Universal Nonsense:
The need for a linguistically fair nonword repetition task

Leela Rao
University of North Carolina at Chapel Hill
April 8, 2016

A thesis presented to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the Bachelor of Science degree with Honors in Psychology

Barbara Davis Goldman, Ph.D.
Committee Chair: Barbara Davis Goldman, Ph.D.

Committee Member: Lucia Binotti, Ph.D.
Committee Member: Kelly Giovanello, Ph.D.

Acknowledgements

First and foremost, I would like to thank all of my family and friends through the last four years at UNC – Chapel Hill. This thesis would not have been possible without all the love and support. Appa, thank you for all the time you took to walk me through my own ideas and for all of the help you provided when it came to all of the computer things. Amma, thank you for showing interest in what I wanted to study and providing real motivation when needed. Arjun, thank you so much for listening to all of my ramblings and providing much needed comic relief! And Nisha, thank you so much for the countless hours of conversation, support, and overall awesomeness.

I would also like to thank all of my mentors at UNC – Chapel Hill in the Departments of Psychology and Neuroscience and Romance Languages. This thesis is truly an integrative project that brings together background knowledge in psychology and linguistics. Specifically, I would like to thank:

- Dr. Barbara Davis Goldman: Thank you, a million times, for introducing me to research in psychology and mentoring me through the process of creating and executing my own study. Your relentless support and enthusiasm definitely kept me motivated to complete this project!
- Dr. Jennifer Arnold: Thank you for sharing your wealth of knowledge in psycholinguistic theory and psychology research in general with me over the years. Being part of your lab has changed the way I plan to approach research in the future!
- Dr. Lucia Binotti: Three classes later and my head is filled with linguistic theories and ideas on how to implement them to make a real difference in the world. ¡Muchísimas gracias por todo su apoyo y ayuda!
- Dr. Kelly Giovanello: Thank you so much for taking the time to help me find a niche for the future. I already love what I intend to study, and I would not have found it without your guidance!

This project was supported by a David Bray Peele Memorial Research Award from the Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill.

Abstract

Nonword, or nonsense word, repetition tasks are an established way of measuring working memory in both children and adults. However, the most common working memory tasks are based on sounds and linguistic rules from the English language, while the number of bilingual speakers in the United States is on the rise. Past research has shown that the wordlikeness, or the similarity of a nonword to a real word in a participant's native language, has a strong impact on the participant's performance on a nonword repetition task. Thus, the present study aimed to create a linguistically more equivalent nonword repetition task for a more diverse range of research participants. A list of 40 nonwords was created by employing linguistic rules and sounds from twelve of the most frequently spoken languages in the US. Undergraduate participants with various language backgrounds were recruited to complete the new task. Results showed that participants with differing language backgrounds performed the same on the new nonword repetition task, but the number of languages participants were exposed to was positively correlated with repetition accuracy. Future directions for further pilot testing are proposed.

Keywords: nonword repetition, nonsense words, working memory, Children's Test of Nonword Repetition, Nonword Repetition Task, bilinguals, memory

Universal Nonsense:

The need for a linguistically fair nonword repetition task

Memory impairment is an integral symptom of many psychological phenomena. Putting aside neurocognitive disorders such as Alzheimer's disease and dementia that directly impact memory, the DSM-5 lists memory impairment as a symptom for a slew of conditions, including psychotic and mood disorders (American Psychiatric Association, 2013). One of the most readily administered types of memory tasks is the nonword, or nonsense word, repetition task. This task consists of a participant repeating individually heard nonsense words back to a researcher. It has been shown to be a fundamental way to measure phonological working memory, or the ability to temporarily store and manipulate phonological (speech sound) information (Deevy, Weil, Leonard, & Goffman, 2010). However, current nonword working memory tasks pose a potential problem when used to measure memory in children and adults who are not native English speakers due to their similarity to real English words. Because many of these tasks are used to determine if a child has specific language impairment (SLI) or the extent of various pathologies in adults, it is imperative to ensure that our measurement of working memory via nonword repetition is not being biased by language exposure.

Working Memory and Nonword Repetition Tasks

The basic goal of a nonword repetition task is to measure working memory, a preliminary step of information retention. One of the prominent theories about working memory divides it into three major parts: the phonological loop, which holds onto verbal information that it can manipulate; the visuospatial sketchpad, which holds onto visual information; and the central executive, which determines which aspects of sights and sounds are remembered and which are forgotten (Baddeley & Hitch, 1994). Nonword repetition tasks aim to measure a person's ability to retain information in their phonological loop, which can be further divided into two parts (Gathercole et al., 1994). The first part of the phonological loop is the phonological short-term store, which immediately holds incoming speech sounds (Gathercole et al., 1994). The second part, the part which makes nonword repetition possible, is the subvocal rehearsal process. This process continually refreshes the information in the short-term store until the information is permanently stored or forgotten (Gathercole et al., 1994). See Figure 1 for a graphic representation of the structure of working memory.

There are two main child-appropriate nonword repetition tasks that are used in clinical practice today. It is important to note that while these tasks are primarily used to measure specific language impairment (SLI) in children, these tasks have been widely used in adult populations as a tool for assessing various disorders as well. The first frequently used nonword repetition task is the Children's Test of Nonword Repetition (CN Rep; Gathercole & Baddeley, 1996), which consists of forty nonwords that can be used to measure nonword repetition in children between four and eight years of age. These words, which are written in the International Phonetic Alphabet (IPA), are divided into four separate groups based on syllable length. Words range from two to five syllables: Each syllable group contains 10 nonwords. Half of the nonwords consist of consonant clusters (e.g. [pɛnl], sounds like "penul") and the other half do not (e.g. [dɪlɔ], sounds like "dilaw"). Of note is the fact that all of the nonwords in the CN Rep are pronounced with traditional English stress patterns (Archibald & Gathercole, 2006).

The second frequently used nonsense word task is the Nonword Repetition Task, which is often used on children between six and twelve years old (NRT; Dollaghan & Campbell, 1998). In contrast to the CN Rep which contains forty nonwords, the NRT consists of 16 nonwords, ranging from two to four syllables. Each nonword is constructed in a consonant-vowel (CV) structure, ensuring that there are no consonant clusters. All of the syllables used in the NRT are from the English language, even though none of the syllables in the nonwords occur in the context that they appear in English.

Lexical Knowledge and Language Exposure

Nonword repetition tasks theoretically measure of phonological working memory independent from long-term lexical memory or any kind of language-based memory. Lexical memory is the "word bank" people have in their minds, which differs from person to person based on experiences and language exposures. Since nonword repetition tasks are not based on entire real words in the English language, it is often assumed that lexical knowledge does not impact nonword repetition performance. However, this does not appear to be the case in practice.

After the publication of the NRT, there was a surge of research on the topic of the "wordlikeness" of nonsense words. Studies revealed that repetition accuracy may not, in fact, be independent of a person's lexical knowledge since nonwords that sounded more similar (and had a high wordlikeness) to real English words elicited a higher repetition accuracy than nonwords with a lower similarity (and a low wordlikeness) to real English words. For example, a study conducted on male children between seven and twelve in an American public school system found that participants recalled nonwords that had constituent syllables that corresponded to real English words significantly more accurately than those that did not. In this study, the word /bæθəsɪs/, pronounced like "bathasis" was repeated significantly more accurately than /fæθəsɪs/, pronounced like "fathasis", even though these two nonwords differed by only one sound ($Z = -1.965$, $p = .05$; Dollaghan, Biber, & Campbell, 1993). Another study by Dollaghan, Biber, and Campbell (1995) found that nonwords with stressed syllable patterns that were similar to real words (as is the case in the nonwords used in the CN Rep) are repeated significantly more accurately than the same nonwords without familiar stressed patterns.

Gathercole, Willis, Emslie, and Baddeley (1991) presented two hypotheses to explain the relationship between phonological working memory and lexical knowledge in nonword repetition tasks. The linguistic hypothesis attests that familiar lexical units (parts of words) in nonwords can be used to *strengthen* a nonword's representation in the phonological working memory system (Gathercole et al., 1991). On the other hand, the memory hypothesis views phonological working memory more as a constraint on the process of learning new words (*interference*) in instances where information from long-term memory lead to the misperception of similar new words (Dollaghan, Biber & Campbell, 1995). Whatever the case may be, the study conducted by Gathercole et al. (1991) found that the repetition accuracy of children aged four through six was independently impacted by both the length of nonwords and their wordlikeness, demonstrating that accuracy on current nonword repetition tasks cannot be considered independent of English lexical knowledge.

Additionally, according to a study conducted on 156 young male participants of various cultural backgrounds (31% White, 67% African American, 1% Asian, and 1% Native American), scores on the NRT did not differ significantly between majority (White) and minority (African American, Asian, and Native American) students between the ages of eleven and fourteen (Campbell, Dollaghan, Needleman, & Janosky, 1997). However, what this study fails to take into account is the impact of lexical knowledge based on a person's native language and language exposure on performance on the task. Given that most of the "minority" participants in the study identified as African American, they grew up hearing and speaking English. Even though African American English and Standard American English are perceived to be different dialects, the dialects still are still comprised of the same lexical foundation. Therefore, it was not possible to detect an impact of lexical knowledge bases formed by knowledge of different languages on NRT performance in this particular sample.

That being said, research in linguistics and speech sciences have shown that the language bias starts much before the lexical level. In linguistics, speech can be divided into various levels: sound information, word-part information, word (lexical) information, phrase information, sentence information, and discourse information. Thus, although the English bias in nonword repetition has been supported by previous literature at the lexical level, there are indications that this bias may even start at the sound (phonetic) level. Studies over the last forty years have shown that infants are able to perceive differences in a vast number of speech sounds before the age of one – a skill that is lost as a native language emerges (Werker & Tees, 1984; Best & McRoberts, 2003; Best, McRoberts, LaFleur & Silver-Isenstadt, 1993). These studies demonstrate that, after the age of one (or sometimes even earlier), monolingual infants lose the ability to differentiate between sounds that are not in their native language(s). Thus, the bias toward native language may be even more engrained in nonword repetition participants than was previously thought.

Language Use in the United States

The number of bilinguals in the United States and around the globe is on the rise. When compared to the US Census in 1980, the 2010 Census showed over a fifty percent increase in the number of people who spoke a language other than English at home, a trend that promises to continue, and likely to proceed more rapidly, as the world becomes increasingly interconnected (U.S., 2013). A more recent report by the US Census Bureau details this trend by explaining that, in addition to English, the following thirteen languages are spoken in close to or more than

600,000 households each across the nation: Spanish, French, French Creole, Italian, Portuguese, German, Russian, Polish, Hindi, Chinese, Korean, Vietnamese, Tagalog, and Arabic (U.S., 2015). These languages, though seemingly quite diverse, can be grouped into various categories based on linguistic evolution and characteristics.

There are two main ways in which linguists visualize the relationship between various languages. The first is the "language tree," where each language creates a branch off the language it stems from (Matthews, 2007). As an example, "Indo-European" would be the trunk of a tree with multiple offshoots, including a "Romance" branch. The "Romance" branch would then further divide into Ibero-Romantic, Italo-Dalmatian, and Gallo-Iberian. Each of these branches would break down further to include the Spanish, Portuguese, Italian, Sicilian, and French languages. The second visualization of languages is the wave model, in which languages are seen as the individual changes that occur within them. Each change in these languages, which occur as part of the natural evolution of spoken language, then ripples out like waves to the languages closest to it, exacting the same change in those (Matthews, 2007).

Regardless of which visualization you choose, the most common languages spoken in the United States can be categorized as seen in Table 1. These categories contain languages with similar origins, scripts, and spoken sounds. However, the categories themselves warrant further explanation. The Austronesian, Germanic, and Romance languages all use the Latin script we are accustomed to in the United States. These languages can easily be broken down into consonants and vowels that are the building blocks of standard consonant-vowel syllables (e.g. ba.na.na, with "." marking syllable boundaries). In contrast, the Afro-Asiatic, Austro-Asiatic, Indo-Iranian, Slavic, Sino-Tibetian, and Koreanic languages do not use the Latin script. These languages are written and spoken differently from English. For example, the Sino-Tibetian and Koreanic languages are considered *tonal* languages in which the tone of a spoken word determines its meaning. These languages have no specific alphabet, but instead use a pictorial writing system. Afro-Asiatic languages, such as Arabic and Hebrew, are consonant-based languages, meaning vowel sounds are merely super- or sub-scripts to the written script. Hindi is similar, but contains a more diverse vowel repertoire. Thus, languages spoken in the US can be divided into three main categories: consonant-vowel languages (e.g. English), tonal languages (e.g. Mandarin), and consonant languages (e.g. Arabic).

The International Phonetic Alphabet

Given the diversity in language families around the globe, the International Phonetic Alphabet (IPA) was developed by linguists to help communicate linguistic information more clearly across language boundaries. The alphabet, pictured in Figure 2, is based on language sounds instead of traditional orthography. While many of the symbols are based on traditional Latin script (/b/, /g/, /t/, /p/), there are others that represent sounds that are not common in the United States (the velar fricative /ɣ/, is commonly known as the “French ‘r’ sound”; IPA). For more information on the IPA or to hear the sounds, please refer to: http://www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds/

Present Study

According to Dollaghan et al. (1993), the fact that almost all of the nonsense words that are used in nonword repetition tasks today are derived from the English language would impact a person’s nonword repetition accuracy if the lexical knowledge a person possesses is activated during the repetition of the word, as predicted in the linguistic hypothesis proposed by Gathercole et al. (1991). This could, according to the researchers, “re-introduc[e] the confounding effects of lexical knowledge that nonsense-word tasks were originally designed to avoid" (Dollaghan et al., 1993). In sum, there appears to be a lack of nonword repetition tasks that can be used to assess phonological working memory function in participants who have been exposed to more than just the English language. Wordlikeness and, in turn, lexical knowledge, have been shown to impact performance on nonword repetition tasks, and humans have shown a bias toward sounds and prosody from their native language starting at a very young age. Due to these biases and the prevalence of English-based nonword repetition task, it is important to devise a more linguistically fair task for use in more diverse populations.

Thus, the present study aimed to create a working memory task that could be used to measure working memory deficits in people with varying language backgrounds. Because the process of creating and standardizing a task requires a vast number of participants, the present study serves as a pilot test to determine if employing linguistic rules and sounds from various languages can reduce the impact that language exposure might have on nonword repetition performance. If the tool is effective in measuring phonological working memory separate from language exposure, it is hypothesized that the scores from the new task will be positively correlated with the scores from a digit span task. Digit span tasks are often used to measure working memory in an attempt to separate memory from lexical knowledge, even though the digits are generally restricted to one language. Because this simple task is a quick measure of working memory, it functions as a manipulation check in the present study. More significantly, if the new tool measures memory without the confound of specific lexical knowledge, there will not be a significant difference in the average performance on the new task between groups of participants with various language backgrounds.

Method

Participants

48 participants were recruited from the Introduction to Psychology Participant Pool and various language departments and multi-cultural clubs on the UNC – Chapel Hill campus. All participants were fluent or native English speakers, but being so was not a requirement for participation. Additionally, all but one participant (98%) had exposure to more than one language (M = 3.4 languages, SD = 1.7 languages, *Range*: 1-9). Participants were all between the ages of 18 and 25. A complete breakdown of participants' language exposures can be seen in Figure 2 and Figure 5.

Materials

Three separate tasks were created for this study: a nonword repetition task, a digit span task, and an extensive language exposure questionnaire.

The Universal Nonsense: A nonword repetition task. The main task used in this study was a nonword repetition task, entitled The Universal Nonsense. This task was developed by selecting 14 of the most frequently spoken languages in America: English, Spanish, French, German, Polish, Russian, Chinese, Korean, Vietnamese, Tagalog, Italian, Portuguese, Hindi, and Arabic. All had greater than 600,000 native speakers in America in 2011. Of these languages, the two tonal and pictoral languages, Chinese and Korean, were omitted due to differences in morphology and syllabification. The remaining 12 languages created the foundation for the Universal Nonsense nonwords.

One-hundred of the frequently used words for each of these 12 languages, for a total of 1200 words, were translated into the International Phonetic Alphabet (IPA) by the researcher, who is trained in phonetic transcription. When possible, translations were cross-checked with online text-to-IPA converters. For languages such as Arabic where vowels are not in the written script, transcriptions are based on a combination of linguistic rules and sound samples. After the 1200 words were translated into IPA, they were divided into syllables using traditional linguistic syllabification rules. When there was any ambiguity about the location of syllable boundaries, traditional English syllabification rules were used. The divided words yielded a list of 523 syllables from the 12 languages

This list of syllables was analyzed for word position and sounds. A list of phonotactic, or language, rules was compiled for the twelve languages from the World Phonotactic Database. This database included information such as whether each language permitted voiceless stops (/p/, /t/, /k/) at the coda, or the end, of a word or whether velar nasals (/ŋ/, which is pronounced similar to the –ng ending present in English gerunds) were permitted at the beginning of words. In order to create a linguistically fair nonword task, the rules that occurred in the majority of the twelve languages (e.g. voiceless stops in codas are not permitted) were “averaged” to create a phonotactic rule list for the new series of nonwords.

Using these rules, the syllables from the languages were divided into *onsets* (permitted word beginnings), *codas* (permitted word endings), *others* (permitted in syllable positions other than the onsets and codas), and *forbidden*. Because the long-range goal of the new task is for use in school-aged children, forbidden syllables were based on the “late-eight” sounds that produce the most articulatory difficulty in children through the age of eight: /s/, /z/, /l/, /r/, /ʃ/, /ʒ/, /θ/, and /ð/ (Shriberg, 1993). Any syllables with these sounds were categorized as forbidden and were not included in the nonwords. All repeat syllables were also categorized as forbidden, for ease in scoring. See Figure 4 for the percentage of languages included in the final version of The Universal Nonsense task.

After categorizing all of the syllables, a computer program was created to randomize the syllables based on their categories. The program produced forty nonwords – 10 each of two-, three-, four-, and five-syllable words. Each nonword was double checked for forbidden sounds. Syllables were moved around between the nonwords to ensure that there was an equal spread of sounds across words of different lengths. In the end, The Universal Nonsense measure consisted of words such as: /fi.nɔ/ (sounds like "fee.no"), /hua.en.nʲi/ (sounds like "hua.en.nyi"), /dɑ̃.de.noŋ.ɪç/ (sounds like "do.de.nong.ish"), and /hæv.kɛu.me.gən.fi/ (sounds like "have.keu.may.gun.fi"). See Table 1 for the complete list of nonwords. All the nonwords were recorded on a digital voice recorder. There was a brief pause in-between the syllables of each nonword, and a three-second pause in-between the nonwords to allow time for participant responses.

Digit span task. A digit span memory task, which is generally administered in the English language to measure participants' phonological working memory, was created to check the construct validity of The Universal Nonsense measure. Deevy et al. (2010) posited that in order to establish the diagnostic accuracy of a new tool, one must compare performance on the new task with a "gold standard" of the field. For working memory, this gold standard is considered to be the backward digit span recall task, in which participants hear a string of numbers and are asked to repeat the digits back to the tester, backwards. However, since nonword repetition tasks generally do not require manipulation of information in a way that the backward digit span does, – nonword repetition tasks require holding information in memory long enough to articulate the sounds – a forward digit span recall task was created and used to serve as a manipulation check for this study. Additionally, the length of the number series generally increases until the participant can no longer repeat the string of numbers accurately (usually around seven digits). Because the length of the nonword syllables was capped at five units, the length of the number series was capped at five units as well.

An online random number generator was used to determine the order of digits to be included in the task. There were five trials each of two-, three-, four-, and five-digit trials, resulting in 20 total trials. Each trial was recorded with a one-second gap in between its constituent digits. There was a three-second gap between trials on the recording to allow time for participant repetitions. See Table 2 for the digit span task.

Language questionnaire. The last component of this study was an extensive language questionnaire programmed on the Qualtrics platform. The aim of this questionnaire was to collect data on any and all language exposure participants had before participating in the study. Participants were asked to choose their primary native language and any secondary native languages from a list of language families. They were then asked if they were indirectly exposed to any languages before or after the age of eight via their surroundings or the media. Finally, participants were asked if they had, in any point, learned any languages in an academic setting, either in person or online. Participants were asked to rate their proficiency in reading, writing, speaking, and comprehension in all languages they were exposed to. The list of language families the participants could choose from can be found in Table 3.

Procedure

There were four procedural components to the present study: obtaining consent, completing two different working memory tasks, and completing the language questionnaire. Recruited participants were read a description of the study and were verbally asked if they gave their consent to participate in the study. Participants who gave verbal consent were administered the Universal Nonsense nonword repetition task and the digit span recall task created for this study. For both tasks, participants heard a spoken stimulus (either a nonsense word or a series of numbers) and were asked to repeat it. Both tasks were recorded on a digital voice recorder. If the participants were not familiar with the names of the English digits (0-9), the digit span task was to be administered to them in their native language, but need for this accommodation did not arise.

The order of the working memory tasks was counterbalanced to control for the potential impact of memory fatigue. In half of the participants, the digit span task was administered before the nonsense word repetition task and in the other half of the participants, the nonsense word repetition task was administered first. Background information regarding the participants' age, gender, and language history was then collected using an online language questionnaire on Qualtrics. Language exposure was measured on a five point Likert scale (0: *No exposure* to 4: *fluent/native speaker*). At the end of the questionnaire, participants were was given a bag of candy to thank them for their time. The entire study session took fifteen minutes to complete.

Scoring. Participants' Qualtrics survey results were saved and the memory tasks were scored. The Universal Nonsense task was scored twice for each participant – once during live data collection and once afterwards using the audio recording. Both times, the participants' responses were transcribed into IPA. Any inconsistencies in the sounds produced compared to the target sounds were marked as incorrect, except for inter-syllable insertions. Each participant received a "syllables correct" and a "words correct" score for the two-, three-, four-, and five-syllable word groups. The total "syllables correct" score was the main score for each participant to allow for more precise scoring (Gathercole et al., 1994). If the speaker consistently mispronounced a sound more than 50% of the time (for example, consistently producing a /θ/, or "th" sound instead of a /s/), the mispronunciation was counted as correct to account for articulation differences.

The digit span recall tasks were scored on a simple correct/incorrect scale, as is the norm for similar tasks – participants who misspoke one number in a string of numbers were scored to be as inaccurate as participants who misspoke all of the digits in a string. This method of scoring was chosen over an individual digit scoring method due to the complex nature of scoring digit series. The types of errors participants could make (e.g. switching numbers in a series with one another, eliminating numbers, or adding numbers) would have made the scoring procedure more complicated without offering much additional information about performance on the task. See Appendix for the complete Universal Nonsense score sheet.

Analyses

After collecting the data, results for the 48 participants were analyzed using SPSS. In this study, the independent variables were considered to be the verbal and nonverbal skills each participant had for each of twenty language groups (see Table 3 for a complete list of the languages). Verbal language skills were calculated by averaging participants' self-ranked oral comprehension and speaking skills for each language, while print-oriented language skills were the averages of participants' self-ranked reading and writing skills for each language. None of the participants cited any exposure or proficiencies in the Baltic, Uralic, or Austro-Asiatic languages, and all participants cited a fluent/native-speaker proficiency in all aspects of English. In sum, a total of 49 semi-independent variables (language exposures) were collected for this study. The dependent variables were considered to be the percentage of syllables participants correctly repeated on The Universal Nonsense task (out of 140 total syllables), as well as participants' overall performance on the digit span recall task (out of 20 series of numbers).

First, a series of correlations were carried out to test the relationship between the number of languages participants were exposed to and their performance on The Universal Nonsense. The number of languages participants were exposed to had a moderately positive correlation to the number of syllables accurately repeated in the two-, three-, and five-syllable words, $r(46) = .328$, $p < .05$; $r(46) = .341$, $p < .05$; $r(46) = .346$, $p < .05$, respectively. In other words, the number of languages participants were exposed to had a stronger relationship with the five-syllable nonwords than the two-syllable nonwords. There was no relationship between language exposure and the four-syllable nonwords on the new task, $r(46) = .182$, $p < .05$. See Figure 6 for a scatterplot of demonstrating the relationship between the number of languages participants were exposed to and their scores on the new task.

Further analysis into participants' performance on The Universal Nonsense revealed no significant differences between the repetition accuracy of the different syllable-length words. However, a trend did arise. The two- ($M = 81.87\%$, $SD = 8.16\%$) and three- ($M = 81.25\%$, $SD = 11.66\%$) syllable repetition accuracies were very similar, but there was a greater difference between the repetition accuracy on these words and the four- ($M = 67.34\%$, $SD = 16.14\%$) and the five- ($M = 45.58\%$, $SD = 14.43\%$) syllable words. There was also a difference between the repetition accuracies of the four- and five- syllable words. See Figure 7 for a visual representation of the accuracy differences across syllable length. Finally, an independent samples t-test revealed that there were no order differences present between the participants who completed the new task first and those who first completed the digit span task, $t(46) = 0.84$, $p = 1.674$.

The first hypothesis of the present study posited there would be a positive correlation between participants' performance on The Universal Nonsense and their performance on the digit span recall task. All participants scored perfectly (five out of five points) on the two- and three-digit recall tasks and 97% of participants scored perfectly on the four-digit recall task, so correlations were not carried out for this part of the measure. There was a good range of scores on the five-digit number series, so a correlation was carried out between these scores and performance on The Universal Nonsense as a manipulation check of the validity of the new task. A moderately significant positive correlation was found between overall performance on The Universal Nonsense and the five-digit recall scores, $r(46) = .366, p < .05$.

Finally, a one-way ANOVA was conducted to test the second hypothesis of the study: If the task did not differentiate between participants with diverse language backgrounds, then language exposure to the different language groups that were included in the task would not impact performance on The Universal Nonsense task. Initially, it seemed that a logistic regression would be the best way to analyze the data due to the sheer number of semi-independent variables and the complex relationships that occurred between them in the participants. However, this type of analysis requires a large sample size (the recommended size for the present study, given the number of independent variables, would have been around 440 participants).

Thus, a comparison of means was conducted to compare the effect of exposure to languages included within the nonwords of The Universal Nonsense task on performance on the task. Exposure to languages included in the task was considered to be a two-level independent variable of whether or not participants had exposure to languages contained within the language groups that were part of the nonwords in The Universal Nonsense. Only three participants did not have exposure to any languages that were included in The Universal Nonsense other than English, while the remaining 45 had exposure to at least one language other than English that was included in the new task. An independent samples t-test was conducted to test the differences of the means of these two groups. Even though the sample sizes were very different, there was no significant difference found between the two groups of participants, $t(46)=1.70$, $p = .096$. Finally, a one-way ANOVA showed that there was no difference between the two groups of participants, $F(1, 46) = 2.885$, $p = .096$.

Discussion

The purpose of this study was to create and pilot test a new nonword repetition task to measure phonological working memory in populations that have been exposed to more than just English in their lifetime. In many research situations, people who have been exposed to languages other than English are often excluded from participation for a variety of reasons, including a lack of standardized non-language specific memory tasks. This study aimed to make some progress in dealing with this issue by positing two hypotheses. First, if the new task was a valid measure of phonological working memory, participants' nonword repetition performance on would be positively correlated with performance on a digit span recall task. The second hypothesis theorized that if the task was indeed "linguistically fair" across people with various language exposures, people with various language backgrounds would perform similarly.

The first hypothesis, which explored the relationship between The Universal Nonsense task and the digit span recall task, was partially supported by the data that were collected from the 48 college participants. All of the participants performed perfectly (scoring five out of five points) for the two- and three- digit number series, as did the majority of the participants in the four-digit number series, indicating the need for a more difficult nonword repetition task. That being said, a moderately positive correlation was found between performance on the five-digit span task and The Universal Nonsense, supporting the idea that The Universal Nonsense measures some of the same constructs as the “gold standard” for phonological working memory. Additional support for the construct validity for The Universal Nonsense comes from the pattern of participants’ performance on the new task alone. Performance for the Universal Nonsense task was better for the four-syllable nonwords than the five-syllable nonwords. However, like on the performance on the two- and three- digit series, performance on the two- and three- syllable nonwords were very similar across participants. Gathercole et al. (1991) found a main effect of syllable length on repetition accuracy on a nonword repetition task, such that one-syllable words were repeated more accurately than four-syllable words, two-syllable words were repeated more accurately than three-syllable words, and three-syllable words were repeated more accurately than four-syllable words.

These data suggest two courses of action in the implementation of the next round of pilot testing for The Universal Nonsense task. First, the digit span recall task that was used as a manipulation check for the new task needs to be made more difficult – potentially by starting the task with strings of five-digit numbers and ending it with strings of eight-digit numbers. The second modification applies to the new task. Seeing that participants performed similarly in both the two- and the three-syllable nonwords, it would be beneficial to do a further analysis into the construction of the nonwords to figure out how to better differentiate the two-syllable nonwords from the three-syllable ones. Potential future analysis could identify the lexical neighborhoods of syllables that were included in the task and the wordlikeness of the final words in the task in order to strengthen the construct validity of the nonwords themselves.

The second hypothesis, which looked at the relationship between language exposure and performance on The Universal Nonsense working memory task, was also supported. Performance on the new task was not found to be significantly different across participants who had exposure to languages that were used to make the task and participants who did not, supporting the idea that the new task was indeed “linguistically equivalent” for participants in this sample. See Figure 6 for a comparison of these participants’ scores. However, the sample size for these two populations were greatly skewed, as only three participants had no exposure to the languages included in the new task (excluding English), and the spread of language exposure among the participants was very diverse (see Figure 3).

Additional investigation into the relationship between language exposure and performance on The Universal Nonsense task revealed a strengthening correlation between the number of languages participants were exposed to at the time of the study and the number of syllables in the nonwords. It is possible that this is due to the resources needed to listen to and maintain phonemic information in different languages, but further research into the differences in language groups and their impacts on repetition accuracy is needed. The next round of pilot testing for Universal Nonsense should also incorporate a larger sample size with speakers of a more diverse selection of languages. As noted earlier, 92% of participants in this study had exposure to a Romance language at some point in their lives. Testing the new task on a more diverse population would cast a better light on any language group biases that need to be accounted for. A larger and more diverse sample would also allow a more in-depth item analysis of the nonwords, which would be needed to eventually standardize performance on the task. Finally, future pilot testing also needs to incorporate more diverse phonetic transcribers in order to make the scoring of the new task more reliable.

The promising nature of the findings of the present study and the ever-growing multilingual population around the globe demonstrate the potential for a linguistically fair nonword repetition task to measure phonological working memory. A tool such as The Universal

Nonsense, once fully pilot tested in both children and adults and standardized, could aid in specific language impairment (SLI) detection in multilingual young children and could be a step toward allowing the participation of a more diverse population in memory studies across the United States. Previous studies have shown that phonological working memory tasks aid in detecting word learning and expressive skill deficiencies, in addition to possible deficiencies in morphological and phonological learning abilities (Gathercole & Baddeley, 1989; Adams & Gathercole, 1993; Nelson, 1987). As the multilingual population around the globe is growing, it is important to allow the tools we use to measure such deficiencies to adapt to our needs.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Archibald, L. M., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, *49*(5), 970-983.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*(4), 485.

Best, C. C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and speech*, *46*(2-3), 183-216.

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant behavior and development*, *18*(3), 339-350.

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing Bias in Language Assessment Processing-Dependent Measures. *Journal of Speech, Language, and Hearing Research*, *40*(3), 519-525.

Deevy, P., Weil, L. W., Leonard, L. B., & Goffman, L. (2010). Extending use of the NRT to preschool-age children with and without specific language impairment. *Language, Speech, and Hearing Services in Schools*, *41*(3), 277-288.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136-1146.

Dollaghan, C., Biber, M., & Campbell, T. (1993). Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech, Language, and Hearing Research*, *36*(5), 1051-1054.

Dollaghan, C. A., Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics*, *16*(02), 211-222

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, *2*(2), 103-127.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, *12*(03), 349-367.

Gathercole, S.E. & Baddeley, A. D. (1996). *The Children's Test of Nonword Repetition.* London: Psychological Corporation.

Matthews, P. H. (2007). *The concise Oxford dictionary of linguistics*. Oxford University Press.

Shriberg, L. (1993). Four new speech and voice-prosody measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech, Language, and Hearing Research*, 36, 105–140.

U.S. Census Bureau (2013): *Language use in the United States: 2011: American community survey reports.*

U.S. Census Bureau (2015): *Language use in the United States: 2009 - 2013: American community survey.*

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, *7*(1), 49-63.

Appendix
Universal Nonsense Phonetic Analysis and Score Sheet

Universal Nonsense Scoring Sheet 3

Phonetic Analysis:

Note the accuracies of the following consonants from the above nonwords. If participants have a less than 50% accuracy on any given sound, do not count the sounds as inaccurate in the next section.

b:	/7	g:	/7	m:	/20	ɲ:	/5	w:	/8
d:	/11	h:	/14	n:	/36	p:	/4		
d^j:	/2	j:	/8	n^j:	/3	t:	/18		
f:	/7	k:	/13	ŋ:	/6	v:	/7		

Accuracies

Two syllable words:

Words correct: ________/10 Percent: _________

Syllables correct: ________/20 Percent: _________

Three syllable words:

Words correct: ________/10 Percent: _________

Syllables correct: ________/30 Percent: _________

Four syllable words:

Words correct: ________/10 Percent: _________

Syllables correct: ________/40 Percent: _________

Five syllable words:

Words correct: ________/10 Percent: _________

Syllables correct: ________/50 Percent: _________

Total words correct: ________/40 Percent: _________

Total syllables correct: ________/140 Percent: _________

Tables

Language group	Language(s)
Afro-Asiatic	Arabic
Austro-Asiatic	Vietnamese
Austronesian	Tagalog
Germanic	English, German
Indo-Iranian	Hindi
Romance	French, Italian, Portuguese, Spanish
Slavic	Polish, Russian
Sino-Tibetian	Chinese (Mandarin and Cantonese)
Koreanic	Korean

Table 1. The 14 most common languages spoken in the United States can be divided into nine language groups.

Two Syllable	Three Syllable	Four Syllable	Five Syllable
fi.nɔ	də.ien.gɔ	an.wi.nʲi.hə	aʊf.di.aj.ŋu.də
ga.hæv	hua.en.nʲi	dɑ̃.de.noŋ.ɪç	hæv.kɛu.me.gən.fi
dʲin.meɪ	pɐ.vɑ.nɑɔ	gən.kʊ.pɑ.ken	hɔ.fi.ɤu.do.hɛ
mɑg.puw	ɪf.dʲin.en	gɔ.ma.mot.ien	jæ.dɐ.bət.nɛt.gɐ
mə.bʊ	joʊ.hai.mʲɛ	gɐ.ɪn.nɑ.ɑn	ma.bwat.aiɲ.wən.ʊn
nit.an	kɔm.mʲetʲ.toi	ma.ɲɯŋ.vɑ.æj	mɛ.dɨɜk.ŋa.hæd.ɤu
on.miɲ	kʊ.imʊ.hi	mɪt.nit.meɪ.jɐ	tɯ.kə.wu.iŋ.keɪf
pɐ.baj	tɑi.ɲɯŋ.dɛ	nɑɔ.æn.baɪ.tɑi	hai.wa.kɑɲ.ga.mɑj
tʌk.wu	tɔ.kə.æn	neu.ɪç.nɛ.moi	æn.nʲe.ʌn.əv.ten
vɛk.du	bwat.hua.hɔ	vo.tɔ.hə.foj	tɛm.kɔm.ho.bən.nɛ

Table 2. Universal Nonsense nonwords. All words are written in the International Phonetic Alphabet (IPA). Syllable boundaries are demarcated with /./.

Two Digit	Three Digit	Four Digit	Five Digit
1 4	7 9 9	1 4 2 3	3 5 9 4 6
3 6	9 1 4	4 2 8 9	9 6 6 2 4
3 4	2 4 2	1 5 8 7	5 9 8 1 5
6 8	9 3 7	5 6 9 3	7 8 2 3 2
1 7	7 8 5	4 3 8 6	8 6 9 2 7

Table 3. Digit span recall task. Participants were asked to engage in recall of each series of numbers. The task was presented with a one-second delay between digits of a series, and a three-second delay between each series.

Language Family Choices Included in the Language Questionnaire
Romance (i.e. French, Spanish, Portuguese, Italian, Romanian)
Germanic (i.e. German, Dutch, Danish, Swedish, Icelandic, Norwegian)
Celtic (i.e. Welsh, Irish, Breton, Scottish Gaelic, Cornish)
Slavic (i.e. Russian, Ukranian, Czech, Slovak, Polish, Slovenian, Bulgarian)
Baltic (i.e. Latvian, Lithuanian)
Indo-Iranian (i.e. Hindi, Urdu, Bengali, Punjabi, Marathi, Gujarati, Bhojpuri, Rajasthani, Nepali, Persian, Pashto, Kurdish, Balochi)
Afro-Asiatic (i.e. Arabic - all dialects, Hausa, Oromo, Somali, Modern Hebrew, Modern Aramaic (all dialects)
Niger-Congo (i.e. Bambara, Soninke, Wolof, Fula, Dogon, Kru, Senufo, Gur, Chamba Leko, Sango, Kwa, Ewe, Swahili, Fang, Zulu, Kongo)
Turkic (i.e. Turkish, Azerbaijani, Turkmen, Tatar, Bashkir, Kazakh, Kyrgyz, Uzbek)
Dravidian (i.e. Tamil, Malayalam, Kannada, Telugu)
Uralic (i.e. Finnic, Hungarian, Estonian)
Japonic (i.e. Japanese, Ryukyuan)
Austro-Asiatic (i.e. Khmer)
Austronesian (i.e. Fijian, Filipino/Tagalog, Indonesian, Javanese, Sundanese)
Tai-Kadai (i.e. Hlai, Ong Be, Kam-Sui, Kra, Tai)
Koreanic (i.e. Korean)
Sino-Tibetian (i.e. Chinese - all dialects, Karen, Burmese - all dialects)
Other (please specify): ____________________

Table 4. List of language families included as choices for all language-related questions in the Universal Language questionnaire.

Figures

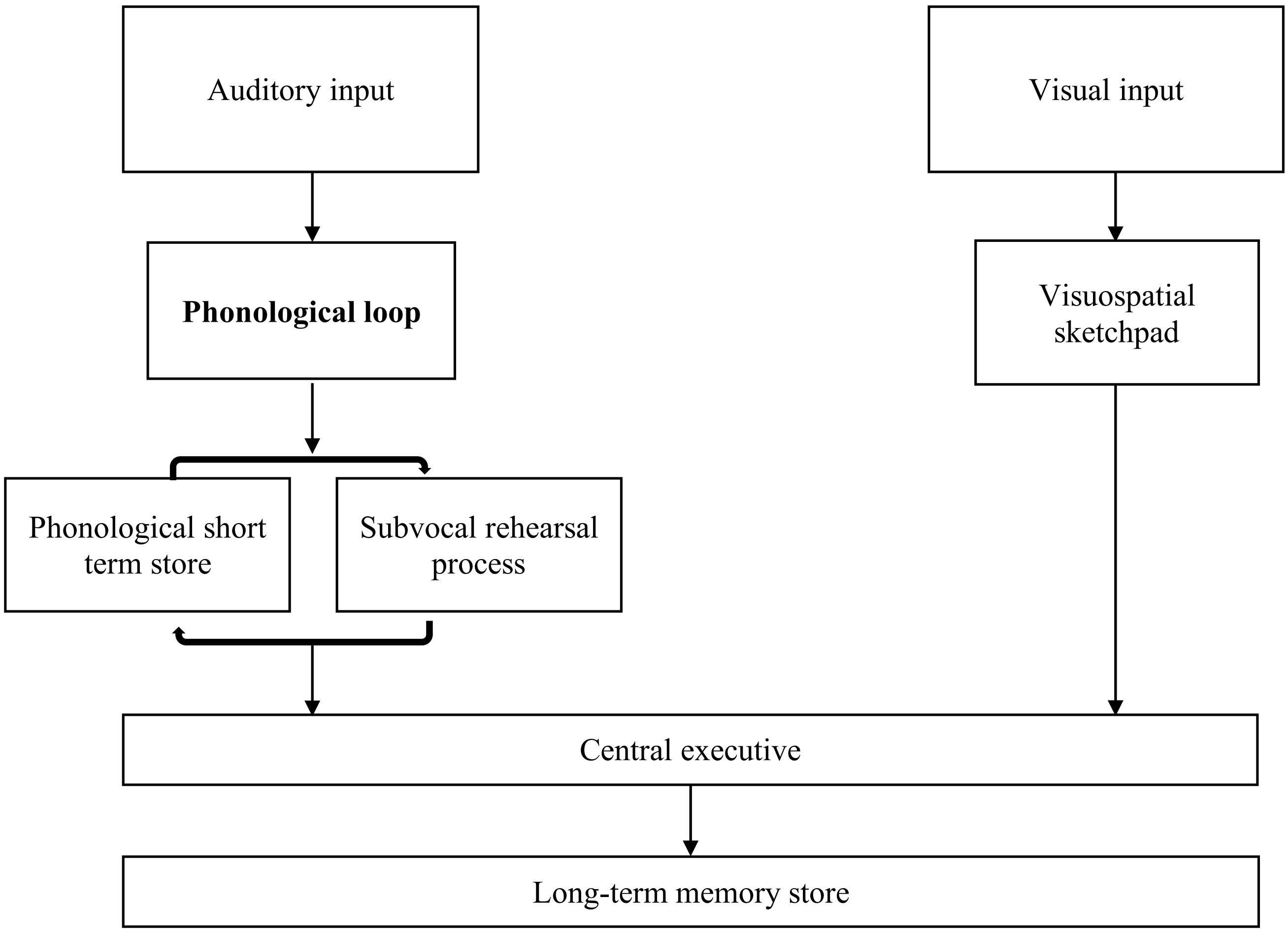

Figure 1. Visual representation of the working memory structure. Nonword repetition tasks aim to measure functioning of the phonological loop, which includes the phonological short-term store and the subvocal rehearsal process.

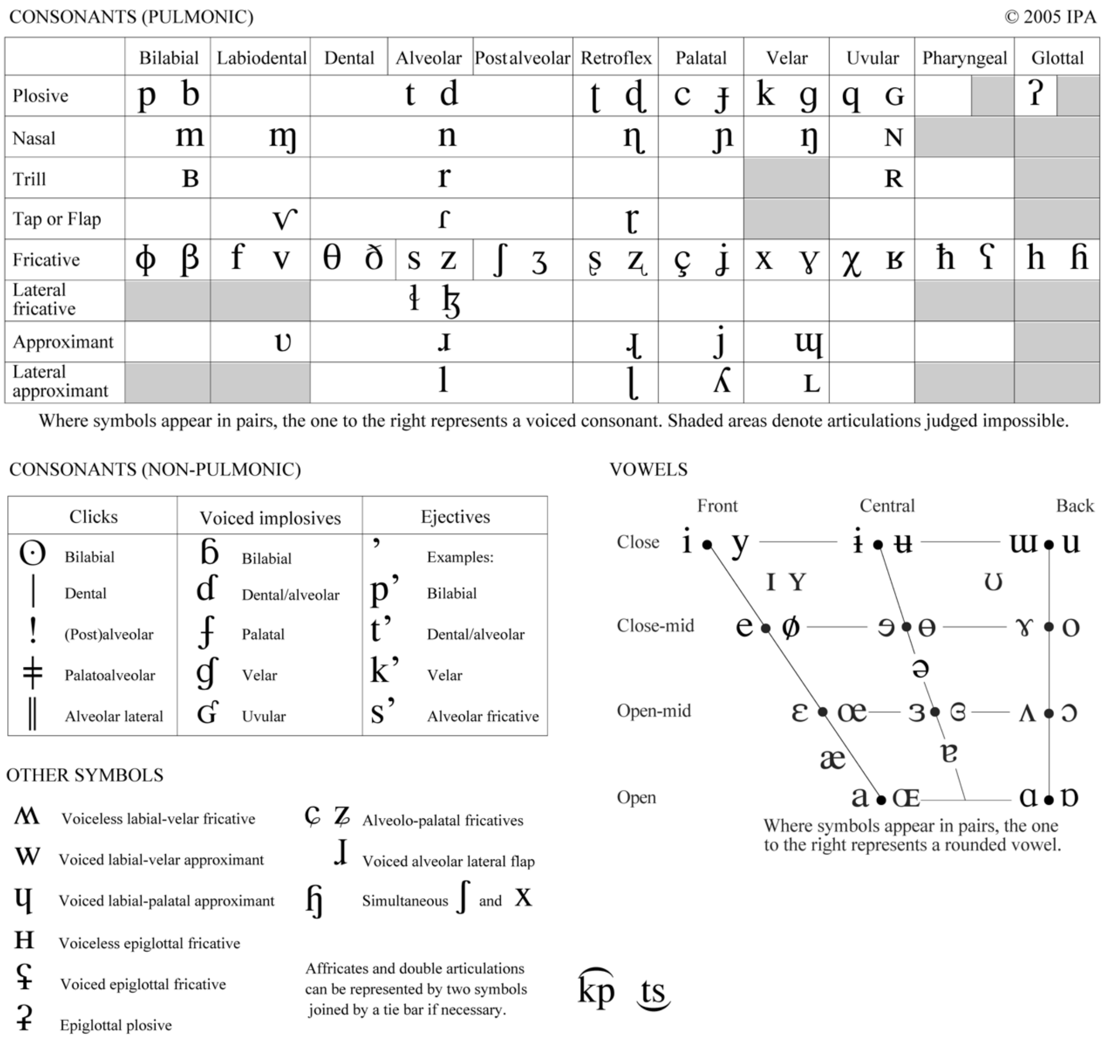

Figure 2. The complete chart of the International Phonetic Alphabet includes consonants and vowels present in almost all languages across the globe (IPA). For consonants, the mode of articulation is on the left-most column, while the place of articulation is on the top row.

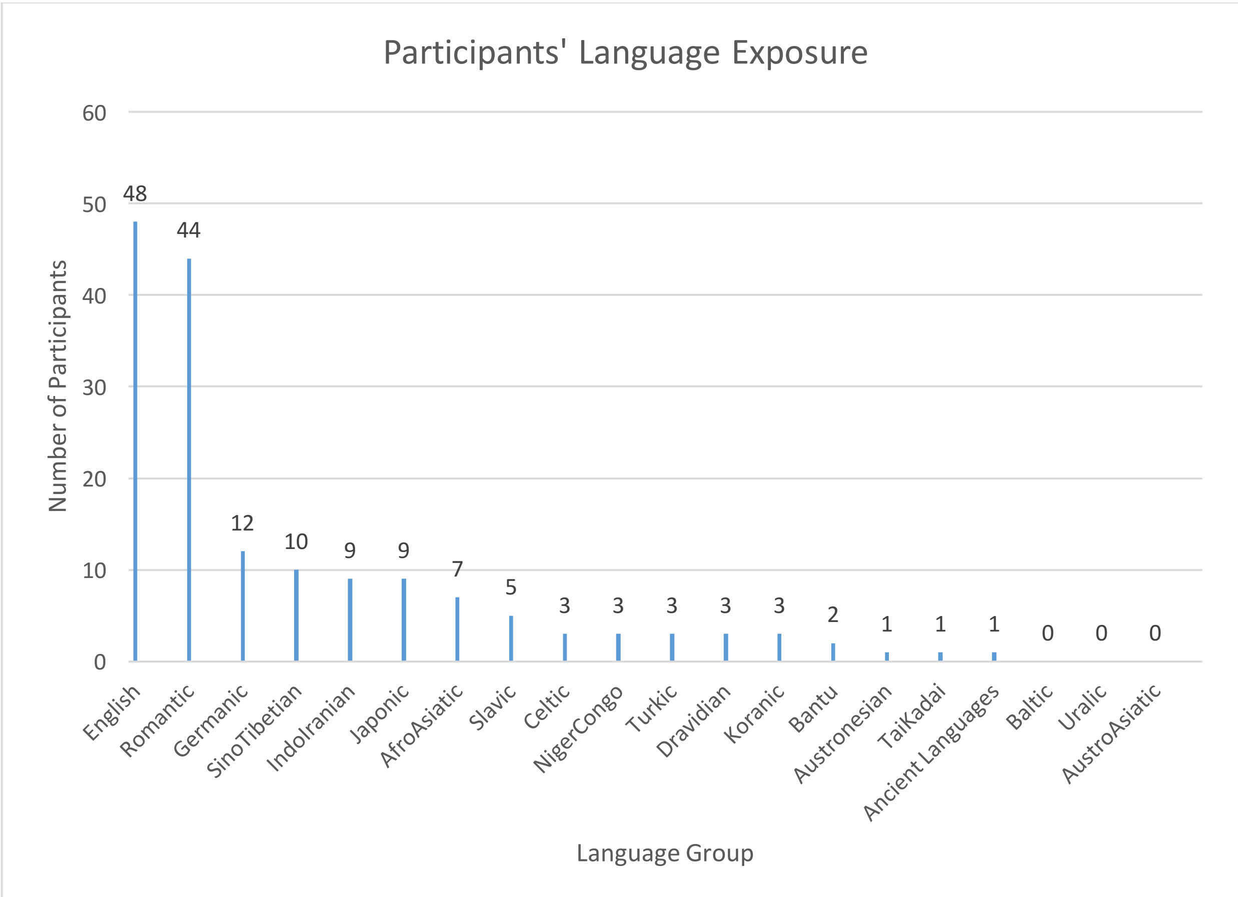

Figure 3. Distribution of the 48 participants' language family exposures.

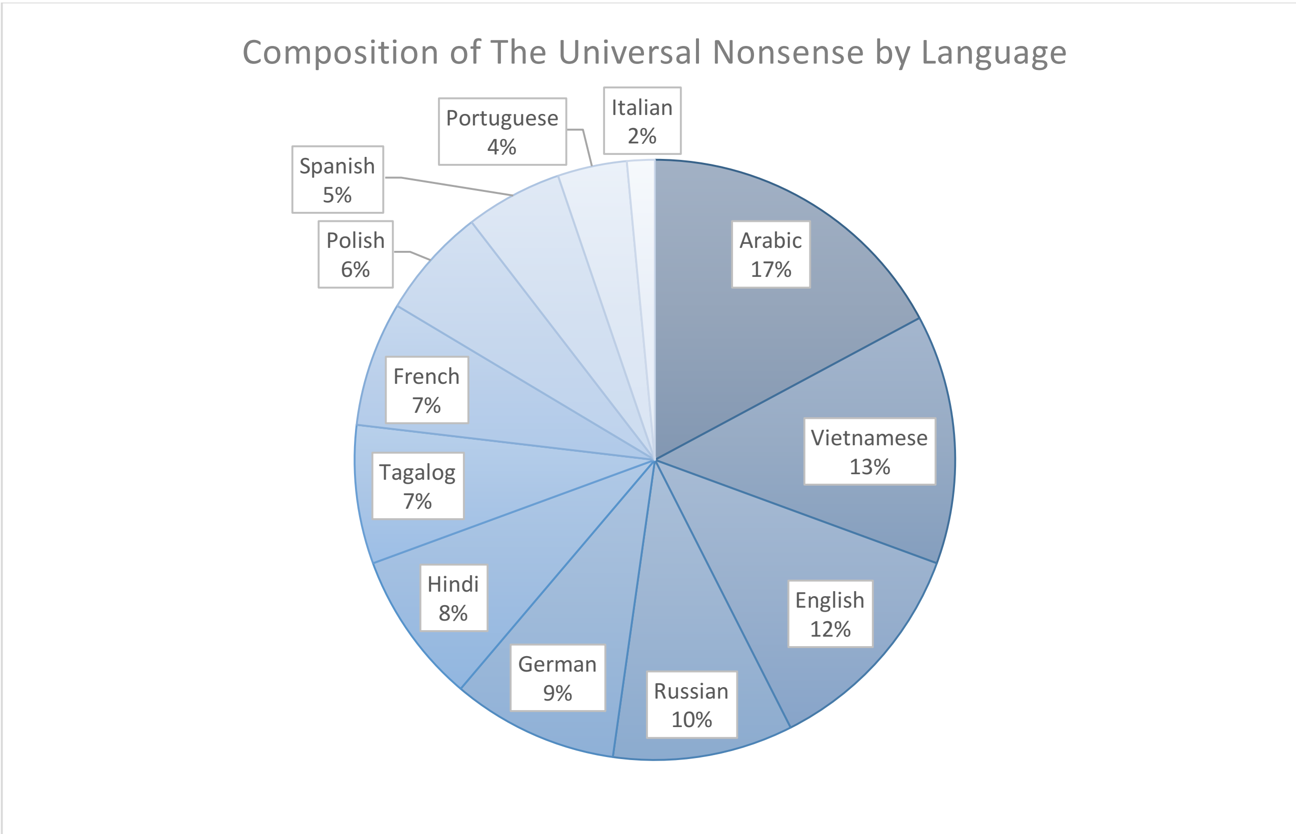

Figure 4. Origins of the syllables used in the final product of The Universal Nonsense. 134 unique syllables were used to create a task that contained a total of 140 syllables.

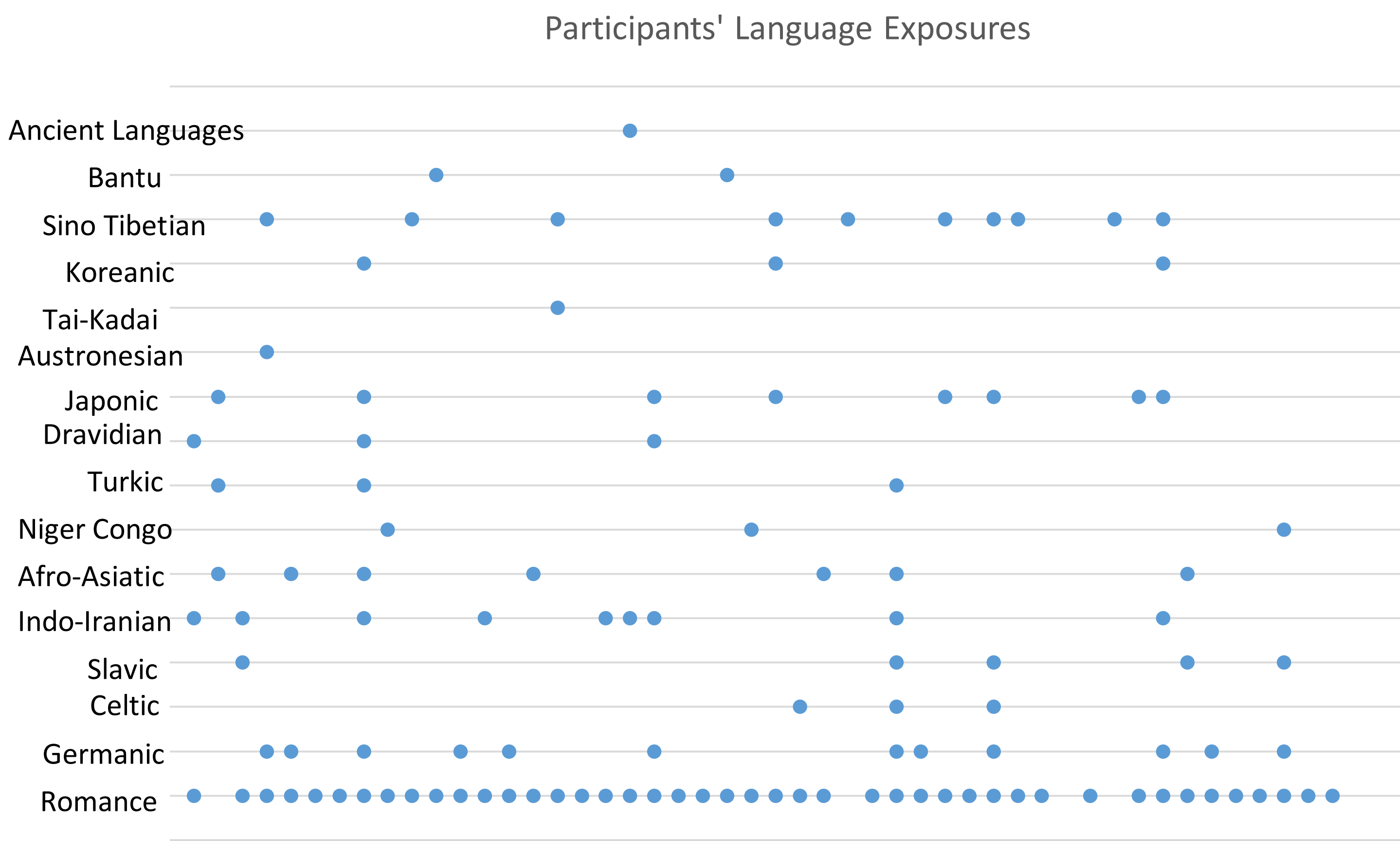

Figure 5. A depiction of the pattern of language exposures participants exhibited. All participants were native or fluent English speakers, while 92% had exposure to a Romance Language at some point in their life.

Participant

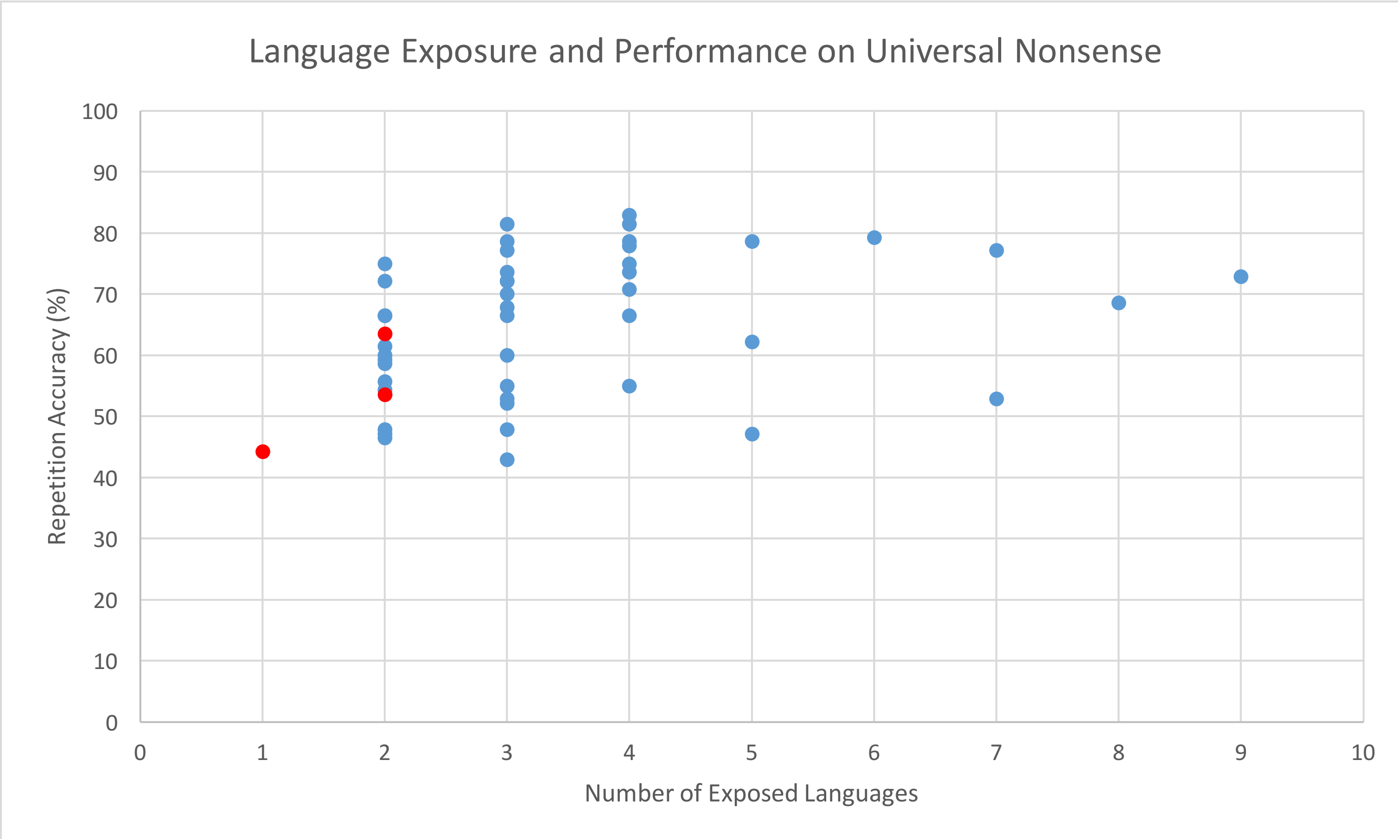

Figure 6. Relationship between the number of languages participants were exposed to and their performance on The Universal Nonsense. Points marked in red are participants who did not have exposure to languages included in Universal Nonsense besides English. Note that the number of exposed languages measured on the x-axis of this chart is a measure of exposure to language groups, not individual languages. It is possible that participants have been exposed to more individual languages than indicated above.

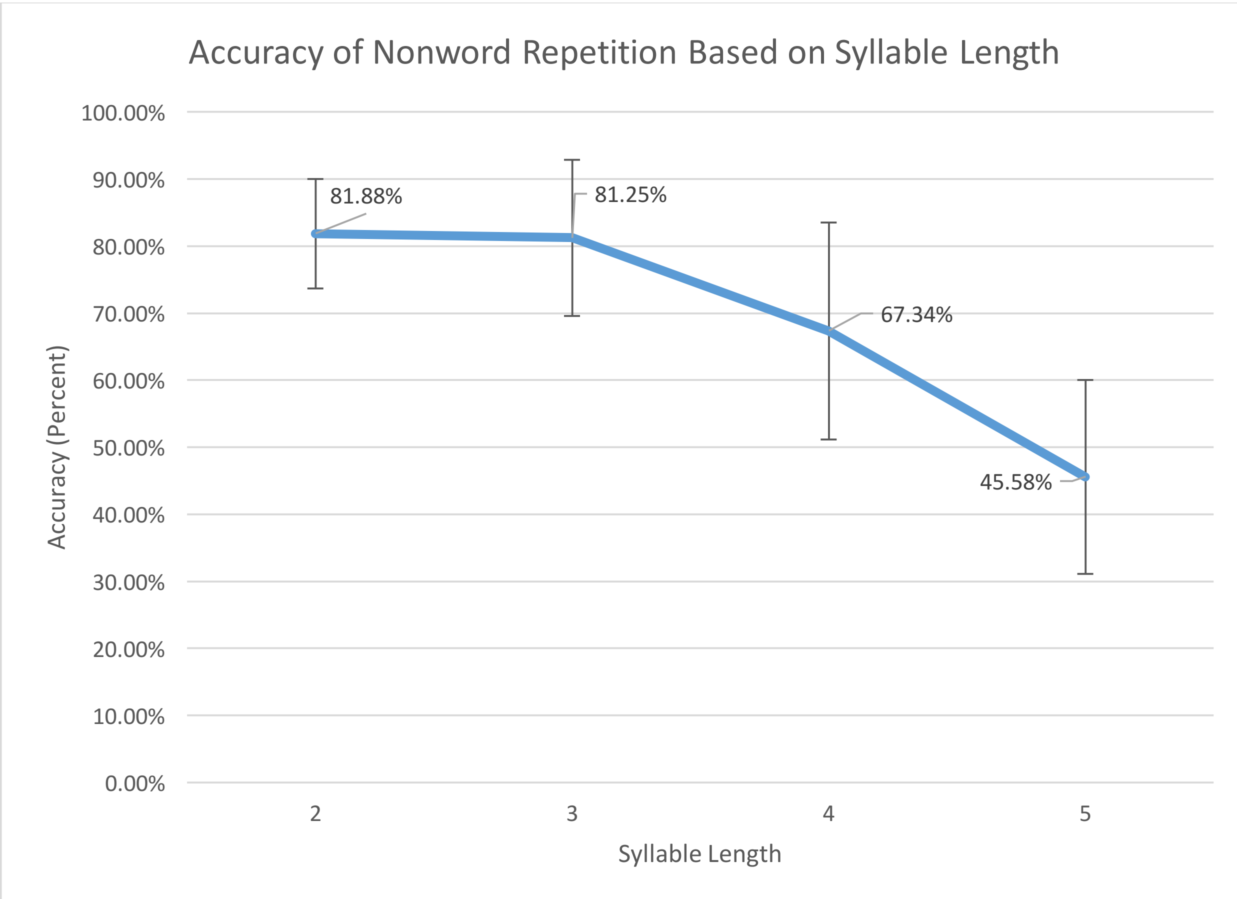

Figure 7. Although none of the means are significantly different, the trend of nonword repetition performance on The Universal Nonsense demonstrates a trend in which repetition accuracy decreases as syllable length of the word to be repeated increases.