Published in *Journal of Speech, Language, and Hearing Research, 59, 38-49*  
(http://jslhr.pubs.asha.org/article.aspx?articleid=2471591&resultClick=3)

English Language Learners’ Nonword Repetition Performance: The Influence of Age, L2 Vocabulary Size, Length of L2 Exposure and L1 Phonology

Tamara Sorenson Duncan and Johanne Paradis  
University of Alberta
Correspondence concerning this article should be addressed to Tamara Sorenson Duncan, Department of Linguistics, University of Alberta, Edmonton, AB, T6G 2E7. Email: tamara.sorensonduncan@ualberta.ca.
Abstract

**Purpose:** This study examined individual differences in English language learners’ (ELLs) nonword repetition (NWR) accuracy, focusing on the effects of age, English vocabulary size, length of exposure to English and first language (L1) phonology.

**Method:** Participants were 75 typically-developing ELLs (mean age 5;8) whose exposure to English began on average at age 4;4. Children spoke either a Chinese language or South Asian language as an L1 and were given English standardized tests for NWR and receptive vocabulary.

**Results:** While the majority of ELLs scored within or above the monolingual normal range (71%), 29% scored below. Mixed logistic regression modeling revealed that a larger English vocabulary, longer English exposure, South Asian L1, and older age all had significant and positive effects on ELLs’ NWR accuracy. Error analyses revealed the following L1 effect: onset consonants were produced more accurately than codas overall but this effect was stronger for the Chinese group whose L1s have a more limited coda inventory compared to English.

**Conclusion:** ELLs’ NWR performance is influenced by a number of factors. Consideration of these factors is important in deciding whether monolingual norm-referencing is appropriate for ELL children.

*Keywords:* nonword repetition, English language learners, child second language acquisition
Nonword repetition (NWR) tasks – tasks that require children to repeat nonsense words – have been used to identify language impairment in a number of different populations of children. For example, researchers have shown that these tasks differentiate between groups of monolingual English learners with and without language impairment (e.g., Archibald & Gathecole, 2006; Dollaghan & Campbell, 1998; Gallon, Harris & van der Lely, 2007; Laing & Kamhi, 2003), including school-aged children from diverse socio-economic and varied English dialect backgrounds (Campbell, Dollaghan, Needleman, & Janoksy, 1997; Ellis Weismer, Tomblin, Zhang, Buckwalter, Chynoweth & Jones, 2000). Additionally, the accuracy of these tasks as clinical assessments has been demonstrated cross-linguistically (e.g., Girbau & Schwartz, 2007 for Spanish; Rispens & Parriger, 2010 for Dutch; Sahlen, Reuterskiöld-Wagner, Nettlebladt & Radeborg, 1999 for Swedish).

NWR tasks have also been recommended as language assessments for culturally and linguistically diverse learners. It is argued that these tasks are more suitable for diverse learners because they are less influenced by children’s prior linguistic experience than assessments that rely on accumulated language-specific knowledge (Armon-Lotem & Chiat, 2012; Campbell, et al., 1997; Ellis Weismer, et al., 2000; Thordardottir & Brandeker, 2013). Research with varied samples of bilingual children, however, has produced mixed results in terms of the clinical effectiveness of NWR tasks, i.e., the ability to differentiate between bilingual children with and without language impairment (e.g., Engel de Abreu, Baldassi, Puglisi & Befi-Lopes, 2013; Girbau & Schwartz, 2008; Guiberson & Rodríguez, 2013; Gutierrez-Clellen & Simon-Cereijido,
With respect to bilingual children who speak a minority first language (L1) and are learning a majority second language (L2), researchers have reported that L2 children’s NWR performance is depressed compared to that of their monolingual peers (Engel de Abreu et al., 2013; Kohnert, Windsor & Yim, 2006; Paradis et al., 2013; Windsor et al., 2010). Consequently, using NWR in English could lead to over-identification of language impairment amongst children who are learning English as an L2 – English language learners (ELL) (Kohnert et al., 2006; Paradis et al., 2013; Windsor et al., 2010). Yet, in diverse social contexts, English-language testing might be the only available option (Paradis et al., 2013). Further research into the factors that influence bilingual, and especially ELL, children’s performance on NWR tasks would enhance our understanding of their effectiveness in clinical assessment and under what circumstances it might be appropriate to use NWR with bilingual children. Accordingly, the goals of this study were to determine (a) how a group of typically-developing ELLs performed on a standardized NWR measure with respect to monolingual norms and (b) how age, English vocabulary size, exposure to English and L1 influenced the variation (individual differences) in children’s scores.

**Factors Influencing Children’s Nonword Repetition Performance**

Research with varied samples of children suggests that individual differences in NWR accuracy may be influenced by: (a) cognitive maturity or age (e.g., Alloway, Gathercole, Willis & Adams, 2004; Edwards, Beckman & Munson., 2004; Gathercole, 1999), (b) vocabulary size (Masoura & Gathercole, 1999; Thordardottir & Brandeker, 2013; Thorn & Gathercole, 1999); (c) amount of L2 exposure (e.g., Paradis et al., 2013; Summers, Bohman, Gillam, Peña & Bedore,
2010; Thordardottir & Brandeker, 2013; Thorn & Gathercole, 1999; Windsor et al., 2010), and (d) L1 background (Sorenson Duncan, 2010). However, to date, these influences have not been considered simultaneously within a single study.

Age. NWR tasks were designed to measure phonological short-term memory capacity, which is known to increase with age (e.g., Case, Midian Kurland & Goldberg, 1982; Dempster, 1978). As expected, research with monolingual children has revealed a significant connection between age and NWR performance, with older children displaying better performance (e.g., Alloway, et al., 2004; Edwards et al., 2004; Gathercole, 1999). However, in these monolingual studies, older in terms of age also means greater experience with the target language (e.g., Gathercole et al., 1992; Gathercole, 2006). ELL children provide an opportunity to study the effects of age more independently from exposure because within a sample of ELL children, children of different ages could have comparable amounts of exposure to English (e.g., both a five year-old and six year-old with 12 months of exposure). To our knowledge, age has not been systematically evaluated as a factor in ELL children’s NWR productions.

Vocabulary Size. Research with monolingual children has demonstrated that NWR performance is influenced by children’s receptive vocabulary size (e.g., Baddeley, Gathercole & Papagno, 1998, Edwards, et al., 2004; Gathercole, 2006; Gathercole, Willis, Emslie & Baddeley, 1991; Gathercole, Willis, Emslie & Baddeley, 1992; Thorn & Page, 2008), with children with larger vocabularies having higher accuracy on NWR tasks (e.g., Edwards et al., 2004; Gathercole et al., 1992). For example, Edwards et al. (2004) reported that vocabulary size accounted for 29% of the variance in NWR accuracy in a sample of English-speaking monolingual children (aged 3;2-8;10, years;months). Studies with both simultaneous bilinguals (Thordardottir & Brandeker, 2013) and foreign language learners (Masoura & Gathercole, 1999; Thorn &
Gathercole, 1999) have also suggested that L2 vocabulary development is closely related to NWR performance for bilingual children. In fact, Thorn and Gathercole (1999) suggested that L2 vocabulary development is more important in predicting L2 NWR performance than L1 vocabulary is for predicting L1 NWR performance. They examined NWR abilities in children (aged 4;01 – 8;0) who were English-speakers learning French as a foreign language. They reported that L2 (French) vocabulary knowledge accounted for 38% of the variance in children’s French-based NWR abilities; whereas L1 (English) vocabulary knowledge only accounted for 9% of the variance in children’s English-based NWR abilities. Similarly, Masoura and Gathercole (1999) examined NWR abilities in Greek-L1, English-L2 children (mean age 8;8). They reported that children’s foreign (English) receptive vocabulary size was highly correlated with their repetition of foreign-like (English) nonwords. Thus, receptive vocabulary size appears to be an important individual difference factor that can impact NWR performance in bilingual children, particularly in the early stages of language learning. To date, this predictive relationship has not been studied in ELLs.

**Amount of L2 Exposure.** Currently, the largest body of evidence suggesting a connection between amount of exposure and bilingual/L2 children’s NWR productions comes from studies that compare these children with monolingual children. Researchers have noted that L2 children have lower NWR performance when compared to their monolingual peers (Engel de Abreu, et al., 2013; Kohnert, et al., 2006; Paradis et al., 2013; Windsor et al., 2010). Since monolingual children would have had greater frequency of exposure, on average, to the target language (e.g., English) than ELL children, amount of exposure could, at least in part, underlie the group differences. Another relevant finding is that bilingual children perform better on nonwords that conform to the phonotactics of their L1 (the language with greater exposure) than
those constructed on the basis of L2 phonotactics (the language with less exposure) (Masoura & Gathercole, 1999; Thorn & Gathercole, 1999).

More direct evidence for an effect of exposure can be found in studies that compare NWR performance amongst groups of bilingual/L2 children with varying degrees of exposure to the target language. Emerging research suggests that bilingual/ELL children’s accuracy on NWR tasks increases with exposure time to the target language (Paradis et al., 2013; Summers et al., 2010; Thordardottir & Brandeker, 2013). This finding, however, is not undisputed. For example, Thordardottir and Brandeker (2013) reported a significant association between simultaneous bilingual children’s amount of exposure to English and their performance on an English-based NWR task, but no significant association was found between these same children’s amount of exposure to French and their performance on a French-based NWR task. As such, further investigations are needed to determine the contexts where exposure can impact NWR performance.

**L1 Phonology.** Since bilingual children speak two languages, and would be dominant in their L1 at the early stages of L2 acquisition, it is important to consider the role of L1 in L2 NWR performance. The presence and extent of crosslinguistic transfer on bilingual children’s NWR performance has not been well-investigated to date. The phonological system of the L1 may influence NWR in the L2 because research with young ELLs has shown that L1 phonological properties are transferred into children’s production of known L2 words (Fabiano-Smith & Barlow, 2010; Tessier, Sorenson Duncan & Paradis, 2013). In a pilot study of eight children, Sorenson Duncan (2010) found that ELL children whose L1s had restricted coda inventories were less accurate with coda consonants on an English-based NWR task compared to children whose L1s had rich coda inventories. Therefore, further research with a larger sample
of ELL children that investigates the impact of L1 phonology on ELLs’ NWR performance is warranted.

We aimed to test the hypothesis that children’s L1s are relevant in understanding their English NWR performance; however, a full contrastive analysis on the basis of children’s entire L1 phonological inventories would go beyond the scope of this paper. Thus, we opted to focus on how one particular property of children’s L1s might impact their NWR performance. Specifically, coda consonants were selected as the basis of this L1 analysis because: (a) there are multiple opportunities to observe children’s coda productions on the CTOPP (Wagner, Torgesen & Rashotte, 1999); (b) coda consonants are noted as a potential source of difficulty (i.e., less accurate) for learners, including ELL children (Morrow, Goldstein, Gilhool & Paradis, 2014; Rattanasone & Demuth, 2014; Sorenson Duncan, 2010); and (c) languages vary with respect to the extent that coda consonants are permitted.

The L2 children in this study spoke L1s from one of two typologically-distinct language groups: (1) Chinese (Cantonese or Mandarin) or (2) South Asian (Hindi, Punjabi or Urdu). These particular languages are commonly spoken by North American children learning English as an L2 (Statistics Canada, 2011; United States Census, 2010). Importantly, these languages differ in syllable structure. The vast majority of words in Chinese languages consist of one open syllable (Duanmu, 2007 for Mandarin; Matthews & Yip, 1994 for Cantonese). In fact, the only coda consonants reported for Mandarin are [n] and [ŋ], (Duanmu, 2007). In Cantonese, only unreleased voiceless stops are permitted in coda position: [p˞] [t˞] [k˞] (Matthews & Yip, 1994). In contrast, South Asian languages have extensive coda inventories (Akram, 2002 for Urdu; Bhatia, 1993 for Punjabi; Broselow, Chen & Huffman, 1997; Gordon, Jany, Nash, & Takara,
2008 for Hindi). For example, Hindi allows up to 33 consonants in coda position (Gordon, et al., 2008), as well as some consonant clusters.

**The Present Study**

The present study addressed the following research questions:

(1) How do ELL children compare to monolingual expectations on an English-based NWR task? If this sample of ELL children follows the same patterns observed in previous studies with ELLs (e.g., Kohnert et al., 2006; Paradis et al., 2013; Windsor et al., 2010), they will show depressed NWR scores compared to monolingual age-based expectations. This finding would suggest that the use of NWR tasks with ELL children carries a risk of over-identification of language impairment in this population. The alternate hypothesis, following from studies suggesting that NWR tasks offer reduced bias in assessment in bilinguals, is that ELLs would have performance in line with monolingual expectations (Arnon-Lotem & Chiat, 2012; Campbell, et al., 1997; Ellis Weismer et al., 2000; Thordardottir & Brandeker, 2013).

(2) Are differences in NWR accuracy influenced by individual differences among ELLs in age, English vocabulary size, exposure to English and L1? We predicted that each of these factors will significantly influence children’s performance. While prior research has shown evidence for each of these factors playing a role, the factors were examined in isolation of each other, and in some cases, in bilingual children other than ELLs (for vocabulary: Masoura & Gathercole, 1999; Thorn & Gathercole, 1999; for exposure: Kohnert, et al., 2006; Paradis et al., 2013; Summers et al., 2010; Thordardottir & Brandeker, 2013; Thorn & Gathercole, 1999; Windsor et al., 2010; for age: Gathercole, 1999; Gathercole et al., 1992; for L1: Sorenson...
Duncan, 2010). Thus, the present study has the potential to also reveal interactions between these factors.

(3) Is the syllable position of consonants related to errors in producing consonants? Does L1 background influence consonant accuracy-by-position? These questions are meant to more specifically examine the extent to which the structure of the nonwords (i.e., phonotactics) and children’s L1 background impact performance. Based on L2 phonological developmental trends (Morrow et al., 2014), ELLs were predicted to have greater difficulty with coda consonants, regardless of their L1. Furthermore, the Chinese group was predicted to have depressed accuracy with coda consonants when compared to the South Asian group because of the crosslinguistic differences in word syllable structure described above.

**Method**

**Participants**

Participants were 75 typically-developing ELLs whose exposure to English began on average at 4;4 (standard deviation, SD = 9 months). Children had an average of 16.73 months of exposure to English in preschool/school (SD = 8.40 months, range = 2-37 months) and their mean age at testing was 5;8 (SD = 5.67 months, range = 5;0-6;9). According to parental report, all children had normal hearing and there were no concerns about speech or language impairments (as these children were selected from a larger database of ELL children, details of their classification as typically-developing can be found in Paradis, Emmerzael and Sorenson Duncan, 2010). Children were living in a majority-English Canadian city, in households where both parents were foreign-born and L2 learners of English. Forty-three percent of the children were also foreign-born. The Canadian-born children were not exposed to English on a consistent
basis until they started a preschool or school program. Having two foreign-born parents and being exclusively or primarily exposed to the native language before English learning began were criteria for inclusion in the study in order to exclude simultaneous bilinguals from the sample. On average, children arrived in Canada at age 18.21 months ($SD = 24.30$ months). There was little variation in this sample in terms of socio-economic status, as measured by mother’s level of education (average schooling = 14.47 years, $SD = 3.34$ years). The majority (54/75) of mothers interviewed had completed some post-secondary school, with many of the mothers (40/75) having at least one university degree. Children spoke an L1 that fit into one of two typologically-distinct language categories: Chinese languages ($n = 36$) or South Asian languages ($n = 39$). In terms of maternal level of education, age, exposure to English, age of exposure and initial exposure to English, there were no statistically-significant differences between the children in these language groups (see Table 1).

The children in this study comprised an overlapping but not identical sample to the participant groups in Blom and Paradis (2013), Paradis et al. (2010), Paradis (2011), Paradis et al. (2013), and Tessier, Paradis and Sorenson Duncan (2013). Factors influencing NWR performance were not analysed in these studies. Ethics approval for conducting this research was granted to the second author of this study from the Health Research Ethics Board at the University of Alberta, Canada.

**Procedures**

Parents and children completed several tasks either at school or in their homes. Tasks were administered by a native-speaker of Canadian English. The details of each task are outlined below.
A parental interview was conducted with the help of a cultural broker/interpreter to obtain relevant background information. Specifically, information was gathered about mother’s level of education, child’s age, age of arrival in Canada, and cumulative English exposure, quantified as the number of months that the child had received “consistent and sustained” exposure to English in a daycare, preschool or school setting (Paradis, 2011, p.222). Overheard English speech or English from media sources was not considered to be consistent and sustained exposure.

**Peabody Picture Vocabulary Test** (PPVT-IIIR, Dunn & Dunn, 1997). The PPVT-IIIR is a measure of receptive vocabulary. Children are shown a set of four pictures and asked to find the picture that corresponds to a word spoken by the research assistant. Raw scores are calculated as the number of correct responses before the ceiling item. Age-referenced standard scores can be calculated but were not necessary for the purposes of this paper. The motivating factors for using raw scores in the regression analyses were: (a) raw NWR counts comprised the outcome variable in the regression analyses and thus, raw vocabulary scores were more appropriate for the predictors, (b) raw scores adequately provide an estimate of children’s relative vocabulary size and (c) chronological age was included as a separate predictor in the model, and thus, age differences were considered in the analyses.

**Comprehensive Test of Phonological Processing – Nonword Repetition Task** (CTOPP: Wagner et al., 1999). The CTOPP includes a NWR task based on the phonotactics of English and nonwords increase in complexity (i.e., syllable length) as the task progresses. This particular NWR task was selected because it is normed with North American not British English, and is used clinically across Canada. The minimal word length is one syllable and the maximum is seven syllables. Each nonword contains at least one coda.
Children repeated each nonword after hearing it played on a computer and their responses were recorded for later transcription and scoring. Raw scores were calculated on the basis of word-by-word scoring and were the sum of correct responses before ceiling (a score of zero on three consecutive nonwords). Following the scoring guidelines of Wagner et al. (1999), if the child produced the nonword with one or more deleted, substituted or additional phonemes, the nonword was scored as incorrect. Age-matched standard scores were based on English monolingual norms, with the average standard mean equaling 10 and a SD of 3 (1 SD range = 7-13). These monolingual norms are based on a sample of 1,656 American children from 30 different states, with the majority of participants (82%) identified as “White” (Wagner et al., 1999, p.63).

In order to address research question (3), segment-level scoring was also conducted. Specifically, onsets and codas were identified following the syllable boundaries provided by Wagner et al. (1999). Across the eighteen nonwords on the CTOPP, there are seventeen word-final codas and five word medial codas. The proportion of correct onsets was calculated for each child by dividing the total number of correct onsets across all attempted nonwords by the total onsets in these nonwords. This procedure was repeated for coda consonants. To maximize the opportunity to observe coda consonants, word-medial and word-final codas were tallied together. Nonword targets were excluded from this analysis when the child did not provide an answer (21/781 nonwords, across 9/75 children).

To assess the reliability of the transcription and scoring, 10% of the sample was transcribed and scored by an independent research assistant. Scoring reliability was measured as the percentage of nonwords per child that received the same score. Any discrepancies were
settled by consensus. For example, if the child produced 13 nonwords and the two scorers assigned 12 of these 13 nonwords the same score, the scoring reliability for this child’s score was 92% (12/13). The mean score agreement was excellent at 91% (SD = 12%). Transcription reliability was measured as the percentage of agreement for consonants per child. Again reliability was assessed by averaging across the reliability for each child. Excellent inter-rater agreement was achieved (mean transcription agreement = 94%, SD = 6%). These reliability measures are commensurate with other studies in the field (e.g., Windsor et al., 2010).

**Results**

**ELLs’ NWR Performance and Monolingual Norms**

The L2 children’s average raw NWR score was 5.43 (SD = 2.61, range = 1-12), resulting in an average standard score of 7.81(SD = 1.84, range = 4-14). Standard scores were used to compare ELLs’ performance with age-based monolingual expectations; the CTOPP has a mean standard score of 10 with a SD of 3. Twenty-two children (29%) obtained standard scores of 6 or less (i.e., below the normal range for monolingual children the same age) (see Figure 1). Twelve children (16%) obtained standard scores of 10 or higher (i.e., at or above the monolingual mean). Therefore, while the majority of these typically developing L2 children scored within normal limits, not all children did.

**The influence of vocabulary size, L2 exposure, age and L1 on NWR Accuracy**

Mixed logistic regression modeling was used to estimate the effect of a number of factors on children’s NWR scores. Children’s responses for each nonword were coded as correct (1) or incorrect (0) and then analyzed in R (R Core Team, 2014) using a logistic link function for
correct production of nonwords and binomial variance. This statistical technique was chosen because it has advantages over other techniques, including robustness, inclusion of multiple factors simultaneously and inclusion of random effect variables (Baayen, 2008). English vocabulary size (PPVT-IIIIR), cumulative exposure to English in months (MOE), L1 (L1 Group), age (Age) and age of arrival (AOA), were included as fixed effects in the mixed effect regression models. AOA was included in the modeling to test for any potential effects of casual English exposure that the children may have received prior to the onset of English exposure through daycare, preschool or school. Both participant and nonword were included in all models as random effects. The outcome variable was the log odds ratio of the probability of correct nonwords over the probability of incorrect nonwords.

The relationship between fixed effects was evaluated to determine if there were any issues of collinearity. Pearson’s correlations were used to test the relationship between continuous variables. No continuous variables were strongly correlated (.75 or higher), and as such they could all be entered together in the models. Welch two-sample t-tests were used to evaluate the relationship between the continuous variables and L1 (a nominal variable). There was no significant relationship between L1 and the continuous variables.

Our regression analysis involved two steps. First, we used the technique of backwards elimination. A full model that overfit the data was created and then, in a step-wise fashion, all non-significant variables were removed. AOA was removed from the model because it was not a significant predictor of NWR accuracy. Second, the optimal model was selected by comparing nested models using maximum likelihood ratio tests (Baayen, 2008), comparing the Akaike information criterion (AIC) generated for each model (Baayen, 2008) and evaluating the goodness of fit of each model with the Concordance Index C (Chatterjee & Hadi, 2006).
ranges between 0.5 and 1; a $C$ value above 0.8 indicates that the model performs well. The model containing MOE, L1, Age, PPVT-IIIR and an interaction between MOE and L1 was found to be optimal as the maximum likelihood ratio test revealed it to be the statistically preferred model (e.g., comparing this model to a nested model with all of the variables except PPVT-IIIR, $\chi^2$ Difference $= 4.083, p = 0.04$). Furthermore, it had the lowest AIC value of the models (912.9) and an appropriate $C$ value (0.82).

The findings from the optimal model are provided in Table 2 and are described here. Cumulative English exposure (MOE) had a positive and significant effect on NWR performance ($\beta = 0.040, z = 2.548, p = 0.011$), that is, children with more exposure to English had greater accuracy with nonword productions. L1 Group had a significant effect on NWR performance ($\beta = 0.930, z = 2.223, p = 0.026$). The Chinese group was the reference level for this effect. As such, the positive $\beta$-value indicates that children in the South Asian group had higher performance than children in the Chinese group. Age had a significant and positive effect on nonword accuracy ($\beta = 0.041, z = 2.214, p = 0.027$), with older children having greater accuracy. English receptive vocabulary size (PPVT-IIIR) had a positive and significant effect on nonword accuracy ($\beta = 0.010, z = 2.017, p = 0.044$), that is, larger vocabularies were associated with more accurate productions. A significant interaction was also found between cumulative English exposure (MOE) and L1 Group ($\beta = -0.042, z = -1.917, p = 0.055$). As the Chinese group was the reference level, the negative $\beta$-value for this interaction suggests that the effect of exposure is reduced in the South Asian group compared to the Chinese group. Put differently, L1-based differences were more pronounced when children had less exposure to English, with the Chinese group showing decreased performance compared to the South Asian group at low levels of exposure.
Effects of Syllable Position and L1 Phonology on NWR Consonant Accuracy

In terms of consonant accuracy by syllable position, children produced onsets significantly more accurately than codas \( t(65) = 6.75, p = 4.78 \times 10^{-9} \), with an average of 89\% \( (SD = 9\%) \) correct onsets and 76\% \( (SD = 14\%) \) correct codas. There was no difference in onset accuracy between children on the basis of L1 \( t(52.24) = 0.43, p = 0.67 \). By contrast, there was a significant difference in coda accuracy between the two L1 groups \( t(53.81) = -2.76, p = 0.008 \) \( d = -0.74 \) (medium), with children in the South Asian languages group producing 79\% \( (SD = 12\%) \) correctly and children in the Chinese Language group producing 70\% \( (SD = 15\%) \) correctly. Figure 2 provides a graphical representation of children’s consonant accuracy by syllable position and L1.

Discussion

This study had two main goals. First, we sought to provide insight into typically-developing ELLs’ performance on a standardized NWR task, specifically with respect to monolingual norms. Second, we examined multiple factors influencing individual differences in ELLs’ NWR performance. Seventy-five typically-developing ELLs who spoke either a Chinese or a South Asian language at home completed the NWR task from the CTOPP (Wagner et al., 1999), a receptive vocabulary measure, PPVT-IIIR (Dunn & Dunn, 1997) and their parents completed an interview, providing relevant background information.

In this study, 29\% of the ELLs scored below the normal range for monolinguals, even though all children were typically-developing L2 learners. These findings are consistent with other studies that have compared monolingual and ELLs’ NWR performance (Kohnert et al.,
2006; Paradis et al., 2013; Windsor et al., 2010). However, these findings appear to conflict with those of some other studies (e.g., Campbell, et al., 1997; Ellis Weismer et al., 2000; Thordardottir & Brandeker, 2013), where NWR tasks are recommended as unbiased assessments. We propose that differing results are likely the consequence of differences in the populations that have been studied. Alternatively, differing results could arise because of differences in the NWR tasks that have been used (Chiat, in press). However, the parallels between our findings and Windsor et al. (2010), who used a different NWR task (Dollaghan and Campell’s Nonword Repetition Task), suggest that our findings are likely not simply a by-product of the NWR task we selected. As such, we conclude that differences between the children are a more likely source of the differing results. In this study, children were sequential bilinguals who began their L2 (English) acquisition when they entered a preschool or school program. Campbell et al. (1997) and Ellis Weismer et al. (2000) examined children from diverse socio-economic status, cultural, and possibly dialectal, backgrounds but all children were monolingual English-speakers. In Thordardottir and Brandeker (2013), bilingual children had been learning both languages (French and English) from birth. Therefore, the extent of bias that exists with NWR tasks could depend on the type of child being assessed. That is, NWR tasks may contain more bias when used with ELLs compared to monolingual children from diverse backgrounds or simultaneous bilingual children.

Linear mixed regression modeling revealed age, English vocabulary size, amount of English exposure and L1 as significant predictors of ELL children’s NWR performance. In regards to age, the modeling revealed a significant main effect of age on NWR scores. That is, the older children in the sample had greater NWR accuracy than the younger children. We note this finding here because ELL children provide an opportunity to examine age in a way that
studies with monolingual children cannot. In a sample of monolingual children, older also means more experience with the language, but in this sample of ELL children, age and exposure are more independent of each other. As such, this study demonstrates that NWR performance improves with age irrespective of the amount of exposure children have received and the size of their English vocabularies. Thus, we interpret the significant result here as support for the long-standing claim that phonological short-term memory capacity increases with cognitive maturity, i.e., age (e.g., Case, Midian Kurland & Goldberg, 1982; Dempster, 1978).

The modeling also revealed a significant effect of English vocabulary size on NWR scores. This significance parallels findings of previous studies with monolingual children, simultaneous bilingual children and children learning a foreign language (Edwards et al., 2004; Gathercole et al., 1992; Masoura & Gathercole, 1999; Thordardottir & Brandeker, 2013; Thorn & Gathercole, 1999). In studies with monolingual children, the connection between these measures has been interpreted as an indication that some reference to existing words in memory occurs during NWR tasks (Edwards et al., 2004; Gathercole et al., 1991; Gathercole et al., 1992; Gathercole, Frankish, Pickering & Peaker, 1999). As such, this study provides further evidence that a vocabulary effect is robust across different populations of language learners and it suggests that existing linguistic knowledge (in this case English vocabulary) is utilized by ELL children when completing a NWR task. The significance of English vocabulary size is also important because it provides insights as to why ELLs do not have comparable NWR performance to their age-matched monolingual peers. Vocabulary size is a noted difference between ELL children and monolinguals, with ELL children having smaller English vocabularies (e.g., Oller, Pearson & Cobo-Lewis, 2007; Paradis et al., 2013; Poulin-Dubois, Bialystok, Blaye, Polonia & Yott, 2013; Windsor & Kohnert, 2004). For example, Paradis et al. (2013) reported that 38% of the
typically-developing ELL children in their sample scored below age-matched monolingual norms for receptive vocabulary. Thus, our results suggest that smaller vocabulary may be one root cause of the depressed NWR performance of ELLs.

Cumulative exposure to consistent English input (e.g., through schooling), was a significant predictor of NWR accuracy, with children who had more exposure to English performing better on the NWR task, and this finding is in parallel with prior research (Kohnert, et al., 2006; Masoura & Gathercole, 1999; Thorn & Gathercole, 1999; Paradis et al., 2013; Summers et al., 2010, Thordardottir & Brandeker, 2013; Windsor et al., 2010). From our analysis, we consider the significant effect of exposure to be important for understanding differences in NWR accuracy across different populations. For example, since monolingual and simultaneous bilingual children would have greater frequency of exposure to English than sequential bilinguals, like ELLs, amount of exposure could be another root cause of the depressed NWR performance of ELLs.

It is important to point out that our measure of exposure is broad in that it only measures L2 input quantity. We did not examine diversity of words, or quality of input properties known to influence L2 learning (Grüter & Paradis, 2014). However, it is worth considering what aspects of the input may influence NWR performance. One possibility is that the influence of greater exposure is connected to vocabulary size. That is, as children’s L2 exposure increases, so too does their L2 vocabulary size (e.g., Golberg, Paradis & Crago, 2008; Oller & Eilers, 2002). This connection is especially evident with foreign-language learners where length of study and vocabulary scores are highly correlated, making it difficult to determine which factor is influencing NWR performance (Masoura & Gathercole, 1999). However, for ELLs, children with comparable amounts of exposure can have different L2 vocabularies and vice versa,
children with similar L2 vocabulary scores can have different amounts of L2 exposure (e.g., Golberg et al., 2008). This point is borne out in the data presented in the current paper, where there was only a weak correlation between vocabulary size and exposure to English \( (r (73) = 0.36, p = 0.002) \). As such, although vocabulary size and amount of exposure were not entirely independent of each other, in this study they each contributed to understanding unique variance in NWR performance. Therefore, it is necessary to consider additional influences that could be related to increased exposure, and in turn, contribute to NWR performance. For example, research with monolinguals has shown that greater familiarity with the phonotactic patterns of English (Edwards et al., 2004) and increased articulatory knowledge (Munson, Edwards & Beckman, 2005) are relevant for NWR accuracy. Following from these studies, future research should look into the details of what specific properties of the input influence ELLs’ NWR accuracy.

In regards to L1, the results from the regression modeling indicated that children with a South Asian L1 were more accurate than children with a Chinese L1. However, the significant interaction between L1 phonology and exposure showed that differences related to L1 phonology were most pronounced for children in the early stages of English exposure. A relevant finding to understanding this interaction is the rate at which ELLs acquire the L2 phonology. Previous research has found rapid mastery of English phonemes by ELLs (e.g., Anderson, 2004; Morrow et al., 2014). For example, using proportion of whole word proximity, a measure of whole-word accuracy, Morrow et al. (2014) reported that by 21 months of exposure, ELL children with diverse L1 backgrounds were producing 93.4% of words in an English spontaneous speech sample correctly. Therefore, in the present study, it is likely that the Chinese L1 ELLs had begun
to develop more target-like English phonologies with longer exposure, and in so doing, L1 interference was reduced in their production of English nonwords.

In order to understand these L1 phonology effects in greater detail, we examined consonant productions by syllable position. As was predicted, coda consonants were a greater source of difficulty for children in the Chinese group compared to the South-Asian group. This specific example of crosslinguistic transfer based on L1 phonotactics suggests that nonwords that are not word-like in the L1 are more difficult for ELL children who are relatively new to English (i.e., within the first three years of their English exposure). As L1 transfer is not possible with monolingual children, this finding with ELLs is a possibly analogous to the findings from the monolingual literature where highly word-like nonwords are reproduced more accurately than less word-like nonwords (e.g., Edwards et al., 2004; Gathercole et al., 1999). Future research should examine whether other potential sources of L1 transfer impact NWR accuracy (e.g., segment-level as opposed to syllable-level phenomena).

Clinical Implications

There is currently conflict amongst recommendations regarding NWR tasks and bilingual children in clinical practice. Studies with bilingual children with extensive exposure find them to be less biased and appropriate; while, studies with ELLs, the present study included, have reported a risk of over-identification of language impairment. However, this risk should not be taken as evidence for abandoning NWR tasks in the assessment of ELL children. In fact, Paradis et al. (2013) found that compared to a standardized measure of receptive vocabulary and one of grammatical development, a NWR task was less biased (i.e., there was less overlap in NWR scores between ELLs with typical development and ELLs with language impairment, compared
to these other measures). As such, a more practical response might be the development of ELL-specific strategies for interpreting NWR performance.

One possible strategy could be limiting the use of English-based NWR tasks to ELLs who have had greater exposure to English and who come from certain L1 backgrounds. In our data, if we divide the sample by amount of exposure and L1, we find that with less than 18 months of exposure to English, 48% of typically-developing ELLs in the Chinese group scored below the normal range with monolinguals. In contrast, with greater than 18 months of exposure, only 19% of typically-developing ELLs in the South Asian Group scored below the normal range for monolinguals. It is important to note, though, that 19% is not zero. As such, simply limiting NWR use to particular groups of ELL children may not adequately mitigate against the risk of over-identification. Other strategies could include the development of ELL norms for NWR (Paradis et al, 2013), alternative scoring systems that take into account ELLs’ L1 background, or the development of NWR tests that are less language-specific (Chiat, in press). In conclusion, this study indicates that NWR should not be used uncritically with ELLs, but also shows that it could have promise as a clinical measure if certain interpretation strategies are developed.

Acknowledgements

We would like to thank the families who participated and the Edmonton Public School Board and the Toronto District Catholic School Board for permitting us to recruit participants in their schools. We would also like to express our appreciation to the Multicultural Health Brokers Cooperative in Edmonton for their dedicated services as interpreters and their help with recruitment. Finally, we are grateful to the following student assistants who collected and processed the data: Kyla Coole, Kristyn Emmerzael, Ruiting Jia, Katryna Lysay, Dorothy Pinto,
Emily Yiu, and Tatiana Zdorenko. This research was supported through the Vanier Canada Graduate Scholarship Program (awarded to Tamara Sorenson Duncan) and grants awarded by the following agencies: The Canadian Language and Literacy Research Network (awarded to Johanne Paradis and Phyllis Schneider), the Alberta Centre for Child, Family, and Community Research (awarded to Johanne Paradis and Phyllis Schneider), and the Alberta Innovates Health Solutions (awarded to Johanne Paradis).
References


### Table 1

Background Variables by First Language

<table>
<thead>
<tr>
<th></th>
<th>Chinese Group Mean (SD)</th>
<th>South Asian Group Mean (SD)</th>
<th>Welch two sample t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother’s Level of Education (years)</td>
<td>14.19 (2.84)</td>
<td>14.72 (3.77)</td>
<td>t(70.23) = -0.628, p = 0.497</td>
</tr>
<tr>
<td>Age (months)</td>
<td>68.89 (5.64)</td>
<td>68.82 (4.90)</td>
<td>t(69.62) = 0.056, p = 0.956</td>
</tr>
<tr>
<td>Exposure to English (months)</td>
<td>16.44 (8.88)</td>
<td>17.00 (8.04)</td>
<td>t(70.70) = -0.283, p = 0.778</td>
</tr>
<tr>
<td>Age of Arrival in Canada (months)</td>
<td>18.47 (25.11)</td>
<td>17.97 (23.85)</td>
<td>t(71.74) = 0.09, p = 0.93</td>
</tr>
<tr>
<td>Age of initial exposure to English (months)</td>
<td>52.44 (8.11)</td>
<td>51.90 (10.07)</td>
<td>t(67.27) = 0.258, p = 0.797</td>
</tr>
<tr>
<td>TOTAL</td>
<td>36 Children</td>
<td>39 Children</td>
<td></td>
</tr>
</tbody>
</table>
Table 2

Optimal Logistic Regression Model for L2 Children’s NWR Performance

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate (β)</th>
<th>SE</th>
<th>z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-5.00</td>
<td>1.380</td>
<td>-3.625</td>
<td>0.003</td>
</tr>
<tr>
<td>MOE</td>
<td>0.040</td>
<td>0.016</td>
<td>2.548</td>
<td>0.011</td>
</tr>
<tr>
<td>L1Group</td>
<td>0.930</td>
<td>0.418</td>
<td>2.223</td>
<td>0.026</td>
</tr>
<tr>
<td>Age</td>
<td>0.041</td>
<td>0.019</td>
<td>2.214</td>
<td>0.027</td>
</tr>
<tr>
<td>PPVT-IIIR</td>
<td>0.010</td>
<td>0.005</td>
<td>2.017</td>
<td>0.044</td>
</tr>
<tr>
<td>MOE * L1Group</td>
<td>-0.042</td>
<td>0.022</td>
<td>-1.917</td>
<td>0.055</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant</td>
<td>0.062</td>
<td>0.248</td>
</tr>
<tr>
<td>Nonword</td>
<td>2.510</td>
<td>1.584</td>
</tr>
</tbody>
</table>

Note. MOE = Months of Exposure to English; L1Group = Chinese or South Asian (Note: the Chinese group was the reference level); PPVT-IIIR = vocabulary measure.
Figure 1. ELL children’s NWR performance compared to age-matched monolingual expectations.
Figure 2. The proportion of consonants produced correctly by syllable position and L1 Group. The left panel illustrates onset accuracy and the right panel depicts coda accuracy.