Title
A Cognitive Psychometric Investigation of Word Production and Phonological Error Rates in Logopenic Progressive Aphasia.

Journal
American Journal of Speech-Language Pathology, 30(3)

ISSN
1058-0360

Authors
Petroi, Diana
Walker, Grant M
Duffy, Joseph R
et al.

Publication Date
2021-05-18

DOI
10.1044/2021_ajslp-20-00221

Peer reviewed

AJSLP

Research Note

A Cognitive Psychometric Investigation of Word Production and Phonological Error Rates in Logopenic Progressive Aphasia

Diana Petroi,[a,b] Grant M. Walker,[c] Joseph R. Duffy,[a]
Gregory S. Hickok,[c] and Keith A. Josephs[a]

Purpose: This study investigated the relationship between word production rates (WPRs) and phonological error rates (PERs) in generative and responsive tasks in logopenic progressive aphasia (lvPPA). We examined whether a portion of the reduced WPR during generative tasks related directly to phonological impairments affecting PER on all tasks, irrespective of other task differences that contributed to WPR.

Method: Two cognitive psychometric models were hypothesized and fit to the total number of words produced and the number of phonological errors produced by 22 participants on 10 tasks. Bayesian inference was used to construct posterior distributions of participant ability and task difficulty parameters. Model fit statistics were compared. Association strengths for average generative WPR and average responsive PER were also evaluated with linear least-squares regression.

Results: Average generative WPR and average responsive PER were significantly associated ($r = -.77$, $p = .00002$). A cognitive psychometric model that assumed reduced WPR on generative tasks reflects a portion of general phonological impairment yielded better fit than a model that ignored performance differences between generative and responsive tasks. Generative fluency tasks that elicited few phonological errors still reflected phonological impairment, via suppression. Individual participants were estimated to suppress between 62% and 93% of phonological errors on generative tasks that would have emerged on responsive tasks.

Conclusions: Suppression of phonological errors may present as decreased WPR on generative tasks in lvPPA. Failure to account for this suppression tendency may lead to overestimation of phonological ability. The findings indicate the need to account for task demands in assessing lvPPA.

Primary progressive aphasia (PPA) is a language-based, neurologic impairment that is an initial sign of an ensuing neurodegenerative disease, which occurs in the absence of other cognitive deficits (Gorno-Tempini et al., 2011; Mesulam, 2001). One of the subtypes of PPA is the logopenic variant (LPA; Botha et al., 2015). Core clinical features of LPA consist of impaired word retrieval and sentence repetition, along with the occurrence of three or more of the following features: phonological errors in tasks involving naming and spontaneous speech, lack of agrammatism, spared comprehension of single words or object knowledge, and/or intact motor speech abilities (Gorno-Tempini et al., 2011). Neuroimaging research on LPA (Beck et al., 2008; Gorno-Tempini et al., 2008, 2011; Josephs et al., 2010; Krishnan et al., 2017; Madhavan et al., 2013; Mesulam et al., 2009; Rohrer et al., 2013, 2010) reveals left hemisphere abnormalities to be greater than those of the right, with significant lateral temporoparietal cortical atrophy, typically also implicating the frontal lobes and precuneus; there is relative medial temporal lobe sparing.

Phonological errors are one of the salient features that may be present in LPA based on the previously published consensus criteria by Gorno-Tempini et al. (2011), yet they are not required for a person to be diagnosed with LPA. Preliminary data on neuroanatomical correlates of phonological errors in LPA reveal associations in the parietal regions, specifically the supramarginal gyrus and other portions of

[a]Department of Neurology, Mayo Clinic, Rochester, MN
[b]Department of Communication Sciences and Disorders, Biola University, CA
[c]Department of Cognitive Sciences, University of California, Irvine, CA
Correspondence to Grant M. Walker grantw@uci.edu

Editor-in-Chief: Julie Barkmeier-Kraemer
Editor: Anastasia Raymer
Received July 27, 2020
Revision received January 10, 2021
Accepted January 19, 2021
https://doi.org/10.1044/2021_AJSLP-20-00221

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

the inferior parietal lobe (Petroi et al., 2020); greater atrophy in such regions may increase the likelihood of persons with lvPPA making phonological errors, particularly substitutions. Petroi et al. (2014) pointed out that phonological errors are usually vaguely described in the literature, simply noted as being frequently or infrequently present during certain tasks with inconsistent use of terminology, and at times mentioned to involve substitutions, omissions, or additions of sounds that are not distorted (Bonner et al., 2010; Gorno-Tempini et al., 2008, 2004; Leyton et al., 2014, 2015; Mesulam et al., 2009; Rogalski et al., 2011; Wilson et al., 2010). This has made it challenging to compare phonological errors across studies. To our knowledge, three studies have been published that have defined and analyzed phonological errors in detail (Dalton et al., 2018; Henry et al., 2016; Petroi et al., 2014).

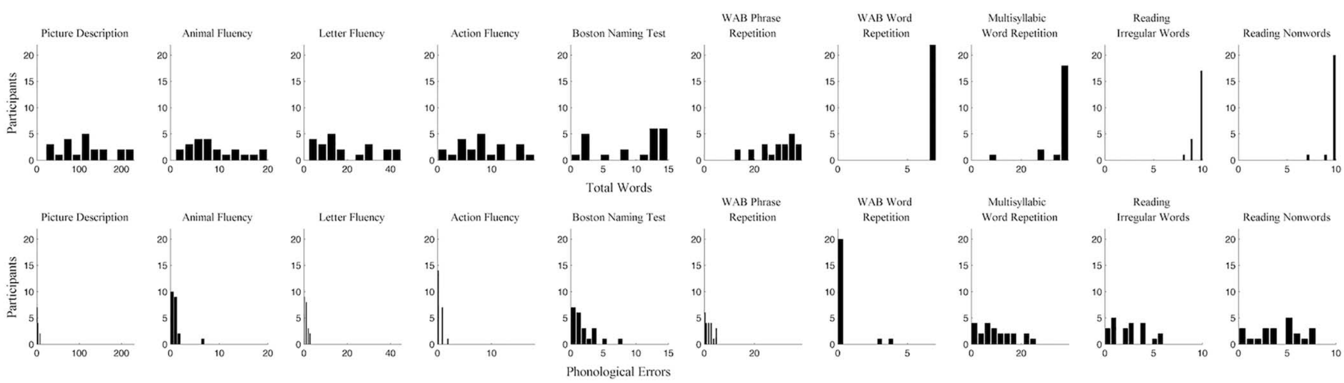
The study by Petroi et al. (2014) found evidence supporting the previously published consensus criteria pertaining to phonological errors (Gorno-Tempini et al., 2011) as well as updated considerations (Botha et al., 2015), revealing that all 22 participants having LPA produced some phonological errors. Based on descriptive statistics, four spoken language tasks were identified as being most likely to elicit a relatively large proportion of phonological errors: reading nonwords and reading irregular words (Western Aphasia Battery–Revised [WAB]; Kertesz, 2007), repetition of multisyllabic words, as well as the 15-item Boston Naming Test (BNT; Lansing et al., 1999). These were referred to as "sensitive tasks" in contrast to those tasks that elicited a smaller proportion of phonological errors, referred to as "insensitive tasks," such as picture description, action and letter fluency, and WAB repetition. The presence or absence of phonological errors fluctuated across tasks, which was thought to be influenced by factors such as task complexity (i.e., the number of syllables per target word potentially increasing task demands or providing greater opportunities for errors), task nature (i.e., the variability in the number of words produced for tasks requiring spontaneous/open-ended responses rather than a target word), the overall severity of aphasia, as well as, potentially, education.

The findings of Petroi et al. (2014) raise additional questions about factors that might be relevant to phonological error assessment. There was noted variance of individual word production rates (WPRs; number of words produced per task, not per minute, as speed was not a focus of this study) within tasks, which influenced the calculation of phonological error rates (PER; number of phonological errors produced per word), and marked group differences in WPR between the PER-sensitive and PER-insensitive tasks. Figure 1 illustrates the frequency histograms for the total number of words and the number of words with phonological errors produced on each task by the 22 participants (originally reported by Petroi et al., 2014). Tasks that elicited lower WPRs (picture description; animal, letter, and action fluency; and WAB phrase repetition) also tended to elicit fewer phonological errors. Tasks that elicited many phonological errors (reading irregular words, reading nonwords, multisyllabic word repetition, and the BNT) also tended to have the least number of word omissions. This suggested the possibility of an interaction needing further exploration. Moreover, while this descriptive study provided useful insights about the presence and nature of phonological errors across different speech production tasks in LPA, it lacked a theory-driven model to motivate task selection or account for phonological errors.

In the current study, we took a confirmatory approach, comparing different explanations for the sources of observed counts of words and phonological errors in the speech production data from the Petroi et al. (2014) study. The purpose of the current study was to test the hypothesis that the phonological deficit in LPA may have a dual presentation in speech production task scores, depending on the demands of the task: The deficit can manifest as phonological errors on some speech production tasks while manifesting as reduced word output on others. We refer to this hypothesized dual presentation as *error suppression,* by which we mean to refer to an effect observed in the data (i.e., counts of words and phonological errors collected from various speech production tasks), such that words that are likely to lead to phonological errors tend to be selectively omitted when task constraints permit it. While this term may suggest that an active, mental process is generating the observed patterns in the data, and this may indeed be so, it is not necessary; an unconscious reflex or some other, complex, mechanistic description may explain the data just as well. The current study remains largely neutral with respect to the underlying neural or psychological mechanisms that give rise to phonological deficits in LPA, given that the aim of our study was not to address this, except insofar as highlighting that there is *something* to be explained regarding the interaction of task demands and phonological error elicitation in this clinical population. In particular, we distinguish between (a) "generative" tasks that allow for strategic word-choices that potentially contribute partial credit to the task scores, which typically require retrieval of phonological forms from long-term memory (e.g., picture description, fluency tasks), and (b) "responsive" tasks that are highly constrained in their potential responses, typically with stimuli that directly and immediately provide the phonological content to be produced (e.g., reading, word repetition).

Use of cognitive psychometrics (Batchelder, 2010) was deemed most appropriate for investigating the relationship between WPR and PER given that the data were multinomial (i.e., each potentially spoken word was categorized as phonologically correct, phonological error, or omission error as defined below), individual response types were hypothesized to be multiply determined (i.e., different omission errors could potentially arise from different mental processes), and participant abilities and task difficulties were hypothesized to be heterogeneous (i.e., some tasks are harder than others, and some participants are more impaired than others; Batchelder, 2010). This is the first study to use cognitive psychometrics to analyze phonological errors in PPA, specifically LPA. However, cognitive psychometric investigations of speech errors in picture naming have been conducted previously in participants with Alzheimer's and vascular dementia

Figure 1. Frequency histograms for the total number of words (top row) and the number of words produced with phonological errors (bottom row) on each task by the 22 participants. WAB = Western Aphasia Battery.



(Chosak Reiter, 2000) and stroke-induced aphasia (Walker et al., 2018).

Method

Data

Cognitive psychometric models were fit to archived data. The data consisted of the total number of words and the number of words containing phonological errors produced by 22 participants with LPA on each of the nine tasks examined in the Petroi et al. (2014) study. The nine tasks included five subtests from the WAB Parts 1 and 2 (Kertesz, 2007)—picture description, animal fluency, repetition, reading irregular words (e.g., debt, yacht), and reading nonwords (e.g., aponster, dosh); the 15-item BNT (Lansing et al., 1999); an action (verb) fluency task (Woods et al., 2005); a letter (F, A, S) fluency task (Loonstra et al., 2001); as well as a multisyllabic word repetition task consisting of participants providing three repetitions of each of 13 words (e.g., specific, aluminum; full list of stimuli can be found in Duffy et al., 2015). In the current study, the trials from the WAB repetition task were divided into two sets with (a) seven, single-word repetition trials (corresponding to Items 1–7) and (b) eight, phrase repetition trials consisting of two to 10 words each (corresponding to Items 8–15) for a total of 39 words. This change in scoring protocol yielded 10 total tasks for analysis in the current study.[1] Using the number of target words for responsive tasks or the maximum number of words produced by any participant for generative tasks, each potential spoken word was categorized as phonologically correct, a phonological error, or an omission error, and the frequencies of the three response types were calculated for each participant on each task. A phonological error was defined as phoneme substitutions, additions, omissions, and/or transpositions produced within recognizable utterances. An omission error was defined as a complete lack of observable utterance. To clarify, if only part of a word was omitted, it was classified as a phonological error; if the whole word was omitted, it was classified as an omission error.

Reliability

Mean intrajudge as well as interjudge reliability indices were calculated across all of the tasks in the original study for four of the 22 participants. Based on a unit-by-unit agreement ratio (Hegde, 2003, pp. 204–208), the intrajudge reliability index was 96% for item-by-item agreement on the presence and absence of phonological errors; it was 89% for the number of items on which phonological errors occurred per task. Similarly, interjudge reliability was 95% for item-by-item agreement for the presence and absence of phonological errors; it was 78% for the number of items on which phonological errors occurred per task. All discrepancies were reviewed and resolved by consensus. See Petroi et al. (2014) for additional details.

Models

Two formal models of the latent processes that generated the data were compared. Using formulations from item response theory and cognitive psychometrics (Batchelder, 2010), both models assumed that the probability of a participant producing a certain number of words or phonological errors on a task resulted from the interaction of the participant's ability (i.e., severity of impairment) and the task's difficulty regarding word selection or phonological production, respectively. The models differed regarding whether or not the number of words produced on generative tasks might

[1]Two supplementary cognitive psychometric analyses were performed, with identical modeling decisions as in the primary analysis, except treating the WAB repetition trials as a single task set, consistent with the Petroi et al. (2014) study. In one analysis, the task was designated as a "generative" task and, in the other analysis, as a "responsive" task, owing to its combined nature of having both single-word and phrase repetition trials. The substantive conclusions from the primary analysis of the current study were unchanged in both supplementary analyses, and the effects on numerical estimations were negligible (aside from the expected difference in difficulty estimates for the WAB repetition tasks), demonstrating the robustness of the current modeling results to relatively minor, arbitrary, scoring decisions.

additionally reflect a latent propensity to (consciously or unconsciously) suppress (i.e., avoid producing) phonological errors that might otherwise emerge in the context of a responsive task. The tree structures applicable to each task are illustrated in Figure 2.

Model A assumed data from all tasks depended on a probability of successful word selection (*a*) and a probability of successful phonological production (*b*), each of which is modulated by the participants' abilities (β_a and β_b) and the tasks' difficulties (δ_a and δ_b). Model B assumed the same tree structure as Model A for responsive tasks, but additionally included a probability of error suppression (*c*) that converted some proportion of phonological errors into complete word omission errors; this probability was modulated by participant ability only, not by task difficulty (i.e., it was assumed that any phonological errors were equally difficult to suppress regardless of the task context). Because the WAB phrase repetition task afforded the opportunity for selective retrieval of words or paraphrasing (e.g., *Pack my box with five dozen jugs of liquid detergent*), it was included in the generative category. As explained in the note, decisions about how to categorize this task did not alter any substantive conclusions.

Model Fitting

Model parameters (participant abilities and task difficulties) were fit to the data using a Bayesian inference procedure (Lee & Wagenmakers, 2014). Specifically, weakly informative prior distributions (i.e., standard normal distributions on a logit scale) for ability and difficulty parameters were updated to posterior distributions via Gibbs sampling using the JAGS software (Plummer, 2003). Four chains of 100,000 samples were run, and convergence was checked by visual inspection (Lee & Wagenmakers, 2014). To elaborate, the data consisted of 440 independent observations of response type frequencies (22 participants × 10 tasks × 2 independent response types, or degrees of freedom). Model A described the data with 64 parameters ([22 participants × 2 abilities] + [10 tasks × 2 difficulties]), while Model B described the data with an additional participant parameter (the proportion of suppressed phonological errors on generative tasks), totaling 86 parameters. Both models therefore had far fewer parameters than degrees of freedom in the data, ensuring model identifiability (Batchelder & Riefer, 1999), except in cases where either only correct responses or only omission errors were made on all tasks; no such cases existed in our sample.

Model Comparison

The models were compared based on how much information in the data could be accurately encoded in their parameters, using the deviance information criterion (DIC) and posterior predictive error model fit statistics for quantitative comparison. The posterior predictive error refers to the difference between the observed frequencies of responses and the frequencies that would be predicted from the model's best fitting parameter values. The DIC refers to the tradeoff in the information contained within the observed data and within the estimated parameters, penalizing the fit of models with more parameters. The mean of the posterior samples was taken as a point estimate for each parameter value. Parameter estimates were compared between models, particularly to understand the implications of the models' assumptions on the inferred phonological processing difficulties of each task.

Linear Regression Analysis

As a further test of the hypothesis that PER on responsive tasks is related to WPR on generative tasks, a least-squares linear regression analysis was carried out to evaluate the strength of the association. The average proportion of the maximum possible words produced on generative tasks was predicted from the average proportion of produced words resulting in phonological errors on responsive tasks.

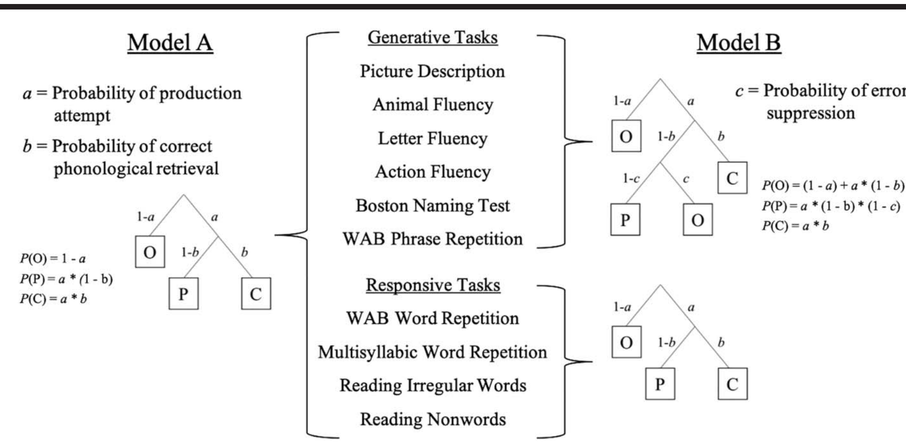


Figure 2. Multinomial processing trees associated with each task for Model A and Model B. Branches are associated with a probability of success or failure, and leaf nodes are associated with response types. Each branch probability depends on a participant's ability and a task's difficulty via a Rasch model equation (not shown; Batchelder, 2010; Rasch, 1960). Equations for the probability of each response type are derived from each tree model. WAB = Western Aphasia Battery); C = phonologically correct; P = phonological error; O = omission error.

Speech Repetition Task Analysis

The task battery included three speech repetition tasks that were presumed to have different task demands: (a) WAB single-word stimuli (high word-choice constraints/low phonological demands), (b) WAB phrase/sentence stimuli of varying length (medium word-choice constraints/medium phonological demands), and (c) multisyllabic-word stimuli to be repeated 3 times consecutively (high word-choice constraints/high phonological demands). The hypothesized deficit in LPA predicts a difference for WPR between single-word repetition versus phrase repetition, but little to no difference in PER, due to balancing of word-choice constraints and phonological demands via error suppression. There is also a predicted relationship between the PER statistic from multisyllable word repetition, where the word-choice constraints and phonological demands can reveal compromised phonological abilities, and the WPR statistic from phrase repetition, where the phonological deficit is expected to manifest as error suppression. These two tasks have the same number of maximum trial words, enabling direct comparison of the impact of their task demands on the collected counts of words and phonological errors. The average number of words produced on the two tasks were compared with a paired, two-sample *t* test ($\alpha = .05$). The relationship between the number of phonological errors on the responsive task and the reduction in the number of words produced on the generative task relative to the responsive task was assessed with simple linear regression, testing for a significant correlation ($\alpha = .05$). Finally, the number of words produced on the generative task was predicted simultaneously from the number of words produced and the number of phonological errors produced on the responsive task, using multiple linear regression, testing for significant coefficients for both predictors ($\alpha = .05$).

Results

Model Fit

A posterior predictive distribution was generated for the total number of words produced and the number of words with phonological errors for each participant; the posterior predictive means were taken as a point estimates of response frequencies, and the highest density intervals (HDI) containing 95% of the posterior predictive distributions were taken as interval estimates of the response frequencies. In other words, the expected data were regenerated from each model's (i.e., Model A or Model B) distribution of best fitting parameters to evaluate how well each model fit the observed data. R^2_{df} is the variance accounted for in observed WPR or PER by model predictions. HDI-α is the proportion of observations falling outside the HDI (expected to be 0.05; lower is better). The DIC balances model fit against degrees of freedom for Bayesian model selection; lower DIC indicates better fit (Spiegelhalter et al., 2002). Table 1 provides a summary of the model fits in terms of the WPR and PER posterior prediction accuracies and overall DIC fit statistics. Table 2 summarizes the posterior predictive mean absolute error in terms of frequency (i.e., the average number of words or phonological errors by which model predictions differed from observed values) across each of the 10 tasks.

Table 1. Summary of the model fits in terms of the WPR and PER prediction accuracies, and overall DIC fit statistics.

Model	WPR	PER	DIC
Model A	$R^2_{218} = .71$ HDI-α = .25	$R^2_{216} = .59$ HDI-α = .05	2,507
Model B	$R^2_{218} = .85$ HDI-α = .13	$R^2_{216} = .54$ HDI-α = .09	2,041

Note. WPR = word production rates; PER = phonological error rates; DIC = deviance information criterion; HDI-α = highest density interval prediction error rate.

Parameter Estimates

Model A estimates of phonological difficulty of tasks matched the previous descriptive results (Petroi et al., 2014). In this case, PER "sensitive" tasks ranked as the most difficult, as expected, because they yielded the most phonological errors. These "sensitive" tasks consisted of reading nonwords, multisyllabic repetition, reading irregular words, and the BNT. Table 3 provides a summary of the Model A parameter estimates for task difficulties. Although task difficulty is measured on a logit scale ranging from −infinity to infinity, for interpretation, the logit values have been converted to proportion values, assuming the average ability level of the participants in the sample. For example, a participant with average abilities would be expected to attempt 68% of the BNT trials, with 19% of those attempts resulting in a phonological error. In contrast, Model B captured more information about WPR than Model A, at the expense of overpredicting PER in a handful of cases. DIC values indicated Model B's increased complexity over Model A was warranted by the amount of additional information in the data that was explained overall. In other words, the evidence supports Model B's assumption that there is a relationship between the PER on responsive tasks and the WPR on generative tasks in participants with LPA. If it is accepted that the WPR on generative tasks reflects a portion of general phonological impairment, Model B further reveals that the tasks that elicit the most phonological errors are no longer deemed to be the most phonologically challenging. Animal and letter fluency tasks were deemed the most phonologically challenging, followed by nonword reading and the BNT. Table 4 provides a summary of the Model B parameter estimates for task difficulties. Again, the logit values have been converted to proportion values, assuming the average ability level for the participants; the proportions in parentheses are the expected proportions after accounting for phonological error suppression. For example, without phonological error suppression, a participant with average ability might be expected to attempt 95% of the BNT trials, leading to a 40% PER; however, assuming an average frequency of error suppression on generative tasks, this participant is

Table 2. Posterior predictive mean absolute error (frequency) across tasks.

	Posterior predictive mean absolute error (frequency)										
Model	Picture description	Animal fluency	Letter fluency	Action fluency	BNT	WAB phrase repetition	WAB word repetition	Multisyllabic word repetition	Reading irregular words	Reading nonwords	Task sum total
Max.:	230	20	45	18	15	39	7	39	10	10	433
A: Words	19.6	4.0	9.0	2.3	3.5	4.1	0	2.7	0.3	0.2	45.8
B: Words	6.5	2.1	5.2	2.0	3.1	3.0	0	3.1	0.3	0.1	25.5
A: Phon Errs	1.2	0.9	0.8	0.4	1.2	1.1	0.3	2.9	1.0	1.9	11.7
B: Phon Errs	1.5	0.8	1.9	0.4	1.4	1.4	0.3	3.4	1.0	1.7	13.7

Note. Words = total words produced; Phon Errs = phonological error frequency; BNT = Boston Naming Test; WAB = Western Aphasia Battery.

only expected to overtly attempt 64% of the BNT trials, leading to a 9% observed PER.

There were 19 of 22 participants (86%) who made enough phonological and/or omission errors across tasks to plausibly estimate their proportion of suppressed errors (i.e., the posterior 95% HDI for the suppression ability parameter was less than two logits wide, a 50% reduction in uncertainty). Among these participants, point estimates for error suppression ranged from 62% to 93%, with an average of 82% of potential phonological errors being suppressed on generative tasks.

Strength of Association Between Generative WPR and Responsive PER

The average PER on responsive tasks was strongly correlated with the average WPR on generative tasks ($r_{20} = -.77$, $p = .00002$). The linear regression model estimated that a 10% increase in the average PER on responsive tasks is associated with a 10.2% decrease in the average WPR on generative tasks.

Comparison of Speech Repetition Tasks

No words were omitted by any participant with LPA on WAB word repetition, a striking contrast to WAB phrase repetition, which had about 26% word omissions for the average participant (range: 0%–69%); both tasks had similarly low PERs (phrases: 9% average, range: 0%–42%; words: 5% average, range: 0%–57%). From this evidence alone, one might be tempted to conclude that phonological errors played no role in the difference between single-word and phrase repetition performances. However, considering these scores in the context of other test scores suggests otherwise: Error suppression provides a more comprehensive explanation of the full data pattern. On average, participants produced seven fewer words (out of 39) on WAB phrase repetition than on

Table 3. Model A estimated task difficulty parameters (posterior mean), expressed as the expected proportion of the maximum possible words produced (WPR) and the expected proportion of the produced words that result in phonological errors (PER) for a hypothetical participant with logopenic progressive aphasia who has average abilities.

Model A		
Task	WPR	PER
Reading nonwords	.98	.41
Multisyllabic word repetition	.95	.26
Reading irregular words	.98	.23
Boston Naming Test	.68	.19
Animal fluency	.42	.12
Letter fluency	.39	.06
WAB phrase repetition	.79	.06
Action fluency	.47	.06
WAB word repetition	.99	.04
Picture description	.50	.01

Note. WAB = Western Aphasia Battery.

Table 4. Model B estimated task difficulty parameters (posterior mean), expressed as the expected proportion of the maximum possible words produced (WPR) and the expected proportion of the produced words that result in phonological errors (PER) for a hypothetical participant with logopenic progressive aphasia who has average abilities.

Model B		
Task	WPR	PER
Animal fluency	.91 (.47)	.57 (.12)
Letter fluency	.83 (.43)	.51(.11)
Boston Naming Test	.95 (.64)	.40 (.09)
Reading nonwords	.98	.40
Action fluency	.77 (.50)	.35 (.07)
Multisyllabic word repetition	.96	.24
Reading irregular words	.98	.21
WAB phrase repetition	.95 (.80)	.19 (.04)
Picture description	.56 (.53)	.04 (.01)
WAB word repetition	.99	.04

Note. The expected proportions that would be observed after accounting for phonological error suppression are shown in parentheses.

multisyllable word repetition (p = .00002), and there was a significant correlation between the number of phonological errors produced on multisyllable word repetition and the reduction in total number of words produced on WAB phrase repetition (r_{20} = .43, p = .045), as predicted. Finally, both the total number of words produced (β = 0.63, p = .0007) and the number of phonological errors produced (β = −0.42, p = .012) on the responsive task were significant, independent predictors of the total number of words produced on the generative task (F = 14.6, R^2 = .61, p = .0001). This means that, as participants produced more words on the responsive task, they also produced more words on the generative task, unless the words that were produced on the responsive task contained phonological errors, in which case, the increases in word production on the generative task were cancelled out. Thus, on further inspection, word omissions during phrase repetition were related to phonological errors, despite a lack of overt increase in their frequency during this task.

Discussion

This is the first study to implement cognitive psychometrics to analyze phonological errors in PPA, in this case in LPA. As proposed in the Petroi et al. (2020) study, we sought to implement a more theory-driven investigation of the patterns observed in the phonological errors associated with LPA to better understand the interactions of behavioral attributes (i.e., participants' word retrieval abilities and task difficulty). Two models were compared: Model A assumed each participant had a word selection ability and phonological ability that was modulated by task difficulties, and Model B assumed an additional participant tendency to suppress a proportion of phonological errors on generative tasks. While Model A confirmed previous descriptive results (Petroi et al., 2014), Model B yielded additional information that would not have otherwise been gleaned, by considering another way in which phonological impairments may be exhibited. Specifically, suppression of phonological errors may manifest as decreased WPR on generative tasks in persons with LPA. This association was confirmed using classical linear regression analysis; higher average PER on responsive tasks was associated with lower average WPR on generative tasks. The association was also confirmed in direct comparisons of speech repetition tasks with varying degrees of word-choice constraints and phonological demands; a reduction in the number of words on a task with weak word-choice constraints was related to the production of phonological errors on a task with strong word-choice constraints. Consistent with the analyses of the psychometric model of individual task effects and the linear model of grouped task effects, the comparison of these specific, speech repetition task effects supported the hypothesized interplay between task demands and elicitation of phonological errors. The cognitive psychometric analysis provided more information about the effects of each of the 10 different speech production tests on WPRs and PERs, as well as providing individual estimates of error suppression rates on generative tasks (ranging from 62% to 93%), rather than a group-level association strength.

One might wonder if there is a simple way to explain these findings in terms of varying activation levels rather than proposing an error suppression phenomenon: For example, if a task and a participant combine to produce low mental/neural activation, then fewer total words would be expected and fewer phonological errors by extension. Thus, if generative tasks induce less activation than responsive tasks, this might explain the observed patterns. While this may be true regarding overall frequencies, this explanation fails to address the observed reduction in the *proportion* or *rate* of phonological errors in generative tasks, not just the overall frequency. By analogy, we might be surprised if a faulty radio speaker coincidentally functioned properly whenever the radio tuner had poor reception. Fewer words are coming out, but the ones that are coming out are surprisingly well formed. As the model comparison confirms, aligning the tasks along a single dimension (e.g., of task-induced activation) does not explain the data as well as assuming that there is a second source of word omissions related to the task demands and the phonological impairments of the participants. The evidence thus supports the claim that there is a qualitative difference, not merely a quantitative difference, in how participants perform on these types of tasks with respect to phonological error production.

This is particularly relevant when considering the diagnostic criteria for subtypes of PPA, which are based on clinical findings that include the use of tasks intended to draw out spoken language and other impairments. Failing to account for this suppression tendency may lead to overestimation of phonological ability. That is, the errors that participants with LPA make in spoken language tasks may extend beyond those of phonological errors. Our aim was to point out that there may be more than one general reason to account for reduced WPR. The error suppression inferred to occur based on the patterns observed in the data from the current study provides this alternative, theory-driven approach that has not previously been discussed in PPA, including LPA. Thus, consideration of task demands (generative versus responsive) and using a variety of spoken language tasks as was referenced in this and other studies (e.g., see Petroi et al., 2014, for a detailed description of these and other tasks) to capture these nuances and the potential suppression effect are recommended when evaluating and differentially diagnosing persons with PPA, particularly those suspected to have LPA.

Moreover, given that attention in recent years has been aimed at maintaining and/or improving communication function in LPA, the present findings have implications for current clinical practice and future research. While limited research has been conducted on treatment of LPA (e.g., Beeson et al., 2011; Henry et al., 2013), these preliminary findings, along with recent efforts aimed at addressing word retrieval (e.g., Henry et al., 2019), have demonstrated positive outcomes for naming in LPA and the semantic variant, including gains in maintenance and generalization. Our findings suggest that interventions geared toward addressing suppression of phonological errors may be needed given that this suppression is thought to present as a reduction in WPR on certain tasks. Incorporating generative tasks into therapy and implementing strategies to combat suppression effects may facilitate improvement of WPR. While the aim of this study was not to examine neuroanatomical correlates of suppression or phonological errors, it has previously been documented that LPA consists of left greater than right hemispheric asymmetry, particularly with marked lateral temporoparietal cortex atrophy, with less atrophy in the right hemisphere and typical sparing of the medial temporal lobe (Beck et al., 2008; Gorno-Tempini et al., 2008, 2011; Josephs et al., 2010; Krishnan et al., 2017; Madhavan et al., 2013; Mesulam et al., 2009; Rohrer et al., 2013, 2010). Because the deficits in LPA develop slowly over time, participants may tap into relatively preserved regions to develop conscious or unconscious compensatory strategies to "conceal" their phonological deficits. However, the avoidance of phonological errors in certain tasks by reducing overall word output may itself serve as a marker of the disease, or a useful strategy. It may be that some persons with LPA have already developed strategies on their own to compensate for their impairments, which can be reinforced or further enhanced by clinicians if they prove to be helpful. It may be just as acceptable to have reduced WPR as it would be to make phonological errors or to use other word retrieval strategies (e.g., semantic feature); this has yet to be explored and may need to be determined on an individual patient basis. Counseling patients and families about strategies to enhance communication contexts related to these aspects and training them to implement the most useful strategies, based on individual needs and interventions that have shown to be effective for persons with LPA (e.g., Henry et al., 2019), remains paramount. Further research is needed to explore which tasks and which strategies are best suited to address phonological impairments, with the goal of maintaining and/or improving not only communication function but also quality of life.

Acknowledgments

This work was supported in part by the National Institute on Deafness and Other Communication Disorders under Grants R01 DC010367 (awarded to Keith Josephs) and P50 DC014664 (awarded to Gregory Hickok). Thank you to Adriana Jones for her assistance with data entry.

References

Batchelder, W. H. (2010). Cognitive psychometrics: Using multinomial processing tree models as measurement tools. In S. E. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 71–93). American Psychological Association. https://doi.org/10.1037/12074-004

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*(1), 57–86. https://doi.org/10.3758/BF03210812

Beck, J., Rohrer, J. D., Campbell, T., Isaacs, A., Morrison, K. E., Goodall, E. F., Warrington, E. K., Stevens, J., Revesz, T., Holton, J., Al-Sarraj, S., King, A., Scahill, R., Warren, J. D., Fox, N. C., Rossor, M. N., Collinge, J., & Mead, S. (2008). A distinct clinical, neuropsychological and radiological phenotype is associated with progranulin gene mutations in a large UK series. *Brain, 131*(3), 706–720. https://doi.org/10.1093/brain/awm320

Beeson, P. M., King, R. M., Bonakarpour, B., Henry, M. L., Cho, H., & Rapcsak, S. (2011). Positive effects of language treatment for the logopenic variant of primary progressive aphasia. *Journal of Molecular Neuroscience, 45,* 724–736. https://doi.org/10.1007/s12031-011-9579-2

Bonner, M. F., Ash, S., & Grossman, M. (2010). The new classification of primary progressive aphasia into semantic, logopenic, or nonfluent/agrammatic variants. *Current Neurology and Neuroscience Reports, 10,* 484–490. https://doi.org/10.1007/s11910-010-0140-4

Botha, H., Duffy, J. R., Whitwell, J. L., Strand, E. A., Machulda, M. M., Schwarz, C. G., Reid, R. I., Spychalla, A. J., Senjem, M. L., Jones, D. T., Lowe, V., Jack, C. R., & Josephs, K. A. (2015). Classification and clinicoradiologic features of primary progressive aphasia (PPA) and apraxia of speech. *Cortex, 69,* 220–236. https://doi.org/10.1016/j.cortex.2015.05.013

Chosak Reiter, J. (2000). Measuring cognitive processes underlying picture naming in Alzheimer's and cerebrovascular dementia: A general processing tree approach. *Journal of Clinical and Experimental Neuropsychology, 22*(3), 351–369. https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT351

Dalton, S. G. H., Schultz, C., Henry, M. L., Hillis, A. E., & Richardson, J. D. (2018). Describing phonological paraphasias in three variants of primary progressive aphasia. *American Journal of Speech-Language Pathology, 27*(1S), 336–349. https://doi.org/10.1044/2017_AJSLP-16-0210

Duffy, J. R., Strand, E. A., Clark, H., Machulda, M., Whitwell, J. L., & Josephs, K. A. (2015). Primary progressive apraxia of speech: Clinical features, and acoustic and neurologic correlates. *American Journal of Speech-Language Pathology, 24*(2), 88–100. https://doi.org/10.1044/2015_AJSLP-14-0174

Gorno-Tempini, M. L., Brambati, S. M., Ginex, V., Ogar, J., Dronkers, N. F., Marcone, A., Perani, D., Garibotto, V., Cappa, S. F., & Miller, B. L. (2008). The logopenic/phonological variant of primary progressive aphasia. *Neurology, 71*(16), 1227–1234. https://doi.org/10.1212/01.wnl.0000320506.79811.da

Gorno-Tempini, M. L., Dronkers, N. F., Rankin, K. P., Ogar, J. M., Phengrasamy, L., Rosen, H. J., Johnson, J. K., Weiner, M. W., & Miller, B. L. (2004). Cognition and anatomy in three variants of primary progressive aphasia. *Annals of Neurology, 55*(3), 335–346. https://doi.org/10.1002/ana.10825

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogar, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Dronkers, N. F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B. L., Knopman, D. S., Hodges, J. R., Mesulam, M. M., & Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology, 76*(11), 1006–1014. https://doi.org/10.1212/WNL.0b013e31821103e6

Hegde, M. N. (2003). *Clinical research in communication disorders: Principles and strategies.* Pro-Ed.

Henry, M. L., Hubbard, H. I., Grasso, S. M., Dial, H. R., Beeson, P. M., Miller, B. L., & Gorno-Tempini, M. L. (2019). Treatment for word retrieval in semantic and logopenic variants of primary progressive aphasia: Immediate and long-term outcomes. *Journal of Speech, Language, and Hearing Research, 62*(8), 2723–2749. https://doi.org/10.1044/2018_JSLHR-L-18-0144

Henry, M. L., Rising, K., DeMarco, A. T., Miller, B. L., Gorno-Tempini, M. L., & Beeson, P. M. (2013). Examining the value of lexical retrieval treatment in primary progressive aphasia:

Two positive cases. *Brain and Language, 127*(2), 145–156. https://doi.org/10.1016/j.bandl.2013.05.018

Henry, M. L., Wilson, S. M., Babiak, M. C., Mandelli, M. L., Beeson, P. M., Miller, Z. A., & Gorno-Tempini, M. L. (2016). Phonological processing in primary progressive aphasia. *Journal of Cognitive Neuroscience, 28*(2), 210–222. https://doi.org/10.1162/jocn_a_00901

Josephs, K. A., Duffy, J. R., Fossett, T. R., Strand, E. A., Claassen, D. O., Whitwell, J. L., & Peller, P. J. (2010). Fluorodeoxyglucose F18 positron emission tomography in progressive apraxia of speech and primary progressive aphasia variants. *Archives of Neurology, 67*(5), 596–605. https://doi.org/10.1001/archneurol.2010.78

Kertesz, A. (2007). *Western Aphasia Battery (Revised)*. PsychCorp.

Krishnan, K., Machulda, M. M., Whitwell, J. L., Butts, A. M., Duffy, J. R., Strand, E. A., Sanjem, M. L., Spychalla, A. J., Jack, C. R., Jr., Lowe, V. J., & Josephs, K. A. (2017). Varying degrees of temporoparietal hypometabolism on FDG-PET reveal amyloid-positive logopenic primary progressive aphasia is not a homogeneous clinical entity. *Journal of Alzheimer's Disease, 55*(3), 1019–1029. https://doi.org/10.3233/JAD-160614

Lansing, A. E., Ivnik, R. J., Cullum, C. M., & Randolph, C. (1999). An empirically derived short form of the Boston Naming Test. *Archives of Clinical Neuropsychology, 14*(6), 481–487. https://doi.org/10.1093/arclin/14.6.481

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. http://www.amazon.com/Bayesian-Cognitive-Modeling-Practical-Course/dp/1107018455

Leyton, C. E., Ballard, K. J., Piguet, O., & Hodges, J. R. (2014). Phonologic errors as a clinical marker of the logopenic variant of PPA. *Neurology, 82*(18), 1620–1627. https://doi.org/10.1212/WNL.0000000000000387

Leyton, C. E., Hodges, J. R., McLean, C. A., Kril, J. J., Piguet, O., & Ballard, K. J. (2015). Is the logopenic-variant of primary progressive aphasia a unitary disorder? *Cortex, 67,* 122–133. https://doi.org/10.1016/j.cortex.2015.03.011

Loonstra, A. S., Tarlow, A. R., & Sellers, A. H. (2001). COWAT metanorms across age, education, and gender. *Applied Neuropsychology, 8*(3), 161–166. https://doi.org/10.1207/S15324826AN0803_5

Madhavan, A., Whitwell, J. L., Weigand, S. D., Duffy, J. R., Strand, E. A., Machulda, M. M., Tosakulwong, N., Senjem, M. L., Gunter, J. L., Lowe, V. J., Petersen, R. C., Jack, C. R., Jr., & Josephs, K. A. (2013). FDG PET and MRI in logopenic primary progressive aphasia versus dementia of the Alzheimer's type. *PLOS ONE, 8*(4), e62471. https://doi.org/10.1371/journal.pone.0062471

Mesulam, M. M. (2001). Primary progressive aphasia. *Annals of Neurology, 49,* 425–432. https://doi.org.10.1002/ana.91

Mesulam, M. M., Wieneke, C., Rogalski, E., Cobia, D., Thompson, C., & Weintraub, S. (2009). Quantitative template for subtyping primary progressive aphasia. *Neurology, 66*(12), 1545–1551. https://doi.org/10.1001/archneurol.2009.288

Petroi, D., Duffy, J. R., Borgert, A., Strand, E. A., Machulda, M. M., Senjem, M. S., Jack, C. R., Jr., Josephs, K. A., & Whitwell, J. L. (2020). Neuroanatomical correlates of phonologic errors in logopenic progressive aphasia. *Brain and Language, 204*. https://doi.org/10.1016/j.bandl.2020.104773

Petroi, D., Duffy, J. R., Strand, E. A., & Josephs, K. A. (2014). Phonologic errors in the logopenic variant of primary progressive aphasia. *Aphasiology, 28*(10), 1223–1243. https://doi.org/10.1080/02687038.2014.910591

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. http://www.ci.tuwien.ac.at/Conferences/DSC-2003/

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielson & Lydiche.

Rogalski, E., Cobia, D., Harrison, T. M., Wieneke, C., Weintraub, S., & Mesulam, M. M. (2011). Progression of language decline and cortical atrophy in subtypes of primary progressive aphasia. *Neurology, 76*(21), 1804–1810. https://doi.org/10.1212/WNL.0b013e31821ccd3c

Rohrer, J. D., Caso, F., Mahoney, C., Henry, M., Rosen, H. J., Rabinovici, G., Rossor, M. N., Miller, B., Warren, J. D., Fox, N. C., Ridgway, G. R., & Gorno-Tempini, M. L. (2013). Patterns of longitudinal brain atrophy in the logopenic variant of primary progressive aphasia. *Brain Language, 127*(2), 121–126. https://doi.org/10.1016/j.bandl.2012.12.008

Rohrer, J. D., Ridgway, G. R., Crutch, S. J., Hailstone, J., Goll, J. C., Clarkson, M. J., Mead, S., Beck, J., Mummery, C., Ourselin, S., Warrington, E. K., Rossor, M. N., & Warren, J. D. (2010). Progressive logopenic/phonological aphasia: Erosion of the language network. *NeuroImage, 49*(1), 984–993. https://doi.org/10.1016/j.neuroimage.2009.08.002

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353

Walker, G. M., Hickok, G., & Fridriksson, J. (2018). A cognitive psychometric model for assessment of picture naming abilities in aphasia. *Psychological Assessment, 30*(6), 809–826. https://doi.org/10.1037/pas0000529

Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., Miller, B. L., & Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain, 133*(7), 2069–2088. https://doi.org/10.1093/brain/awq129

Woods, S. P., Scott, J. C., Sires, D. A., Grant, I., Heaton, R. K., & Tröster, A. I. (2005). Action (verb) fluency: Test–retest reliability, normative standards, and construct validity. *Journal of the International Neuropsychological Society, 11*(4), 408–415. https://doi.org/10.1017/S1355617705050460