

Supplemental Material S1. Procedures, language profiles, and decision processes used to classify children as DLD or TD.

The language samples examined in the current study have been analyzed as part of several studies and as part of several groupings and subgroupings depending on the research questions and samples available at the time of analysis. Labels used to classify the children across studies have varied, with SLI often used to describe the children classified in this study as DLD. Criteria and cut scores (e.g., at or below -1 *SD* vs. at or above -1 *SD*) for individual tests for the DLD and TD groups also have varied across studies. For the age- and language-matched groups, some studies matched children individually, but as we began matching children based on their dialects and combining samples across studies, matching was at the group level, with follow-up tests completed to confirm the group-based matches. Finally, before 2010, either raw score on a vocabulary measure or MLU in morphemes was used to identify the language-matched controls based on the research questions and views of these measures. Post hoc, the two measures are correlated ($r = .58$, $p < .001$), and we often focus on MLU as the matching variable (for additional support for this decision, see DeThorne et al., 2005).

Nevertheless, the samples were elicited during three distinct time periods. Below, we describe the procedures, language profiles, and decision processes used to classify children as DLD or TD for each period, with added focus on the children with a test score that placed them at the upper or lower border of their assigned group (i.e., a score at -1 *SD*). During recruitment and initial testing, care was taken to accurately classify children as either DLD or TD because misclassifications would make it extremely difficult to identify group differences. Post hoc, all studies conducted have identified group differences between those classified as either DLD or TD.

First Time Period of Data Collection: 1994–1998 (93 Language Samples)

The language samples were collected for a set of word learning studies and a study on past tense (Horohov, 1999; Horohov & Oetting, 2004; Oetting, 1999, 2005; Oetting & Horohov, 1997; see also, Oetting et al., 1996, 1997). Of the 93 children, 31 were classified as DLD, 31 as age-matched TD6 controls, and 31 as language-matched TD4 controls (60% male, 40% female; 40 AAE speakers, 53 SWE speakers).

All children completed the Peabody Picture Vocabulary Test–Revised (PPVT-R; Dunn & Dunn, 1987) and the Goldman-Fristoe Test of Articulation (GFTA; Goldman & Fristoe, 1986). All children but two in the TD4 group completed the Columbia Mental Maturity Scale (CMMS; Burgemeister et al., 1972), and all but 12 in the TD4 group completed the syntax subtest of the Test of Language Development Primary–Second Edition (TOLD-P2; Newcomer & Hammill, 1988). Those who did not complete the TOLD-P2 were too young for the test.

To safeguard against misdiagnosis, we tested children in the spring semester to allow for correction of any clinical errors that may have occurred at the beginning of the school year and to give children time to acclimate to school and testing formats. We also interviewed the children's teachers and speech-language pathologists (SLPs) to confirm that they had language concerns for those with DLD relative to their classroom peers and the children's teachers to confirm that those classified as TD were performing

within the average range relative to their classroom peers. As reported by Oetting and Cleveland (2006), 25 (21%) children who were recruited for the TD groups and whose teachers had no concerns earned a standard score significantly below -1 SD on either the PPVT-R ($n = 16$), TOLD-P2 ($n = 16$), and/or CMMS ($n = 6$); these children were not selected for a TD group, and we interpreted their results as consistent with previous descriptions of assessment tools as culturally and linguistically biased. Data for the 93 children by dialect and group can be found in Table 1 of Oetting and McDonald (2001).

All Children Met the Following Criteria

- Lived in a native, monolingual English-speaking home per teacher report.
- Passed a hearing screening within 6 months of data collection.
- Did not present frank neurological impairments or socioemotional deficits per teacher report.
- Earned above the 16th percentile and demonstrated adequate articulation of /s, z, d, t/ in singletons and in consonant clusters in the word final position as measured by the GFTA.

DLD Group. All children received services for language by an SLP at school. All earned a standard score at or above -1 SD on the CMMS, and all scored below 85 (-1 SD) on the PPVT-R and TOLD-P2, except four (2 AAE, and 2 SWE). The four exceptions included two who earned an 85 on the PPVT-R and three who earned either an 85, 89, or 91 on the TOLD-P2. Three of these children earned a Developmental Sentence Score (DSS) that was at or below the 10th percentile based on normative data provided by Lee (1974). The DSS score of the fourth child fell between the 10th and 25th percentile, and two of his TOLD-P2 subtest scores were below -1 SD. These children were classified as DLD based on these test data and the teacher and SLP interviews.

TD6 Group. None of the children were receiving services and did not present a history of services by an SLP. They were enrolled in a classroom with a child with DLD, and matching was completed at the individual level by age (± 5 months). All earned a standard score at or above -1 SD on the CMMS, and all scored above 85 (-1 SD) on the PPVT-R and TOLD-P2, except one SWE-speaking child who earned 85 on the PPVT-R (and a 98 on the TOLD-P2) and one SWE-speaking child who earned an 83 on the TOLD-P2 (and a 92 on the PPVT-R). Both children were classified as TD given the higher score of the two tests and teacher interviews.

TD4 Group. None of the children were receiving services and did not present a history of services by an SLP. They were enrolled in Head Start or childcare centers that fed into the public schools of those classified as DLD and matched to a child with DLD by either raw PPVT-R score or MLU in morphemes depending on the study. All but two completed the CMMS, and all who completed it scored at or above -1 SD. All children also earned a standard score above 85 (-1 SD) on the PPVT-R, and of the 19 who were old enough to complete the syntax subtest of the TOLD-P2, all earned a standard score above 85 (-1 SD), except one AAE-speaking child who earned a 79 (and a 99 on the PPVT-R) and one SWE-speaking child who earned an 85 (and a 90 on the PPVT-R). Both children were classified as TD given the higher score of the two tests and teacher interviews.

Second Time Period of Data Collection: 2005–2007 (47 Language Samples)

The language samples were collected as part of two dissertations focused on either regular and irregular past tense (Pruitt, 2006; Pruitt & Oetting, 2009, 2011) or auxiliary *am*, *is*, and *are* (Garrity, 2007; Garrity & Oetting, 2010). The Pruitt study included 45 AAE-speaking children (15 classified as TD from a low SES background, 15 as TD from a middle SES background, selected to be an age match to a TD child in the low SES group, and 15 as TD from a middle SES background selected to be a language match to a TD child in the low SES group). Samples from the TD low SES group were not included in the current study because we did not have an equivalent SWE group. The Garrity study included 30 AAE-speaking children (10 classified as DLD, 10 as TD age-matched controls, and 10 as TD language-matched controls). Also, 13 children participated in both studies, which led to 47 language samples for the current analysis. All 47 children were classified as speakers of AAE (34% male, 66% female). Data for these children by group can be found in Table 1 of Oetting et al. (2010).

All children completed two nonverbal subtests of the Leiter International performance Scale-Revised (Leiter-R; Roid & Miller, 1998), the PPVT-3 (Dunn & Dunn, 1997), TOLD-P3 (Newcomer & Hammill, 1997), and an articulation screener. The Pruitt study included the TOLD-P3 for descriptive purposes, whereas the Garrity study included the TOLD-P3 to help classify the children as DLD or TD.

To safeguard against misdiagnosis, the 10 with DLD came from a pool of 37 children receiving services by an SLP at school and who returned a consent form; 27 of these children were excluded from the DLD group based on other diagnoses, low scores on the Leiter-R and/or the articulation screener, and/or high scores on the PPVT-3 and TOLD-P3. Children classified as TD across studies came from a pool that included more than 150 children; not all children were tested because testing was discontinued when the TD matches were identified (for details, see appendices in original dissertations).

All Children Met the Following Criteria

- Passed a hearing screening within 6 months of data collection.
- Classified as a speaker of AAE.
- Earned a standard score at or above -1 SD on the nonverbal subtests of the Leiter-R.
- Demonstrated adequate articulation of /m, z, r/ and/or /d, t/ in the word final position as measured on screeners created for the original studies.

DLD Group. All children received services for language by an SLP at school. All children also scored below 85 (-1 SD) on both tests, except one who earned 87 on the PPVT-3 (and 57 on the TOLD-P3). This child remained classified as DLD given his TOLD-P3 score and discussion with his SLP.

TD6 Group. None of the children were receiving services and did not present a history of services by an SLP. At the time of the study, the children were selected if they presented an age that was within 4 months of a child with DLD (Garrity, 2007) or within 2 months of a child in the low-SES TD group (Pruitt, 2006). All children also scored above 85 (-1 SD) on the PPVT-3 and TOLD-P3.

TD4 Group. None of the children were receiving services and did not present a history of services by an SLP. Raw score (± 7 points) on the PPVT-Third Edition (PPVT-3; Dunn & Dunn, 2005) served as the matching variable for the Pruitt study, and MLU (± 0.66 morphemes) served as the matching variable for the Garrity study. All children earned a standard score above 85 (-1 SD) on the on the PPVT-3 and TOLD-P3, except three from the Pruitt study who earned an 83, 83, and 85 on the TOLD-P3 (with PPVT-3 scores of 91, 91, and 93). Recall that the TOLD-P3 was not used to classify children in the Pruitt study so they were classified as TD based on the PPVT-3 scores. For other analyses and these here, these children remained classified as TD based on their PPVT-3 scores.

Third Time Period of Data Collection: 2010–2014 (106 Language Samples)

The language samples were collected for a study examining children's tense and agreement systems. The children included 70 who spoke AAE and 36 who spoke SWE (49% male, 51% female). Data for these children by their dialect and group can be found in Table 1 of Oetting et al. (2016).

All children completed the Primary Test of Nonverbal Intelligence (PTONI; Ehrler & McGhee, 2008), the syntax subtest of the Diagnostic Evaluation of Language Variation: Norm-Reference (DELV-NR; Seymour et al., 2005), PPVT-4 (Dunn & Dunn, 2007), GFTA-3 (Goldman & Fristoe, 2000), and the phonological probe of the Test of Early Grammar Development (TEGI; Rice & Wexler, 2001), with the PPVT-4 administered for descriptive purposes. For additional information about the children's dialects, all children also completed the Diagnostic Evaluation of Language Variation – Screener (DELV-ST; Seymour et al., 2003). Finally, during the end of the spring semester of each year, teachers were asked to complete the Teacher Rating of Oral Language and Literacy (TROLL; Dickenson et al., 2003); we ended up not receiving many of them for the 106 participants, but we had these data for some and examined them at a later date with other children (Gregory & Oetting, 2018).

To safeguard against misdiagnosis, the children's clinical status was determined through a review of standardized test scores, family/school histories, and discussions with the school SLPs and teachers. Data collection spanned five years and included consent forms returned for over 600 children; not all of these children were tested, met the eligibility criteria, and/or were selected given that match design of the study.

All Children Met the Following Criteria

- Passed a hearing screening within 6 months of data collection.
- Earned a standard score at or above 82 (-1.2) on the PTONI.
- Earned a standard score above 85 (-1 SD) on the GFTA-3 and adequate articulation of /s, z, d, t/ in the word final position as measured by the phonological probe of the TEGI.

DLD Group

All children earned a standard score at or below 7 (-1 SD) on the syntax subtest of the DELV-NR. Given that a standard score of 7 reflects a -1 SD on the syntax subtest of the DELV-NR, we also considered the PPVT-4 scores of these children. Of the five children with DLD (2 AAE; 3 SWE) who earned a 7, their PPVT-4 scores were

71, 79, 84, 85, and 86. One of these children received services for language by a school SLP and reported a positive family history for speech or language impairment, two received services by a school SLP but did not report a positive family history of speech or language impairment, and two were not receiving services by a school SLP but reported a positive family history of speech or language impairment. On the DELV-ST, three of these children scored high risk for impairment and one scored medium risk. The TOLD-P3 was added to the battery toward the end of the data collection period and four of these children had TOLD-P3 standard scores, which were 78, 80, 80, and 86. From these data and discussions with the SLPs and teachers, these children were classified as DLD. Finally, although not considered at the time of the original studies, four of these children had TROLLs, and their standardized z scores on this tool were -1.37 , -1.66 , -0.30 , and 0.11 using the mean from the children in the study with these scores (Gregory & Oetting, 2018).

TD6 Group. None of the children were receiving services and did not present a history of services by an SLP. All earned a standard score at or above 8 (-0.67 SD) on the syntax subtest of the DELV-NR, and produced a dialect that matched that of an AAE- or SWE-speaking child with DLD, and within each dialect then matched a DLD child on age, PTONI score, and when possible maternal education. All children also scored above 85 (-1 SD) on the PPVT-4 (Dunn & Dunn, 2005) except one AAE-speaking child who earned an 85 (and a DELV-NR score of 9). This child did not have a positive family history of language impairment, scored low risk on the DELV-ST, earned a 109 on the TOLD-P3, and the child's teacher was not concerned about her language abilities relative to classroom peers. This child was classified as TD6.

References

- Burgmeister, B., Blum, H., & Lorge, I. (1972). *Columbia Mental Maturity Scale*. Psych Corp.
- DeThorne, L. S., Johnson, B. W., & Loeb, J. W. (2005). A closer look at MLU: What does it really measure? *Clinical Linguistics and Phonetics*, 19(8), 635–648. <https://doi.org/10.1080/02699200410001716165>.
- Dickinson, D. K., & McCabe, A., & Sprague, K. (2003). Teacher Rating of Oral Language and Literacy development (TROLL): Individualizing early literacy instruction with a standards-based rating tool. *The Reading Teacher*, 56(6), 554–569.
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised*. American Guidance Service.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third edition*. American Guidance Service.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test-Fourth Edition*. Psych Corp.
- Ehrler, D. J., & McGhee, R. L. (2008). *Primary Test of Nonverbal Intelligence*. Pro-Ed.
- Garrity, A. W. (2007). *BE use in African American English speakers as a function of language ability* [Unpublished doctoral dissertation]. Louisiana State University.
- Goldman, R., & Fristoe, M. (1986). *Goldman Fristoe Test of Articulation*. American Guidance Service.
- Goldman, R., & Fristoe, M. (2000). *Goldman Fristoe Test of Articulation-Second edition*. American Guidance Service.
- Gregory, K. D., & Oetting, J. B. (2018). Use of teacher ratings when screening nonmainstream English-speaking kindergartners for language impairment. *Language, Speech, and Hearing Services in Schools*, 49(2), 218–231. https://doi.org/10.1044/2017_LSHSS-17-0045
- Horohov, J. E. (1999). *Input manipulations, working memory, and word learning abilities of children* [Unpublished doctoral dissertation]. Louisiana State University.
- Horohov, J., & Oetting, J. B. (2004). Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Journal of Applied Psycholinguistics*, 25(1), 43–65. <https://doi.org/10.1017/S0142716404001031>

- Newcomer, P. L., & Hammill, D. D. (1982). *Test of Language Development-Primary*. Pro-ed.
- Newcomer, P. L., & Hammill, D. D. (1988). *Test of Language Development-Primary* (2nd edition). Pro-Ed.
- Newcomer, P. L., & Hammill, D. D. (1997). *Test of Language Development-Primary* (3rd edition). Pro-Ed.
- Oetting, J. B. (1999). Children with SLI make use of argument structure cues to learn novel verbs. *Journal of Speech, Language, and Hearing Research*, 42(5), 1261–1274. <https://doi.org/10.1044/jslhr.4205.1261>
- Oetting, J. B. (2005). Assessing language in children who speak a nonmainstream dialect of English. In M. Ball (Ed.), *Clinical sociolinguistics* (pp. 180–192). Blackwell.
- Oetting, J., Cantrell, J., & Horohov, J. (1999). A study of specific language impairment (SLI) in the context of nonstandard dialect. *Clinical Linguistics and Phonetics*, 13(1), 25–44. <https://doi.org/10.1080/026992099299220>
- Oetting, J. B., & Cleveland, L. H. (2006). The clinical utility of nonword repetition for children in the rural South of the United States. *Clinical Linguistics and Phonetics*, 20(7-8), 553–561. [https://doi.org/10.1044/1058-0360\(2010/09-0064\)](https://doi.org/10.1044/1058-0360(2010/09-0064))
- Oetting, J., & Horohov, J. (1997). Past tense marking by children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 40(1), 62–74. <https://doi.org/10.1044/jslhr.4001.62>
- Oetting, J. B., Horohov, J., & Costanza, A. (1996). Two aspects of past tense marking by children with specific language impairment. In T. Powell (Ed.), *Pathologies of speech and language: Contributions of clinical phonetics and linguistics* (pp. 65–74). ICPLA.
- Oetting, J. B., & McDonald, J. L. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research*, 44(1), 207–223. [https://doi.org/10.1044/1092-4388\(2001/018\)](https://doi.org/10.1044/1092-4388(2001/018))
- Oetting, J. B., Newkirk, B. L., Hartfield, L. R., Wynn, C. G., Pruitt, S. L., & Garrity, A. W. (2010). Index of Productive Syntax for children who speak African American English. *Language, Speech, and Hearing Services in Schools*, 41(3), 328–339. [https://doi.org/10.1044/0161-1461\(2009/08-0077\)](https://doi.org/10.1044/0161-1461(2009/08-0077))
- Oetting, J. B., Rivière, A. M., Berry, J. R., Gregory, K. D., Villa, T. M., & McDonald, J. L. (2021). Marking of tense and agreement in language samples by children with and without specific language impairment in African American English and Southern White English: Evaluation of scoring approaches and cut scores across structures. *Journal of Speech, Language, and Hearing Research*, 64(2), 491–509. https://doi.org/10.1044/2020_JSLHR-20-00243
- Pruitt, S. L. (2006). *Grammatical morphology of children reared in poverty: Implications for specific language impairment* [Unpublished doctoral dissertation]. Louisiana State University.
- Pruitt, S. L., & Oetting, J. B. (2009). Past tense marking by African American English-speaking children reared in poverty. *Journal of Speech, Language, and Hearing Research*, 52(1), 2–15. [https://doi.org/10.1044/1092-4388\(2008/07-0176\)](https://doi.org/10.1044/1092-4388(2008/07-0176))
- Rice, M. L., & Wexler, K. (2001). *Rice/Wexler Test of Early Grammatical Impairment*. University of Kansas. Psych Corp.
- Seymour, H. N., Roeper, T. W., & deVilliers, J. (2003). *Diagnostic Evaluation of Language Variation-Screening Test*. Ventriss Learning.
- Seymour, H. N., Roeper, T. W., & deVilliers, J. (2005). *Diagnostic Evaluation of Language Variation-Norm Referenced*. Ventriss Learning.