International Journal of Health Technology and Innovation
An Official Journal of **Kalam Institute of Health Technology** | ISSN - **2583-8547**

RESEARCH ARTICLE

The Maze Task as a Sentence Formulation Treatment for Aphasia: A Feasibility Study

Erin L. O'Bryan[1*], Huabo Lu[2], Imran Musaji[1], Alexandra Olmstead[3]

[1]*Wichita State University, Department of Communication Sciences and Disorders Wichita, KS, USA.*
[2]*Wichita State University, School of Computing, Wichita, KS, USA.*
[3]*Department of Physics and Astronomy, The University of New Mexico, Yale Blvd NE, Albuquerque, USA.*

Received: 15th August, 2024; Revised: 12th September, 2024; Accepted: 26th November, 2024; Available Online: 05th December, 2024

ABSTRACT

This study aimed to examine the feasibility and acceptability of the maze task, a tool commonly used to investigate the time course of sentence processing, as a treatment for sentence formulation in people with aphasia. Five participants with chronic aphasia completed seven sessions of the computerized maze task. A quasi-experimental pretest-posttest design was used to examine whether participants improved at the task with repeated practice. In each session, participants made three independent attempts per sentence, followed by an attempt with verbal and visual cues provided by a clinician. Task accuracy and response time data were collected to measure improvement. At the end of each session, the participants completed a subjective experience questionnaire to examine the acceptability of performing the task. The results indicated that all participants demonstrated improvement in both accuracy and response time. These improvements were statistically significant, and the large effect sizes indicated potential therapeutic significance. The participants rated the task with high acceptability scores across all seven sessions. The use of the maze task to improve sentence formulation appeared to be feasible and acceptable. Future research is needed to examine the characteristics of people with aphasia who would benefit from this task and the additional effects of using it.

Keywords: Aphasia, Sentence processing, Computer-based, Syntax, Reading.

International Journal of Health Technology and Innovation (2024)

How to cite this article: O'Bryan EL, Lu H, Musaji I, Olmstead A. The Maze Task as a Sentence Formulation Treatment for Aphasia: A Feasibility Study. International Journal of Health Technology and Innovation. 2024;3(3):4-13.

Doi: 10.60142/ijhti.v3i03.02

Source of support: This work was supported by a University Research/Creative Projects Award from Wichita State University, grant #200182.

Conflict of interest: None

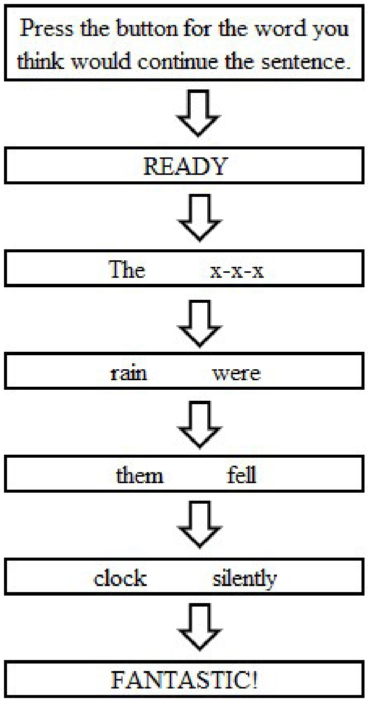
INTRODUCTION

Many people with aphasia (PWA) following stroke or brain injury experience difficulties with sentence processing, including both sentence production and comprehension. Sentence processing impairments are often categorized into three groups: difficulties with structure, difficulties with verb processing, and difficulties with grammatical morphology.[1-3] The current paper focuses on difficulties with structure, specifically word order and grammatical sequencing. Treatments that address these challenges are typically based on linguistic and sentence-processing models of how adults produce and comprehend sentences.

The mental processes involved in sentence production and comprehension have been studied in adults with typical language abilities for more than fifty years.[4,5] A variety of psycholinguistic tasks have been developed to test hypotheses about knowledge representations and the timing of decision-making as sentences are processed or formulated.[6] These psycholinguistic tasks can serve as useful tools for examining sentence processing in PWA.[7,8] Some of these tasks may be useful as treatment tools.

The 'maze task' is a computer-based language processing tool originally designed by Kenneth Forster and colleagues for the purpose of investigating questions about lexical processing and sentence processing.[9-11] The task requires the participant to select the correct continuation word from sequences of word pairs to build a grammatical sentence. First, the participant sees the first word of the sentence on the screen. Next, they are presented with two words, from which they must select the next word that can continue the sentence, as shown in Fig 1. The participant is asked to press the left or right button to indicate which of the two words can continue the sentence. Each time the correct word is selected, two more words appear, encouraging the participant to construct a grammatical sentence one word at a time. Forster and colleagues refer to this type of maze task as a 'G-maze' or 'grammaticality maze' because the incorrect choice is a real word, although it is one that would result in an ungrammatical sequence. The 'G-maze'

**Author for Correspondence: erin.obryan@wichita.edu*



Example sequence seen by participant. Each rectangle represents the text seen on the computer screen. The participant's response moves the sequence to the next screen. In this example, the correct responses would be LEFT (The), LEFT (rain), RIGHT (fell), RIGHT (silently).

Fig. 1: Example screen sequence in the maze task

is an alternative to an ‘L-maze’ or ‘lexicality maze’, in which the incorrect choice is a nonword.[9]

Research on this version of the maze task in adults with typical language abilities has shown that each word must be integrated with the previous words in order to select the correct next word and that the time it takes to make each choice is sensitive to the complexity of the grammatical structure, predicate event structure, transitivity, and thematic role assignment.[10,12-15] Longer response times typically indicate processing difficulty or reanalysis of the sentence fragment up to that point.[10]

Sentence production models typically include levels or stages in which the earlier ones involve formulating the sentence prior to phonetic encoding and then verbal articulation of the sentence.[5,2] The maze task requires an essential component of sentence production, the formulation of the word sequence, but it does not require the entire sentence production process.

The maze task has been previously adapted for use as a language training tool for second-language learners of Spanish.[16] Enkin modified the original maze task described by Forster and colleagues[10] by providing immediate feedback and multiple attempts to complete the sentence. Enkin found that learners who used the maze task for training improved in sentence production tasks and grammaticality judgment tasks. Additionally, these second language learners rated the maze task as ‘fun’ and ‘helpful.’ It is possible that it might provide a similarly effective and enjoyable task for PWA targeting sentence production.

It is reasonable to expect that PWA will process sentences differently than neurotypical adults and that there will be variation depending on the characteristics and severity of each PWA’s impairment. This task allows researchers to test hypotheses about the strategies that a PWA uses while reading and processing sentences.

In addition to assessment, the maze task appears well suited to adaptation as a treatment task, particularly when targeting accurate sequencing of words by grammatical category. Specifically, the maze task decreases the cognitive burden of lexical retrieval by providing a field of two words to choose from, allowing participants to focus primarily on the sequence of the words. The maze task can also be performed quickly, allowing multiple sentence exemplars to be presented in one session. Prior research on the performance of the maze task by adults with typical language and the use of the maze task as a training tool for second language learners raises the question of whether people with aphasia could use the maze task to improve the sentence formulation component of sentence production.

Comparison to other Syntax Treatments

Other treatment tasks have been developed to target sentence production and comprehension. These treatments differ from the maze task in mode of presentation, mode of participant response, and type of scaffolding. For example, Treatment of Underlying Forms (TUF) is a syntax treatment supported by considerable evidence that targets improved production of grammatical sentences.[17] TUF aims to improve understanding of the thematic role of each word in a sentence, leading to improved understanding of the syntactic structure. The ‘thematic role’ refers to the role that each noun phrase plays in the sentence in relation to its verb. For example, in the sentence ‘the girl saw the dog,’ ‘the girl’ takes the thematic role of agent or ‘do-er’ of the verb ‘saw,’ and ‘the dog’ takes the role of a theme (sometimes called the ‘undergoer’ or patient). In the TUF treatment protocol, the participant starts with the predicate and its arguments in the canonical order: ‘agent verb theme.’ Participants then move the words to their spoken surface positions, following the steps of the syntactic derivation.[17] The maze task differs from TUF in the mode of sentence presentation: participants practice formulating the spoken sentence structure word-by-word incrementally, starting with the first word of the sentence.

The Helm Elicited Language Program for Syntax Stimulation (HELPSS)[18] and the more recent version of HELPSS, Sentence Production Program for Aphasia (SPPA),[19] target the production of grammatical sequences by training whole sentences. Both the maze task and SPPA involve practicing sentences with all words in their target order. Both treatment tools also involve practicing many examples of each sentence type in order to increase the neural connections supporting sentence production tasks.[15,20,21] Key differences between SPPA and the maze task include modes of participant response and scaffolding. In the maze task, the participant selects the words in the sentence by pressing buttons, in contrast with SPPA, in which the participant says the words aloud. Regarding scaffolding, the maze task scaffolds the sentence formulation process by presenting two words to choose from. In SPPA, the process is scaffolded by hearing a recent model of the full spoken target sentence.

Similar to the maze task, SentenceShaper™ is a software program that facilitates sentence formulation. SentenceShaper™ enables a user to record individual words and phrases and then reorder and combine them to build sentences or multi-sentence narratives.[22] SentenceShaper™ differs from the maze task in that the user must retrieve and provide the lexical items from an open field. The maze task only requires selecting one word from two choices. Another difference is that SentenceShaper™ builds an auditory representation of a sentence, while the maze task builds a text version of a sentence. SentenceShaper™ may be used either as a communication aid or a therapy tool. To date, the maze task has not been adapted as a communication aid.

The aim of the current study was to determine the feasibility and acceptability of the maze task as a potential treatment activity for PWA. Feasibility was examined by investigating whether participants improved at the maze task with repeated practice. Acceptability was examined by asking participants to give feedback about their experience of doing the task. This pre-efficacy study focused on three research questions:

- Did participants' accuracy on formulating grammatical sequences increase from the first session to the last session?
- Did participants' response times for making correct word choices decrease from the first treatment session to the last treatment session?
- Did participants rate the treatment task as acceptable?

MATERIALS AND METHODS

Participants

Volunteers were recruited using a flyer presented in an aphasia group at a university speech-language and hearing clinic and distributed to clinical supervisors at the same clinic. Participants met the inclusion criteria if they were diagnosed as having aphasia by the *Western Aphasia Battery-Revised (WAB-R)*[23] and had normal or corrected-to-normal vision. Participants were excluded if the onset of aphasia was less than six months prior to the study. Twelve participants began the study. Five participants were unable to complete the treatment sessions before the clinic shut down in response to COVID-19. Two other participants dropped out of the study after five treatment sessions, one for medical reasons and the other because of scheduling issues. The remaining five participants completed all seven treatment sessions.

Participants included two males and three females, ages 30-69 years (M = 57.4 years). All participants were in the chronic phase with time post-onset of aphasia ranging from 3 to 8 years (M = 5.4 years). All participants were monolingual English speakers who reported normal or corrected-to-normal vision and hearing and no history of language or learning disorders other than acquired aphasia. All participants were right-handed prior to onset, and all reported having no left-handed family members. Four of the five participants had

Table 1: Participant demographics

Participant	*Sex*	*Age*	*Years post-onset*	*Etiology*	*AQ*[a]	*Sub-type*[b]	*Lesion location*
K01	F	53	7	CVA[c]	67.2	Broca's	Large left MCA[d] infarct
E04	F	68	5	CVA	94	Anomic	Ischemia left parietal lobe, minimal intraparenchymal hemorrhage
A06	M	30	8	TBI[e]	80.6	Anomic	Subarachnoid hemorrhage left fronto-temporo-parietal craniectomy
C07	F	69	4	CVA	48.5	Broca's	Ischemia left MCA, PCA[f], and superior cerebellar artery, petechial hemorrhage left basal ganglia
M08	M	67	3	CVA	25.4	Broca's	Left MCA infarct involving left temporal and posterior frontal lobes

Note. [a]AQ is used to abbreviate aphasia quotient from the *Western Aphasia Battery-Revised (WAB-R)* out of 100. [b]Sub-type is used to abbreviate the aphasia sub-type indicated by the *WAB-R*. [c]CVA = cerebrovascular accident. [d]MCA = middle cerebral artery. [e]TBI = traumatic brain injury. [f]PCA = posterior cerebral artery.

aphasia following a cerebrovascular accident (CVA), and one participant had aphasia following a traumatic brain injury with a left hemisphere lesion caused by a motor vehicle accident.

All five participants continued to participate in the weekly aphasia support group throughout the study. Four of the five participants continued to receive individual speech therapy once per week during the study in accordance with pre-existing treatment plans. However, treatment goals during these independent sessions were confirmed to not include any focus on sentence-level production, sentence reading, or formulation of grammatical sequences. Treatment goals focused instead on pictured object or object naming, speech production of individual words, and/or augmentative and alternative communication. The fifth participant received no individual speech therapy during the study. Table 1 provides descriptions of the participants.

Participants signed consent forms after the researchers provided written and spoken descriptions of the study and answered clarification questions. The university Institutional Review Board approved the study under IRB #s 4596 and 4889. The *Western Aphasia Battery-Revised (WAB-R;* Kertesz, 2007)[23] was administered the week prior to beginning treatment. Three of the five participants were classified as Broca's type according to the *WAB-R*, and the other two participants were classified as anomic type. *WAB-R* aphasia quotients (AQs) ranged from 25.4 (severe) to 94 (mild).

Materials

Equipment

Participants sat in front of a laptop computer with a keyboard modified so that only three keys were visible: the space bar marked with red tape, the left shift key, and the right shift key. The other keys were covered by a thin mat. All five participants chose to press the buttons with their non-dominant left hand because all participants had some loss of function in the dominant right hand due to stroke or brain injury. The task was presented using DMDX experimental software.[24] The DMDX software recorded response time and accuracy data for every key press. Response time was measured as the number of milliseconds, to at least one-hundredth of a millisecond, from the time that the word choices appeared on the screen until the left or right shift key was pressed.

Treatment task

Forster and colleagues originally described the use of the maze task as a psycholinguistic assessment tool.[10] Investigators made several changes to the task to adapt it as a treatment task for aphasia therapy. As per Enkin's[16] second language learning adaptation, participants were given immediate feedback on word choice accuracy and were provided multiple attempts to complete each sentence. Additional scaffolding was introduced in the current study so that when a participant made three incorrect independent attempts, the clinician provided support through a verbal cue (said the correct word aloud) and a visual cue (pointed to the correct answer and correct key). All cueing was provided by a graduate student clinician supervised by the first author. The task was modified to display text using the New Courier font with a larger font size of 50 points.

At the start of each session, the clinician provided the task instructions verbally with supportive pointing as follows:

In this task, you will build a sentence. You will use three buttons: the red button (pointing to the space bar marked with red tape), the left shift key (pointing), and the right shift key (pointing). You will see two words come up on the screen. You should choose which word makes sense to continue a sentence. You will push the left shift key (pointing) to pick the word on the left side of the screen (pointing) and the right shift key (pointing) to pick the word on the right side of the screen (pointing). We will do the first sentence together. After that, you will try each sentence three times. If you need help after three tries, I will help you by saying the correct word and pointing to the correct word and the correct key.

At the beginning of the task, the participants saw a screen that stated, 'Push the red button (space bar) when you are ready to begin a new sentence.' They were then shown the first word and then prompted to select the correct next word from a field of two. If an ungrammatical continuation word (incorrect response) was selected, then the text on the screen said 'Try again!' and then returned to the beginning of the sentence. If the participant successfully navigated to the end of the sentence with all correct selections, then the text on the screen said 'Fantastic!' As in the original maze task, only the two-word choices were presented on the screen at each step. Previously selected words did not remain visible to the participant as the sentence was formulated. This presentation of the sentence in which words are temporarily available has a working memory load similar to other sentence-processing tasks, such as self-paced reading, which has been used to study comprehension in people with aphasia.[7]

After an ungrammatical continuation word was selected three times, the clinician began cueing the participant. Cueing involved giving the participant a verbal cue (saying the correct selection aloud) and a visual cue (pointing to the correct answer and correct key), allowing the participant to make the correct word selection with the clinician's help.

Participants were first introduced to the treatment task during individual familiarisation sessions. Familiarisation sessions were the same as treatment sessions except that the sentence stimuli were different, consisting of 10 simple sentences, and the clinician provided verbal and visual cues on every trial of the first sentence and as many as requested during the remaining nine sentences. Participants did not read the selected words aloud during the task because during the familiarisation sessions, it was found that verbalization made the task much more difficult. Data from the familiarisation sessions were not included in the analyses. The treatment was then administered one time per week for seven weeks. Each treatment consisted of two 20-minute blocks separated by a 5-minute break halfway through the session. All treatment sessions were conducted in a private room in the university speech-language-hearing clinic.

Target sentences

The same 88 target sentences generated by the first author were presented in the same order during each treatment session. However, the number of sentences that each participant attempted was determined by the number they completed during the treatment time (two twenty-minute blocks). The researchers decided to limit the length of treatment sessions based on time rather than number of sentences completed because of the variability in participants' skill levels.

The sentences ranged in length from five to 13 words, with a mean length of 9.92 words (SD = 1.93 words). The grammatical complexity of the sentences ranged from simple active sentences (e.g., 'Many seals live in Antarctica') to unreduced relative clause sentences (e.g., 'The actress that the writer spotted left in a hurry'). Grammatically complex sentences were interspersed between simple active sentences, with the first grammatically complex sentence occurring as the twelfth sentence in the list. This sentence order was held constant for all treatment sessions. All treatment sessions began with the same sentences regardless of which sentences the participant completed in the previous session. The incorrect word selections (referred to as non-continuation words) were words that would result in an ungrammatical sentence fragment. For example, following the initial word 'The' in Fig 1, 'rain' would be a correct continuation, but 'were' would be ungrammatical. Additional examples are shown in Supplementary Fig 1.

The presentation of the correct word choice on the left versus right side of the screen was randomized; however, the side that the correct word occurred on was always the same for a given word pair. Thus, when a participant attempted a sentence multiple times, the correct choices were presented in the same locations as on previous attempts.

Post-session Subjective Experience Questionnaire

At the end of each treatment session, participants completed a questionnaire to describe the degree to which they found the task fun, helpful, difficult, and stressful (see Supplementary Fig 2 for the full questionnaire). The clinician-administered each question by reading it aloud while pointing to the words. The participants responded by circling a number on a 5-point Likert scale with a happy face and sad face at either end, modeled after the visual scale used in the *Assessment for Living with Aphasia*.[25] For example, for the question: 'Did you find this task to be fun?', the participants marked a number on a continuum labeled 'not fun' (sad face) on one end and 'fun' (happy face) on the other end with the numbers 1 through 5 in between. The clinician read the words at each end of the scale aloud while pointing to help the participants understand the questionnaire.

Statistical Analysis

Three key categories of data were collected, cleaned, and analyzed. The first category of data included each participant's accuracy score, calculated as the percentage of correct word choices out of all independently selected word choices on the first attempt of the sentence during the 40-minute treatment session. These data corresponded to research question 1: did the participant's accuracy improve (increase) from the first treatment session to the last treatment session?

The second category of data included response time measured as the number of milliseconds, to a hundredth of a millisecond, for making a correct and independently selected word choice on the first attempt of a sentence. Response times for incorrect choices and for repeated attempts of the sentence were not included in statistical analyses. These data corresponded to research question 2: did participant response time improve (decrease) from the first treatment session to the last treatment session?

The third category of data included responses to the qualitative experiences of the participants. This included participants' subjective ratings (i.e., fun, helpful, difficult, stressful) of the experience of doing the treatment task. Analysis of this data was primarily descriptive, with means, standard deviations, medians, and ranges calculated for participant ratings on each question. These data corresponded to research question 3: did participants rate the treatment task as acceptable?

After participants completed the maze task, accuracy and response time data were extracted from the DMDX experimental software output files using Python code written by the second and fourth authors. The code used a lookup table to specify the item, sentence position, verb type, and correctness of each response and arranged this information in a data matrix. The matrix was used to prepare the data for statistical analysis.

Data from categories 1 (response accuracy) and 2 (response time) were cleaned and analyzed using IBM SPSS Statistics (Version 27). To better support any statistically significant findings, only analysis of the predetermined (*a priori)* metrics (accuracy, response time, qualitative experience) was conducted, tests of normality were conducted, and effect sizes were calculated and reported. Whenever more than one statistical test was available, the most conservative test appropriate for small sample sizes was selected. For example, the Shapiro-Wilkes test was used to assess normality, and effect sizes were estimated utilizing Cohen's *d* with Hedge's correction. It is important to note that for both dependent variables (response time and response accuracy), every subject demonstrated improvement on both measures, reducing the chance of stochastic errors from averaging or from outliers. Because treatment effects of interest were in one direction, and all individuals showed improvement for both dependent variables, one-tailed tests were indicated.

For data categories 1 (response accuracy) and 2 (response time), both parametric and non-parametric results were reported and compared. According to performed tests of normality (see results section), these data did not appear to violate the assumption of normality so parametric tests were conducted for comparisons (paired sample *t*-tests). However, the reliability of tests of normality decreases with smaller sample sizes. There is a lack of consensus over whether assuming non-normal distribution and conducting non-parametric tests in such situations improves statistical analysis, or instead introduces new sources of error.[26,27] As a compromise, after completing the parametric tests, data analyses were repeated with non-parametric comparisons (related samples Wilcoxon Signed-Rank Tests).

RESULT

When working with small samples, it is important to test assumptions of normality to determine appropriate statistical methods. Our sample size (n = 5) is small so the more conservative Shapiro-Wilk test was performed. The results of this test showed no evidence of non-normality in the distributions of the dependent variable's pretest response time (W = 0.99, p = 0.99), posttest response time (W = 0.94, p = 0.68), pretest accuracy (W = 0.93, p = 0.56) or posttest accuracy (W = 0.83, p = 0.15), and so we retain the null hypothesis of normality at the more conservative $\alpha > 0.1$ threshold. Based on this outcome, the use of a parametric test was supported, and means were used to summarise the variables. Mean task accuracy and response times for the first and last maze task sessions are shown in Table 2.

Accepting the assumption of normality, we compared the means from the first and last sessions using paired-sample *t*-tests. Because treatment effects of interest were in one direction and all participants showed improvement for both dependent variables, one-tailed tests were indicated. A paired samples *t*-test found statistically significant reductions in response time after intervention (M = 2402.08 ms, SD = 1019.97 ms) as compared with before intervention (M = 2685.76 ms, SD = 1087.82 ms; t(4) = -3.37, p = 0.014). On average, response time dropped 284 milliseconds, which corresponded to a large estimated effect size (Cohen's *d*, based on differences, with Hedges correction = 1.36).

A paired samples *t*-test found statistically significant improvements in response accuracy after intervention (M = 84.08%. SD = 17.48%) as compared with before intervention (M = 79.99%, SD = 15.63%; t(4) = 2.41, p = 0.037). On average, accuracy improved by 4.09%, which corresponded to a large estimated effect size (Cohen's *d*, based on differences, with Hedges correction = 0.973).

For the repeated non-parametric analysis, a Wilcoxon signed-rank test determined that there was a statistically significant improvement in median response accuracy after intervention (83.9%) compared with the pretest median response accuracy (93.3%), z = 2.023, p = 0.043. A Wilcoxon signed rank test also determined that there was a statistically significant reduction in median response times (2499.2 ms) compared with the pretest median response times (2736.4 ms), z= -2.023, p = 0.043. The results of both the parametric and non-parametric statistical tests were in agreement.

Subjective Experience Questionnaire Results

Table 3 shows the ranges, medians, means, and standard deviations of the ratings on the subjective experience questionnaire for each question. Overall, the majority of ratings were in the favorable range ('fun,' 'helpful,' 'not difficult'), but examination of the range column reveals some unfavorable ratings such as 'not fun' and 'difficult.' In response to the question, 'Would you perform this task at home on your own time?', the median answer was '5', indicating 'yes.'

DISCUSSION

The purpose of this study was to examine the feasibility and acceptability of using a computerized maze task to help people with aphasia improve at formulating grammatical sentences. The first research question asked whether participants' accuracy would improve at the maze task with repeated practice. The

Table 2: Maze task accuracy and response times

Participant	*% accuracy*		*Mean response time in milliseconds*	
	First Tx	*Last Tx*	*First Tx*	*Last Tx*
K01	83.91%	93.33%	3077.93	2499.18
E04	96.35%	98.10%	2091.40	1782.68
A06	87.64%	94.42%	2736.36	2524.24
C07	77.09%	78.86%	4212.81	3957.38
M08	54.95%	55.70%	1310.30	1246.91

Table 3: Subjective experience questionnaire results from all five participants and all treatment sessions

Question	*Range of responses*	*Median*	*Mean*	*SD*
Fun (5) vs. not fun (1)	2-5	5	4.30	0.98
Helpful (5) vs. not helpful (1)	3-5	4	4.31	0.82
Not difficult (5) vs. difficult (1)	1-5	4	3.75	0.91
Not stressful (5) vs. stressful (1)	3-5	4	4.09	0.84
Would do it at home on my own time (yes=5)	2-5	5	4.13	0.97
Wanted to try again on missed trials (yes=5)	3-5	5	4.24	0.87
More fun than traditional individual therapy (yes=5)	2-5	3	3.54	0.85

results indicated a positive, statistically significant increase in accuracy for all participants. The second research question asked whether participants' response time would improve with repeated practice. The results indicated that participants made correct choices more quickly during the last treatment session compared with the first treatment session. While the observed changes appear small, the effect sizes were relatively large. In sentence processing research, statistically significant shorter response times are widely accepted as an indicator of reduced processing load and greater ease.[6]

The third and final research question asked about the subjective experience of participants. Overall, participants rated the task as fun, helpful, not difficult, and not stressful. Rarely do participants rated a task as 'difficult' and 'not fun.' These ratings could simply be the result of a person having 'better days' or being in a 'better mood' on some days compared to others.

Limitations and Future Directions

An important thing to consider about the participants' improvement on the maze task is that some aspects of the task likely made it easier. As noted in the method section, participants received the same sentences in the same order. This result can be seen as a direct treatment effect since participants showed improvement in the training sentences. We cannot rule out the possibility that participants were memorizing the sentences rather than processing grammatical structure, but it should be noted that memorizing sentences is difficult for most people with aphasia. In future research using the maze task, memorization of target sentences should be prevented definitively by presenting different sentences in each training session and randomizing the order of the sentences.

Participants may have also shown more improvement over subsequent trials because stimuli word pairs were always presented to participants in the same visuospatial field. Although the location of the correct word varied randomly throughout each sentence, these locations remained consistent in subsequent treatment sessions. It is possible that some participants may have improved their performance by memorizing these locations, decreasing the language processing necessary to respond. In future research using the maze task, the location of the correct word should be randomized for every treatment session to prevent participants from memorizing the location of the correct words.

These aspects of the task presentation warrant the addition of important qualifiers to the claim that participants improved at the task: participants improved at the task when the same stimuli were presented in the same visuospatial field over subsequent sessions. However, it is worth considering that making the task easier in this way could be appropriate and beneficial for some participants in their first exposure to the maze task. Two of the participants performed as low as 55% and 77% on the task. When a task is too difficult, participants may become frustrated and discouraged. Perhaps once participants improve at the task with the same presentation, they could move on to a second stage in which novel stimuli are presented with correct words in randomized locations. Then, at this more advanced stage, participants might be encouraged to process the sentences more deeply.

Some aspects of the maze task likely made it more challenging for PWA. In the current study, the previously selected words did not remain visible on the screen as the sentence progressed. For this reason, the task required holding the previously selected words in memory. This memory load is similar to the self-paced reading and moving-window tasks that have been used to investigate sentence comprehension in people with aphasia.[7] The memory load is also similar to that involved in auditory comprehension of sentences where the words are fleeting. Future research should also address whether the task is more efficacious if previously selected words remain on the screen to decrease the memory load required.

An obvious limitation of this study was the quasi-experimental research design. The design could be improved in future studies by including three or more baseline probes prior to treatment, as recommended for single-subject design studies like those described by Beeson and Robey.[28] In order to make the baseline probe phase differ from the treatment phase, participants would receive no feedback or cueing and would attempt each sentence only once. Another way to strengthen the study would be to include a control group of participants who completed the task at the beginning and end of the study with no intervening treatment sessions and no cueing. A between-group design such as this would require more participants.

The usual concern about not having a baseline or control group is that some other factor unrelated to the treatment could have resulted in the change from one pre-treatment data point to one post-treatment data point. Recall that the dependent variables that showed significant changes from pre- to post-treatment were the percentage of accurate choices and the response times for making correct choices. Although it is possible that some other factor such as aphasia group participation or health factors could have influenced the results, these factors seem less likely to result in improved task performance compared with practicing the task itself for seven sessions. For this reason, the researchers suspect that the task improvement was related to practicing the task with the same sentences.

In the current study, participants received the same treatment sentences in every session. A future study might include a list of randomized untrained and novel sentences to assess if improvements are generalized to novel stimuli. This would also address concerns about whether participants memorized the sentences or the location of correct responses. Future research is also warranted to investigate whether the maze task treatment will generalize to improvement in sentence production outside of the maze task, to outcome metrics on standardized tests of language performance, and to connected speech. Standardized tests of language performance would also allow questions about participants' processing strategies to be investigated. For example, the *Western Aphasia Battery-Revised*[23] part two reading subtest or the *Reading*

Comprehension Battery for Aphasia-2[29] could illuminate specific reading strategies impacting participant performance.

Statistical analysis of small intervention studies can be problematic. A common refrain in the 'future studies' section of aphasiology literature is that larger sample sizes should be utilized in the future. However, to date, the majority of published aphasia treatment studies have small sample sizes[30] or are single-subject design studies.[28] It is true that small sample sizes reduce the power of statistical analyses and may introduce increased opportunities for bias and error. However, when dealing with novel therapeutic approaches with a clinically significant but relatively rare population, there are strong ethical and logistical arguments justifying the use of small sample sizes.

In the current study, variability in aphasia severity among participants may have increased the variance in the results. For example, participant M08 exhibited the most severe aphasia resulting from a large left MCA infarct affecting the left temporal and posterior frontal lobe, which may be related to the finding that his accuracy improved only 0.75%, and his response time decreased only an average of 63 milliseconds from the beginning to the end of treatment. In contrast, participant E04 exhibited the least severe aphasia resulting from an infarct in the left parietal lobe, which may be related to the finding that her accuracy improved by 1.75%, and her response time decreased by an average of 309 milliseconds. These results suggest that the milder participant exhibited much larger gains from the treatment. This variance could be reduced in future studies by including a greater number of participants with more restrictive inclusion criteria to increase statistical power. Further research is needed to determine how characteristics and severity of aphasia may interact with this treatment's efficacy.

Clinical Implications

More research is needed to determine implications for clinical practice. However, if future research shows that the maze task results in an improvement of grammatical sequencing with untrained sentences, it has promise as an aphasia therapy tool for addressing sentence formulation. Additionally, most of the time, participants in this study indicated that they would be willing to perform this task at home on their own time. Thus, if the treatment is found to result in meaningful gains, the maze task may become a valuable home practice activity.

Another direction to consider is that the maze task can be used as a tool for investigating how PWA processes language. Although variation between individuals with aphasia can be expected, future research could answer questions about whether a particular individual is sensitive to syntactic category information, morphological endings, function words, and many other features of written language. These findings could inform a customized treatment plan for sentence processing interventions.

Whether the maze task involves determining which word forms a grammatical sequence or recalling a complete sentence, including function words, grammatical competence, or processing skills may be relevant to task performance. Further research is needed to determine how the participants are doing the task as well as whether some participants are doing one thing and other participants are doing another.

CONCLUSION

In conclusion, the current study was a first step in examining the feasibility and acceptability of the maze sentence processing task as an aphasia treatment tool. The results showed that participants learned to perform the task, demonstrated improved accuracy, and demonstrated faster response times given repeated practice with a closed set of stimulus sentences and clinician cueing as needed. Additionally, most of the participants tolerated doing the task and rated it as fun and helpful much of the time. These preliminary results suggest that additional studies are warranted to determine whether this approach could be useful as an aphasia treatment tool.

ACKNOWLEDGMENT

This work was supported by a University Research/Creative Projects Award from Wichita State University, grant #200182. The authors have no other financial or non-financial interests to declare. The authors would like to thank the research participants, their families, Dr. Shari DeVeney, Dr. Sean Hess, and Dr. Cynthia Richburg for scientific advice and helpful comments, Jonathan Forster for technical support of the DMDX experimental software, graduate student Addison Powell for administering treatment, and graduate student Daphne Keese for administering assessments.

REFERENCES

1. Faroqi-Shah Y, Baker AL. Agrammatic aphasia. In: Coppens P, ed. *Aphasia Rehabilitation: Clinical Challenges*. 2017:101-141.
2. Marshall J. Disorders of sentence processing in aphasia. In: Papathanasiou I, Coppens P, eds. *Aphasia and Related Neurogenic Communication Disorders, Third Edition*. 2022:295-318.
3. Poirier SÈ, Fossard M, Monetta L. The efficacy of treatments for sentence production deficits in aphasia: a systematic review. *Aphasiology*. 2023;37(1):122-142. https://doi.org/10.1080/02687038.2021.1983152
4. Bever TG. The cognitive basis for linguistic structures. In: Hayes JR, ed. *Cognition and the Development of Language*. 1970:279-362. https://doi.org/10.1093/acprof:oso/9780199677139.003.0001
5. Garrett M. Processes in language production. In: Newmeyer FJ, ed. *Linguistics: The Cambridge survey*. 1988:69-96. https://doi.org/10.1017/CBO9780511621062.004
6. Ferreira F, Çokal D. Sentence processing. In: Hickock G, Small SL, eds. *Neurobiology of Language*. 2016:265-274. https://doi.org10.1016/B978-0-12-407794-2.00022-5
7. DeDe G. Reading and listening in people with aphasia: effects of syntactic complexity. *American Journal of Speech-Language Pathology*. 2013;22(4):579-590. https://doi.org/10.1044/1058-0360(2013/12-0111)
8. Kim ES, Suleman S, Hopper T. Cognitive effort during a short-term memory (STM) task in individuals with aphasia. *Journal of Neurolinguistics*. 2018;48:190-198. https://doi.org/10.1016/j.jneuroling.2017.12.007
9. Forster KI. Using a maze task to track lexical and sentence

processing. *The Mental Lexicon*. 2010;5(3):347-357. https://doi.org/10.1075/ml.5.3.05for

10. Forster KI, Guerrera C, Elliot L. The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*. 2009;41(1):163-171. https://doi.org/10.3758/BRM.41.1.163
11. Freedman SE, Forster KI. The psychological status of overgenerated sentences. *Cognition*. 1985;19:101-131. https://doi.org/10.1016/0010-0277(85)90015-0
12. Nicol JL, Forster KI, Veres C. Subject–verb agreement processes in comprehension. *Journal of Memory and Language*. 1997;36:569-587. https://doi.org/10.1006/jmla.1996.2497
13. O'Bryan E, Folli R, Harley H, et al. Evidence for the use of verb telicity in sentence comprehension. In: Folli R, Sevdali C, Truswell R, eds. *Syntax and Its Limits*. Oxford University Press; 2013:80-104. https://doi.org/10.1093/acprof:oso/9780199683239.003.0005
14. Qiao X, Shen L, Forster K. Relative clause processing in Mandarin: Evidence from the maze task. *Language and Cognitive Processes*. 2012;27(4):611-630. https://doi.org/10.1080/01690965.2011.578394
15. Witzel N, Witzel J, Forster K. Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research*. 2012;41:105-128. https://doi.org/10.1007/s10936-011-9179-x
16. Enkin E. The Maze Task: Training Methods for Second Language Learning. *Journal of Second Language Acquisition and Teaching*. 2012;19:56-78.
17. Thompson CK, Shapiro LP. Treating agrammatic aphasia within a linguistic framework: Treatment of Underlying Forms. *Aphasiology*. 2005;19(10-11):1021-1036. https://doi.org/10.1080/02687030544000227
18. Helm-Estabrooks NA. *Helm Elicited Language Program for Syntax Stimulation (HELPSS*. Exceptional Resources; 1981.
19. Helm-Estabrooks N, Nicholas M. *Sentence Production Program for Aphasia*. Pro-Ed, Inc; 2000.
20. Fink RB, Schwartz MF, Rochon E, et al. Syntax stimulation revisited: An analysis of generalization of treatment effects. *American Journal of Speech-Language Pathology*. 1995;4:99-104. https://doi.org/10.1044/1058-0360.0404.99
21. Helm-Estabrooks N, Ramsberger G. Treatment of agrammatism in long-term Broca's aphasia. *British Journal of Disorders of Communication*. 1986;21:39-45. https://doi.org/10.3109/13682828609018542
22. Fink RB, Bartlett MR, Lowery JS, et al. Aphasic speech with and without SentenceShaperR: Two methods for assessing informativeness. *Aphasiology*. 2008;22:679-690. https://doi.org/10.1080/02687030701800792
23. Kertesz A. *Western Aphasia Battery Revised examiner's manual*. Pearson; 2007.
24. Forster KI, Forster JC. DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*. 2003;35(1):116-124. https://doi.org/10.3758/BF03195503
25. Kagan A, Simmons-Mackie N, Victor JC, et al. *Assessment for Living with Aphasia*. Aphasia Institute; 2011.
26. Dwivedi AK, Mallawaarachchi I, Alvarado LA. Analysis of small sample size studies using nnon-parametricbootstrap test with pooled resampling method. *Statistics in Medicine*. 2017;36(14):2187-2205. https://doi.org/10.1002/sim.7263
27. Mundry R, Fischer J. Use of statistical programs for nnon-parametrictests of small samples often leads to incorrect Pvalues: examples from animal behaviour. *Animal Behaviour*. 1998;56(1):256-259. https://doi.org/10.1006/anbe.1998.0756
28. Beeson PM, Robey RR. Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review*. 2006;16(4):161-169. https://doi.org/10.1007/s11065-006-9013-7
29. LaPointe LL, Horner J. *Reading Comprehension Battery for Aphasia*. Second edition. Examiner's manual. Pro-Ed; 1998. https://doi.org/10.1080/02687030544000227
30. Kintz S, Wright HH. Discourse measurement in aphasia research. *Aphasiology*. 2018;32(4):472-474. https://doi.org/10.1080/02687038.2017.1398807

Supplementary Fig 1

Examples of Maze Task Sentences

In the sentences below, the non-continuation word (wrong choice) is in parenthesis. The characters' x-x-x' always appear opposite the first word of the sentence. The position of the correct choice (left or right) was randomized.

- Sharon (x-x-x) / met (key) / the (pry) / woman (their) / by (is) / the (put) / window (around).
- The (x-x-x) / bracelet (these) / by (him) / the (fib) / perfume (become) / counter (stretch) / costs (mean) / three (he) / hundred (begin) / dollars (bottom).

Supplementary Fig 2

Post-Session Subjective Experience Questionnaire

Below is a questionnaire in which you can give me your opinion on your experience with the maze task today! Please rate your experience from 1 to 5, with 1 being the worst rating and 5 being the best rating.