Supplemental Material S1. Power analysis.

It is useful to have a sense of the potential power and detectable effect sizes given the sample size used. We used a simulation-based power analysis to gain insight into this, resembling simulated sample size pathways used for similar data analysis challenges in other literatures (Moineddin et al., 2007; Rusticus & Lovato, 2014; Teare et al., 2014). We blended post-hoc and sensitivity techniques to get a sense of the likelihood of whether our analysis pathway would yield accurate clustering results given (1) a small set of possible effect sizes and (2) a range of potential cognitive covariates. This likelihood was assessed by creating a large set of simulated datasets, all created using random variables with a prespecified covariance structure corresponding to the hypotheses expressed in the introduction. If the sample size and data analysis spelled out previously can accurately pick out the prespecified structure, it seems likely they are appropriate to characterize the dataset that we uncovered. If, on the other hand, the pathway chosen only rarely uncovers the structure enforced in the dataset, we might have reason to be skeptical the picture painted in the experiment is complete.

Methods

Our primary hypothesis of interest was the one linking the four phonetic plasticity measures that we created. In our simulated datasets, we ensured the four measures were always correlated. Secondarily, we were interested in the idea that some subset of the other cognitive measures would be correlated with the four phonetic plasticity measures. However, we did not come in with strong hypotheses about what other measures might be important in understanding the shared variance in the measures of phonetic plasticity. As such, in our power modeling, we allowed the number of correlated cognitive measures to vary randomly from zero to seven; in some runs, only a handful of cognitive measures correlated with phonetic plasticity, while, in others, many did. Since we had no hypotheses about which would be more correlated with phonetic plasticity than others, these were chosen randomly for each run.

The stipulated relationships above were combined with known information about covariation in the measures taken from the NIH Toolbox. For covariation within the six measures coming from the cognition assessment (Weintraub et al., 2013), covariation could be reconstructed from a published factor analysis of the cognitive assessments (Mungas et al., 2014). This is not true for the Words-in-Noise measure from the audition assessment (Zecker et al., 2013); for that one, we assumed that the correlations with the other cognitive measures would be zero. Although this is probably a simplistic assumption, no previous datasets were available to add more detail to the models.

Finally, the key correlations added to the models—the ones within the measures of phonetic plasticity and from those measures to the relevant sets of cognitive measures—were set at three possible levels, reflecting a small, moderate, and large effect sizes. A small effect size, we believed, would have a true correlation of .3; a moderate one .5; and a large one .7. Given that this is a study of individual differences, these correlations would be attenuated by the reliability of the measures. Although assessing test-retest reliability was challenging for reasons mentioned in the discussion of the main experiment, pilot data, where each measure was run twice with different stimulus sets, did yield measures corresponding to construct validity, which could provide a benchmark that could be used for power simulations. These could be used to approximate the extent to which correlations within individuals would be attenuated by measurement error using the Spearman correction formula. The correction formula suggests the

https://doi.org/10.1044/2021_JSLHR-21-00004

idea that the maximum possible observed correlation between two variables should be roughly proportional to the square root of the product of their reliabilities. Although the formula itself has been questioned (Nimon et al., 2012), it allowed us a rough standard to use for judgment.

The pilot data that assessed construct validity was used to generate an attenuation factor reflecting the extent to which the proposed small, moderate, and large effect sizes would be weakened by error within participants. Our highest within-participant estimate of validity from the pilot dataset was .74 (for explicit learning) and the lowest was .48 (for incidental learning); the attenuation factor reflected the square root of the product of those lower- and higher-end estimates, .59. Thus, we modeled observed correlations of .18, .30, and .42 for our predicted small, moderate, and large effect sizes in our simulated datasets. Although reporting of effect sizes is not common in studies of individual differences in phonetic plasticity, cases where correlation coefficients are available show Pearson’s correlations near the moderate (Banks et al., 2015) and high (McLaughlin et al., 2018) levels chosen for these simulations. For each simulation, a covariance matrix like in Table S1 was used to generate simulated covariances.

	ExplLrng	RateLrng	AcntLrng	Vocab	WdsInNoise	Switching	InhbCtrl	ProcSpd	WorkMem	EpisMem
InciLrng	.30	.30	.30	.30	0	.30	0	0	0	.30
ExplLrng		.30	.30	.30	0	.30	0	0	0	.30
RateLrng			.30	.30	0	.30	0	0	0	.30
AcntLrng				.30	0	.30	0	0	0	.30
Vocab					0	.14	.14	.13	.24	.10
WdsInNoise						0	0	0	0	0
Switching							.53	.48	.51	.49
InhbCtrl								.46	.49	.47
ProcSpd									.44	.43
WorkMem										.59

Table S1. Sample correlation matrix used to populate a simulated dataset. This is a sample with a moderate effect size and correlations between the measures of phonetic plasticity and episodic memory, task switching, and vocabulary.

Note that some aspects of the matrix displayed in Table S1 work against finding a consistent factor structure in the dataset. For example, given the amount of covariance within the cognitive measures, it would be surprising for the measures of phonetic plasticity to correlate with each other, task switching, and episodic memory but not with working memory, inhibitory control, or processing speed. This approach was adopted for the sake of simplicity. For the large effect size, some combinations of the phonetic plasticity measures with the other cognitive measures were impossible to model, as the structure of variance similar to the approach depicted above is not mathematically possible. To pick a simple example, consider the instance of a model where all seven cognitive measures are truly tied to all four phonetic plasticity measures, and therefore given an observed correlation of .41 between phonetic plasticity and those measures. It is mathematically impossible to have such a large correlation between all those

https://doi.org/10.1044/2021_JSLHR-21-00004

measures while also stipulating that the correlations between the words-in-noise measures and the other cognitive measures is truly zero. Rather than improvising a correction to the model in these circumstances, these simulations were discarded.

Next, the covariance table above was converted into a set of outcomes for 80 participants using the mvrnorm command in the MASS R package (Ripley et al., 2020), allowing for variability due to sampling. Table S2 shows what the correlation table above translates to in terms of covariance for a randomly generated set of 80 modeled participants. These simulated datasets were then used for subsequent analyses.

	ExplLrng	RateLrng	AcntLrng	Vocab	WdsInNoise	Switching	InhbCtrl	ProcSpd	WorkMem	EpisMem
InciLrng	.21	.26	.44	.25	.19	.33	.06	-.04	-.01	.24
ExplLrng		.30	.19	.16	.02	.23	-.02	-.02	.10	.37
RateLrng			.38	.34	-.06	.34	-.06	.07	0	.27
AcntLrng				.20	.06	.36	.14	.05	0	.31
Vocab					-.20	.33	.06	.22	.11	.06
WdsInNoise						.06	-.17	-.10	.01	.04
Switching							.40	.53	.52	.28
InhbCtrl								.47	.46	.53
ProcSpd									.60	.49
WorkMem										.60

Table S2. Simulated correlations stemming from the table given in Table S1. Note that the correlations generally reflect the patterns of Table S1 but differ in some details due to sampling variance.

Analysis

A total of 10000 datasets were generated for the small and moderate effect sizes using the parameters outlined above. Simulated datasets like those found in Table S2 were used to run a data analysis resembling the one in the experimental methods. Although it would be plausible to imagine that the methods for determining the number of clusters described above would sometimes lead simulated datasets to have, say, one cluster or three clusters present in the data, we decided to set the number of clusters at two, as much of cluster analysis requires individualized intuition that would be impossible at the level of tens of thousands of simulated datapoints. We used the same cutoff criterion for determining whether a variable belonged with the factor that was described; if its loading was greater than 0.29, it was included. For the large effect size, about 4500 simulations were performed, based on the mathematical constraints of possible covariance matrices.

We came up with four criteria that reflected a “successful” cluster analysis, described below from roughly least to most restrictive. Given our primary interests, we were most curious whether each cluster analysis successfully extracted all four phonetic plasticity measures as a single cluster; this was coded as successful either of the two measured clusters included all four plasticity measures. However, there were other ways to quantify a “successful” simulation that may also have been of note. For example, we could see how often one of the clusters picked out only relevant variables, even if not all of them were a part of the cluster. We were also curious how often one of the clusters resulting from the analyses performed on the simulated datasets included all of the relevant variables (plasticity + the relevant cognitive measures), even if they picked up other variables that were not truly correlated. Finally, the most restrictive criteria we used required one of the factors to include all and only the factors of interest.

	InciLrng	ExplLrng	RateLrng	AcntLrng	Vocab	WdsInNoise	Switching	InhbCtrl	ProcSpd	WorkMem	EpisMem
Cluster 1							✓	✓	✓	✓	✓
Cluster 2	✓	✓	✓	✓	✓		✓				

Table S3. Resulting clusters from the dataset depicted in Table S2. Checks indicate variables that related to each factor.

Table S3 shows examples of the clusters that resulted from the dataset displayed in Table S2. Cluster 1 does not include any of the plasticity measures. Cluster 2, however, includes all of them, as well as the cognitive factors of vocabulary and task switching. As such, it fulfills the first criterion ("is there a factor that includes all four phonetic plasticity measures?") and the second ("is there a factor that includes only phonetic plasticity and other relevant cognitive factors, even if it not all of them?"), but not the third ("is there a factor that includes all the relevant measures, even if it includes some of the others?") or the fourth ("is there a factor that includes all of and only the relevant measures?"). We anticipated that the fourth criterion, especially, would be a high bar to clear, given our sample size and the complexity of the measures chosen.

Results

Figure S1 shows the results of the simulations that we performed. All simulations included the second criterion; there was always a factor that included at least some of the plasticity measures, even if it did not include all of them. Criteria 3 and 4 also ended up being interchangeable in the simulations that we performed; if a simulation included *all* the relevant factors, it also included *only* those relevant factors. As such, Figure S1 shows the most basic criterion ("was there a factor uncovered that included all four phonetic plasticity measures?") and the strictest ("was there a factor uncovered that included all of and only the relevant plasticity and cognitive measures?"). Note that, for the reasons spelled out previously, the possible configurations in the large effect size were constrained mathematically; as such, there were no available simulations where all seven cognitive variables were correlated with phonetic plasticity, and simulations with a high number of relevant cognitive variables reflected a smaller sample of possible correlations (e.g., the only mathematically feasible simulations with six cognitive measures included had either vocabulary or the words-in-noise measure as the odd measure out).

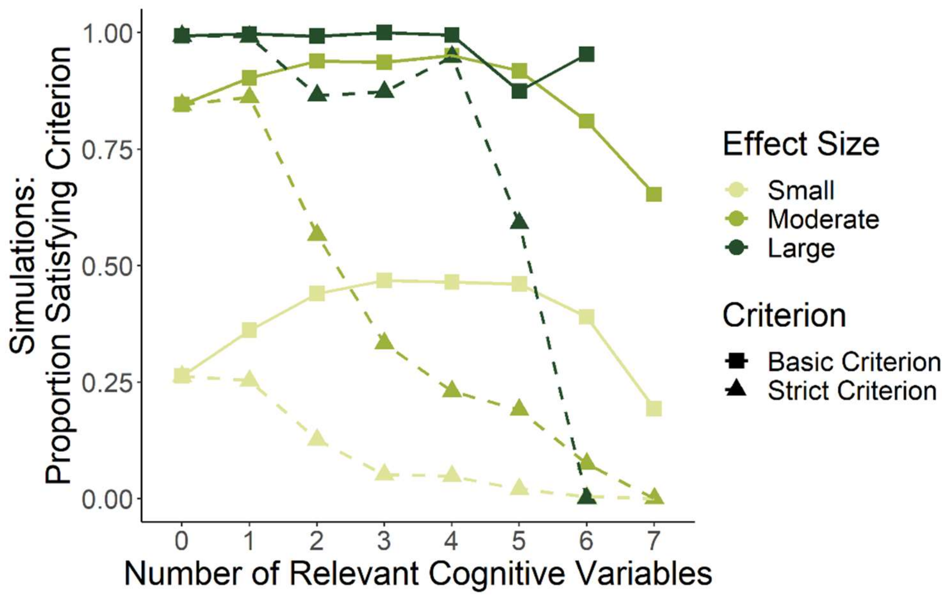

Figure S1. The likelihood of fulfilling the two criteria examined with a simulated small (unattenuated *r* = .30), moderate (unattenuated *r* = .50), and large (unattenuated *r* = .70) effect sizes according to the number of additional relevant cognitive measures included. The basic criterion indicates the percentage of models with a factor that includes all four measures of phonetic plasticity; the strict criterion indicates the proportion of models with a factor that includes those four measures plus all of and only the other relevant cognitive measures.

As is apparent from Figure S1, the likelihood of successfully reaching a criterion of success depends on the criterion, the number of relevant other variables, and the effect size. Even a moderate effect size led to a high likelihood of fulfilling the basic success criterion for all but the largest number of cognitive factors using the number of participants that were used in the present study. 95% of all simulations performed with four relevant cognitive factors led to a factor analysis that included a single factor with all four phonetic plasticity measures, correctly uncovering the underlying factor structure. However, the same was not true for uncovering a factor that linked all and only relevant cognitive factors to our measures of phonetic plasticity; this was only reliable for large numbers of relevant cognitive factors at a large effect size. This came despite the fact this modeling work made several limiting assumptions that worked against the likelihood of success. For instance, the likelihood of success was limited by the random selection of sets of relevant cognitive factors despite what is known about the structure of those factors in real-world contexts. As such, if we assume a moderate effect size, it seems likely that we can say with some certainty that the fact that incidental learning was not linked to the other phonetic plasticity measures is not the result of the sample size nor the modeling approaches chosen. However, it also seems likely, unless the true effect size linking these factors is large, that the factor analysis performed in the main experiment may be missing some of the cognitive factors that explain variance in phonetic plasticity.

References

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *Journal of the Acoustical Society of America*, *137*(4), 2015–2024. https://doi.org/10.1121/1.4916265

McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., & Van Engen, K. J. (2018). Coping with adversity: Individual differences in the perception of noisy and accented speech. *Attention, Perception, and Psychophysics*, *80*(6), 1559–1570. https://doi.org/10.3758/s13414-018-1537-4

Moineddin, R., Matheson, F. I., & Glazier, R. H. (2007). A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology*, *7*, 34. https://doi.org/10.1186/1471-2288-7-34

Mungas, D., Heaton, R., Tulsky, D., Zelazo, P. D., Slotkin, J., Blitz, D., Lai, J. S., & Gershon, R. (2014). Factor structure, convergent validity, and discriminant validity of the NIH toolbox cognitive health battery (NIHTB-CHB) in adults. *Journal of the International Neuropsychological Society*, *20*(6), 579–587. https://doi.org/10.1017/S1355617714000307

Nimon, K., Zientek, L. R., & Henson, R. K. (2012). The assumption of a reliable instrument and other pitfalls to avoid when considering the reliability of data. *Frontiers in Psychology*, *3*, 102. https://doi.org/10.3389/fpsyg.2012.00102

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2020). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. https://cran.r-project.org/web/packages/MASS/index.html

Rusticus, S. A., & Lovato, C. Y. (2014). Impact of sample size and variability on the power and type I error rates of equivalence tests: A simulation study. *Practical Assessment, Research and Evaluation*, *19*, 11.

Teare, M. D., Dimairo, M., Shephard, N., Hayman, A., Whitehead, A., & Walter, S. J. (2014). Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. *Trials*, *15*, 264.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., … Gershon, R. C. (2013). Cognition assessment using the NIH toolbox. *Neurology*, *80*(Suppl 3), S54–S64.

Zecker, S. G., Hoffman, H. J., Frisina, R., Dubno, J. R., Dhar, S., Wallhagen, M., Kraus, N., Griffith, J. W., Walton, J. P., Eddins, D. A., Newman, C., Victorson, D., Warrier, C. M., & Wilson, R. H. (2013). Audition assessment using the NIH Toolbox. *Neurology*, *80*(Suppl 3), S45–S48. https://doi.org/10.1212/wnl.0b013e3182872dd2