Article

CHILD LANGUAGE TEACHING & THERAPY

Child Language Teaching and Therapy
29(1) 77–89

DOI: 10.1177/0265659012465208
clt.sagepub.com
SAGE

Can a novel word repetition task be a language-neutral assessment tool? Evidence from Welsh–English bilingual children

Kathryn M Sharp
Cardiff University, UK

Virginia C Mueller Gathercole
Bangor University, UK; Florida International University, USA

Abstract

In recent years, there has been growing recognition of a need for a general, non-language-specific assessment tool that could be used to evaluate general speech and language abilities in children, especially to assist in identifying atypical development in bilingual children who speak a language unfamiliar to the assessor. It has been suggested that a non-word repetition task (NWRT) may be a suitable candidate to fill this role, as it does not rely on knowledge of particular words for performance, and it may be possible to devise non-words that are not specific to any given language. The current study reports performance on a Welsh non-word repetition task by typically developing Welsh–English bilingual children with varying levels of exposure to Welsh in the home (Only Welsh at Home, Only English at Home, or Welsh and English at Home). The focus of the study was on repetition of initial consonants and consonant clusters in novel words. Both quantitative and qualitative differences were found across groups, according to level of exposure to Welsh, on sounds unique to Welsh, but not on sounds shared by Welsh and English. The data suggest that level of knowledge of the language tested has an important impact on children's performance on non-word repetition and that the use of the NWRT as a universal speech and language assessment tool should be adopted with caution.

Keywords

Non-word repetition task, bilinguals, exposure, speech and language assessment, Welsh

1 Introduction

In recent years, there has been an increasing search for a language-neutral assessment tool that could be used to assist in identifying atypical language development – e.g. specific language impairment (SLI) – especially for use with bilingual children. One candidate is the

Corresponding author:
Kathryn M Sharp, WISERD, Cardiff University, 46 Park Place, Cardiff, CF10 3BB, Wales, UK.
Email: SharpKM@cardiff.ac.uk

non-word repetition task (NWRT). Performance on the NWRT has been shown to be affected by non-linguistic and linguistic skills such as short-term memory (STM) and morphophonological processing (e.g. Armon-Lotem, 2011), both of which have been shown to be impaired in children with language difficulties (e.g. Montgomery, 2003). In a NWRT paradigm, children are asked to repeat non-existent 'words' in their language. As it involves non-words, the task is often considered to be non-language-specific, and thus potentially appropriate for assessing not only monolinguals, but also bilinguals learning a variety of languages. This opens the possibility that such a task could be used to test for SLI universally, based on the fact that children with SLI have been shown to perform significantly differently from typically developing peers on NWRTs (Armon-Lotem and Chiat, 2009).

Initial research has suggested that performance on NWRTs in first language (L1) versus second language (L2) populations of children is approximately equivalent. This has been reported for L1 Hebrew versus Arabic–Hebrew bilingual children (Armon-Lotem and Jaber, 2011), L1 Russian versus Russian–Hebrew and Russian–German children (Armon-Lotem and Chiat, 2009) and L1 English versus Turkish–English bilingual children (Marinis, 2010). At the same time, however, there is also evidence that language experience affects performance on these tasks. Roy and Chiat (2010), for example, discovered that low SES (socioeconomic status) typically developing children exhibited similarly low scores on a NWRT (at least at an early age) when compared to clinically referred children with language impairment of the same age. It is typical for low SES children to have lower levels of linguistic input (Gathercole and Hoff, 2007), so this result can possibly be interpreted as an effect of differential input.

The effects of language experience are observable even at very young ages. A recent study by Keren-Portnoy et al. (2010) assessed the role of language experience in a group of 26-month-old children. These authors assessed children's abilities via a NWRT, which included real words, non-words and, importantly, non-words that were specifically developed to include sounds already emerging in the individual children's repertoires. Children who exhibited consistent and stable production of certain consonants earlier were able to repeat non-words containing those consonants with a higher level of accuracy than children who did not have those consonants in their repertoires. The authors argue that this was because children with earlier mastery had had more experience with (i.e. practice) producing those sounds.

The current study explores the performance of typically developing Welsh–English bilingual children's performance on a Welsh NWRT. The children varied in their level of experience with Welsh, and, in turn, the amount of practice they had with the phonology of Welsh. The study examines whether these differences in exposure affect performance on a NWRT in these simultaneous and early sequential bilingual children.

II Method

The non-word repetition task was designed as a background measure to test children's articulation of word-initial consonants that participate in the process of soft mutation (SM) in Welsh (see Ball and Muller, 1992; Gathercole and Thomas, 2005; Thomas and Gathercole, 2007), for comparison with their performance on a series of mutation tasks, as part of the first author's PhD work. As a result, the design focused only on performance on word-initial consonants. Nevertheless, performance on these word-initial sounds in non-words provides evidence relevant to the debate on the usefulness of NWRTs for assessing bilingual populations.

1 Design

The experiment presented consonant phonemes in word-initial position in a set of 108 novel words, which were treated as nouns, as they were presented as names of picture characters. Novel nouns were carefully formed so that they were phonotactically legal and followed the general phonological patterns of Welsh (see Awbery, 1984). One hundred and eight novel picture characters were drawn up as referents of the novel nouns. The non-words were presented in the context of two games: a board game and a card posting game. These games involved eliciting the names of characters from participants, by asking them to repeat the names of those characters after the experimenter said them.

2 Linguistic stimuli

The 108 invented nouns involved the 13 consonantal phonemes of the Welsh language that participate in the SM process. Nine of the 13 phonemes are subject to word-initial mutation: /p/, /t/, /k/, /b/, /d/, /ɡ/, /m/, /ɬ/ and /rʰ/. The remaining four phonemes, /v/, /ð/, /r/ and /l/, occur word initially as products of the SM. The consonants were balanced for frequency of occurrence across the set of 108 novel nouns. Each consonant appeared as an initial singleton consonant in four different invented nouns (13 × 4 = 52). Additionally, some novel nouns were devised to reflect Welsh-relevant initial consonant clusters: stop + nasal and stop + liquid. This resulted in 56 additional invented words with 14 different consonant clusters. Table 1 outlines the full list of initial consonants and consonant clusters.

The choice of vowels appearing after a given initial consonant/consonant cluster was strictly controlled. Seven different vowels appeared after the word-initial consonants: /a/, /ɛ/, /i/, /ɔ/, /ɨ/, /ʊ/, and /ə/. The four novel nouns with the same initial consonant/consonant cluster all had a different vowel. The use of the seven vowels across all 108 invented words was balanced as evenly as possible, with each of the seven vowels appearing in 15 to 16 different nouns. The number of nouns and the number of vowel types could not allow for perfect balancing (108/7 = 15.4), but the closest approximation to this was achieved.

Table 1. Full list of consonants and consonant clusters that featured in novel words.

Individual consonants	Consonant clusters
/p/	/pl/ /pr/
/t/	/tl/ /tr/
/k/	/kl/ /kr/ /kn/
/b/	/bl/ /br/
/d/	/dl/ /dr/
/g/	/gl/ /gr/ /gn/
/v/	n/a
/ð/	n/a
/m/	n/a
/ɬ/	n/a
/l/	n/a
/rʰ/	n/a
/r/	n/a

Nouns were further balanced for number of syllables. Half of the nouns were monosyllabic and the remaining nouns were disyllabic. This balancing was applied systematically across all of the sub-sets of four nouns for each singleton consonant and consonant cluster.

3 Non-linguistic stimuli

Each invented noun was paired with a visual stimulus: an invented picture character. These picture characters provided referents for the nouns. Characters were developed in such a way as to make the task as fun and child-friendly as possible. They were colourful and varied in shape, colour and other characteristics. Each character appeared within the context of one of the two games: 54 characters appeared in the board game, and the remaining 54 characters appeared in the card posting game.

The first game was a board game in which the participant could move from one spot to the next, revealing a character by removing a circle from each spot. The participant gradually moved around the board until all the characters had been revealed. The second game involved two coloured boxes – one red, one blue – and the set of 54 cards. The cards with the novel characters on them had either a blue or red background colour, which indicated which box each card was to be inserted in.

4 Participants

Participants were recruited from a dozen Welsh-medium schools in the counties of Gwynedd and Anglesey; these areas have similar Welsh-speaking demographics (see Welsh Census, 2001). Forty-five bilingual children aged between 4;11 and 5;11 participated. Twenty-one were female, and 24 were male. Participants were classified according to home language group based on information provided by parents in a detailed questionnaire. This questionnaire included questions regarding the language(s) the children spoke to and heard from their parents and other significant relatives and the language(s) used in school and in social settings; the questionnaire thus provided a comprehensive picture of the language(s) heard by each individual child. Children came from three home language backgrounds:

- Only Welsh at Home (OWH), defined as 80%–100% input in Welsh;
- Welsh and English at Home (WEH), defined as 40% to 60% input in each language; and
- Only English at Home (OEH), defined as 80%–100% input in English.

There were 15 participants from each home language type, with mean ages and age ranges as follows:

- OWH: 5;7 (5;2–5;11);
- WEH: 5;7 (4;11–5;11); and
- OEH: 5;6 (5;0–5;11).

5 Procedure

Each participant played the board game first. A pilot run of the games made it clear that the posting game was the more engaging and entertaining of the two, and so this was left until last.

For the board game, the experimenter (E) explained that the child would be taking away circular counters from the board, one at a time, revealing funny little people, all of whom had special

names. E would say the names, and then the child would need to play 'copycat' and repeat each name.

Before the game began participants were given the opportunity to practise repeating real words, to familiarize them with the requirements of the task. After several practice trials the target trials began. E uttered each invented word clearly and distinctly; then, immediately afterwards, the child was asked to repeat exactly the word they had just heard.

Immediately after the board game task the participant was presented with the card posting game. The participant's task was to 'post' each card into the box corresponding to the background colour of that card (i.e. red background card into red box, blue background card into blue box). E again 'named' each character as it appeared. Before posting each card the participant was required to repeat the name of the character on the card.

The experimenter repeated the model on request if prompted by the child. There were no cases in which a child asked the experimenter to repeat a model more than once. In some cases, children made several attempts after hearing a given model. The attempt that most closely approximated the target was accepted. Trials for both tasks were audio recorded. E later re-listened to all recordings several times to ensure that each phone uttered in every trial was identified and transcribed correctly.

A random selection of 10% of the data recordings, with selections from every child included, was sent to an impartial Welsh-speaking phonetician, who scored correctness of responses and identified error types independently. Inter-rater reliability was at 96% for accuracy and at 97% for error identification.

III Results

Results are presented in two main sections. First the accuracy of children's consonant productions is reported, and then types of errors children made are presented. To score for accuracy, articulated repetitions were scored as 1 when correct targets were produced, and as 0 if incorrect, non-target consonants or consonant clusters were produced. When incorrect, the child's actual articulation was transcribed for later analysis.

1 Accuracy of consonant articulation/production

Some of the sounds and clusters tested are common to Welsh and English, and some are unique to Welsh. Data were analysed in groups by consonant and consonant cluster type, and according to whether they were common to the two languages or unique to Welsh. In each case, an initial ANOVA collapsing performance across the sounds in question compared the performance of the three home language groups; this was followed up by an ANOVA in which performance on each phonological target (i.e. initial consonant or consonant cluster) was entered separately; accuracy of production scores was the dependent variable in all cases.

a Singletons: Consonants common to Welsh and English. The singleton consonants common to Welsh and English are /p, t, k, b, d, ɡ, m, v, l, ð/. An initial ANOVA examined overall performance on these common singletons (out of 40 trials) by the three home language groups. There was no significant difference across groups, $F(2,42) = 2.34$, $p = 0.109$.

To explore performance on these phonemes in more detail, a repeated measures ANOVA was conducted in which phoneme type and home language were treated as independent variables, and performance accuracy on each phoneme was treated as the dependent variable. Analyses revealed

significant effects of phoneme, $F(9, 45) = 63.178$, $p < 0.001$, but, as in the initial analysis, not of home language, $F(2, 45) = 2.341$, $p = 0.109$. Pair-wise comparisons revealed significant differences in performance between /ð/ and all other targets (all ps < 0.001). Performance was relatively poor for the former, but uniformly high for all other sounds. Details of performance are shown in the top two rows in Table 2.

b Singletons: Consonants unique to Welsh. The consonants unique to Welsh are the liquids /ɬ/, /r/ and /rʰ/. An initial ANOVA examining overall performance on these Welsh-specific singletons (out of 12 trials) across the three home language groups revealed a significant difference across groups, $F(2,42) = 4.86$, $p = 0.013$. Scheffé's multiple comparisons revealed that the OEH and the OWH groups performed significantly differently, $p = 0.014$, with OWH showing a mean of 9.2 correct and OEH 6.3 correct, with WEH children between these two, at 8.1 correct.

Levels of target response rates for each of the three sounds are shown in the second section of Table 2. To explore the results in more detail, a repeated measures ANOVA in which phoneme type and home language were treated as independent variables and accuracy scores on each phoneme as the dependent variable revealed that there was a significant difference overall in performance across these three phonemes, $F(2, 44) = 6.135$, $p = 0.003$, with performance rates for /r/ being significantly higher than performance rates for /rʰ/ ($p = 0.001$) (and /ɬ/ between them). As in the initial analysis, a significant main effect of home language was also present, $F(2, 44) = 4.860$, $p = 0.013$, and, again, Tukey's b revealed that the OEH children performed more poorly than the other groups overall, $p = 0.046$. No significant interaction of Phoneme Type × Home Language was found, $F(4, 44) = 1.588$, $p = 0.185$. The lower performance rates of the OEH group seem to reflect their lack of experience with, and exposure to, these Welsh-specific sounds.

c Clusters common to Welsh and English. The only cluster type common to Welsh and English is that of stop + /l/. Four of the six stop + /l/ clusters included in the task are common to Welsh and English. An initial one-way ANOVA on the general performance on these common clusters (out of

Table 2. Accuracy rates by consonant/consonant cluster (some individual, some grouped).

Consonant type	Common or Welsh-specific?	Consonant / cluster	Accuracy rates (by group)			Home language differences
			OWH	WEH	OEH	
Singletons	Common	/p/, /t/, /k/, /b/, /d/, /g/, /m/, /v/	96.5	94.7	96.7	n.s.
		/ð/	57.8	62.5	60.0	
	Welsh	/ɬ/	63.1	79.6	56.2	OEH < OWH,
		/r/	87.5	80.4	59.4	WEH, $p = 0.046$
		/rʰ/	80.3	48.2	43.3	
Clusters	Common	/pl/, /kl/, /bl/, /gl/	78.3	86.2	78.6	n.s.
		/pr/, /tr/, /kr/, /br/, /dr/, /gr/	72.7	75.6	60.4	
	Welsh	/kn/	79.7	64.3	65.0	OEH < OWH,
		/gn/	53.1	41.1	30.0	$p = 0.059$
		/tl/	50.0	39.3	38.3	
		/dl/	26.5	28.5	16.7	

16 trials) revealed no significant difference in performance across the home language groups, $F(2,42) = 0.465$, $p = 0.631$.

To explore performance further, a repeated measures ANOVA with cluster type and home language as independent variables and accuracy scores for each cluster type as the dependent variable confirmed that there were no significant effects of phoneme for these common stop + /l/ clusters, $F(3, 44) = 0.540$, $p = 0.656$, nor of home language, $F(2, 44) = 0.456$, $p = 0.631$. Levels of target response rates for all four consonant clusters involving /l/ common to Welsh and English were high, ranging from 76% to 80%. It may be worth noting, however, that for all such C + /l/ clusters, there was a tendency for the WEH children to outperform the other two groups (see Table 2).

d Clusters unique to Welsh. There were three main types of clusters unique to Welsh, stop + /r/ (6 sets), alveolar stop + /l/ (2 sets), and velar stop + /n/ (2 sets). An initial one-way ANOVA to examine overall performance (out of 40 trials) by the three home language groups revealed a significant effect of home language, $F(2,42)= 3.38$, $p = 0.043$. Scheffé's multiple comparisons showed a near-significant difference between the OEH and OWH children's performance, at $p = 0.059$, with the OWH children giving a mean of 26.1 correct responses, and the OEH children 20.5, with WEH children in between at 25.1 correct.

To explore performance further, a repeated measures ANOVA in which these clusters and home languages were treated as independent variables and accuracy scores on each cluster type as the dependent variable revealed main effects of both cluster type, $F(9, 44) = 23.857$, $p < 0.001$, and of home language, $F(2, 44) = 3.382$, $p = 0.043$. Post-hoc analyses revealed best performance on C + /r/ and /kn/ clusters, least good performance on /dl/, and intermediate performance on /tl/ and /gn/ (all *p*s <.001).

As in the initial analysis, the main effect of language group observed in this analysis reflected a marginally significant difference in the performance rates of the OWH group versus the OEH group ($p = 0.054$, planned Tukey comparisons). No significant interaction of phoneme type and home language was observed $F(18, 44) = 1.196$, $p = 0.260$. These results indicate that the OEH group consistently performed below the OWH group for Welsh clusters overall, and that the OWH group generally performed highest, with the WEH group falling between the two groups. Closer inspection of the data suggested considerable variation in the performance of the WEH group; in some cases, they tended to perform best of all groups (/pr/, /br/) (pr: WEH − 87.5%, OWH − 68.75%, OEH − 53.33; br: WEH − 80.36%, OWH − 68.75%, OEH − 66.67%), in some cases their performance was similar to that of the OWH children (/kr/, /dl/), (kr: WEH − 75%, OWH − 76.56%, OEH − 56.67%; dl: WEH − 28.57%, OWH − 26.56%, OEH − 11.67%), and in some cases they fell between the OWH and OEH children (/gn/) (WEH − 41.07%, OWH − 53.13%, OEH − 30%) or were more like the OEH children (e.g. /tl/, /kn/) (tl: WEH − 39.29%, OWH − 50%, OEH − 38.33%; kn: WEH − 64.29%, OWH − 79.69%, OEH − 65%). For further details, see the last section of Table 2.

Thus, language exposure and experience seemed to play a role in the children's ability to articulate Welsh clusters accurately, in that the OWH children performed best, the OEH children performed worst, and the WEH children's performance generally fell between the two. However, children from all groups exhibited some difficulty in articulating some clusters, especially /tl/, /dl/ and /gn/, suggesting that clusters of these types are generally difficult for children to produce at this age.

2 Summary of accuracy data

Overall, no differences across home language groups emerged for those sounds common to both Welsh and English; performance across groups was equivalent on sounds shared by the two

languages. As noted, however, there seemed to be a tendency for the WEH group to show better mastery of stop + /l/ clusters, common to both languages. This group have been exposed to both languages from birth and so may have profited from more practice with these clusters that are present in both languages.

For Welsh-specific sounds, the OEH group showed an overall disadvantage when compared to the others, especially the OWH group. This difference across groups presumably stems from the fact that the OEH group have had more limited exposure to Welsh, particularly prior to attending school.

With the exception of the common stop + /l/ clusters, children from all home language groups struggled to produce clusters accurately. This was most evident for the Welsh alveolar stop + /l/ and stop + /n/ clusters. In particular, children struggled most with articulating /dl/. The co-articulation of /d/ and /l/, both voiced and both alveolar, may be especially difficult.

3 Error types

In a NWRT, not only the accuracy of performance is relevant, but the types of errors children make can also be instructive. In order to further explore potential differences associated with distinct levels of exposure to Welsh, the types of errors made by the children from the three home language groups were examined. The sounds for which there were extremely low errors rates – the stops, /m/, and /v/ – were not examined further for errors.

For all other sounds, the types of errors that were produced by each group of children were examined for commonalities and differences. Some consonants/clusters showed homogeneous error type patterns across home language groups, while others showed marked differences. We have grouped the patterns into three types, according to the level of commonality across home language groups.

In the first group, there is a high level of agreement across groups in the sound substitutions made. For each phoneme in this group, the three home language groups shared the same single sound substitution in at least 60% of their errors. Thus, the children in all groups substituted [v] for /ð/ over 68% of the time, [x] for /ɬ/ over 73% of the time, and just the stops for the clusters /tr/, /dr/, /tl/, and /gn/ at least 60% of the time. When the second most common sound substitution for each of these phonemes is added, the home language groups shared substitutions in at least 70% of their responses. For example, the common secondary substitution types were [d] for /ð/, [s] for /ɬ/, and so forth. These primary and secondary substitutions are shown in Table 3. Beyond these first two substitution types, the sound substitutions were limited; the most common third substitutions for each phoneme are also shown in Table 3. Overall, these substitutions reflected simplifications of difficult consonants and clusters.

A second group, shown in Table 4, showed intermediate agreement across home language groups. These sounds showed less than 60% agreement on the top substitution for each sound, but approximately 60% agreement or more across groups in the two most common substitutions used. Thus, for example, for /rʰ/, all groups substituted [h] or [r] at least 65% of the time; for /pl/, they substituted [pr] or [p] in at least 75% of their errors; for /dl/ they substituted [d] or [gl] at least 58.7% of the time; and so forth. Beyond these two primary substitutions, there was more variation than was the case in the first group of sounds shown in Table 3. The most common third type of substitution for each phoneme in Table 4 is shown at the bottom, but each home language group showed a range of additional substitutions beyond these in many cases. Thus, for example, for /rʰ/, beyond the common substitutions of [h], [r], and [pr], the WEH and OEH groups used 7 and 8 additional forms, respectively. This suggests that the latter two groups experienced particular difficulty with processing or reproducing [rʰ]. For /kn/, the OWH and OEH groups both favoured

Table 3. Details for error types for consonants/clusters with a high level of agreement in substitutions across home language groups.

Target consonant / consonant cluster	/ð/	/ɬ/	/tr/	/dr/	/tl/	/gn/
Most common substitution:	[v]	[x]	[t]	[d]	[t]	[g]
Mean	72.9	74.6	77.6	63.6	70.5	66.0
OWH	70.3	73.9	93.3	70.6	70.6	75.9
WEH	80.0	75.0	73.3	61.1	65.7	60.0
OEH	68.6	75.0	86.7	61.3	75.0	64.3
Second most common substitution:	[d]	[s]	[tɹ]	[tr]	[tr]	[n]
Mean	13.1	9.5	8.2	15.2	16.2	18.9
OWH	10.8	13.0	6.7	0.0	14.7	13.8
WEH	5.7	0.0	6.7	16.7	17.1	34.3
OEH	22.9	10.7	13.3	22.6	16.7	9.5
Third most common substitution:	[f]	[h]	[tl]	[dl]	[sl,kl]	[gr,k,gl,kn]
Mean	3.7	6.3	8.2	7.6	4.8,4.8	all 2.8
OWH	0.0	0.0	0.0	17.6	5.9,2.9	3.4,3.4,0,0
WEH	5.7	4.3	13.3	5.6	2.9,11.8	2.9,0,2.9,0
OEH	5.7	13.0	13.3	3.2	5.6,0	2.4,4.8,4.8,7.1

substitution of [k] (64% and 76% of errors in each case) while the WEH group rarely substituted [k] (10% of all errors), and favoured [n]-substitution more (50% of all their errors).

For the last group of phonemes, shown in Table 5, there were marked differences in error types across home language groups, with no single sound substitution shared by the three home language groups in more than 28% of articulations; e.g. for /pr/ all home language groups used [p] at least 27.8% of the time, but their substitutions differed radically beyond this. Table 5 shows the top three error types made by each group, and the proportion of errors accounted for by different error types. It also shows the number of variants produced for each phoneme by each home language group. The number of types of substitutions in all home language groups, but especially in the OEH group, is again suggestive of struggling with these sounds.

To bring out specific cases, for /br/-initial novel nouns, the OEH group showed errors spread across 14 types of substitutions. This suggested great difficulty in the OEH group in the perception and/or articulation of /br/. The OEH group used [bɹ] in some cases, again reflecting an influence of English. In contrast, approximately 90% of error types in the OWH group were accounted for by three error types, including the Welsh cluster [pr]. The WEH group fell somewhere in between, with 44.44% of errors being accounted for by [b] and [bl] substitution, and the remainder consisting of a mix of five different error types.

For /gr/, the OWH group favoured [g]-substitution, (54% of all errors), the WEH children favoured [r]-substitution (29% of all errors) but also substituted [g], [gl] and [br] (each type accounted for 21% of all errors), while the OEH group showed a wider variety of error types (9 in all), including [gɹ]-substitution in 14% of their errors, suggesting they had more difficulty with this sound than the other groups.

The substitutions for /r/ showed that both the OWH and WEH groups favoured [l]-substitution (OWH − 37.5%, WEH − 54.6%), while the OEH group substituted [ɹ] (65.4%), a substitution much less prominent in the WEH group (18.2%), and not occurring in the OWH children at all. This is clearly carryover from English, as /ɹ/ does not occur at all in Welsh. The OWH group also substituted the Welsh cluster [gr] frequently (37.5%), which did not occur at all in the other groups.

Table 4. Details of error types for consonants/clusters with intermediate levels of agreement in substitutions across home language groups.

Target consonant / consonant cluster	/rʰ/	/pl/	/kl/	/bl/	/kr/	/dl/	/kn/
Most common substitution:	[h]	[pr]	[k]	[b]	[k]	[d]	[k]
Mean	54.9	57.6	45.2	37.0	59.6	40.7	57.8
OWH	52.6	62.5	38.5	55.6	57.1	37.0	64.3
WEH	55.2	62.5	66.7	40.0	60.0	50.0	10.0
OEH	55.9	50.0	41.7	23.1	60.7	37.3	76.2
Second most common substitution:	[r]	[p]	[kr]	[br]	[kl]	[gl]	[n,gn]
MEAN	15.9	24.2	32.3	33.3	12.3	29.6	13.3,13.3
OWH	31.6	25.0	38.5	22.2	21.4	21.7	7.1,14.3
WEH	10.3	25.0	0.0	20.0	6.7	31.6	50.0,0
OEH	11.8	25.0	41.7	46.2	11.11	35.3	0,19.0
Third most common substitution:	[pr]	[bl]	[kw]	[gl]	[pr]	[dr]	[kr]
MEAN	7.3	12.1	9.7	11.1	5.4	14.1	6.7
OWH	5.3	6.3	23.1	11.1	0.0	23.9	7.1
WEH	6.9	11.1	0.0	20.0	20.0	10.5	20.0
OEH	8.8	22.2	0.0	7.7	0.0	7.8	0.0

The sounds for which the home language groups differed the most in their substitutions primarily involved the voiced trill, /r/. For /r/, /br/ and /gr/, the OWH children in their errors substituted Welsh-specific sounds, often involving [r] itself; e.g. [br] and [gr] for /r/, [pr] for /br/. The OEH children often substituted some form of [ɹ], but, as noted, also showed use of a wide variety of sounds, suggesting extra difficulty. The WEH children substituted [l] for /r/ and [bl] for /br/ (much like the OWH group) and substituted some Welsh-specific sounds, sometimes including /r/ e.g. [br] for /gr/. They also showed similarities to the errors of the OEH group, sometimes substituting a form of [ɹ], but they did not exhibit as many error types as the OEH group, suggesting that they struggled less with /r/ and clusters involving /r/ than the OEH group.

4 *Summary of error data*

Overall, differences across home language groups were observed primarily for those consonants/consonant clusters that were unique to Welsh. The voiced trill /r/ and clusters involving /r/ (i.e. /br/ and /gr/) showed particularly diverse patterns across language groups. Carryover effects from English were evident in the OEH group in the form of stop + /ɹ/ clusters. Additionally, the OWH children's substitutions were sometimes more Welsh-like (e.g. /rʰ / > [r]).

Overall, the results of the consonant errors reflect both effects of language exposure, as evidenced by the WEH & OEH groups' poorer performance on Welsh-specific consonants (e.g. /rʰ/), and language-specific effects, as shown by English-like errors such as [ɹ] for /r/ in the OEH group, and Welsh-like errors such as [gr] for /gl/ in the OWH group.

IV Discussion

The results of the current study shed light on whether or not a NWRT can be used to universally assess children for language impairments. Effects of language exposure and cross-linguistic effects were apparent in these data. The OEH children exhibited lower performance rates for several singleton consonants and consonant clusters unique to Welsh. These findings reveal a role

Table 5. Details of error types for consonants/clusters with low levels of agreement in substitutions across home language groups.

Target consonant / consonant cluster	/r/	/pr/	/gr/	/br/	/gl/
OWH:					
Number of variants	3	6	6	5	3
First	[l] 37.5	[pl] 50.0	[g] 53.8	[bl] 42.1	[gr] 42.9
Second	[gr] 37.5	[p] 27.8	[gw,kr] 15.4 (each)	[pr] 26.3	[gw] 35.7
Third	[br] 25.0	[pw,br,f,pɹ] 5.6 (each)	[r, gl] 7.7 (each)	[b] 21.1	[g] 21.4
WEH:					
Number of variants	5	5	5	7	4
First	[l] 54.5	[pl] 36.4	[r] 28.6	[bl] 22.2	[l] 33.3
Second	[ɹ] 18.2	[p] 36.4	[gl,g,br] 21.4 (each)	[b] 22.2	[g] 33.3
Third	[kr,w,gl] 9.1 (each)	[br,tr,bl] 9.1 (each)	[gw] 7.1	[pr,vr,r,dr,l] 11.1 (each)	[gr] 22.2
OEH:					
Number of variants	7	7	9	14	3
First	[ɹ] 65.4	[p] 38.1	[kr] 23.8	[bɹ] 17.4	[l] 50.0
Second	[br] 11.5	[kr] 14.3	[g] 19.0	[bl,gr] 13.0 (each)	[gr] 35.7
Third	[w] 7.7	[pɹ] 14.3	[gɹ] 14.3	[r,dr,pl] 8.7 (each)	[g] 14.3

of relative level of experience, and exposure to Welsh, since the OEH children do not hear Welsh at home.

This effect of exposure on accuracy of performance is further highlighted in children's error types. The OEH children, especially, showed some carry-over from English, and the OEH and WEH children showed a greater variation in substitutions than the OWH children, especially for /r/, /r^{h}/, /pr/, /br/, /gr/, and /gn/. The fact that the WEH group sometimes patterned more like the OEH children with these substitutions shows that even 2L1 children can perform differently from their more Welsh-dominant OWH peers.

Chiat (2011) states that it is 'impossible to create a universal NWR test … [because a] NWR test … measures phonological processing and memory which give rise to phonological representations' (p. 2). Phonological representations are of course language-specific, and the language experience of a child in the language specific to the task is bound to affect their processing of and subsequent reproduction of novel words in that task. For this reason, children perform differently according to their experience and exposure to the language.

The results of the current study are consistent with this position, as the OEH group showed poorer performance overall relative to children who have a greater level of Welsh input. The results also suggest that differences can manifest even in children who have been hearing Welsh from birth, but with distinct levels of exposure (OWH versus WEH).

Even if a NWRT were designed to be universal – in that the non-words created would include sounds and sound combinations common to almost all of the world's languages – the experience of a child in their given language is still likely to have an impact on their reproduction at the phonological level. Thus, these data suggest that a fully language-independent NWRT is impossible, at least if it does not take into account factors such as language proficiency, relative language exposure, and relative language experience.

The attraction of NWRT tasks, or repetition tasks in general (e.g. Chiat et al., in press), for the assessment of bilingual children has been that they might be a key tool for identifying problems in children whose first language the tester does not know. The hope has been that such a test could be designed in a way that would bypass assessment of the child's proficiency in the particular language and provide a window directly into the child's language-independent language abilities. This study makes it clear that designing a language-neutral task along these lines is virtually impossible. What options does that leave for researchers and therapists who wish to find an efficient, yet reliable, means for assessing such children's language abilities, in a way that would bypass performance necessarily being contingent on the child's proficiency in the language being tested? One option that some researchers have been exploring and therapists have begun using is dynamic assessment or response to intervention (e.g. Genesee et al., in press). By examining children's abilities to learn language by exposing them to or teaching them new aspects or constructions in the L2, and observing their progression over periods of time and a series of stages, one may be able to gain an accurate picture of the child's language-learning capacities. Such an approach holds promise; only further research and exploration will help provide further good solutions.

Acknowledgements

The authors would like to thank Dr Hans Stadthagen-González for his work developing the novel characters for the task. We would also like to thank the schools, parents and pupils for their cooperation and participation.

Funding

This research was supported in part by ESRC and WAG/HEFCW grant RES-535-30-0061, to support the ESRC Centre for Research on Bilingualism in Theory and Practice, and the first author's PhD studies.

Conflict of interest statement

The authors declare that there is no conflict of interest.

References

Armon-Lotem S (2011) How do sequential bilingual children perform on non-word repetition tasks? Unpublished paper presented at the 36th Annual Boston University Conference on Language Development (BUCLD), Boston, MA, November 2011.

Armon-Lotem S and Chiat S (2009) A Russian-Hebrew NWR task tested with bilinguals with TDL and SLI: Rationale and findings. *COST Meeting for Action IS0804*, Amsterdam, The Netherlands, 16–18 November 2009.

Armon-Lotem S and Jaber A (2011) NWR tasks in Hebrew and Arabic: Principles and design. *COST Meeting for Action IS0804*, Eskisehir, Turkey, 25–27 May 2011

Awbery G (1984) Phonotactic constraints in Welsh. In M. Ball & G. Jones (Eds.), *Welsh Phonology*. Cardiff: University of Wales Press.

Ball MJ and Muller N (1992) *Mutation in Welsh*. London: Routledge.

Chiat S (2011) Principles for constructing COST non-word repetition tests: Choices and decisions. *COST Meeting for Action IS0804*, Eskisehir, Turkey, 25–27 May 2011.

Chiat S, Armon-Lotem S, Marinis T, Polišenská K, Roy P, and Seeff-Gabriel B (in press) The potential of sentence imitation tasks for assessment of language abilities in sequential bilingual children. In: Gathercole VCM (ed.) *Bilinguals and assessment: State of the art guide to issues and solutions from around the world*. Clevedon: Multilingual Matters.

Gathercole VCM and Hoff E (2007) Input and the acquisition of language: Three questions. In: Hoff E and Shatz M (eds) *The handbook of language development*. New York: Blackwell, 107–27.

Gathercole VCM and Thomas EM (2005) Minority language survival: Input factors influencing the acquisition of Welsh. In: Cohen J, McAlister K, Rolstad K, and MacSwan J (eds) *ISB4: Proceedings of the 4th International Symposium on Bilingualism*. Somerville, MA: Cascadilla Press, 852–74.

Genesee F, Savage R, Erdos C, and Haigh C (in press) Identification of reading difficulties in students schooled in a second language. In: Gathercole VCM (ed.) *Bilinguals and assessment: State of the art guide to issues and solutions from around the world*. Clevedon: Multilingual Matters.

Keren-Portnoy T, Vihman MM, DePaolis RA, Whitaker CJ, and Williams NM (2010) The role of vocal practice in constructing phonological working memory. *Journal of Speech, Language, and Hearing Research* 53: 1280–93.

Marinis T (2010) Measuring English non-word repetition in typically developing Turkish-English children. *COST Meeting for Action IS0804*, Newcastle, UK, 28 May 2010.

Montgomery JW (2003) Working memory and comprehension in children with specific language impairment: What we know so far. *Journal of Communication Disorders* 36: 221–31.

Roy P and Chiat S (2010) The pre-school repetition test: Evidence of typically developing and clinic samples in the UK. *COST Meeting for Action IS0804*, Newcastle, UK, 28 May 2010.

Thomas EM and Gathercole VCM (2007) Children's productive command of grammatical gender and mutation in Welsh: An alternative to rule-based learning. *First Language* 27: 251–78.