2*	(.99)	12.1	(.89)	14.8	(1.4)	*17.4*	(1.3)

Note: Values in italics are significantly different from L1 baseline. Fluency is words word-processed per minute, including those deleted from the final draft. Quality is a holistic composite based on an average of two peer raters' scores. TP is the total number of pauses in writing that were greater than 5 seconds in length. Standard errors are given in parentheses.

L2 skill wrote more fluently across essay conditions (12.4 wpm) than did those below the median (10.0 wpm), $F(1, 40) = 3.63$, $p < .06$. Neither topic nor topic order was found to be reliable in this or any subsequent analyses.

Baseline L1 essays were more fluent than baseline L2 essays. Pairwise comparisons revealed a reliable effect of L1 fluency (11.4 wpm) over L2 fluency (10.3 wpm), $F(1, 40) = 5.78$, $p < .02$, but no interaction with L2 composite. No irrelevant speech effect was found since baseline L1 condition (11.4 wpm) was not significantly different from L1 with irrelevant speech (11.9 wpm), $F = 1$. No cross-language irrelevant speech effect was found, $F < 1$. There was no 6-digit load effect on fluency, $F < 1$. L1 fluency was 11.4 wpm in L1 baseline and 11.0 wpm in L1 writing with a digit load.

Pause frequency, duration, and location

The total number of pauses longer than 5 seconds was significantly affected by essay context, $F(4, 30) = 4.31$, $p < .007$. There were fewer such pauses in L1 (12.9) than in L2 (15.2), $F(1, 34) = 4.55$, $p < .04$, but no irrelevant speech effect ($F < 1$) or any cross-language irrelevant speech effect ($F = 1.6$). But in contrast to fluency, pause frequency did show a reliable 6-digit load effect, with 12.9 pauses in L1 and 17.4 pauses in L1 with a digit load, $F(1, 33) = 8.84$, $p < .005$. Average sentence length, mean pause duration, and pause location were not significantly affected by essay context, Fs = 1.

Overall writing quality

Overall writing quality was significantly affected by essay context, $F(4, 32) = 8.29$, $p < .0001$. Moreover, those with higher L2 composite scores wrote better essays, $F(1, 35) = 3.84$, $p < .05$. Quality was significantly higher in L1 (95.5%) than in L2 (84.3%), $F(1, 37) = 21.44$, $p < .0001$. As with the writing fluency measure, there was no irrelevant speech effect nor 6-digit load effect on overall

quality scores. Irrelevant speech did reduce quality for L2 essays. Essays written in L2 with L1 irrelevant speech were significantly poorer (84.0%) than those written in L1 alone (95.5%), $F(1, 36) = 24.46$, $p < .0001$. There was also a significant difference between essays written in L2 with L1 irrelevant speech (84.0%) and those written in L1 with L1 irrelevant speech (93.0%), $F(1, 37) = 4.73$, $p < .05$.

DISCUSSION

Previous research with over 100 monolinguals indicated that irrelevant speech in the background of a writing task reduces writing fluency, but that maintaining a concurrent 6-digit load reduces both fluency and quality, as well as other temporal measures (Ransdell et al., 1996). These variables have far fewer deleterious effects on Estonian students, who speak, on average, four languages fluently and as many as eight. Clearly this is an exciting result for those who believe that multilingualism confers some advantages in WM processing (i.e., Cook, 1997; Grosjean & Miller, 1994). These multilinguals spoke many languages, which may form a kind of domain-specific LT-WM knowledge, but they were not significantly better than monolinguals on measures of verbal or nonverbal skill or on college grade point average (Ransdell, Hawkins, & Adams, in press).

Brookings (1990) found that dual-task performance is unrelated to psychometric ability factors in monolinguals. But little research has been done to link the WM flexibility found in successful dual-task performance with high second language proficiency (SLP). Sasaki (1996) did find SLP to be distinct from, but correlated with, a general factor of cognitive ability in Japanese–English bilingual college students. Her verbal protocols revealed better strategy use in high-SLP students than in low-SLP students, but no better IQ. Nayak et al. (1990) found no overall achievement differences between monolinguals and bilinguals, but bilinguals exhibited greater flexibility in their use of strategies to accommodate a given situation.

The link between greater WM control and stronger reading and writing skills in monolinguals is clear. For example, monolingual college students who score higher on reading comprehension measures are better able to allocate resources to either storage or processing tasks, whereas less skilled readers have no such flexibility (Ransdell & Levy, 1999). Perhaps the constant activation and inhibition of language of input required of bilinguals has produced a specialized form of LT-WM knowledge. Experiment 2 was designed to extend the generality of these effects to a larger, American sample and to bilinguals who know L2 languages that are more similar to English. Estonian is not an Indo-European language; Spanish (a Romance language) and Polish (a Slavic language) are more closely related to a Germanic language like English.

EXPERIMENT 2

A total of 40 Spanish–English bilinguals and 40 Polish–English bilinguals were tested using the same paradigm as in Experiment 1 in order to test the generality of the findings. Bilingual participants should show similar, but perhaps not as

extensive, results of potentially more efficient LT-WM relative to multilingual writers. It was hypothesized that these bilinguals would show limited or no irrelevant speech and 6-digit load effects while writing, both in terms of writing quality and fluency.

METHOD

Participants

These bilinguals spoke only two languages on average, relative to the four, on average, spoken by the multilinguals in Experiment 1. The Spanish sample rated themselves as 85% ($SD = 14.7$) in composite English skill and the Polish sample, 79% ($SD = 13.6$), relative to the 63% of the Estonian sample. Nonverbal (Cattell Culture-Fair test scores) were comparable across groups: the Estonian sample average was 34%, the Spanish, 41% ($SD = 8.3$), and the Polish, 37% ($SD = 10.3$) (the monolinguals in a previous study scored 49% and were not significantly different from the Estonian sample). Average reading comprehension scores were 72% ($SD = 16.5$) in the Spanish sample and 54% ($SD = 14.2$) in the Polish sample, compared to 64% in the Estonian sample and 74% in the monolingual group. It is difficult to compare reading comprehension scores directly because of translation differences. Reading comprehension was significantly lower for the multilinguals (64%) and the Polish–English bilinguals (58%), relative to either the Spanish–English bilinguals (72%) or the monolinguals (74%). The latter two groups did not differ, presumably because they were much less likely to be affected by any translation problems in the topics covered in the reading comprehension test. For example, one of the eight passages was about the World Series. This topic is very familiar to both English native speakers and most Spanish native speakers living in the United States. However, the Estonian multilinguals had no idea what the World Series was, and the Polish–English bilinguals were likely to have had considerably less exposure than the Spanish–English bilinguals to American baseball.

Nonverbal performance, on the other hand, does not have potential translation problems and did not show any differences across the different samples. Cattell Culture-Fair test performance was not significantly different across groups: multilinguals, 34%; bilinguals, 44%; and monolinguals, 49%. There was a nonsignificant trend for the multilinguals to do more poorly on the nonverbal test than the bilinguals or the monolinguals. Even so, the monolinguals performed more poorly on the secondary tasks than either of the other groups.

The Spanish sample's average age was 26 and the Polish sample's, 39. Both were older, especially the Polish sample, than the Estonian participants, whose mean age was 21. The monolinguals from previous research were comparable in age to the Spanish sample, with a mean age of 27 (Ransdell et al., 1996). In general, older participants may have less efficient access to LT-WM, but since the order of average age (Estonian, Spanish, English, and Polish) is different from the expected effects, this should not present a problem in interpretation. It is also possible that LT-WM skills improve with age, but the present experiments were not designed to test this interesting idea.

Design and procedure

The design and procedure were identical to that of Experiment 1 except for the following differences. Participants were recruited from an American public university. The pattern of language dominance was such that the labels "L1" and "L2" did not easily apply to the design in terms of stronger and weaker languages. For example, for 15 (38%) of the Spanish-speaking participants, the L1 was Spanish, but the dominant language became English before the age of 11. For 8 (20%) in the Spanish sample, English was the L1, and it remained so. For the other 17 (42%) Spanish speakers, Spanish was the L1, and it remained so. In the Polish sample, all were dominant speakers of Polish and claimed English to be their L2. Language shift analyses are included in the results section to help disentangle these subject category differences.

As in Experiment 1, all essays were translated into English for writing quality assessment. The raters were blind as to the condition and whether the essay was originally written in Spanish, Polish, or English. The translators made an effort to provide accurate translations, preserving "errors" and indicating in subscript the type of error that was made in the native language in the few situations in which it would have been difficult to recreate the error in translation.

All of the Spanish language essays were scored for writing quality in Spanish. One of the raters who also scored all of the essays translated into English from Spanish was a native Spanish speaker (the second author) and provided another blind scoring approximately one year after her original assessments of the translated English essays. Baseline essay quality in the original essays correlated significantly with the translated versions (Cronbach's alpha = .97) as did the essays in the irrelevant speech (alpha = .96) and the 6-digit load (alpha = 94) conditions. These high reliability coefficients indicate that the translated essays preserved the dimensions that affected quality ratings in the original native language essays.

RESULTS

As in Experiment 1, MANOVAs were conducted on each dependent variable with a composite score (average of speaking, comprehending, reading, and writing) self-report of fluency in L2 as the covariate followed by a priori pairwise comparisons. No other covariates were significantly related to essay context. Table 2 shows the means and standard errors for significantly affected dependent measures in each essay context. Finally, ANCOVAs were conducted on fluency and quality scores for the digit load condition as compared to the baseline, controlling for digit recall performance. An analysis of covariance with language shift on fluency and quality scores in the Spanish sample was conducted in order to control for differences between those who shifted from one language as dominant to another.

Writing fluency

On the data combined for the Spanish and Polish samples, an omnibus MANOVA with L2 composite scores as a covariate (median split) showed a main effect of essay context, $F(4, 57) = 22.2$, $p < .001$. An interaction was also

Table 2. *Effects of essay context on writing performance in Experiment 2 (bilinguals)*

	Essay context condition									
Measure	L1		L2		L1 w/ speech		L2 w/ speech		L1 w/ 6-digit	
Fluency	10.0	(.66)	*12.3*	(.76)	10.4	(.62)	*12.4*	(.79)	*7.5*	(.47)
Quality (%)	79.0	(2.8)	*90.0*	(2.7)	83.0	(2.9)	91.6	(2.7)	73.6	(2.4)
TP	14.4	(.99)	14.8	(.95)	15.8	(.96)	15.2	(1.0)	*20.1*	(1.3)
PLOC (%)	21.1	(3.2)	22.8	(3.6)	22.4	(3.6)	24.0	(3.7)	*14.5*	(2.2)
SENTL	17.0	(1.3)	*20.0*	(1.6)	17.7	(1.4)	19.1	(1.3)	15.8	(1.0)

Note: Values in italics are significantly different from L1 baseline. Fluency is words word-processed per minute, including those deleted from the final draft. Quality is a holistic composite based on an average of two peer raters' scores. TP is the total number of pauses in writing greater than 5 seconds in length. PLOC is the percentage of pauses located at grammatical boundaries relative to all pauses. SENTL is the average sentence length. Standard errors are given in parentheses.

reliable between essay context and L2 composite, $F(4, 57) = 2.62$, $p < .04$, with those above the median writing more fluently than those below, especially in L2 conditions.

In contrast to Experiment 1 results, L2 essays (English) were more fluent (12.3 wpm) than L1 essays (10.0 wpm), $F(1, 64) = 20.8$, $p < .001$, but there was no interaction with L2 composite. Recall that for 58% of the Spanish participants English was reported as the stronger language. And for both samples, "school" writing in English was more typical than in their native languages. However, for the purposes of these analyses, L1 refers to Spanish or Polish in the first condition and L2 refers to English in the second. The results are, in fact, all the more powerful, given the heterogeneity of L1 and L2 baseline conditions. As was reported, L1 baseline was still more fluent than writing with irrelevant speech or a digit load.

As in Experiment 1, no irrelevant speech effect was found since baseline L1 condition (10.0 wpm) was not significantly different from L1 with irrelevant speech (10.4 wpm), $F = 1.7$. A main effect of L2 composite was found; those above the median were more fluent (12.2 wpm) than those below (7.9 wpm), $F(1, 63) = 12.1$, $p < .001$. There was no reliable interaction with L2 composite, $F < 1$.

An ANCOVA with language shift as the covariate within the Spanish sample revealed an interaction between irrelevant speech and language dominance shift, $F(1, 31) = 7.22$, $p < .01$. Simple effect tests showed that those who did not shift language dominance wrote significantly more fluently in the baseline condition (13.3 wpm) relative to the irrelevant speech condition (12.4 wpm) ($p < .05$). Those who did shift from English dominance to Spanish, or vice versa, showed no irrelevant speech effect: baseline (12.1 wpm) versus irrelevant speech (12.5 wpm).

Moreover, no cross-language irrelevant speech effect was found. L1 irrelevant speech had a significantly facilitating effect on L2 writing (12.4) relative to L1 writing alone (10.0), $F(1, 62) = 15.6$, $p < .001$. The interaction between L2 composite and essay context was not reliable, $F < 1$. There was also a main effect of L2 composite, with those above the median being more fluent (13.0 wpm) than those below (9.1 wpm), $F(1, 62) = 8.75$, $p < .004$.

In contrast to Experiment 1, there was now a 6-digit load effect on fluency, $F(1, 62) = 25.79$, $p < .001$. Baseline L1 fluency was 10 wpm, compared to 7.5 wpm for L1 writing with a digit load. An interaction between essay context and L2 composite was found, $F(1, 62) = 7.19$, $p < .009$. Writers above the median were more fluent in the no-digit load condition (12.05) than in the 6-digit load condition (8.6). Those below the median were equally poor in fluency (7.7 and 6.7 wpm, respectively). As before, those above the median in L2 composite were more fluent (10.3 wpm) than those below (7.2 wpm), $F(1, 62) = 8.99$, $p < .004$.

An analysis of covariance showed that there was no interaction between the number of digits recalled and the effect of the digit load condition, $F(1, 71) = 2.70$, $p = .11$, using either a lax criteria of recall (correct digit recalled in any order) or a strict criteria of recall (correct digit recalled in exact order), $F(1, 71) = 1.57$, $p = .21$. There was also no significant correlation between digit recall and fluency in the digit load condition, using either a lax criteria, $r = -.17$, $p = .15$, or a strict criteria, $r = -.16$, $p = .17$. These analyses suggest that there was no trade-off between writing fluency and digit recall in the concurrent digit load condition.

An ANCOVA with language shift as the covariate within the Spanish sample revealed an interaction between digit load effect and language dominance shift, $F(1, 31) = 13.33$, $p < .001$. Simple effect tests showed that, while there was a digit load effect for both groups, there was a greater digit load effect for those who did not shift language dominance (baseline, 13.3, vs. digit load, 8.1, a difference of 5.3 wpm). Those who did shift dominance wrote 12.1 wpm in the baseline condition and 8.5 wpm in the digit load condition, a difference of 3.6 wpm.

Writing quality

A reliable main effect of essay context on quality scores was found in an omnibus MANOVA, $F(4, 45) = 3.71$, $p < .01$, but there was no interaction with L2 composite. And of all the pairwise comparisons reported here, only the comparison of baseline L1 and L2 was significantly different in quality. L1 essays were 79% in overall quality and L2 essays were 90%, $F(1, 56) = 3.80$, $p < .05$. There were no interactions with L2 composite scores.

An analysis of covariance showed that there was no interaction between the number of digits recalled and the effect of the digit load condition on quality scores, $F(1, 57) = 2.11$, $p = .15$, using a lax criteria of recall (correct digit recalled in any order) or a strict criteria of recall (correct digit recalled in exact order), $F(1, 57) < 1$. There was also no significant correlation between digit recall and quality in the digit load condition: lax criteria, $r = .15$, $p = .26$, or strict

criteria, $r = .18$, $p = .18$. These analyses suggest that, just as with fluency, there was no trade-off between writing quality and digit recall in the concurrent digit load condition.

Furthermore, repeated-measures ANOVAs conducted on essay quality using the quality scores generated from the native language (in Spanish) yielded the same outcomes as those found with the translated essays. Namely, the irrelevant speech condition was not significantly different from the baseline condition, $F(1, 39) = 1.20$, $p = .28$, nor was the 6-digit load condition, $F(1, 39) = 2.42$, $p = .12$.

Sentence length, total pauses, and mean pause duration and location

An omnibus MANOVA revealed a significant main effect of essay context on average words per sentence, $F(4, 57) = 3.32$, $p < .01$, but no interaction with L2 composite scores. Only one pairwise comparison was significant: that between baseline L1 (17.0 words per sentence) and baseline L2 (20.0), $F(1, 64) = 5.35$, $p < .02$.

The omnibus MANOVA for total number of pauses (over 5 seconds in length) was reliable, $F(4, 38) = 6.33$, $p < .001$. The interaction with L2 composite scores was also significant, $F(4, 38) = 2.58$, $p < .05$. Those above the median paused less often (15.4 pauses) than those below (16.8). L1 essays contained no more pauses than L2 essays, but there was a reliable interaction in this comparison with essay context, $F(1, 52) = 6.93$, $p < .01$. Those with fewer pauses in L2 baseline condition were above the median in L2 composite skill.

There was also a 6-digit load effect on total number of pauses, as in Experiment 1. Baseline L1 essays contained 14.4 pauses, and those written with a digit load contained 20.1 pauses, $F(1, 49) = 15.72$, $p < .001$. No other pairwise comparisons or interactions were reliable for total number of pauses.

Mean pause duration yielded no omnibus effect nor reliable pairwise differences. Finally, mean number of pauses located at grammatical boundaries was significantly affected by essay context, $F(4, 36) = 4.27$, $p < .005$. Those with high L2 composite skill were more likely to pause at grammatical boundaries (26.7% of the time) compared to those below the median (15.2%), $F(1, 39) = 3.89$, $p < .05$. Moreover, as with fluency and total number of pauses, there was a 6-digit load effect, $F(1, 47) = 11.59$, $p < .001$. Pause location at grammatical boundaries was 21.1% during baseline L1 essays but only 14.5% during L1 essays with a digit load.

DISCUSSION

Multilingual and bilingual writers showed far fewer irrelevant speech effects across all six dependent variables, despite large sample sizes and diverse language backgrounds. In fact, only one bilingual subgroup – the Spanish bilinguals who had not shifted language dominance later in life – showed any irrelevant speech effect. Multilinguals experienced a single 6-digit load effect on total number of pauses. Bilinguals wrote less fluently with a greater number of pauses under a 6-digit load. For bilinguals, pauses occurred less frequently at grammati-

cal boundaries during the 6-digit load condition. In three previous experiments, comparable monolingual writers showed irrelevant speech effects for fluency and total number of pauses and a 6-digit load effect for all six dependent variables (Ransdell et al., 1996). Just as was found in Ericsson and Kintsch (1995), dual-task interference effects were very much limited in an "expert" population.

Other expert populations have shown dramatic effects of practice in short-term memory tasks like digit span. For example, with 50 hours of practice, a "normal" digit span of about 5 to 9 digits can be expanded to over 20 digits (Chase & Ericsson, 1982). Superior memory for chess positions (but not randomly placed pieces) in world chess masters has been shown to be relatively unaffected by a secondary task like calculating a running total of random digits (Charness, 1976). More recently, evidence has suggested that expert performance is a result of slow adaptation to the demands of the skill domain, over years or even decades of deliberate practice (Ericsson, Krampe, & Tesch-Romer, 1993). Fabbro et al. (1991) showed that professional interpreters of at least 10 years' experience were superior in detecting semantic errors in a dichotic listening task, even compared to bilingual students training to be interpreters.

Ericsson and Delaney (1999) compared three general approaches to the study of WM: a basic capacity approach, a transient storage approach, and an expert performance approach. While these preliminary results only suggest, but do not test, the idea, we argue that the skill required to inhibit the demands of concurrent memory loads during transient storage may be akin to that of inhibiting nonactive language of input in expert performance. In the present study, skill in a second language may have served as a kind of LT-WM with which to maintain primary task performance. Furthermore, self-rated L2 composite scores were more often a significant effect than essay context, with those above the median performing better than those below. Secondary tasks that were capacity-limiting in monolinguals of comparable L1 verbal and nonverbal ability were not capacity-limiting in the multilinguals of Experiment 1 nor, to a lesser extent, in the bilinguals of Experiment 2.

There are some limitations to the conclusions of this study. A direct test of better suppression skill in bilinguals was not conducted. We made the inference that a lack of interference effects in several dual-task situations, particularly for multilinguals, suggests that some general cognitive mechanisms are at work. Future longitudinal research needs to investigate the development of LT-WM skills related to second language training. Most assumptions in the literature suggest that second language acquisition is facilitated by good phonological and general WM resources (Service & Kohonen, 1995). We propose that the reverse should also be considered. That is, second language training and experience may actually provide a kind of expert LT-WM knowledge that can be used in general dual-task situations to improve performance.

If a lifetime of practice in selectively attending to language of input improves general cognitive ability or access to an expert kind of LT-WM (as with the professional interpreters in the Fabbro et al., 1991, study), this will have very important implications. The first implication is that bilingualism is a positive adaptation to the environment, beyond increased communicative competence in a second language and greater cultural understanding. For bilingual parents who

are contemplating whether to teach their children more than one language, the message is clear. Bilingualism can be a valued precursor of general as well as lingual flexibility. A second, more general, implication is that second language experience can change structural and functional properties of mind. Studies of bilingual aphasia indicate differences in resource allocation among the two or more languages, but not gross distinctions in areas contributing to performance (Obler & Gjerlow, 1999). Finally, when bilinguals are equated for socioeconomic status, immigration status, basic health, and other factors that potentially impact learning and school performance, these individuals may exceed monolingual performance even outside the domain of language. As the world becomes increasingly more bilingual, it is imperative that the cognitive advantages of bilingualism be brought to the forefront of investigation.

ACKNOWLEDGMENTS

The authors would like to thank Kaisa Niit for translating Estonian materials and Ulle Indermitte, Helen Griin, Monika Siiraki, Sandra Sandokow, and Tim Torres for assistance in data collection. The authors are also grateful to Catherine Snow, Juhani Lehto, and two anonymous reviewers for comments on an earlier draft and to Toomas Niit for hosting the first author's stay in Estonia. Portions of this research were presented in 1998 at the Psychonomics Society's 39th Annual Conference in Dallas, TX. Estonian data were collected while the first author was a Visiting Fulbright Scholar at Tallinn Pedagogical University, Tallinn, Estonia.

REFERENCES

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

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics, 19,* 65–85.

Brookings, J. B. (1990). A confirmatory factor analytic study of time-sharing performance and cognitive abilities. *Intelligence, 14,* 43–59.

Brown, J. I., Fishco, V., & Hanna, G. (1993). *Nelson–Denny reading test: Part II comprehension test.* Chicago, IL: Riverside Publishing.

Cattell, R. B., & Cattell, A. K. (1963). *Handbook for the individual or group, Scale 3, Form A.* Champaign, IL: Institute for Personality and Ability Testing.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 641–653.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 6, pp. 1–68). New York: Academic Press.

Cook, V. (1997). The consequences of bilingualism for cognitive processing. In A. M. D. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 279–299). Mahwah, NJ: Erlbaum.

Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 257–297). Cambridge: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406.

Fabbro, F., Gran, B., & Gran, L. (1991). Hemispheric specialization for semantic and syntactic components of language in simultaneous interpreters. *Brain and Language, 41,* 1–42.

Freedle, R., & Kostin, I. (1999). Does the text matter in a multiple choice test of comprehension? *Language Testing, 16*, 2–32.

Gathercole, S., & Baddeley, A. D. (1993). *Working memory and language.* London: Erlbaum.

Grosjean, F., & Miller, J. L. (1994). Going in and out of languages: An example of bilingual flexibility. *Psychological Science, 5,* 201–206.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. E. Ransdell (Eds.), *The science of writing* (pp. 1–27). Mahwah, NJ: Erlbaum.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. E. Ransdell (Eds.), *The science of writing* (pp. 57–71). Mahwah, NJ: Erlbaum.

Levy, C. M., & Ransdell, S. E. (1995). Is writing as difficult as it seems? *Memory and Cognition, 23,* 767–779.

McLaughlin, B. (1990). The relationship between first and second languages: Language proficiency and language aptitude. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency* (pp. 158–174). Cambridge: Cambridge University Press.

Naiman, N., Frohlich, M., Stern, H., & Todesco, A. (1977). *The good language learner.* Toronto: OISE Publications.

Nayak, N., Hansen, N., Krueger, N., & McLaughlin, B. (1990). Language-learning strategies in monolingual and bilingual adults. *Language Learning, 40,* 221–244.

Obler, L. K., & Gjerlow, K. (1999). *Language and brain.* Cambridge: Cambridge University Press.

Ransdell, S., Hawkins, C., & Adams, R. (in press). Predicting academic performance in American and Estonian college students. Research monograph to appear in the *International Journal of Educational Research.*

Ransdell, S. E., & Levy, C. M. (1999). Writing span, reading span, and working memory: The importance of strategy. In M. Torrance & G. C. Jeffery (Eds.), *High and low level writing processes: Managing cognitive demands* (pp. 99–113). Amsterdam: Amsterdam University Press.

(1996). Working memory constraints on writing quality and fluency. In C. M. Levy & S. E. Ransdell (Eds.), *The science of writing* (pp. 93–105). Mahwah, NJ: Erlbaum.

Ransdell, S. E., Levy, C. M., & Kellogg, R. T. (1996). *The effects of attended irrelevant speech on writing quality and fluency.* Paper presented at the Psychonomics Society 37th Annual Conference, Chicago, IL.

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude and intelligence.* New York: Peter Lang.

Service, E., & Kohonen, V. (1995). Is the relationship between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16,* 155–172.

Vaid, J., & Lambert, W. (1979). Differential cerebral involvement in the cognitive functioning of bilinguals. *Brain and Language, 8,* 92–110.