Supplement Article

Impact of Personal Relevance on Acquisition and Generalization of Script Training for Aphasia: A Preliminary Analysis

Leora R. Cherney,^{a,b} Rosalind C. Kaye,^a Jaime B. Lee,^a and Sarel van Vuuren^c

Purpose: The importance of personalization in script training in aphasia has been assumed but never tested. This study compared acquisition and generalization of personally relevant versus generic words or phrases appearing in the same scripts.

Method: Eight individuals (6 men; 2 women) with chronic aphasia received 3 weeks of intensive computer-based script training. For each participant, 2 scripts, a trained and an untrained generalization script, were embedded with 4 personally relevant word choices and 2–4 generic items that were similar across participants. Scripts were probed for accuracy at baseline and posttreatment. Significance testing was conducted on baseline and posttreatment scores, and for gains in personally relevant versus generic items. Effect sizes were computed.

Results: Both personally relevant and generic items improved significantly on trained scripts. Improvements on untrained scripts were smaller, with only personally relevant items reaching significance. There was no significant difference on gains made on personalized versus generic items for trained scripts ($p = .059$), but the effect size was large ($d = 0.90$). For generalization scripts, this effect was small ($d = 0.25$) and nonsignificant.

Conclusions: Personally relevant words and phrases were acquired, although not generalized, more successfully than generic words and phrases. Data supports the importance of personalization in script training, but the degree of that importance requires further investigation.

Script training is a social approach to aphasia treatment consistent with the life participation philosophy that focuses on improving communication in everyday activities (LPAA Project Group, 2000; Simmons-Mackie, 2008). It typically involves the repeated practice of phrases and sentences embedded within a monologue or dialogue individualized to the person with aphasia (PWA). The rationale for script training is derived from the instance theory of automatization (Logan, 1988), which suggests that automaticity of skills is achieved by retrieving memories of complete, context-bound, skilled performances. These memories are formed by repeatedly practicing a particular task. The goal is for PWAs to achieve “islands of automatic speech” (i.e., the specific words and phrases trained within the script) that can be produced fluently for use in real-life

discourse (Youmans, Holland, Munoz, & Bourgeois, 2005). A number of studies have evaluated and found support for script training for PWAs (Bilda, 2011; Cherney, Halper, Holland, & Cole, 2008; Cherney, Halper, & Kaye, 2011; Goldberg, Haley, & Jacks, 2012; Lee, Kaye, & Cherney, 2009; Youmans et al., 2005). Objective assessment of change as well as patient-reported outcomes by participants and family members, using psychometrically valid scales and qualitative interview data, provide evidence for script training in aphasia (Cherney et al., 2011; Manheim, Halper, & Cherney, 2009).

Although the specific methods of script training have varied across studies (e.g., delivered by clinician vs. computer vs. video, using repetition vs. choral reading), an important common feature has been the use of scripts that are personalized in some way for the PWA. For example, in Youmans et al.’s (2005) study, participants generated personally relevant topics and gave their input on the wording only after the scripts were written. The scripts were limited to three or four sentences, and contained opportunities for alternative words to be practiced interchangeably, such as preferred items in the grocery story. Goldberg et al. (2012) used communication supports, including clip art pictures, photographs, and the Life Interests and Values cards

^aRehabilitation Institute of Chicago, Center for Aphasia Research and Treatment, Chicago, IL

^bNorthwestern University, Chicago, IL

^cInstitute of Cognitive Science, University of Colorado-Boulder

Correspondence to Leora R. Cherney: lcherney@ric.org

Editor: Anastasia Raymer

Associate Editor: Gina Youmans

Received September 16, 2014

Revision received February 13, 2015

Accepted May 13, 2015

DOI: 10.1044/2015_AJSLP-14-0162

Disclosure: The authors have declared that no competing interests existed at the time of publication.

(Haley, Womack, Helm-Estabrooks, Caignon, & McCulloch, 2010), to help two participants select topics for their personalized scripts. After the topics were selected, participants were involved in a collaborative process in which they reviewed and made changes to the scripts with the researchers. In a computer-based study of script training using the AphasiaScripts software, the researchers describe codeveloping scripts “from topics of interest and situations in which participants expressed a desire for more effective communication, such as ordering in a restaurant, making a doctor’s appointment, or talking to grandchildren” (Lee et al., 2009, p. 888). These studies, as well as others, stress the importance of the script topic being meaningful and relevant to the PWA.

Personalization

Though relatively understudied in the field of communication disorders, personalization is an important aspect of rehabilitation that has the potential to affect patient engagement, satisfaction, and treatment outcomes (Cohen & Sloan, 2007; Cordova & Lepper, 1996). For PWAs, the use of personalized therapy materials is also supported by empirical studies. For example, Wallace and Canter (1985) investigated the effect of personally relevant language materials on the language performance of 24 participants with severe aphasia. The participants showed statistically better performance on personally relevant items than on nonpersonal items on four language tasks including auditory comprehension, repetition, naming, and reading comprehension. Because their personal and nonpersonal lexical items were not matched for frequency of occurrence in everyday oral communication, the researchers cautioned that the personalized material may have been easier because of its higher frequency of occurrence. They also suggested that participants might have performed better on the personalized items because of “greater motivation to respond to information that is more familiar and relevant” (Wallace & Canter, 1985, p. 388). In a more recent study, McKelvey, Hux, Dietz, and Beukelman (2010) evaluated the impact of personal relevance on the preferences and word-picture matching accuracy of people with severe aphasia and found that participants both preferred personally relevant materials and performed more accurately when presented with personalized target words and images. However, neither of these studies was conducted within a treatment paradigm whereby the impact of personalization on treatment outcomes was evaluated.

Personalization and Script Training

A primary reason for script personalization is based on the premise that script learning is item specific. Increased production of specific words and phrases from a target script have been reported in participants following script training (e.g., Cherney et al., 2008; Youmans et al., 2005). Therefore, for script training to have the greatest impact on a PWA for whom verbal expression is effortful, it is important to

include specific content (items within the script) that is relevant to and functional for that individual.

Personally relevant scripts are also assumed to be more engaging and motivating for participants, and several studies have emphasized the importance of these factors in script training (Cherney et al., 2011; Holland, Halper, & Cherney, 2010; Lee & Cherney, 2008). Holland et al. (2010) examined the content of 100 scripts co-constructed by PWAs and a clinician and concluded that PWAs choose to speak about their family, interests, and life experiences. The researchers suggested that regardless of how the content is used in treatment, “materials should emphasize matters of high personal relevance to those treated” (Holland et al., 2010, p. 198). They also noted that motivational differences related to script topics could affect the amount of practice participants with aphasia choose to complete. Lee and Cherney (2008) proposed that “more important than the topic itself is its relevance and usefulness in participants’ lives and frequently encountered situations” (p. 19).

A related issue is whether personalization results in more successful acquisition of script content or generalization of learned items to other contexts. Whereas previous research indicates that participants with aphasia successfully acquire trained conversational scripts, generalization has been more limited and variable (Youmans et al., 2005). For two participants, Goldberg et al. (2012) reported improved grammatical morpheme use and increased rate of speech in generalization probes that included topic-related conversations. Increased use of specific words and phrases trained within a script to novel contexts, such as a restaurant or the gym, has also been reported (Cherney et al., 2011). It is possible that the words and phrases that generalize to novel contexts are those that are personally relevant to the PWA.

As previously discussed, scripts can be personalized in different ways and to different degrees. At one end of the continuum, the PWA may identify the script topic and contribute to the development of its content with the assistance of a clinician or family member (e.g., Goldberg et al., 2012; Lee & Cherney, 2008; Youmans, Youmans, & Hancock, 2011). In contrast, script training of conversational dialogues has been successfully conducted without much personalization of the content (Bilda, 2011; Nobis-Bosch, Springer, Radermacher, & Huber, 2011). Although personalization has been assumed by many to be an essential component of script training, to date, there is no experimental evidence that personalization results in more successful acquisition or generalization of script content. The purpose of this study was to examine the role of personal relevance in the acquisition of specific words and phrases within a trained script and their generalization to an untrained script about a different topic.

A number of possibilities have been proposed for why personal relevance may be important in learning and generalization. Wallace and Canter (1985) suggested there is greater motivation when material is familiar and more easily recognizable, noting increased attentiveness to that material. McKelvey et al. (2010) also suggested ease of

recognition, meaningfulness, and higher motivation. Given that previous experiments have shown that those with aphasia perform better with personally relevant materials (McKelvey et al., 2010; Wallace & Canter, 1985), we hypothesized that there would be better acquisition during training and generalization of the personally relevant words and phrases compared to generic words and phrases. We define *generic* words and phrases as those that are common and useful, but not specific or personalized to the participant relative to the selected topic. To be specific, we addressed the following questions:

- Are personally relevant words and phrases acquired more successfully than generic words and phrases in a trained conversational script?
- Do personally relevant words and phrases generalize more successfully than generic words and phrases to an untrained script about a different topic?

Method

Experimental Design

To answer these questions, we analyzed data that had previously been collected as part of a randomized controlled cross-over design examining the effects of high- and low-cue training conditions on the acquisition and maintenance of conversational script production (Cherney, Kaye, & van Vuuren, 2014; van Vuuren & Cherney, 2014). Participants practiced one script for 3 weeks under either the high- or low-cue condition. Then following a 3-week wash-out period, they practiced a second script for 3 weeks under the other cue condition. For our current purposes, we examined how well PWAs used specific personally relevant and generic words and phrases in the scripts that were trained under the high-cue condition, regardless of the order in which they were practiced. In the high-cue condition, participants were provided with a variety of cues prior to any independent verbalization of the script sentences in an attempt to minimize the production of any errors. We selected the high-cue condition because it is more typical of the script training that has been described in the literature and that is usually done in clinical practice. Furthermore, the original cross-over study had shown that there were no order effects affecting the data (Cherney et al., 2014). We also examined the performance of these same specific personally relevant and generic words and phrases in an untrained script on a different topic. The untrained script was developed to assess generalization and was matched for length and grammatical complexity to the trained script. The study was approved by the Institutional Review Board of Northwestern University.

Participants

Eight individuals (6 men; 2 women) with chronic aphasia due to a single left hemisphere stroke with resulting aphasia participated. All but one of the participants (AMBD) presented with a nonfluent aphasia. To be eligible

for the study, participants had to be native speakers of English, pass a pure-tone hearing screening at 25 dB in the better ear, have no history of significant psychological problems or neurological conditions other than stroke, and receive no other speech-language therapy services for 1 month prior to the study or during the study. Participants ranged in age from 25 to 66 years ($M = 52.0$, $SD = 14.0$); time post stroke onset was between 8 and 59 months ($M = 26.4$; $SD = 19.2$); aphasia severity was based on the Western Aphasia Battery–Revised Aphasia Quotient (Kertesz, 2007) and ranged from 28.1 to 80.1 ($M = 58.0$; $SD = 18.5$); and education ranged from 11 to 18 years ($M = 14.3$; $SD = 2.3$). Table 1 shows the demographic data for each participant. It also shows the topics of the trained and the matched generalization scripts that were analyzed in this study.

Scripts

Participants' scripts were based on templates we had developed in advance. Each template was a dialogue of 10 turns about a common topic, with the PWA as the responder in each turn. Templates were developed in pairs, with one serving as the template for the trained script and the other serving as the template for the untrained script used to assess generalization to a different context. The Restaurant (trained) template was always paired with the Cooking (generalization) template; the Catching Up (trained) template was always paired with the Making Plans (generalization) template. The templates were closely matched on number of words, sentences, syllables, verbs, morphemes, average words per sentence, average syllables per word, grammatical complexity (ratio of morphemes to words), and the Flesch Reading Ease Score (Flesch, 1948), a formula combining words per sentences and syllables per word.

The templates were designed so that eight items (either single words or phrases) were shared by each pair of trained and generalization scripts. Four shared items were generic—common and useful words or phrases such as “sounds good,” “leftovers,” “what’s new,” and “afternoon.” It was planned that the other four shared items would be personally relevant to the particular participant. Four blank slots appeared in each template; in the scripts derived from the templates, these slots were filled with words and phrases that were generated by each participant in response to open-ended questions, and were deemed to be personally relevant.

Figure 1 is an example of a script pair—Restaurant (trained) and Cooking (generalization)—that was used for one participant. Both contain the same four personally relevant items (*grilled chicken*, *gnocchi*, *iced tea*, *cheesecake*) and the same four generic items (*sounds good*, *dinner*, *leftovers*, *forty-five minutes*).

Personally relevant responses were obtained through brief interviews (10–20 min) with a speech-language pathologist (SLP), who was trained in using supported conversation. If the participant with aphasia had difficulty providing his or her desired response, the clinician offered several spoken and written possibilities on the basis of whatever

Table 1. Demographic and language characteristics of participants.

Participant	Age (years)	Gender	Handedness	TPO (months)	Education (years)	WAB-R AQ	Script topic	
							Trained	Generalization
ABEJO	51.8	M	R	48	16	68.8	Catching Up	Making Plans
AMBDE	61.6	M	R	21	11	74.5	Restaurant	Cooking
CAVCH	25.0	F	L	13	13	54.4	Catching Up	Making Plans
PIESH	66.4	F	R	59	14	67.6	Restaurant	Cooking
SMIDA	44.5	M	R	8	18	28.1	Restaurant	Cooking
SMISC	59.3	M	L	10	16	35.2	Catching Up	Making Plans
STEOM	42.9	M	R	15	12	55.0	Restaurant	Cooking
WELED	64.5	M	R	37	14	80.1	Catching Up	Making Plans
<i>M</i>	52.0			26.4	14.3	58.0		

Note. TPO = time postonset; WAB-R = Western Aphasia Battery-Revised; AQ = Aphasia Quotient; M = male; F = female; R = right; L = left.

information the PWA was able to generate. In an iterative process of question, response, clarification, and verification, the PWA's specific and preferred responses were obtained. For two participants with the most severe aphasia (SMIDA, SMISC), the spouse also was consulted to confirm the participant's responses and preferences. Table 2 shows the four interview questions that guided the clinician during the

interview, and each participant's final responses to those questions.

The participants in the original cross-over study for which the scripts were created (Cherney et al., 2014) varied greatly in severity of aphasia. For those participants who were more severe (Aphasia Quotient less than 55), it was necessary to simplify the overall difficulty level of their

Figure 1. Sample of a paired trained and untrained generalization script. The four personalized choices in both scripts are *grilled chicken*, *gnocchi*, *iced tea*, and *cheesecake*. The four generic items in both scripts are *sounds good*, *dinner*, *forty-five minutes*, and *leftovers*. The script templates are matched for number of sentences, words, syllables, morphemes, and verbs, as well as words per sentence, syllables and morphemes per word, and the Flesch Reading Ease Score.

Trained script (Restaurant)	Untrained script (Cooking)
Pat: Welcome to Jameson's. How are you tonight? You: I'm fine, thanks. Is grilled chicken on the menu tonight? Pat: Yes, it is. You: SOUNDS GOOD . I'll get that. What comes with it? Pat: Either soup or salad. You: I'm in the mood for a garden salad. May I have that? Pat: Sure. You: And could you put in lots of cucumbers? Pat: No problem. You: Actually, I changed my mind about my order. I'd rather have gnocchi . Pat: Anything to drink? You: I'll have iced tea with DINNER . Pat: How about dessert? You: Let me look at the menu a minute. I think I'll have cheesecake . Pat: Sorry, we ran out. You: Then I won't get dessert, thanks. But can I have a decaf? With no milk or sugar. Pat: Sure. You: I'm going to a movie in FORTY-FIVE MINUTES , so I'm in a hurry. Can you take my credit card now? Pat: Of course. You: Thanks. And can I have a container for the LEFTOVERS ?	Pat: What time's dinner? You: At six. Do you want to go out or eat here? Pat: Let's eat at your place You: SOUNDS GOOD . I think I have enough for both of us. Pat: Great. What do you have? You: There's some food in the fridge that I could take out. Pat: Like what? You: I can heat up some gnocchi from Jameson's. It's a great place. Pat: Yum. You: Or I can make grilled chicken if we don't want LEFTOVERS . Do you like mashed potatoes? Pat: I love them! You: Then I'll make those too. They're pretty easy. Pat: Do they take long? You: They take about FORTY-FIVE MINUTES . I'll start them right now. Pat: Should I bring something to drink? You: I have iced tea . But can you bring some dessert? Pat: what would you like? You: How about cheesecake ? I also rented The Notebook. It's one of my favorite movies. We could watch after DINNER . Pat: How about a walk, too? You: Good idea. Let's take a walk around the block later.

Table 2. Interview questions and each participant's personally relevant responses.

Interview questions	Participants' answers			
	AMBDE	PIESH	SMIDA	STEOM
Script topic: Restaurant–Cooking				
What is a restaurant you go to often? What is a dish you might order there?	The chicken	Chicken	Burritos	Shrimp fried rice
What is another dish you might order there?	Patty melt	Gnocchi	Shrimp with rice	Sweet and sour chicken
What is your favorite cold beverage to eat with a meal? Examples: <i>iced tea, soda/pop, juice, water</i>	Iced tea	Have iced tea	Juice	Mountain Dew
What is a dessert you like?	Ice cream sundae	Cheesecake	Crème brûlée	Ice cream
Script topic: Catching up–Making plans	ABEJO	CAVCH	SMISC	WELED
What is an INDOOR activity you enjoy doing with a friend? Examples: <i>TV, movies, bowling</i>	Go to the movies	Go shopping	Watch TV	Talk [in the] afternoon
What is an OUTDOOR activity you enjoy doing with a friend? Examples: <i>ball game, golf, walk, boating, swimming, beach</i>	For a long walk	Walk	Boating on the Fox River	Park Grill
What is the name of someone important to YOU that someone might ask about? Examples: <i>family, pet, relative, friend</i>	Cindy's	Lorelei's	Laura's	Mike
What is something that you are trying to do more of? Examples: <i>exercising, eating right, getting out more, reading more</i>	To exercise more	I try to get out a lot	Getting out a lot	Walking and losing weight

scripts. In the process of making these adjustments to the individual participant's scripts (i.e., by modifying number of words, sentences, syllables, verbs, morphemes, words per sentence, syllables, and morphemes per word, and the Flesch Reading Ease Score), some of the shared items were removed. This affected only the generic items. For example, in STEOM and SMIDA's Restaurant–Cooking script pairs, “forty-five minutes” and “dinner” were excluded. Therefore, although all participants in our study had four or more personally relevant items in their scripts, the number of generic items shared across script pairs ranged from two to four.

Table 1 lists each participant and the script pair he or she received. As noted above, each participant's trained and generalization scripts shared identical items. Table 3 compares personally relevant and generic items in each script for number of items, number of words, and complexity

as measured by words per item, syllables per word, and morphemes per word. Overall, when items in all scripts were combined, personally relevant items had significantly more words per item than generic items (2.5 as compared to 1.8, $p < .05$), whereas generic items had significantly more morphemes per word (1.6 as compared to 1.3, $p < .05$).

Treatment Intervention

The trained script was practiced at home for 90 min a day, 6 days a week for 3 weeks on a loaned laptop using a newer version of the AphasiaScripts software called AphasiaRx. The 90 min of daily practice were usually divided into three 30-min practice sessions. The software uses an anthropomorphically accurate “digital” therapist capable of visually modeling speech (see Figure 2). The digital

Table 3. Number and complexity of targeted items presented to each participant.

Participant	Personally relevant					Generic				
	No. of items	No. of words	Complexity measures			No. of items	No. of words	Complexity measures		
			Words per item	Syllables per word	Morphemes per word			Words per item	Syllables per word	Morphemes per word
ABEJO	4	12	3.0	1.3	1.2	3	8	2.7	1.1	1.3
AMBDE	4	9	2.3	1.3	1.1	4	6	1.5	2.0	2.0
CAVCH	4	11	2.8	1.3	1.2	3	5	1.7	1.6	1.6
PIESH	4	6	1.5	1.5	1.3	4	6	1.5	2.0	2.0
SMIDA	4	7	1.8	1.4	1.1	2	3	1.5	1.7	1.7
SMISC	3	12	4.0	1.3	1.2	3	8	2.7	1.1	1.1
STEOM	4	11	2.8	1.3	1.1	4	6	1.5	1.3	1.5
WELED	5	8	1.6	1.3	1.3	3	5	1.7	1.6	1.6
<i>M</i>	4.0	9.5	2.5	1.3	1.2	3.3	5.9	1.8	1.6	1.6

Note. Words per items were significantly higher for personally relevant than for generic items; morphemes per word were significantly higher for generic than for personally relevant items ($p < .05$).

Figure 2. Screenshots of personalized script training (a) and (b) and oral reading probes (c). (a) At the beginning of each treatment session, the participant listens to the entire script read aloud by the digital therapist while the words are highlighted on the screen. (b) The participant practices each sentence by listening and pointing to each word as it is read aloud and highlighted, by repeated choral reading, and by independent reading. (c) During probes, the digital therapist says her sentence and the participant with aphasia reads the response aloud independently without cues.

therapist guides the PWA through the three parts of the program that include: (a) listening silently to the whole script; (b) repeatedly practicing each turn of the script first with choral reading in unison with the digital therapist and then independently; and (c) practicing the entire script while taking turns with the digital therapist as would occur in a real conversation. During each part of the program, the sentences of the script are written on the screen. The digital therapist “speaks” the words with mouth movements similar to that of a real person, and the words are highlighted as they are spoken. General details about the AphasiaScripts software, as well as specific information about its delivery in this study, have been described elsewhere (Cherney et al., 2008, 2014). Delivering the treatment via computer ensured fidelity of treatment administration within and across participants. The computer program captured log-on and log-off times and every keystroke made by the participant during treatment so that compliance with the intensive treatment regimen could be monitored. The logs were downloaded and checked during a weekly visit to the clinic. If participants did not meet the required 540 min of practice for the week, they were instructed to add the missing minutes to the following week’s practice schedule. At the end of the 3 weeks, all participants had practiced the required total amount of time. The weekly visits also allowed a research SLP to observe a 30-min practice session to ensure that the participant was correctly attempting each part of the program.

Data Collection and Analysis

Probes

Trained and generalization scripts were probed sequentially on 3 different days at baseline, and on 1 day at the end of the 3-week treatment. Trained scripts were also probed twice weekly during the treatment phase. Similar to the treatment sessions, probes were administered by the computer. Treatment and probe sessions were set up by the research SLP on predetermined days by a calendar function integrated into the software program. When probe and treatment sessions were scheduled for the same day, this

function ensured that the probes always preceded practice. Probes were elicited with instructions from the digital therapist to “read the sentences aloud” and “press the space bar as soon as you are done.” Each script turn, consisting of the digital therapist’s line and the PWA’s response, appeared on the screen, as illustrated in Figure 2. After the digital therapist spoke her turn, the PWA read aloud his or her part without any cues from the digital therapist. The participant then pressed the space bar to go to the next turn. High-quality audio recordings of the probes were captured by the computer program. If any problem with the recording function was detected prior to or during the probe sessions, the probes were not initiated or they were stopped until the problem (e.g., turning on the microphone) could be fixed.

The primary outcome measure was the accuracy of production of the specific personally relevant and generic words and phrases obtained during probe sessions. The .wav format files were accessed, and from the recordings, each word of the probe sentences was scored on the Naming and Oral Reading for Language in Aphasia (NORLA-6) scale (Gingrich, Hurwitz, Lee, Carpenter, & Cherney, 2013). The scale ranges from 0 (no response) and 1 (unintelligible or unrelated response) to 4 (accurate but delayed or self-corrected response) and 5 (accurate and immediate response). Semantic or phonologic paraphasias are scored as 2, whereas appropriate and intelligible responses with minor errors such as the omission of a grammatical morpheme are scored as 3. A multidimensional scoring system such as the NORLA-6 is more sensitive to identifying potential change than a binary scoring system that only captures correct and incorrect responses. Percent accuracy was the NORLA-6 score of each item as a percent of the maximum accuracy of that item. Percent accuracy was determined for each item separately, whether it was a single- or multiword item. For example, if a single-word item were “rice,” then the maximum NORLA-6 score would be 5. If the PWA’s production received a score of 3, the percent accuracy would be 60% (3/5). If a multiword item were “fried rice,” the maximum NORLA-6 score would be 10 (i.e., 5 + 5). If the PWA’s production received a score of 1 for “fried” and 4 for “rice,” the percent accuracy would be 50% (i.e., (1+4)/10).

Interrater Reliability

As previously reported, point-to-point interrater reliability (Pearson's r) between two trained SLPs for scoring the script probes was .94 (Cherney et al., 2014). For the purposes of this study, the scores for each generic and personally relevant item were reviewed by a third trained SLP and reliability of 97.9% was achieved. Discrepancies were resolved by consensus. Examples of the few cases where there was a discrepancy included scoring a word as a 5 instead of a 4 on the basis of whether or not there was response latency, and scoring a word as a 1 instead of a 2 on the basis of whether or not a phonologic error changed a response to an unrelated response.

Analyses

Analyses were conducted in Excel. One-tailed, paired t tests were used to determine whether there were significant gains from baseline to posttreatment accuracy for personally relevant and generic items, in both the trained and generalization scripts. For both the trained and generalization scripts, a paired, one-tailed t test was calculated to explore whether the gain from baseline to posttreatment on personally relevant items was significantly different from the gain from baseline to posttreatment on generic items. In addition, the effect size (ES) of personally relevant gains over generic gains, at a confidence level of .95, was computed using Cohen's d (Cohen, 1988). Gains from baseline to posttreatment were calculated separately for personally relevant and generic items in both trained and generalization scripts. Cohen's d was calculated by subtracting the mean gain for the generic items from the mean gain for the personally relevant items, and then dividing by their pooled standard deviation. Effect sizes were interpreted using benchmarks for interpreting Cohen's d , whereby 0.2 equates to a small effect, 0.5 equates to a medium effect, and 0.8 equates to large effects (Cohen, 1988).

Results

Question 1

Are personally relevant words and phrases acquired more successfully than generic words and phrases in a trained conversational script? Figure 3a shows the percent accuracy at baseline and posttreatment in the trained scripts for both personally relevant and generic items. Before treatment, baseline accuracy was essentially the same for both personally relevant (54.4%) and generic (51.4%) items in the trained scripts. A paired, two-tailed t test indicated no significant difference between the two baselines ($p = .68$).

There was a significant gain in acquisition of both personally relevant and generic words in the trained scripts. Paired, one-tailed t tests indicated a significant increase from baseline to posttreatment for both personally relevant items (from 54.4 to 84.2 mean percent accuracy, with a mean gain of 29.8, $p < .001$) and generic items (from 51.4 to 73.0 mean percent accuracy, with a mean gain of 21.6, $p < .005$).

A paired, one-tailed t test that compared the gain made on personally relevant items to the gain made on generic items was not significant ($p = .059$), but the effect size was large (Cohen's $d = 0.9$).

Question 2

Do personally relevant words and phrases generalize more successfully than generic words and phrases to an untrained script about a different topic? Figure 3b shows the percent accuracy at baseline and posttreatment in the generalization scripts for both personally relevant and generic items. Before treatment, baseline accuracy was essentially the same for both personally relevant (55.4%) and generic (50.2%) items in the generalization scripts. Note that these items were identical to those in the trained scripts, but probed in the context of the generalization script. A paired, two-tailed t test indicated no significant difference between the two baselines ($p = .39$).

For the generalization scripts, paired, one-tailed t tests indicated a significant increase from baseline to posttreatment for personally relevant items (from 55.4 to 67.9 mean percent accuracy, with a mean gain of 12.6, $p < .05$). However, the increase from baseline to posttreatment for generic items (from 50.2 to 58.4 mean percent accuracy, with a mean gain of 8.2, $p = .067$) was not significant.

A paired, one-tailed t test between the gain made on personally relevant items (12.6%) versus the gain made on generic items (8.2%) was not significant ($p = .32$) and the effect size was small (Cohen's $d = 0.25$).

Discussion

In this preliminary study, eight individuals with chronic aphasia received intensive computer-based script training for 3 weeks. Embedded in both the trained script and the untrained generalization script were four personally relevant words and phrases that each participant had selected themselves, and two to four generic words and phrases that were selected by the investigators. Training resulted in significant gains on the trained scripts from baseline to immediately posttreatment in accuracy of production for both personally relevant and generic words and phrases. Significant gains were also evident in the untrained generalization script for production of the personally relevant words and phrases, but not the generic words and phrases. When the gain in performance accuracy for the personally relevant items was compared to the gain in performance for the generic items, clinically meaningful but not statistically significant differences were shown for the trained scripts.

The finding of significant improvement from baseline to posttreatment during training lends support to the growing evidence for script training. For acquisition, although the difference in the gain scores for personally relevant items as compared to generic items was not significant, the large effect size indicates that there may be an advantage of using personally relevant items during script training. Of note, baseline productions of personally relevant and generic

Figure 3. Gains in accuracy from baseline to posttreatment in personally relevant versus generic items. (a) Acquisition in a trained script. Personal and generic items showed a significant gain from baseline to posttreatment. A large effect size (Cohen's $d = 0.90$) indicated a preference for acquisition of personalized versus generic items, but this was not significant ($p = .059$). (b) Generalization to an untrained script. Personal, but not generic, items showed a small but significant gain from baseline to posttreatment. A small and nonsignificant effect size (Cohen's $d = 0.25$) indicated little difference between personalized versus generic items in the generalization script.

items were not significantly different. Therefore, it is not the case that participants had better performance on personally relevant items prior to treatment due to preference or familiarity, for example. It appears that they more effectively produced the personally relevant items as a result of the script training. This finding is consistent with literature that indicates personalization is critical for engagement in learning (Cordova & Lepper, 1996).

The fact that the difference between gains in personally relevant versus generic items for the trained scripts was not significant (but did reach a p value of .059) may likely reflect the small sample size. Included in this reanalysis of data from a previous study were only eight participants and only six to eight items per participant. Nevertheless, results suggest that further investigation of the impact of personalization of specific words and phrases on script acquisition may be warranted.

For the generalization scripts, the gains made from baseline to posttreatment were smaller than the gains made in the trained script. As an example, CAVCH practiced a personalized phrase “go shopping” in the sentence “Remember how we used to **go shopping** all the time?” Production of this phrase improved from a baseline accuracy of 40% to a posttreatment accuracy of 100% in the trained script, but performance on this item did not improve in the untrained generalization script (which included the sentence “If it’s too cold, we can **go shopping**”). AMBDE practiced a personalized phrase “**the chicken**” in the sentence “I’m fine, thanks. What’s **the chicken** special tonight?” Production of this phrase went from a baseline accuracy of 47% to a posttreatment accuracy of 100% in the trained script but only increased from a baseline accuracy of 57% to a

posttreatment accuracy of 70% in the untrained generalization script (which included the sentence “I can heat up **the chicken** from IHOP. It’s a great place.”). Although small, the gains for the personally relevant items were significant in the generalization script, whereas the gains from baseline to posttreatment for the generic items were not significant. Because one of the main purposes of script training is to acquire words and phrases that will generalize to different linguistic and situational contexts, the finding of significant gains with personally relevant items in an untrained generalization script is noteworthy. This finding is also consistent with Goldberg et al.’s (2012) report of generalization of script-related words to an untrained context.

It is also possible that participants may require more practice on items before a significant advantage of personally relevant items over generic items is evident on statistical testing. For this study, scripts were fairly lengthy (10 turns) and although participants accomplished the required training time, the personalized and generic items within the scripts may not have been practiced in the same way or for the same amount of time. Without monitoring every treatment session in its entirety, we cannot be certain that participants were fully compliant with each component of the script training protocol. Successful acquisition and generalization of a word or phrase may require consolidation of many more instances of practice of the word/phrase than were experienced in this study. The theory of overlearning may be relevant in this regard, although the quantity of exposures that would define overlearning is unknown in script learning (Driskell, Willis, & Cooper, 1992). It may also be necessary to train the same item in a variety of contexts (as demonstrated by Youmans et al., 2005, but

not Goldberg et al., 2012) before it can truly consolidate in memory as an independent, stable item that the PWA can consistently generalize to other contexts.

Although there was a trend of personally relevant items being better acquired than generic items and a large effect size in terms of acquisition, examination of individual participant's performance indicated that some participants showed an opposite pattern. A cursory review of the characteristics of these participants did not reveal any consistent variables such as severity or type of aphasia influencing the gain. Future research with a larger sample might reveal characteristics that are associated with better performance on personally relevant versus generic items.

As described in the Methods, the trained and generalization scripts were modified in our attempt to match the overall level of difficulty of the scripts to the severity of participants' language impairments. As a result, there were differences in the personally relevant and generic items across participants' scripts. The personally relevant items contained more words per item than the generic items, which would lead one to think they would be more difficult to acquire. On the other hand, the generic items were more complex than the personally relevant items, as measured by the number of syllables and morphemes in each word or phrase. It is not clear how these variables within the stimuli affected acquisition and generalization and the relevant weight that they may have contributed to learning. In addition, the stimuli were not controlled for factors such as part of speech, imageability, and frequency, which may influence script acquisition and generalization in addition to personalization. It is important to be aware of these shortcomings as we interpret the results of this study.

Our findings suggest that personalization is an important consideration in script training for aphasia. However, there are a variety of ways to personalize scripts and different degrees to which a script can be personalized. Several studies, for example, describe procedures in which participants selected topics for script training on the basis of situations in their lives in which they desired greater communicative competence, after which scripts were co-constructed by the clinician/researcher and PWA (Goldberg et al., 2012; Lee & Cherney, 2008; Youmans et al., 2011). Other studies, including this one, have incorporated the use of template-based scripts in which personally relevant words are inserted into a pre-established script on a given topic (Cherney et al., 2014; Youmans et al., 2005). Still others have carried out script training of conversational dialogues without personalizing the content for participants to any degree (Bilda, 2011; Nobis-Bosch et al., 2011). Diversity in the way scripts are personalized as well as differences in how performance is measured may account for differences in results across studies with regard to acquisition and generalization. Nevertheless, given the current evidence supporting script training in aphasia, we would anticipate its widespread use in clinical practice. However, script development is a very time-consuming process that busy clinicians may avoid given the time pressures of the clinic. In this study, we used previously developed templates with slots to fill in words or phrases that

were personally relevant to the participant and selected by them. Although the level of personalization may be relatively constrained in a predetermined template script, the significant improvements from baseline to posttreatment that were made in personalized and generic items in the trained scripts, and personalized items in the untrained scripts, support the use of templates in script training. Templates that can be easily modified to an individual participant's specific interests and preferences may be a simple and time-effective way to facilitate the use of personalized scripting within clinical practice.

There are several methodological limitations associated with this study, some of which have occurred because the intent of the original study from which these data were derived was not to examine personal relevance, but to examine the impact of cuing in script training. These limitations have been previously discussed above. First, results and interpretation are limited by the small number of participants and the exploratory nature of the statistical analyses used. In addition, there were too few overlapping personally relevant and generic items in the trained and generalization scripts to detect differences that may have been statistically significant. Although we controlled for the number of words in the personalized versus generic items in the overall scripts, there were other differences in the individual items that were not controlled, including the number of words per item and morphemes per word. Nevertheless, data from this preliminary study highlights the importance of word-level personalization in script training. It may serve as a precursor to a larger controlled study that utilizes more stringent statistical analyses to further assess the impact of personalization.

Acknowledgments

This study was supported by the National Institute on Deafness and Other Communication Disorders, Award No. 1R01 DC011754 (to LRC and SvV). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding organizations. Endorsement by the Federal Government should not be assumed. AphasiaScripts® is registered to the Rehabilitation Institute of Chicago, Chicago, IL. None of the authors have a financial interest in the software product. The authors extend their thanks to Julia Carpenter, Rachel Hitch, and Rosalind Hurwitz who collected and scored data; to Anita Halper and Audrey Holland who served as consultants on the grant; and to Nattawut Ngampatipatpong who modified the AphasiaScripts software for the purposes of this study.

References

- Bilda, K. (2011). Video-based conversational script training for aphasia: A therapy study. *Aphasiology*, 25, 191–201.
- Cherney, L. R., Halper, A. S., Holland, A. L., & Cole, R. (2008). Computerized script training for aphasia: Preliminary results. *American Journal of Speech-Language Pathology*, 17, 19–34.
- Cherney, L. R., Halper, A. S., & Kaye, R. C. (2011). Computer-based script training for aphasia: Emerging themes from

- post-treatment interviews. *Journal of Communication Disorders*, 44, 493–501.
- Cherney, L. R., Kaye, R. C., & van Vuuren, S.** (2014). Acquisition and maintenance of scripts in aphasia: A comparison of two cuing conditions. *American Journal of Speech-Language Pathology*, 23, S343–S360.
- Cohen, J.** (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
- Cohen, M. J., & Sloan, D. L.** (2007). *Visual supports for people with autism: A guide for parents and professionals*. Bethesda, MD: Woodbine House.
- Cordova, D. I., & Lepper, M. R.** (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, 88, 715–730.
- Driskell, J. E., Willis, R. P., & Cooper, C.** (1992). Effect of over-learning on retention. *Journal of Applied Psychology*, 77, 615–622.
- Flesch, R. F.** (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221–223.
- Gingrich, L., Hurwitz, R., Lee, J., Carpenter, J., & Cherney, L. R.** (2013, November). *Quantifying naming and oral reading performance in aphasia: The NORLA-6 scale*. Presented at the Annual Convention of the American Speech-Language-Hearing Association, Chicago, IL.
- Goldberg, S., Haley, K. L., & Jacks, A.** (2012). Script training and generalization for people with aphasia. *American Journal of Speech-Language Pathology*, 21, 222–238.
- Haley, K. L., Womack, J. L., Helm-Estabrooks, N., Caignon, D., & McCulloch, K.** (2010). *Life interests and values cards*. Chapel Hill, NC: University of North Carolina.
- Holland, A. L., Halper, A. S., & Cherney, L. R.** (2010). Tell me your story: Analysis of script topics selected by persons with aphasia. *American Journal of Speech-Language Pathology*, 19, 198–203.
- Kertesz, A.** (2007). *Western Aphasia Battery—Revised*. San Antonio, TX: PsychCorp.
- Lee, J. B., & Cherney, L. R.** (2008). The changing face of aphasia therapy. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, 18, 15–23.
- Lee, J. B., Kaye, R. C., & Cherney, L. R.** (2009). Conversational script performance in adults with non-fluent aphasia: Treatment intensity and aphasia severity. *Aphasiology*, 23, 885–897.
- Logan, G. D.** (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492–527.
- LPAA Project Group.** (2000). Life participation approach to aphasia: A statement of values for the future. *ASHA Leader*, 5(3), 4–6.
- Manheim, L. M., Halper, A. S., & Cherney, L. R.** (2009). Patient-reported changes in communication after computer-based script training for aphasia. *Archives of Physical Medicine and Rehabilitation*, 90, 623–627.
- McKelvey, M. L., Hux, K., Dietz, A., & Beukelman, D. R.** (2010). Impact of personal relevance and contextualization on word-picture matching by people with aphasia. *American Journal of Speech-Language Pathology*, 19, 22–33.
- Nobis-Bosch, R., Springer, L., Radermacher, I., & Huber, W.** (2011). Supervised home training of dialogue skills in chronic aphasia: A randomized parallel group study. *Journal of Speech, Language, and Hearing Research*, 54, 1118–1136.
- Simmons-Mackie, N.** (2008). Social approaches to aphasia intervention. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (5th ed., pp. 290–318). Philadelphia, PA: Lippincott Williams & Wilkins.
- Wallace, G. L., & Canter, G. J.** (1985). Effects of personally relevant language materials on the performance of severely aphasic individuals. *Journal of Speech and Hearing Disorders*, 50, 385–390.
- Youmans, G., Holland, A., Munoz, M., & Bourgeois, M.** (2005). Script training and automaticity in two individuals with aphasia. *Aphasiology*, 19, 435–450.
- Youmans, G., Youmans, S. R., & Hancock, A. B.** (2011). Script training treatment for adults with apraxia of speech. *American Journal of Speech-Language Pathology*, 20, 23–37.
- van Vuuren, S., & Cherney, L. R.** (2014). A virtual therapist for speech and language therapy. In T. Bickmore, S. Marsella, & C. Sidner (Eds.), *Intelligent Virtual Agents 14th International Conference, IVA 2014, Boston, MA, USA, Lecture Notes in Artificial Intelligence 8637* (pp. 438–448). Switzerland: Springer.