Journal of Applied Research in Memory and Cognition 5 (2016) 434–444

ELSEVIER

Journal of Applied Research in Memory and Cognition

Working Memory Capacity as a Predictor of Simultaneous Language Interpreting Performance

Brooke N. Macnamara* and Andrew R.A. Conway

Princeton University, Princeton, NJ, United States

What is the relationship between working memory capacity and skill acquisition among American Sign Language (ASL)-English simultaneous interpreter trainees? ASL-English interpreter trainees were administered measures of working memory capacity, several other cognitive abilities, and simultaneous interpreting performance over two years. We examined amount of training, initial cognitive abilities, change in cognitive abilities, and initial simultaneous interpreting performance as predictors of final simultaneous interpreting performance. We found that amount of training, working memory capacity, and initial simultaneous interpreting performance positively predicted final simultaneous interpreting performance. Several other cognitive ability measures also predicted simultaneous interpreting performance though working memory capacity yielded the strongest and most consistent results among the included measures. Initial simultaneous interpreting performance and working memory capacity accounted for 73% of the variance in final simultaneous interpreting performance. This finding suggests that interpreter programs could measure these two factors at admission to estimate students' likely success in the program.

Keywords: Working memory capacity, Simultaneous interpreting, American Sign Language, Training, Bilingual advantage, Skill acquisition

Simultaneous language interpreting is an extremely demanding cognitive processing task (Christoffels, de Groot, & Waldorp, 2003; Frauenfelder & Schriefers, 1997; MacWhinney, 1997). The interpreter must listen and comprehend the source message in one language while concurrently planning the closest equivalents and producing the message in another language with little to no control over the input rate or content of the message. In other words, simultaneous interpreting requires (1) alternately activating and suppressing production in the two languages (Christoffels et al., 2003), (2) analyzing the speaker's goals, inferences, and subtleties (Cowan, 2000; Moser, 1978; Seal, 2004; Shlesinger, 2003), and (3) making rapid decisions on how to convey the meaning in a different language while taking into account cultural nuances and differing communication rules (Christoffels et al., 2003; Treisman, 1965).

Given the aforementioned interpreting demands – comprehension, planning, task switching, and reasoning – which presumably recruit general cognitive abilities, individual differences in general cognitive abilities such as working memory capacity should be highly predictive of overall interpreting performance, just as they are for second language proficiency (see Linck, Osthus, Koeth, & Bunting, 2014). Along these lines, Gerver, Longley, Long, and Lambert (1989) studied differences between passing and failing interpreter students on a variety of discourse processing and verbal abilities. Relative to their failing counterparts, passing interpreter students had better memory for texts, better comprehension, and better verbal generation. However, the discourse-processing abilities such as those measured by Gerver and colleagues are strongly related to working memory capacity and other domain-general cognitive abilities (Gernsbacher, 1990; Just & Carpenter, 1992; Kintsch, 1988; Kintsch & van Dijk, 1978). Therefore, domain-general abilities may be able to predict performance differences in simultaneous interpreting over and above what can be explained by linguistic skill alone and may explain unique variance even after taking into account domain-specific training or experience.

Author Note

* Corresponding author at: Department of Psychological Sciences, Case Western Reserve University, 11220 Bellflower Rd., Cleveland, OH 44106, United States. Tel.:+1 216 368 2681.

E-mail address: brooke.macnamara@case.edu (B.N. Macnamara).

Indeed, using discriminant function analysis, Macnamara, Moore, Kegl, and Conway (2011) found that a combination of psychomotor speed, task switching speed, risk aversion, and set-shifting ability predicted group membership between non-expert and expert professional American Sign Language (ASL)-English simultaneous interpreters. However, given the cross-sectional design of the study, Macnamara et al. could not examine whether the cognitive profiles of the expert and non-expert interpreters had remained stable since before interpreter training to the time of the study, or whether these abilities changed as a function of interpreting experience, perhaps varying in their amount of change. Given that simultaneous interpreting is a complex bilingual task, this question is relevant not only to individual differences research and skill acquisition and expertise research, but to research conducted on the bilingual advantage.

The Bilingual Advantage

Research suggests that both of a bilingual's languages are continuously active and accessible (e.g., Kroll, Bobb, & Wodniecka, 2006; Marian & Spivey, 2003). This continuous competition between the two language systems presumably requires increased cognitive control in order for bilinguals to (1) resolve conflict between competing alternatives and (2) select the appropriate target language (Bialystok, 1999; Bialystok & Craik, 2010; Green, 1998). If experience engaging additional cognitive control (in order to manage the competing languages) enhances executive functions, then bilinguals should demonstrate cognitive advantages relative to monolinguals. Indeed, multiple studies have observed this "bilingual advantage" in infants, children, and adults (Bialystok, 2009; Bialystok & Craik, 2010) and many recognize that bilingualism provides a fresh perspective into research on memory and language (Kroll, Bobb, & Hoshino, 2014). Cognitive advantages related to bilingualism appear limited to executive control, namely inhibition (Bialystok, 2006; Bialystok & Martin, 2004; Martin-Rhee & Bialystok, 2008), task switching (Prior & MacWhinney, 2010), mental flexibility (Bialystok, 1999), and working memory (Morales, Calvo, & Bialystok, 2013 though see also Bialystok, 2009; Engel de Abreu, 2011).

According to the bilingual advantage hypothesis, bilinguals who speak two languages with a high degree of conflict should demonstrate the greatest cognitive control advantages while bilinguals who speak two languages with little conflict should demonstrate small or no cognitive advantages relative to monolinguals. A few studies have investigated the relationship between inter-language conflict and cognitive abilities. One of the most compelling studies in support of the bilingual advantage (Emmorey, Luk, Pyers, & Bialystok, 2008b) is one in which the researchers found a bilingual advantage for a sample of Vietnamese-English, Cantonese-English, and Italian-English bilinguals, but did *not* find an advantage for ASL-English bilinguals. The authors explained the results in terms of differences in conflict resolution demands and target language selection demands. That is, the two languages of bimodal (signed language-spoken language) bilinguals have distinct motor and perceptual pathways reducing the need for both conflict resolution (Emmorey, Borinstein, Thompson, & Gollan, 2008a) and target language selection (bimodal bilinguals can and frequently do produce features of both languages concurrently; Pyers & Emmorey, 2008). However, research with unimodal (spoken language–spoken language) bilinguals of various language pairs with varying amounts of inter-language conflict has, at least among children, revealed uniform bilingual advantages (Barac & Bialystok, 2012). These results suggest that differences in inter-language conflict alone may not be enough to vary the amount of executive control enhancement.

Other studies have failed to observe a bilingual advantage at all, even among unimodal bilinguals with language pairs previously tested (Morton & Harper, 2007; Paap & Greenberg, 2013). A recent review of the literature observed that over 80% of tests for the bilingual advantage since 2011 yielded null results (Paap, Johnson, & Sawi, 2015) and a recent meta-analysis (de Bruin, Treccani, & Della Sala, 2015) suggests that much of the bilingual advantage may be a publication bias artifact. The controversy surrounding the bilingual advantage is furthered when one considers that the effect may be confounded with various group differences (e.g., Morton & Harper, 2007), though some studies finding a bilingual advantage have controlled for such differences (e.g., socio-economic status; Bialystok, 2009). In a domain as complex as bilingualism, with many different language pairs, language statuses, ages of acquisition, frequency of use of each language, and the contexts in which each language is used, many differences among samples emerge, thus making generalizations difficult.

Given the complexity of bilingualism, one must consider the likelihood that cognitive control demands imposed on bilinguals vary widely. For example, the dominant language on a typical college campus surrounds bilinguals in classrooms, dorms, and through reading. In this case, the environment continuously cues the target language. Bottom-up processes should be engaged and little top-down control would be necessary to select among competing alternatives and appropriately select the target language. On the other hand, if bilinguals frequently function in an environment that cues both their languages and they must manage switching between the languages in this environment, then top-down cognitive control processes should be engaged and necessary for appropriate language selection.

Along these lines, Macnamara and Conway (2014) proposed the bilingual management demand hypothesis, which may account for some of the discrepant findings in the literature. The bilingual management demand hypothesis proposes that enhanced cognitive control depends on the cognitive demands encountered in managing the two languages and the amount of experience with these bilingual management demands. According to this hypothesis, enhancements to cognitive control from bilingualism would be unlikely among samples of bilinguals who function in environments with clear language cues that trigger bottom-up processes. The low bilingual management presumably demanded by the college environment may be the reason Paap and Greenberg (2013) did not observe a bilingual advantage among college student bilinguals. Although some evidence suggests that both languages in bilinguals are activated even in monolingual environments (e.g., Marian & Spivey, 2003), investigations comparing bilinguals who frequently switch languages to those who do not have demonstrated enhanced cognitive control among the frequent language switchers relative to the less frequent language switchers (Calabria et al., 2011; Festman et al., 2010; Prior and Gollan, 2011; Soveri et al., 2011). Preliminary findings on bimodal bilingual interpreter students, who must frequently switch languages within an environment strongly cueing both languages, demonstrated substantial enhancement in cognitive abilities over the course of training (Macnamara & Conway, 2014). Likewise, Hervais-Adelman, Moser-Mercer, and Golestani (2015) found functional brain changes to areas involved in executive control among spoken language simultaneous interpreter students over the span of their interpreter training—changes that did not occur for the control group. These findings support the idea that cognitive enhancements among (some) bilinguals may not just be a result of being bilingual, or of the degree of similarity between languages, but rather (1) the magnitude of the bilingual management demands and (2) the amount of experience managing those demands.

The Present Study

The goal of the present study is to examine relationships among cognitive abilities, training, and performance of ASL-English simultaneous interpreting. Measures were chosen to (1) test potential differences between two types of working memory capacity, and (2) examine the role of multiple cognitive abilities often cited as contributing to the bilingual advantage, which might also predict interpreting performance. The longitudinal design of the present study offers a framework for testing and comparing multiple hypotheses regarding the relationships among training, cognitive abilities, and performance in a complex bilingual processing task. One hypothesis of interest to skill acquisition and expertise researchers focusing on the role of experiential factors is that experience is compensatory. That is, working memory capacity and other cognitive abilities should predict initial performance, but with accumulated training, variance in interpreting performance will converge and cognitive abilities will cease to be predictive. A second hypothesis of interest to individual differences researchers focusing on the importance of cognitive abilities is termed "the rich get richer" (Hambrick & Engle, 2003). That is, working memory capacity and cognitive abilities should predict initial performance, and with accumulated training, performance variance will diverge so that those who begin with higher initial performance will have the steepest learning curves while those with the lowest initial performance will produce the shallowest learning curves. A third hypothesis of interest to bilingual advantage researchers focusing on improvements in cognitive abilities from bilingual experience is that change in working memory capacity and cognitive abilities will best predict skill acquisition rates. That is, individuals with the most enhancement in cognitive abilities will have the steepest learning curves while individual whose working memory capacity and cognitive abilities remain relatively stable over time will produce the shallowest learning curves. Many studies have examined these hypotheses, but rarely have such studies used real-world longitudinal data.

Methods

Participants

Fifty ASL-English simultaneous interpreting students were recruited in the middle of their first semester and again three more times over the course of the two-year programs[1],[2]. Of the 50 participants, 34 provided data at all four time points. Of the sixteen who did not provide data at all four time points, nine did not pass coursework in the training program and were asked by the faculty to either repeat the coursework or leave the program, three left the program for personal reasons but came back to the program or planned to come back to the program, and four finished the program on pace with their peers, but chose not to continue participating in the study. Demographic information is presented in Table 1. As can be seen, the participants who did not provide data at all four time points do not appear qualitatively different from the participants who provided full data.

Measurements[3]

Working Memory Capacity. We administered two types of working memory capacity measures to test (1) whether they would differentially predict simultaneous interpreting performance and (2) whether performance on the measures would differentially change over time. These working memory measure types – maintenance & interference and coordination & transformation – were chosen based on differences in executive control demands (Oberauer, Süβ, Wilhelm, & Wittmann, 2008). Maintenance and interference tasks are designed to tap one's ability to maintain information in the face of a secondary distracting processing task. This approach is often referred to as "processing as storage" and is assessed using complex span tasks (Oberauer et al., 2008). Coordination & transformation tasks are designed to tap one's ability to coordinate and transform information in one's mind (Oberauer et al., 2008). We hypothesized that the process of simultaneous interpreting mirrors the process tapped by the coordination & transformation tasks more so than the maintenance & interference tasks because simultaneous interpreters must continuously maintain a portion of the source message while transforming it into the grammatical

---

[1] One school used a trimester schedule and the other used semesters. We adjusted by recruiting participants on the basis of program advancement (e.g., approximately 20%, 40%, 60% and 80% way through the program). We use the more common term "semester" to refer to the division of terms for both schools.

[2] Data on a subset of these participants – those who began their interpreter training in the first year of data collection and completed their interpreter training in the second year of data collection – are reported in Macnamara & Conway, 2014.

[3] We also administered two affective trait measures: the Behavioral Approach System (BAS) Sensitivity Scale (Carver & White, 1994), which measures reward sensitivity, drive, and fun-seeking traits, and the Behavioral Inhibition System (BIS) Sensitivity Scale (Carver & White, 1994), which measures risk-taking sensitivity and anxiety surrounding aversive stimuli and novelty. No patterns from these measures emerged.

Table 1
Demographic Information at Wave 1

	Sample providing data	Drop-outs	Sample providing data—drop-outs comparison
Sample size			
Institution I	19	8	
Institution II	15	8	
Age			
M (SD)	26.68 (6.88)	30.50 (8.29)	$t(1, 48) = 1.71$, ns
Median (range)	24 (20–56)	28 (22–47)	
CODAs	3	2	
Female	85%	75%	$t(1, 48) = 0.86$, ns
Semesters of college			
M (SD)	8.67 (3.65)	7.56 (3.67)	$t(1, 48) = 1.00$, ns
Median (range)	8 (2–17)	7 (4–16)	
Semesters of ASL			
M (SD)	5.24 (2.34)	4.56 (1.46)	$t(1, 48) = 1.07$, ns
Median (range)	5 (1–12)	5 (1–6)	

Note: CODAs = children of deaf adults (i.e., participants exposed to ASL since birth). All 95% confidence intervals contain 0.

structure and lexical format of the target language. Previous results (Macnamara & Conway, 2014) suggest that interpreting experience differentially improves scores in these two task types (no change with interpreting experience for maintenance & interference performance; improvements with interpreting experience for coordination & transformation).

Maintenance & Interference. To examine working memory as maintenance & interference, we administered two complex span tasks, the automated reading span task and the automated operation span task (Unsworth, Heitz, Schrock, & Engle, 2005). In these tasks participants attempt to recall letters while reading sentences or solving math problems, respectively. These scores were combined to create a composite working memory capacity: maintenance & interference measure.

Coordination & Transformation. To examine working memory capacity as coordination & transformation, we administered a backward digit span task and a letter-number sequencing task (adapted from Wechsler, 1997). In the backward digit span task participants are asked to recall digits in reverse order of presentation. In the letter–number sequencing task participants view pseudorandom strings of digits and letters and then are asked to recall the numbers and letters in numerical and alphabetical sequence (not the order of presentation). These scores were combined to create a composite working memory capacity: coordination & transformation measure.

Fluid Intelligence. We administered two non-verbal reasoning/fluid intelligence tasks—Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 1962) and Cattell's Culture Fair Test (Cattell & Cattell, 1960). In these tasks participants are asked to recognize patterns, reason, and problem solve to the best of their ability. These scores were combined to create a composite non-verbal reasoning/fluid intelligence measure.

Perceptual Speed. We administered two perceptual speed tasks – Letter Comparison and Pattern Comparison (Salthouse & Babcock, 1991). In these tasks participants are asked to compare pairs of letter strings or figures and decide whether they are the same or different as quickly as possible. These scores were combined to create a composite perceptual speed measure.

Psychomotor Speed. We administered two psychomotor speed tasks – Connection A: Numbers and Connection A: Letters (Salthouse et al., 2000). In these tasks participants are asked to draw lines among numbers or letters in sequence as quickly as possible. These scores were combined to create a composite psychomotor speed measure.

Task-switching Speed. We administered two task-switching speed tasks – Connections B: Numbers–Letters and Connections B: Letters–Numbers (Salthouse et al., 2000). In these tasks participants are asked to draw lines by alternating sequences of numbers and letters (e.g., 1-A-2-B-3-C, etc.) as quickly as possible. These scores were combined to create a composite task-switching speed measure.

Mental Flexibility. We administered a computerized version of the Wisconsin Card Sorting Test (Grant & Berg, 1948) to measure one's ability to adapt to changes and to shift among different sets of rules.

Simultaneous Interpreting Performance. Participants listened to an audio-recorded short story about two boys talking over lunch and attempted to simultaneously interpret the story into ASL. The story was designed to be easy to comprehend (readability average = grade 6; Flesch–Kincaid Grade Level [Kincaid, Fishburne, Rogers, & Chissom, 1975] = 4.9, Gunning-Fog Score [Gunning, 1952] = 8.5, Coleman–Liau Index [Coleman & Liau, 1975] = 7.3, SMOG Index [McLaughlin, 1969] = 6, Automated Readability Index [Senter & Smith, 1967] = 4.3) but challenging to interpret. In addition to restructuring the sentences into ASL word order (which varies between Subject–Verb–Object and Object–Subject–Verb), placing adjectives after the nouns, and explicitly indicating spatial information that is generally implicit in English narration, the text included (1) a garden path sentence: "And Joe told him a little white lie will come back to haunt him.", (2) words that are homonyms in English but not in ASL ("run," "room") and in which the most frequent meaning in English would produce the incorrect lexical item in ASL (e.g., "room" in the story means vacancy/availability, not a space in a building with walls and there are different signs for each of these meanings), and (3) terms in which there is no one-for-one translation—"morally superior," "white lie," "haunt," and "have a say." Participants were video recorded while interpreting.

Idea Unit Presence. Each interpretation was scored on idea unit presence performance to measure how equivalent the interpretation was to the source message. To create this measure, the story was divided into 22 idea units to separate each meaningful idea presented in the story (e.g., Joe looks at the food on his tray; Joe makes a face). Each idea unit was scored from 0 to 4. A score of 0 would mean that the idea unit was completely absent from the interpretation. A score of 1 would mean that the idea appeared, but most of the information about this idea was missing or incorrect (for example, if the interpretation mentioned a tray, but not that there was food on it or that Joe looked at it). A score of 2 would mean that about half of the information about this idea was missing or incorrect, and so on, up to a score of 4, which would mean that the idea was completely present in the interpretation. The idea unit presence performance score for the interpretation was the average of the 22 idea unit scores.

Target Language Structure. Each interpretation was scored on target language structure to measure how well the grammatical structure of the target language (ASL) was presented in the interpretation. The sentence structure of ASL differs from English. To score target language structure performance, each of the 22 idea units was scored on a scale of 0–4. A score of 0 would mean that the interpretation of the idea unit was incomprehensible in ASL (for example, an incomprehensible idea unit in English might be "Tray the looks food at his Joe"). A score of 4 would mean that the interpretation of the idea unit was produced in a way most resembling how a native ASL user would have produced it. The target language structure performance score for the interpretation was the average of the 22 idea unit scores.

Production Fluency. Each interpretation was scored on production fluency to measure how much of an "accent" one had in the target language. As with any non-native language, individuals may be able to appropriately produce the words and the sentence structure, but could have such a strong accent that they are difficult to understand. To score production fluency, each interpretation was given an overall score of 0–4. A score of 0 would mean that the production fluency was so poor that it was incomprehensible in ASL. A score of 4 would mean that the interpretation was produced in a way most resembling how a native ASL user would have produced it. The production fluency score was assigned to each interpretation as a whole.

Composite Simultaneous Interpreting Performance. The sum of th e idea unit presence score, target language structure score, and production fluency score were summed to create a composite simultaneous interpreting performance score for each interpretation[4].

Procedure

Participants completed each of the tasks mid-semester for four consecutive semesters (approximately 20%, 40%, 60% and 80% completion through the interpreter training program). Data collection for each wave took place in two sessions. In the first session participants completed Pattern Comparison, Letter Comparison, Connections A: Letters, Connections A: Numbers, Connections B: Numbers–Letters, Connections B: Letters–Numbers, Raven's Advanced Progressive Matrices, and Cattell's Culture Fair Test. In the second session participants completed the Automated Reading Span, the Automated Operation Span, backward digit span, letter–number sequencing, and the Wisconsin Card Sorting Test. In Waves 2, 3, and 4 participants also completed the simultaneous interpreting task at the end of the second session. The interpreting performance measure was not included during the first wave of data collection because the participants had only a few weeks of interpreter training at that point. All participants completed the tasks in the same order.

The length of time between data collection waves (approximately 6 months) coupled with the generally high test-retest reliability of the cognitive ability measures (c.f., e.g., Conway et al., 2005; Lemay, Bédard, Rouleau, & Tremblay, 2004; Unsworth et al., 2005) suggests that changes observed in performance on these measures are due to changes in cognitive abilities and not practice effects. Likewise, change in performance on the simultaneous interpreting measure is likely due to change in interpreting skill and not practice effects: the three administrations of the simultaneous interpreting measure, administered approximately 6 months apart, follow the mandatory wait time rules imposed by the National Council of Interpreting between interpreting tests (that is, an individual must wait 6 months before retaking the National Interpreting Certification test; NAD-RID National Interpreter Certification (NIC) Candidate Handbook, 2011).

Results and Discussion

Simultaneous Interpreting Performance

Seven simultaneous interpreting performance measures from Wave 2 were lost due to technical failure. These scores were replaced with the mean scores for Wave 2. Performance on the simultaneous interpreting performance task was rated by two nationally certified ASL-English interpreters. The first author served as one rater and an outside rater from another state who

---

[4] We also counted the number of lexical errors for each interpretation. An example of a lexical error would be producing the sign that equates to a space in a building with walls (room) instead of the sign that equates to a vacancy (room) when interpreting "They have room on the board . . .." Given the non-normal distribution of errors, and that errors are reflected in the idea unit presence score, we do not conduct analyses with this measure.

Table 2
Intercorrelations among the simultaneous interpreting performance measures

	Initial			Final		
	1	2	3	1	2	3
Initial						
1. Idea unit presence	–					
2. Target language structure	.94	–				
3. Production fluency	.66	.75	–			
Final						
Idea unit presence	.75	.73	.65	–		
2. Target language structure	.66	.69	.74	.93	–	
3. Production fluency	.57	.61	.58	.53	.66	–

Note: All correlations significant at the $p \leq .001$ level. $N = 34$.

did not know any of the participants served as a second rater. The second rater did not have access to participants' cognitive ability measure scores. The second rater was also blind to time in the interpreting program, that is, the videos were shuffled and presented to the second rater in a random sequence so that whether the video was from Wave 2, 3, or 4 was unknown. Inter-rater reliability between the two raters was high: idea unit presence, $r = .89$; target language structure, $r = .80$; production fluency, 84% within 1 score on the scale. The scores from the second rater were used for all analyses. The three simultaneous interpreting performance scores were highly correlated with one another (see Table 2). For simplicity, we report the results of the composite simultaneous interpreting performance score (the sum of the three interpreting measures).

Statistical Power

While the sample size ($N = 34$) is modest and smaller than we anticipated, the longitudinal design of the study afforded us sufficient power to detect moderate effects. For example, post-hoc power analyses indicate that with $N = 34$, alpha = .05, and Cohen's $d = .50$, post-hoc power = .80. For repeated measures ANOVA, with $N = 34$, alpha = .05, and $\eta^2 = .10$, power = .90.

Hypothesis 1: Experience is Compensatory.

We sought to examine whether experience was compensatory. For this hypothesis to be supported, working memory capacity and/or other cognitive abilities would predict initial simultaneous interpreting performance but fail to predict final simultaneous interpreting performance because, with accumulated training, variance in interpreting performance would converge.

This hypothesis was not supported. Both types of initial working memory capacity were important predictors of initial simultaneous interpreting performance and even stronger predictors of final interpreting performance. Fluid intelligence and psychomotor speed also positively predicted final interpreting performance. See Table 3.

Regarding the effect of training, we tested whether amount of accumulated time in the training program predicted interpreter performance. A repeated measures ANOVA revealed that training time was a significant predictor of simultaneous interpreting

Table 3
Correlations between initial cognitive abilities and interpreting performance

	Simultaneous interpreting performance			
	Initial		Final	
	r	95% CI	*r*	95% CI
Working memory capacity: maintenance & interference	**.43***	.11, .67	**.59****	.32, .77
Working memory capacity: coordination & transformation	**.39***	.06, .64	**.60****	.33, .78
Non-verbal reasoning/fluid intelligence	.32†	−.02, .59	**.43***	.11, .67
Psychomotor speed	.31†	−.03, .59	**.34***	.00, .61
Perceptual speed	.26	−.09, .55	.30	−.04, .58
Task switching speed	.19	−.16, 50	.25	−.10, .54
Mental flexibility	.07	−.27, .40	.19	−.16, .50

Note: Bold = significant at the $p < .05$ level.
$N = 34$. 95% CI = 95% confidence interval: lower bound and upper bound.
** $p < .001$.
* $p < .05$.
† $p < .08$.

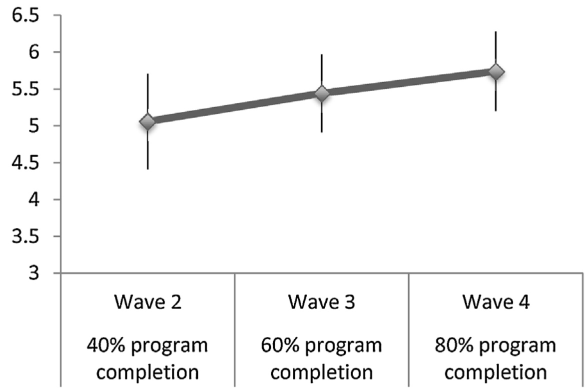

Fig. 1. Simultaneous interpreting performance. Error bars represent 95% confidence intervals. $N = 34$.

performance, $F(2, 66) = 5.97$, $p < .01$, $\eta^2 = .15$, see Fig. 1. However, interpreting performance variance did not converge with training.

Hypothesis 2: The Rich Get Richer.

We sought to examine whether "the rich get richer" (Hambrick & Engle, 2003). For this hypothesis to be supported, working memory capacity and/or cognitive abilities would predict initial performance and skill acquisition rates. That is, with accumulated training, performance variance would diverge so that those who begin with higher initial performance would have the steepest learning curves while those with the lowest initial performance would produce the shallowest learning curves.

We found limited support for this hypothesis. Initial working memory capacity predicted initial simultaneous interpreting performance (see Table 3) and initial simultaneous interpreting performance predicted final simultaneous interpreting performance: $r = .79$, $p < .001$. For example, those who initially performed in the bottom quartile typically performed in the bottom quartile in the final measure of performance.

To test whether initial cognitive abilities predicted skill acquisition rates, the change score was calculated between initial and final simultaneous interpreting performance and correlation analyses were conducted between initial cognitive abilities and the interpreting performance change scores. No cognitive abilities significantly correlated with interpreting performance change scores, all $ps > .13$. That is, contrary to the rich-get-richer hypothesis, individuals with higher working memory capacity did not have steeper learning curves than those with lower working memory capacity. Rates of improvement varied—some individuals improved at rates faster than their peers and some improved more slowly. Working memory capacity did not moderate the relationship between rate of learning and final performance, $\Delta R^2 = .00$. Initial performance negatively predicted change in performance, such that those initially performing the most poorly experienced greater improvements in performance, $\beta = -.55$, $t = -.3.71$, $p = .001$. This finding is counter to the rich-get-richer hypothesis.

Change in Cognitive Abilities: Is There a Bilingual Advantage?

To test change in cognitive abilities occurring from interpreter training, repeated measures ANOVAs were conducted. To correct for multiple comparisons, Bonferroni's correction was applied such that only p values<.007 were considered statistically significant. Improvements in cognitive abilities were observed, see Fig. 2a. Mental flexibility significantly improved with interpreter training, $F(3, 99) = 4.48$, $p = .005$, $\eta^2 = .12$. Psychomotor speed significantly improved with interpreter training, $F(3, 99) = 49.87$, $p < .001$, $\eta^2 = .60$. Perceptual speed significantly improved with interpreter training, $F(3, 99) = 4.63$, $p = .004$, $\eta^2 = .12$. Fluid intelligence/non-verbal reasoning significantly improved with interpreter training, $F(3, 99) = 9.26$, $p < .001$, $\eta^2 = .22$. Task switching speed significantly improved with interpreter training, $F(3, 99) = 13.46$, $p < .001$, $\eta^2 = .29$. The two types of working memory capacity differed as predicted, see Fig. 2b. One's ability to coordinate and transform maintained information (working memory capacity: coordination & transformation) improved significantly with interpreter training, $F(3, 99) = 20.98$, $p < .001$, $\eta^2 = .39$. However, one's ability to maintain information in the face of interference (working memory capacity: maintenance & interference) did not improve with interpreter training, $F(3, 99) = 0.52$, $p = .671$, $\eta^2 = .02$.

As previously discussed, a bilingual advantage has not previously been found for non-interpreting bimodal bilinguals, presumably because the distinct perceptual and motor

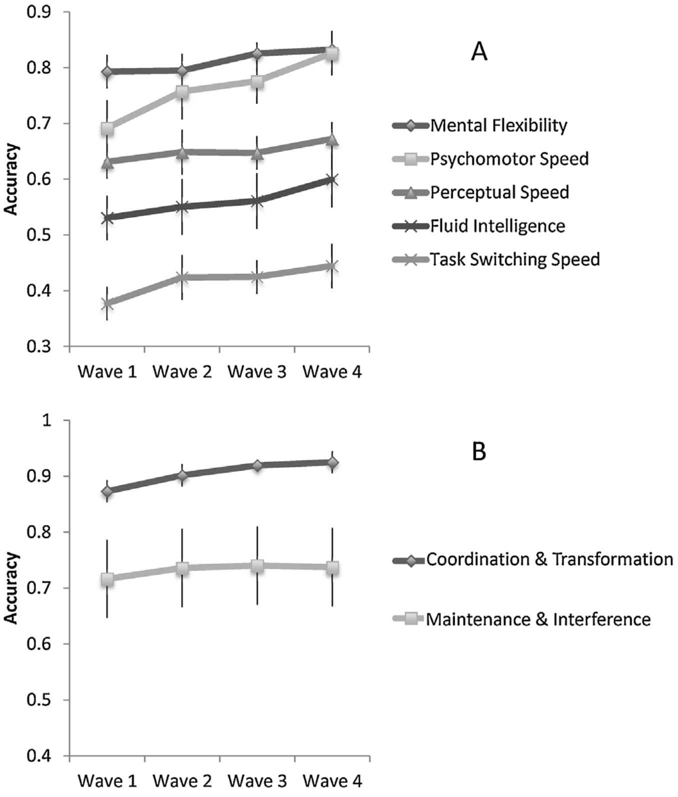

Fig. 2. Panel A: Cognitive ability performance over time. Panel B: Working memory capacity over time. Error bars represent 95% confidence intervals. $N = 34$.

pathways of the two languages decrease conflict and the amount of cognitive control necessary to resolve that conflict. However, simultaneous interpreting demands substantial conflict resolution and target language selection because the environment strongly cues both languages and because each language is being processed concurrently—while one is being perceived and comprehended, the other is being produced. The bimodality of the languages may not contribute to a bilingual advantage, but the intense bilingual management demands of simultaneous interpreting may be the contributing factor in enhancing cognitive abilities. Thus referring to our finding as a "bilingual advantage" may be a misnomer, but there is some evidence for an interpreter advantage or a more general high-bilingual-management-demand advantage.

Hypothesis 3: Change in Cognitive Abilities Predict Change in Interpreting Performance.. We sought to examine whether improvements in cognitive abilities from bilingual experience would predict skill acquisition rates. For this hypothesis to be supported, individuals with the most enhancement in cognitive abilities would demonstrate the steepest learning curves while individuals whose working memory capacity and/or cognitive abilities remained relatively stable over time would produce the shallowest learning curves.

This hypothesis was not supported. To test whether change in cognitive abilities predicted performance, three sets of correlation analyses were conducted. The first set of correlations tested whether early change – change in cognitive abilities from Wave 1 to Wave 2 – predicted early interpreting performance in Wave 2. The second set of correlations tested whether change in cognitive abilities from Wave 1 to Wave 3 predicted interpreting performance in Wave 3. The third set of correlations tested whether change in cognitive abilities from Wave 1 to Wave 4 predicted final interpreting performance. There was no indication that change in cognitive abilities predicted interpreting performance at any time.

Comparing the Two Types of Working Memory Capacity

We hypothesized that coordination & transformation type working memory capacity measures were more related to the process of simultaneous interpreting than maintenance & interference type tasks. Performance on the coordination & transformation measures improved with interpreter training supporting this hypothesis. To test whether coordination & transformation explains any unique variance in simultaneous interpreting performance, hierarchical regression was conducted. The analysis predicting initial simultaneous interpreting performance revealed that the two types of working memory capacity accounted for 19% of the variance, but coordination & transformation did not account for a significant amount of unique variance, $\Delta R^2 = .01$, $F(1, 31) = 0.35$, $p = .558$. The analysis predicting final simultaneous interpreting performance revealed that the two types of working memory capacity accounted for 40% of the variance. Coordination & transformation contributed more unique variance than it did in initial simultaneous interpreting performance, but did not reach statistical significance, $\Delta R^2 = .05$, $F(1, 31) = 2.76$, $p = .107$.

Predicting Performance

Final simultaneous interpreting performance was strongly predicted by working memory capacity and initial simultaneous interpreting performance. In fact, if one takes into account working memory capacity and initial simultaneous interpreting performance, these two measures account for 73% of the variance in final simultaneous interpreting performance, $F(2, 31) = 41.36$, $p < .001$, $R^2 = .73$[5]. Working memory capacity accounts for a significant amount of unique variance in final interpreting performance after accounting for initial interpreting performance, $\Delta R^2 = .10$, $F(1, 31) = 11.93$, $p = .002$.

General Discussion

We did not find that training compensated for cognitive abilities, that cognitive abilities predicted variance in rate of skill acquisition, or that enhancements to cognitive abilities predicted performance. Instead, our findings help to paint a picture of how aspects of these hypotheses and their associated factors contribute to complex performance in a real world task. Specifically, our key findings were that (1) increases in training were associated with increases in performance, (2) initial performance predicted final performance, and (3) working memory capacity predicted initial interpreting performance and was even more predictive of final interpreting performance.

These findings demonstrate the importance of both training and abilities on predicting performance. The longitudinal design allowed us to understand more nuanced patterns of skill acquisition. For example, one's position relative to others initially (e.g., top performer, second best performer, etc.) was a good indicator of one's final relative position: the top performers generally maintained a high level of performance and the bottom performers generally maintained a low level of performance relative to the rest of the group. However, the best performers improved more slowly; the bottom performers benefitted more from training, although training was not enough for the bottom performers to "catch up."

Individuals differed in both their initial performance and their rate of learning. Working memory predicted initial and final performance but not rate of learning. The pattern of results suggest that working memory capacity and training are additive. That is, some individuals high in working memory capacity have initially high simultaneous interpreting performance scores and do not improve as much as others with training. Other high working memory capacity individuals initially have closer to average simultaneous interpreting performance scores but improve quickly such that they perform more similarly to other high working memory capacity individuals by the end of training. Initial simultaneous interpreting performance and working memory capacity accounted for 73% of the variance in final interpreting performance.

---

[5] This is based on working memory capacity: coordination & transformation. A model with working memory capacity: maintenance & interference and initial interpreting performance also predicts final performance well, $F(2, 31) = 36.85$, $p < .001$, $R^2 = .70$.

Accounting for nearly 75% of the variance in final interpreting performance is an impressive result but we would be remiss not to acknowledge a few limitations of the current study. The final sample size was smaller than we set out to obtain. Data collection was labor intensive given the number of tasks, the longitudinal design, and the fact that the interpreter students were tested at their home institutions 1.5 and 2.5 h away respectively, rather than in our own laboratory setting. For these reasons we also had to limit the number of tasks that could be administered. A more ideal approach would be to administer more tasks to a larger sample and conduct latent growth curve analyses. Another limitation is that there was no comparison group. While the high reliability of the tests and the spaced administration leave minimal chance that the effects are due to practice, an added control group would allow for testing of practice effects.

Taken together, our findings clearly indicate that for simultaneous interpreting, working memory capacity is a crucial predictor of performance. This finding can make a substantial impact on the field of simultaneous interpreting where, for both bimodal and unimodal bilingual interpreters, it is estimated that approximately 90% of individuals who enter interpreting programs fail to become interpreters, either by failing to complete the program or by graduating without the skills necessary to work as a proficient interpreter (Humphrey & Alcorn, 1994; Timarová & Ungoed-Thomas, 2008). With only two measures, working memory capacity and initial performance, we captured over 70% of the variance in final performance. This finding suggests that interpreter programs could measure these two factors at admission to estimate students' likely success in the program.

Working memory capacity's interaction with training was more complex than hypothesized: common laboratory findings (e.g., that experience is compensatory, that the rich get richer) did not fully explain the additive interaction between working memory capacity and training on performance. One reason could be the complexity of a real-world task over laboratory-designed tasks or the variety of simple real-world tasks that have previously been studied. For example, Crossman (1959) found that amount of experience strongly predicted cigar rolling skill—working memory capacity may not be highly predictive of performance in such a task. However, many differences between the two tasks are observable, not only in complexity, but in predictability of the input and environment, in whether the dimensions are discrete or continuous, and whether the task proceeds in steps versus concurrent processes (Feltovich, Spiro, & Coulson, 1989, 1993; Hoffman et al., 2013). Future research should include examinations of these task features. Through these investigations we can better understand how task features moderate the impact of training on performance and of cognitive abilities on performance. Across a wide range of real-world tasks, training programs, and professions, this information can help systematically guide admission, recruiting, hiring, and training decisions to improve the caliber of performance.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

Acknowledgements

We thank Eileen Forrestal, Cynthia Williams, David Rivera, Rob Hills, Vanessa Watson, Ashley Graham, Danny MacDougal, and the participants in this study for their time and effort.

References

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child Development*, *83*(2), 413–422.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, *70*, 636–644.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology*, *60*(1), 68.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language & Cognition*, *12*(01), 3–11.

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, *19*(1), 19–23.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*, *7*(3), 325–339.

Calabria, M., Hernández, M., Branzi, F. M., & Costa, A. (2011). Qualitative differences between bilingual language control and executive control: Evidence from task-switching. *Frontiers in Psychology*, *2*, 1–10.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*, 319–333.

Cattell, R. B., & Cattell, A. K. S. (1960). *Handbook for the individual or group culture fair intelligence test*. Savoy, IL: Institute for Personality and Ability Testing.

Christoffels, I. K., de Groot, A. M. B., & Waldorp, L. J. (2003). Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism: Language & Cognition*, *6*(3), 201–211.

Coleman, M., & Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, *60*(2), 283–284.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Cowan, N. (2000). Processing limits of selective attention and working memory: Potential implications for interpreting. *Interpreting*, *5*(2), 117–146.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, *2*(2), 153–166.

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism an example of publication bias? *Psychological Science*, *26*(1), 99–107.

Emmorey, K., Borinstein, H. B., Thompson, R., & Gollan, T. H. (2008). Bimodal bilingualism. *Bilingualism: Language & Cognition*, *11*(01), 43–61.

Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals evidence from bimodal bilinguals. *Psychological Science*, *19*(12), 1201–1206.

Engel de Abreu, P. M. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory*, *19*(5), 529–537.

Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1989). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misconceptions. In D. Evans, & V. Patel (Eds.), *Cognitive science in medicine*. Cambridge: MIT (Bradford) Press.

Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1993). Learning, teaching, and testing for complex conceptual understanding. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 181–217). Hillsdale, NJ: Lawrence Erlbaum Associates.

Festman, J., Rodriguez-Fornells, A., & Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions*, *6*(5), 1–12.

Frauenfelder, U. H., & Schriefers, H. (1997). A psycholinguistic perspective on simultaneous interpretation. *Interpreting*, *2*, 55–89.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gerver, D., Longley, P. E., Long, J., & Lambert, S. (1989). Selection tests for trainee conference interpreters. *Meta*, *34*(4), 724–735.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism*, *1*, 67–81.

Gunning, R. (1952). *The technique of clear writing*. New York, NY: McGraw-Hill.

Hambrick, D. Z., & Engle, R. W. (2003). The role of working memory in problem solving. In J. E. Davidson, & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 176–206). Cambridge: Cambridge University Press.

Hervais-Adelman, A., Moser-Mercer, B., & Golestani, N. (2015). Brain functional plasticity associated with the emergence of expertise in extreme language control. *NeuroImage*, *114*, 264–274.

Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D. H. (2013). *Accelerated expertise: Training for high proficiency in a complex world*. New York: Psychology Press.

Humphrey, J., & Alcorn, B. (1994). *So you want to be an interpreter?: An introduction to sign language interpreting*. Clearwater, FL: H & H Publishers.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel (No RBR-8-75)*. Naval Technical Training Command Millington TN Research Branch.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science*, *23*(3), 159–163.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*, 119–135.

Lemay, S., Bédard, M. A., Rouleau, I., & Tremblay, P. L. (2004). Practice effect and test–retest reliability of attentional and executive tests in middle-aged to elderly subjects. *The Clinical Neuropsychologist*, *18*(2), 284–302.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861–883.

Macnamara, B. N., & Conway, A. R. A. (2014). Novel evidence in support of the bilingual advantage: Influences of task demands and experience on cognitive control and working memory. *Psychonomic Bulletin & Review*, *21*, 520–525. http://dx.doi.org/10.3758/s13423-013-0524-y

Macnamara, B. N., Moore, A. B., Kegl, J. A., & Conway, A. R. A. (2011). Domain-general cognitive abilities and interpreter skill prediction. *Interpreting*, *13*(1), 121–142.

MacWhinney, B. (1997). Simultaneous interpretation and the competition model. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 215–233). Thousand Oaks, CA: Sage.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language & Cognition*, *6*(02), 97–115.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language & Cognition*, *11*(01), 81–93.

McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading*, *12*(8), 639–646.

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *114*(2), 187–202.

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*(6), 719–726.

Moser, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. In D. Gerver, & H. W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 353–368). New York/London: Plenum Press.

National Association of the Deaf, & Registry of Interpreters for the Deaf. (2011). *NAD-RID national interpreter certification (NIC) candidate handbook*. Alexandria, VA: Registry of Interpreters for the Deaf.

Oberauer, K., Süβ, H. M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*(6), 641–652.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society*, *17*(4), 682–691.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language & Cognition*, *13*(02), 253–262.

Pyers, J. E., & Emmorey, K. (2008). The face of bimodal bilingualism grammatical markers in American Sign Language are produced when bilinguals speak to English monolinguals. *Psychological Science*, *19*(6), 531–535.

Raven, J. C., Raven, J. C., & Court, J. H. (1962). *Coloured progressive matrices*. London: HK Lewis.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763.

Salthouse, T. A., Toth, J., Daniels, K., Parks, C., Pak, R., Wolbrette, M., et al. (2000). Effects of aging on efficiency of task switching in a variant of the trail making test. *Neuropsychology*, *14*(1), 102.

Seal, B. C. (2004). Psychological testing of sign language interpreters. *Journal of Deaf Studies and Deaf Education*, *9*(1), 39–52.

Senter, R. J., & Smith, E. A. (1967). *Automated readability index*. Aerospace Medical Research Laboratories.

Shlesinger, M. (2003). Effects of presentation rate on working memory in simultaneous interpreting. *The Interpreters' Newsletter*, *12*, 37–49.

Soveri, A., Laine, M., Hämäläinen, H., & Hugdahl, K. (2011). Bilingual advantage in attentional control: Evidence from the forced-attention dichotic listening paradigm. *Bilingualism: Language and Cognition*, *14*(3), 371–378.

Timarová, Š., & Ungoed-Thomas, H. (2008). Admission testing for interpreting courses. *The Interpreter and Translator Trainer*, *2*(1), 29–46.

Treisman, A. M. (1965). The effects of redundancy and familiarity on translating and repeating back a foreign and a native language. *British Journal of Psychology*, *56*(4), 369–379.

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

Wechsler, D. (1997). *WMS-III: Wechsler memory scale administration and scoring manual*. Psychological Corporation.

Received 30 September 2015;
received in revised form 11 December 2015;
accepted 12 December 2015
Available online 19 December 2015