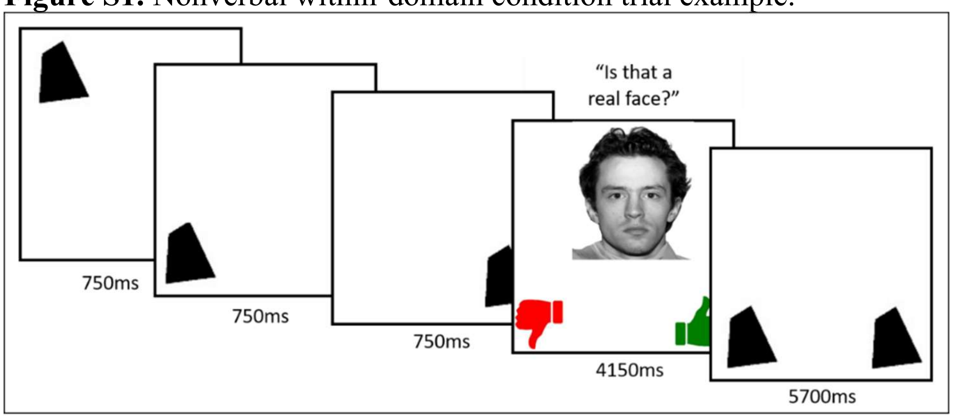
Supplemental Material S1. Additional methods.

Exceptions to Eligibility Criteria

In our participant sample, we included one participant in our DLD group who had a Core Language score of 85, but who also had a history of speech-language and reading difficulty and a scaled score of 6 (9th percentile) on the Recalling Sentences subscale of the Clinical Evaluation of Language Fundamentals, Fifth Edition (CELF-5; Wiig et al., 2013). This subscale is thought to be particularly sensitive to core deficits of DLD (Christensen, 2019; Conti-Ramsden et al., 2001; and see Oetting et al., 2016, for evidence of sentence recall diagnostic accuracy in nonmainstream dialects of English) and a Core Language composite score of 85 has a sensitivity of 1.0 and a specificity of .91. Further, we analyzed performance with and without this participant, and results were not meaningfully different. We included one DLD participant whose nonverbal *t*-score was 27, which was within one standard error measurement of our cutoff score of 30, and we included one TD participant whose nonverbal *t*-score was 32. For the TD participant, a *t*-score of 32 is below the average range of 40–60, but this participant had no history of intellectual disability, speech-language services, or parent reported developmental concerns. Performance patterns for these participants were not meaningfully different from performance of the DLD and TD group, respectively. Note that these participants are included in all group summary statistics.

Figure S1. Nonverbal within-domain condition trial example.



For the nonverbal within-domain condition (Figure S1), recall training involved three presentations of a single abstract shape which appeared in one of four possible pseudorandomized locations (i.e., quadrants) on the screen, with no repeated locations during a given trial's recall training phase. The interference processing phase involved a face/nonface presented for the duration of the interference processing phase and a question prompt ("Is that a real face?"). Participants had 2550 ms after the offset of the question prompt to respond prior to the onset of the test phase (i.e., allowing for variance in *free time*). This duration is 750 ms shorter due to the verbal interference processing phase having a second presentation of the stimulus. Specifically, the visual stimulus – face/nonface – was present throughout the interference phase whereas the verbal stimulus – word/nonword – was inherently fleeting; thus, the verbal interference phase required an additional item presentation in order to equate these tasks to a greater degree. This timing difference also accounts for potential processing-based differences between interference phase stimuli (e.g., processing verbal stimuli may be slower than nonverbal stimuli; Oberauer & Souza, 2020) and patterns observed in task piloting. For the recall test phase, the abstract shape reappeared in two of the three locations presented in recall training simultaneously and participants were to select the item location that occurred earlier in the recall training phase sequence. Participants had 5700ms (i.e., 5050 ms as in the verbal within-domain condition plus the duration of a word stimulus – 650 ms) after the onset of images to respond prior to the onset of the subsequent trial. For the divided-domain conditions, verbal recall was paired with nonverbal interference processing and nonverbal recall was paired with verbal interference processing.

The distributed-domain condition adopted features of the verbal and nonverbal conditions. This condition represents a step toward a more ecologically valid condition (i.e., items in the environment typically have some degree verbal *and* nonverbal features rather than verbal *or* nonverbal features) and a step toward testing domain as a continuum rather than category (e.g., domain is represented via distributed activation patterns in the SOB-CS; Oberauer et al., 2012). Recall training involved three presentations of a single word and associated image sequentially presented in one of four possible pseudorandomized locations on the screen; no location was repeated within a given trial’s recall training phase. The interference processing phase involved a word presented auditorily coming from either the left or right side of the laptop speakers, followed by a question prompt (“Where did you hear it?”), and then a second production of the word from the same side as the first production. Timing for the distributed-domain condition was the same as the verbal within-domain condition. Participants had 3300ms after the word offset to respond before advancing to the test phase, thus allowing for *free time* to vary by trial. For the recall test phase, the word and associated image simultaneously reappeared in two of the three locations presented in recall training and participants were to select the item location that occurred earlier in the recall training phase sequence (i.e., testing item location plus serial position). Participants had 5050ms after the offset of the auditorily presented word to respond prior to the onset of the subsequent trial.

Table S1. Verbal recall stimuli comparisons.

Feature	Verbal-Within	Verbal-Divided	Distributed	
Syllables				VW < VD p = .84
Mean (SD)	1.36 (0.48)	1.38 (0.53)	1.47 (0.52)	VD > D p = .56
Range	1–2	1–3	1–2	VW < D p = .47
Frequency				VW < VD p = .31
Mean (SD)	0.004 (0.005)	0.003 (0.003)	0.003 (0.003)	VD < D p = .84
Range	0.001–0.024	0.001–0.014	0.001–0.010	VW < D p = .24
Category	NA	NA	NA	VW – VD p = .84 VD – D p = .57 VW – D p = .68
Source	NA	NA	NA	VW – VD p = 1.00 VD – D p = .73 VW – D p = .73

Note. Frequency based on the CHILDES database (MacWhinney, 2000; Warren corpora); Categories included: Animals (e.g., rabbit), Toys (e.g., ball), Places (e.g., jungle), Outdoors (e.g., tree), Food (e.g., sandwich), Small Household Items (e.g., camera), Large Household Items (e.g., TV), Body Parts (e.g., head), Clothing (e.g., shirt), People (e.g., nurse), Vehicles (e.g., bus), and Routines (e.g., dinner). Sources included: CHILDES database (MacWhinney, 2000; Warren corpora; Warren-Leubecker & Bohannon, 1984) or the MacArthur-Bates Communicative Development Inventories (Fenson et al., 2007).

Table S2. Verbal interference phase stimuli comparisons.

Feature	**Verbal-Within**	**Nonverbal-Divided**	
Syllables (all)			
Mean (*SD*)	1.33 (0.49)	1.07 (0.26)	VW > NVD p = .07
Range	1–2	1–2	
Frequency (nonwords)			
Mean (*SD*)	259.12(297.92)	666.17(737.83)	VW < NVD p = .18
Range	40–917	0–1741	
Category (true words)	NA	NA	VW-NVD p = .80
Source (true words)	NA	NA	VW-NVD p = .80

Note. Words and nonwords collapsed for syllable count; Frequency for words was only available for two stimuli, thus it was not analyzed, and frequency for nonwords is the summed frequency of phonological neighbors based on the ARC Nonword Database (Rastle et al., 2002). Categories included: Animals, Toys, Places, Outdoors, Food, Small Household Items, Large Household Items, Body Parts, Clothing, People, Vehicles, and Routines. Sources included: CHILDES database (MacWhinney, 2000; Warren corpora; Warren-Leubecker & Bohannon, 1984) or the MacArthur-Bates Communicative Development Inventories (Fenson et al., 2007).

References

Christensen, R. V. (2019). Sentence repetition: A clinical marker for developmental language disorder in Danish. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4450–4463. https://doi.org/10.1044/2019_JSLHR-L-18-0327

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Fensen, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communication Development Inventories*. Brookes.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Erlbaum.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review, 19,* 779–819. https://doi.org/10.3758/s13423-012-0272-4

Oberauer, K., & Souza, A. S. (2020). How fast can people refresh and rehearse information in working memory? *Memory and Cognition, 48,* 1442–1459. https://doi.org/10.17605/OSF.IO/MCXDN

Oetting, J. B., McDonald, J. L., Seidel, C. M., & Hegartya, M. (2015). The development of English as a second language with and without specific language impairment: Clinical implications. *Journal of Speech, Language, and Hearing Research*, *24*(2), 1–14. https://doi.org/10.1044/2015_JSLHR-L-15-0008

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *55*(4), 1339–1362. https://doi.org/10.1080/02724980244000099

Warren-Leubecker, A., & Bohannon, J. N. (1984). Intonation patterns in child-directed speech: Mother-father differences. *Child Development, 55*(4), 1379–1385.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence–Second Edition*. NCS Pearson.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals–Fifth Edition*. Pearson.