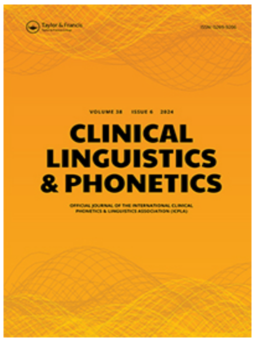

Clinical Linguistics & Phonetics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/iclp20

The grammaticality judgement of inflectional morphology in children with and without Developmental Language Disorder

Samuel D. Calder, Denis Visentin, Mary Claessen, Lillian Hollingsworth, Susan Ebbels, Karen Smith-Lock & Suze Leitão

To cite this article: Samuel D. Calder, Denis Visentin, Mary Claessen, Lillian Hollingsworth, Susan Ebbels, Karen Smith-Lock & Suze Leitão (21 Jul 2023): The grammaticality judgement of inflectional morphology in children with and without Developmental Language Disorder, Clinical Linguistics & Phonetics, DOI: 10.1080/02699206.2023.2236768

To link to this article: https://doi.org/10.1080/02699206.2023.2236768

Published online: 21 Jul 2023.

CLINICAL LINGUISTICS & PHONETICS
https://doi.org/10.1080/02699206.2023.2236768

The grammaticality judgement of inflectional morphology in children with and without Developmental Language Disorder

Samuel D. Calder[a,b], Denis Visentin[a], Mary Claessen[c], Lillian Hollingsworth[d], Susan Ebbels[e,f], Karen Smith-Lock[b], and Suze Leitão[b]

[a]Health Sciences, College of Health and Medicine, University of Tasmania, Launceston, Australia; [b]School of Allied Health, Faculty of Health Sciences, Curtin University, Perth, Australia; [c]Speech Pathology Australia, Melbourne, Australia; [d]Therapy Focus, Perth, Australia; [e]Moor House Research and Training Institute, Moor House School & College, Oxted, UK; [f]Language and Cognition, Division of Psychology and Language, University College London, London, UK

ABSTRACT

To explore the clinical potential of grammaticality judgement tasks, this study investigated whether a Grammaticality Judgment Task (GJT) of inflectional morphology could differentiate between a clinically selected sample of children with DLD and children in mainstream (i.e. regular education) schools. We also explored the relationship between grammaticality judgement and measures of receptive vocabulary, receptive grammar, and nonword repetition. Children with DLD ($n = 30$; age range = 69–80 months) and mainstream children in Pre-primary, Year 1, and Year 2 (n = 89, age range = 61–96 months) were assessed on a GJT of regular past tense, third person singular, and possessive *'s*. The GJT was sensitive to developmental differences in mainstream children and differentiated children with DLD from Year 1 and 2 mainstream children, with DLD results consistent with a one-year delay in performance compared to controls. The GJT was the strongest discriminator of membership to a clinically selected sample of children with DLD (ROC curve analysis, area under the curve = 88%). Receptive grammar, receptive vocabulary, and nonword repetition were related to performance on the GJT. The grammaticality judgement of inflectional morphology shows promise as a reliable indicator of DLD and a measure sensitive to developmental differences in mainstream children. GJTs should continue to be explored for clinical application as a potential tool for both assessment and intervention.

ARTICLE HISTORY
Received 26 January 2023
Revised 8 July 2023
Accepted 10 July 2023

KEYWORDS
Developmental Language Disorder; grammaticality judgement; morphosyntax

Introduction

Roughly 7% of the population in English-speaking countries is affected by Developmental Language Disorder (DLD) (Calder, Brennan-Jones, et al., 2022; Norbury et al., 2016; Tomblin et al., 1997). Children with DLD have a slower pace of language development and difficulty producing and understanding language (Bishop et al., 2016). Their language is characterised by morphosyntax difficulties, including tense-related morphology (e.g. Rice et al., 1999). In fact, difficulty in the acquisition of finite marking on verbs is recognised as a distinct aetiological construct (D. V. Bishop et al., 2006; Rice & Wexler, 1996; Rice et al.,

CONTACT Samuel D. Calder samuel.calder@utas.edu.au Private Bag 19, nipaluna/Hobart, Tasmania 7001, Australia

1995) and predictive marker of language growth for DLD (Rice, 2012; Rice et al., 1998), and as a clinical marker of language disorder from preschool age to adolescence (Dale et al., 2018; Rice et al., 2009; see also Leonard, 2014 for a comprehensive overview). Although the literature is dominated by attention to finiteness marking, studies have indicated that children with DLD also have difficulty with nominal marking (e.g. Biedermann et al., 2021; Leonard et al., 2015), including possessive *'s* in English-speaking children (Calder, Claessen, et al., 2022; Tomas et al., 2015). As expected, difficulties with morphosyntax are a primary source of concern upon referral (Bishop & Hayiou-Thomas, 2008).

Grammaticality judgement tasks

Grammaticality judgement tasks (GJTs) have a long history in child language research, particularly to track the language development of children with DLD (aka Specific Language Impairment (SLI)) longitudinally (e.g. Dale et al., 2018; Rice et al., 1999, 2009). Essentially, GJTs require a listener to determine if a sentence is grammatically *right* or *wrong*. Instead of simply producing or comprehending a sentence, the child is asked to explicitly reflect on the grammatical form of a production. As such, the judgement task draws on a child's meta-linguistic awareness; their ability to reflect upon the structural features of spoken language (Tunmer & Herriman, 1984). Metalinguistic awareness may play a role in the acquisition of language, as it allows a child to self-monitor their output and thereby check and correct their errors (c.f. Marshall & Morton, 1978 for a detailed outline of how this might work). While important for typically developing children, the ability to reflect on the grammaticality of a production is particularly relevant for a child with DLD in a therapy context, where they must note the grammatical features being targeted and monitor their own usage of the targets.

In addition to requiring *meta*-linguistic skill, the ability to judge grammaticality requires linguistic competence. A child cannot judge the grammaticality of an utterance if they do not know the grammar in the first place. There is evidence that performance on GJTs in typical speakers differs with age and the grammatical target being judged (Gottardo et al., 1996; McDonald, 2008), as would be expected as grammatical competence develops. McDonald (2008, p. 254) outlined a rough order of mastery of judgement tasks at school-age (6–11 years): 1) word order (e.g. *The teacher graded the tests* vs *The teacher the tests graded*), 2) present progressive – ing (e.g. *The little girl is playing with her dolls* vs *The little girl is play with her dolls*), 3) omitted determiners and auxiliaries (e.g. *The lady drove the same car for the past twenty years* vs *The lady drove same car for the past twenty years*), 4) agreement (e.g. *The boy jumps whenever he is startled* vs *The boy jump whenever he is startled*) and 5) irregular forms (e.g. *Several of the men decided not to go to the football game* vs *Several of the mans decided not to go to the football game; Last week the pilot flew to Paris* vs *Last week the pilot flied to Paris*). Because GJTs reflect grammatical competence, they have been used to identify individuals with difficulties learning language (e.g. Rice et al., 1999, 2009).

In addition to linguistic competence and metalinguistic awareness, GJTs require the child to hold the target utterance in memory long enough to judge its grammaticality. There is some evidence that phonological working memory is significantly correlated with the ability to judge and correct ungrammatical sentences in eight-year-olds (Gottardo et al., 1996; McDonald, 2008). Furthermore, phonological working memory has also been associated with grammatical measures such as receptive syntax in seven-year-olds (Ellis Weismer et al., 1999). Indeed, tasks that tap into phonological memory, such as nonword repetition tasks, have shown to be useful for identifying children with DLD (e.g. Graf Estes et al., 2007). Therefore, given that children with DLD are known to have difficulty with phonological working memory (Ellis Weismer et al., 1999), we might expect children with DLD to also demonstrate difficulty with GJTs.

Of course, a child cannot judge an utterance if they cannot perceive all the relevant grammatical information. Davies et al. (2017) found typical speakers as young as 24 months demonstrate sensitivity to plural marking as measured by looking behaviour, but only to lexical items marked with [s] (e.g. *cat**s***) and not [z] (e.g. *dog**s***) allomorphs. Davies et al. used acoustic analysis to argue [s] is the most perceptually salient plural allomorph due to its extended frication period; whereas [z] is less salient because vowels are longer before voiced consonants and therefore have a shorter word duration to frication duration ratio (Davies et al., 2017). The syllabic plural allomorph [əz] (e.g. *bus**es***) appeared to be detected only at 36 months (Davies et al., 2020), with sensitivity to plural marking across all allomorphs established at 48 months. Even at 48 months, however, children did not show sensitivity to singular novel words baring no affixes as they did not elicit systematic looking behaviour (Davies et al., 2019). These perceptual data may suggest that even for typical speakers, the ability to judge grammaticality of incorrect sentences (e.g. *There are two bus**) will emerge later than the ability to judge the grammaticality of correct sentences (e.g. *There are two bus**es***).

The use of GJTs in DLD research is widespread (e.g. Dale et al., 2018; Rice, 2012; Rice et al., 1999, 2009; Smith-Lock, 1995; Wulfeck et al., 2004). Smith-Lock (1995) compared 5–7-year-old children with DLD to age-matched and language-matched typical speakers and found children with DLD had grammaticality judgement skills like those of younger children matched for expressive language. She concluded that the ability to judge and repair morphosyntactic errors developed alongside expressive language skills in both typically developing children and children with DLD. In an experimental study, Wulfeck et al. (2004) demonstrated that 7–12-year-old children with DLD performed significantly lower than typically developing children and children with focal brain lesions on a GJT manipulating determiner and auxiliary use, suggesting an underlying pervasive pathology in DLD.

In longitudinal research, Rice et al. (1999) tracked three groups of children ($n = 21$ 6; 0-year-old children with DLD (then SLI); $n = 19$ language-matched children; $n = 21$ aged-matched children) over 2 years, measuring performance on a GJT testing optional infinitives and tense agreement. Findings indicated children with DLD rejected ungrammatical morphosyntax that they themselves could produce correctly but did not reject ungrammatical morphosyntax that reflected errors they were likely to commit. The same groups of children were tracked until 15 years-old (Rice et al., 2009), and findings indicated that children with DLD performed below language- and age-matched peers on a GJT at all nine testing points. Furthermore, the language- and age-matched groups reached ceiling on the task, whereas children with DLD reached a lower asymptote. Rice et al.'s findings suggest that GJTs testing knowledge of finiteness marking may hold promise as a tool to identify DLD in childhood and adolescence. Finally, Dale et al. (2018) demonstrated that a short form of 20 items on a GJT could reliably detect the heritability of DLD in a twin study of 16-year-olds, further contributing evidence for the use of GJTs to identify DLD.

Despite the use of GJTs in experimental settings, translation to clinical practice has been sparse. To this end, we developed a clinical tool to assess the ability to judge the grammaticality of morphosyntactic structures known to be challenging for children with DLD. For the tool to be clinically useful, we need to understand how typically developing children perform on the task, how their performance compares to the performance of children with DLD, and what factors contribute to successful performance. We identified the objectives below.

(1) To determine whether there are age-related differences in the ability to judge grammaticality in preschool and early school-aged children from a mainstream population of children in a regular education setting.
(2) To determine whether there is a difference in the ability to judge grammaticality in children with DLD compared to mainstream children, and if so, to determine whether the difference is related to the morphemes and respective allomorphs being judged.
(3) To determine whether the ability to judge grammaticality predicted membership of a clinically selected sample of children with DLD.
(4) To determine whether measures of receptive vocabulary, receptive grammar, and/or nonword repetition are related to performance on the GJT.

Methods and materials

Ethical approval was obtained from the Curtin University Human Research Ethics Committee (Approval number: HRE2017–0835), the Western Australian Department of Education (Approval number: D190018955), and Catholic Education Western Australia (Approval number: RP2018/52).

Participants

This study used a cross-sectional design sampling children from both selected clinical and unselected mainstream contexts (see Table 1). Participants included $n = 30$ Year 1 children diagnosed with DLD (mean age = 74.70 months, SD = 3.21, range = 69–81 months; 66.67% males) representing a clinically selected sample recruited from specialised educational programmes in Perth, Western Australia. Children attending the specialised programmes met criteria for DLD as outlined by Bishop et al. (2016), including persistent difficulties in the acquisition and use of language where they are unlikely to catch up to their peers; language difficulties create obstacles to communication and/or learning in everyday life; and the language disorder is not associated with another known biomedical condition, such as: sensorineural hearing loss, autism spectrum disorder, intellectual disability. All participants with DLD spoke English as their primary language.

Mainstream controls included $n = 89$ children in a regular education setting from Pre-Primary ($n = 31$, mean age = 67.61 months, SD = 4.00, range = 61–76 months, 35.5% male), Year 1 ($n = 31$, mean age = 76.39 months, SD = 3.96, range = 69–83 months, 54.8% male), and Year 2 ($n = 27$, mean age = 88.59 months, SD = 4.25, range = 82–96 months, 41.2% male) grade levels in Perth, Western Australia. All children were recruited from a catholic primary school. Teachers reported that English was the children's primary language and the

Table 1. Descriptive statistics by group.

Variable	DLD (*n* = 30) Mean (SD)	Pre-primary (*n* = 31) Mean (SD)	Year 1 (*n* = 31) Mean (SD)	Year 2 (*n* = 27) Mean (SD)
Age	74.70 (3.21)	67.61 (4.00)	76.39 (3.96)	88.59 (4.25)
TROG 2	5.83 (2.49)	6.03 (3.16)	10.29 (3.90)	11.37 (4.07)
PPVT 4	95.20 (18.38)	97.19 (16.11)	113.48 (18.19)	127.70 (19.11)
NRT PPC	0.73 (0.10)	0.76 (0.10)	0.82 (0.10)	0.86 (0.10)
GJT *A′*				
Total score	0.53 (0.09)	0.56 (0.15)	0.71 (0.16)	0.78 (0.14)
ED	0.54 (0.10)	0.58 (0.18)	0.73 (0.17)	0.81 (0.17)
3s	0.53 (0.09)	0.66 (0.18)	0.77 (0.19)	0.87 (0.16)
's	0.54 (0.10)	0.52 (0.12)	0.67 (0.17)	0.71 (0.17)
ED [t]	0.53 (0.16)	0.61 (0.21)	0.77 (0.22)	0.84 (0.19)
ED [d]	0.56 (0.14)	0.61 (0.21)	0.78 (0.20)	0.85 (0.18)
ED [əd]	0.50 (0.11)	0.55 (0.17)	0.68 (0.18)	0.78 (0.18)
3s [s]	0.56 (0.14)	0.65 (0.20)	0.76 (0.21)	0.88 (0.18)
3s [z]	0.53 (0.16)	0.69 (0.25)	0.81 (0.24)	0.88 (0.17)
3s [əz]	0.54 (0.11)	0.65 (0.20)	0.75 (0.20)	0.87 (0.18)
's [s]	0.53 (0.11)	0.51 (0.17)	0.67 (0.18)	0.67 (0.16)
's [z]	0.57 (0.15)	0.55 (0.16)	0.70 (0.18)	0.70 (0.19)
's [əz]	0.57 (0.13)	0.54 (0.14)	0.68 (0.20)	0.75 (0.19)

NRT PPC = Nonword Repetition Test Percent Phonemes Correct (Dollaghan & Campbell, 1998); TROG-2 = Test for Reception of Grammar-2 (D. Bishop, 2003); PPVT-4 = Peabody Vocabulary Picture Test-4 (Dunn & Dunn, 2007). All scores are reported as raw scores across measures.

GJT *A′*= grammaticality judgement task *A′*. *ED* = regular past tense; *3s* = third person singular; *'s* = possessive 's. All GJT scores are reported as *A′*.

children had typical language and learning skills. This sample represented an unselected mainstream sample of children, which may have included some children with DLD who were not in receipt of services through the specialised educational programme. No additional information on participant ethnicity in either group was collected.

Measures

Descriptive statistics for all measures across participants are reported in Table 1. Receptive grammar was measured using the Test for Reception of Grammar-2 (TROG-2; D. Bishop, 2003). The TROG-2 was normed on 792 children aged 4–16 years in the UK and Australia, and it has strong internal consistency reliability ($r = .877$) and appropriate construct validity (D. Bishop, 2003). The Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007) was administered as a measure of receptive vocabulary. The PPVT-4 was normed on 3,540 individuals aged 2;6 to 90 years and older. It is reported to have strong internal consistency reliability ($r = .93$) and appropriate construct validity (Dunn & Dunn, 2007). Dolloghan and Campbell's (1998) nonword repetition task (NRT) was used as a measure which draws upon phonological memory (Gathercole, 1995) and other phonological skills representing the complex interplay between phonology and verbal working memory (see Archibald & Gathercole, 2006). The NRT was scored for percentage phonemes correct (PPC). Participants repeated 16 nonwords with four items of each syllable length, ranging from one to four syllables. The nonwords were recorded prior to administration and delivered via laptop to ensure consistency of the verbal stimuli. All scores are reported as raw scores across measures.

Grammaticality judgement task

The structure of the grammaticality judgement task (GJT) was based on the Grammar Elicitation Test developed by Smith-Lock and Leitão (Smith-Lock et al., 2013) and included morphosyntactic structures: regular past tense (*ED*), third person singular (*3s*), and possessive (*'s*). These structures were selected as they are known to be areas of weakness for children with DLD. There was a total of 90 items on the GJT which consisted of three subtests of 30 items for each morpheme. Each subtest contained equal numbers of voiceless, voiced, and syllabic allomorphs. Within each subtest, half (i.e. 15 items) were grammatical, and half were ungrammatical. Similarly, for each allomorph, half (i.e. five items) were grammatical, and half were ungrammatical. Two test sets were created, with items randomly ordered. If an item was grammatical in one set, it was ungrammatical in the other. Half of the participants were randomly assigned to one set, and half to the other. This ensured that no participant received both a grammatical and ungrammatical version of the same stimulus.

For each item, participants were presented with a visual stimulus (video or still) and an accompanying sentence. Audio stimuli were recorded by an Australian English-speaking female blinded to the purpose of the research. Live-action videos accompanied *ED* and *3s* stimuli. For *ED*, the action in the video was carried out, then the frame was frozen, and the audio stimulus began, making a past tense structure felicitous. For *3s* videos, audio was introduced, while the action was taking place on screen. In addition, the video faded to black so the child would not see a frozen image suggestive of a completed action, so a present tense structure was felicitous. For the *'s* subtest, still images of nouns depicting ownership were retrieved from copyright-free image sources. Audio was introduced as soon as the image was shown, and the image remained visible until the child responded. Example items are presented in Table 2, including the mean length of the audio stimuli for both grammatical and non-grammatical pairs.

Procedure. The GJT was administered via PowerPoint through a laptop. All children were asked to judge the grammaticality of the sentences by the experimenter saying, 'I want you to tell

Table 2. Example targets and probes for structures on the GJT including mean length in milliseconds.

Target structures	Allomorphs	Example targets	Grammatical	Ungrammatical
*ED*30 items	10 × [d]: voiced segmental	Yawned	*The man* ***yawned***.	*The man* ***yawn****.
		Clapped	*The man* ***clapped***.	*The man* ***clap****.
	10 × [t]: voiceless segmental	Painted	*The girl* ***painted*** *a picture.*	*The girl* ***paint**** *a picture*
	10 × [əd]: syllabic		Mean length: 1527 ms (range: 1000–2400 ms, SD: 351 ms)	Mean length: 1527 ms (range: 800-2300 ms, SD: 378 ms)
3s 30 items	10 × [z]: voiced segmental	Pours	*The woman* ***pours*** *the water.*	*The woman* ***pour**** *the water.*
		Laughs	*The boy* ***laughs***.	*The boy* ***laugh****.
	10 × [s]: voiceless segmental	Sneezes	*The man* ***sneezes***.	*The man* ***sneeze****.
	10 × [əz]: syllabic		Mean length: 1594 ms (range: 1100–2100 ms, SD: 265 ms)	Mean length: 1612 ms (range: 1200–2200 ms, SD: 303 ms)
's 30 items	10 × [z]: voiced segmental	Frog's	*The frog has a lily pad. This is the* ***frog's*** *lily pad.*	*The frog has a lily pad. This is the* ***frog**** *lily pad.*
		Cat's		
	10 × [s]: voiceless segmental	Mouse's	*The cat is on this chair. It is the* ***cat's*** *favourite chair.*	*The cat is on this chair. It is the* ***cat**** *favourite chair.*
	10 × [əz]: syllabic		*The mouse has cheese. This is the* ***mouse's*** *cheese.*	*The mouse has cheese. This is the* ***mouse**** *cheese.*
			Mean length: 3321 ms (range: 2100– 4200 ms, SD: 447 ms)	Mean length: 3097 ms (range: 2100–4100 ms, SD: 450 ms)

ED = regular past tense; *3s* = third person singular; *'s* = possessive 's.

me if the sentence sounds right to you. If it does, push the "yes" button. If the sentence doesn't sound right, push the "no" button'. All children were provided one demonstration item and five practice trials which included grammatical and ungrammatical versions of the same three sentences. Feedback was provided for practice items. If the child contested that a sentence was grammatical/ungrammatical based on semantic reasoning (e.g. 'The man didn't grab the glass. He picked it up'.), the examiner explained, 'She will always say the right word. Do the sentences sound right to you?' This feedback for semantic errors (only) was allowed for the first four test items. No participants persisted with semantic judgements. Children were permitted to take breaks during this task as needed. Non-specific verbal and non-verbal positive feedback was given every four to five trials.

Participant responses were recorded as 'yes' or 'no', which was noted as 'correct' or 'incorrect'. These scores were then converted to *A'* to account for 'yes bias' using the formula outlined by Rice et al. (1999): $A' = (0.5 + (y - x)\ (1 + y - x)/4y\ (1 - x))$, where y = 'yes' for grammatical items and x = 'yes' for ungrammatical items. Therefore, a score of 1.00 would indicate 'yes' responses to all grammatical items and 'no' response to all ungrammatical items. Whereas a score of 0.50 would indicate 'yes' responses to both grammatical and ungrammatical items suggesting the child performed at chance, and a score below 0.50 indicates a tendency to respond 'no' to items.

Psychometric properties. Since the GJT is an experimental task, we report on brief analyses of reliability. Internal consistency reliability was assessed for both sets across and between morphemes. Cronbach's alpha for the 90-item Total Score on one set GJT was .93, indicating excellent internal consistency. Further calculations indicated that internal consistency for items in the *ED* (.87) and *3s* (.89) subtests were good, and the *'s* (.76) subtest was acceptable. For the second set, Cronbach's alpha for the 90-item Total Score was .95, which is excellent. Internal consistency for the *ed* (.90) and *3s* (.91) subtests were also excellent, and good for the *'s* (.86) subtest.

Intra-class correlation coefficient (ICC) was calculated to assess the test–retest reliability of the GJT by retesting $n = 20$ of the mainstream sample. ICC estimates and their 95% confidence intervals were based on absolute agreement, two-way mixed-effects model. ICC = .91, 95% CI [0.65, 0.97], which is excellent.

Results

Objective 1: To determine if there are age-related differences in the ability to judge grammaticality across mainstream children from Pre-Primary to Year 2

Linear regression was used to determine the relationship between age and performance on the GJT Total Score, and morpheme and allomorph subtests and other measures for the control group participants. Results from the regression for the effect of age across all measures in children without DLD are reported in Table 3. The age-effect for non-DLD children indicated that an increase in age of 12 months corresponds to an increase of 0.125 on the GJT *A'* Total Score ($p <$ 0.001). As the values for GJT *A'* range from 0 to 1, with a score of 0.5 indicating results no better than chance, this increase demonstrates a clinically relevant increase in performance with age. Similar significant age effects were found for the *ED* (0.134), *3s* (0.109), and *'s* (0.110) subtests (all $p < 0.001$).

Table 3. Regression for effect of age (in months) amongst mainstream participants.

Independent variables	*β*	SE	*t*	*p*	±95% CI
TROG 2	0.254	0.041	6.140	<0.001	[0.172, 0.3361]
PPVT 4	1.378	0.196	7.040	<0.001	[0.989, 1.766]
NRT PPC	0.004	0.001	3.270	0.002	[0.001, 0.006]
GJT *A'*					
Total score	0.010	0.002	6.300	<0.001	[0.007, 0.014]
ED	0.011	0.002	5.840	<0.001	[0.007, 0.015]
3s	0.009	0.002	4.570	<0.001	[0.005, 0.013]
's	0.009	0.002	5.400	<0.001	[0.006, 0.013]

NRT PPC = Nonword Repetition Test Percent Phonemes Correct (Dollaghan & Campbell, 1998); TROG-2 = Test for Reception of Grammar-2 (D. Bishop, 2003); PPVT-4 = Peabody Vocabulary Picture Test-4 (Dunn & Dunn, 2007).
GJT *A'*= grammaticality judgement task *A'*. *ED* = regular past tense; *3s* = third person singular; *'s* = possessive 's.

Objective 2: To determine if there is a difference in the ability to judge grammaticality between mainstream children and children with DLD

Linear regression was used to test the between-group differences on all measures as the dependent variables for DLD and control participants with DLD status as the independent variable accounting for age as a covariate. Age adjustment is required as the control participants range from Pre-Primary to Year 2 children. Results are presented in Table 4, where ***β*** values indicate the difference in scores between children with DLD and the mainstream controls adjusted for age, i.e. negative ***β*** values indicate poorer performance by the DLD group. Results from regression analyses adjusted for age indicated that children with DLD scored 0.12 points lower on the GJT Total Score (expressed as *A'*), which is consistent with DLD participants being 1 year behind mainstream controls, with a mean score for the DLD participants close to chance scoring only (mean = 0.53). When results on the GJT are analysed for morphemes and allomorphs, children with DLD scored lower on all subtests and appear to be further behind controls on the *3s* subtest and less behind on the *'s* subtest. The effect sizes of allomorphs within each morpheme are similar, which indicates that differences in performance may not be meaningful, especially when considering the overlap in 95% confidence intervals. Scores from all individual cases on the GJT *A'* Total

Table 4. Age-adjusted regression results for DLD group compared to mainstream controls across all dependent variables.

Dependent variables	*β*	SE	*t*	*p*	±95% CI
TROG 2	−2.72	0.72	−3.76	<0.001	[−4.16, −1.29]
PPVT 4	−13.92	3.80	−3.66	<0.001	[−21.45, −6.38]
NRT PPC	−0.07	0.02	−3.36	0.001	[−0.11, −0.03]
GJT *A'*					
Total score	−0.12	0.03	−4.36	<0.001	[−0.18, −0.07]
ED	−0.13	0.03	−4.08	<0.001	[−0.20, −0.07]
3s	−0.20	0.03	−6.06	<0.001	[−0.27, −0.14]
's	−0.07	0.03	−2.35	0.020	[−0.13, −0.01]
ED [t]	−0.18	0.04	−4.36	<0.001	[−0.26, −0.10]
ED [d]	−0.15	0.04	−3.85	<0.001	[−0.22, −0.10]
ED [əd]	−0.15	0.03	−4.38	<0.001	[−0.21, −0.08]
3s [s]	0.18	0.04	−4.67	<0.001	[−0.26, −0.10]
3s [z]	−0.24	0.04	−5.30	<0.001	[−0.32, −0.15]
3s [əz]	−0.20	0.04	−5.27	<0.001	[−0.27, −0.12]
's [s]	−0.06	0.03	−1.85	0.067	[−0.13, 0.00]
's [z]	−0.06	0.04	−1.72	0.088	[−0.13, 0.01]
's [əz]	−0.06	0.03	−1.75	0.082	[−0.13, 0.01]

β values indicate the difference in score between the DLD group and mainstream controls adjusting for age.

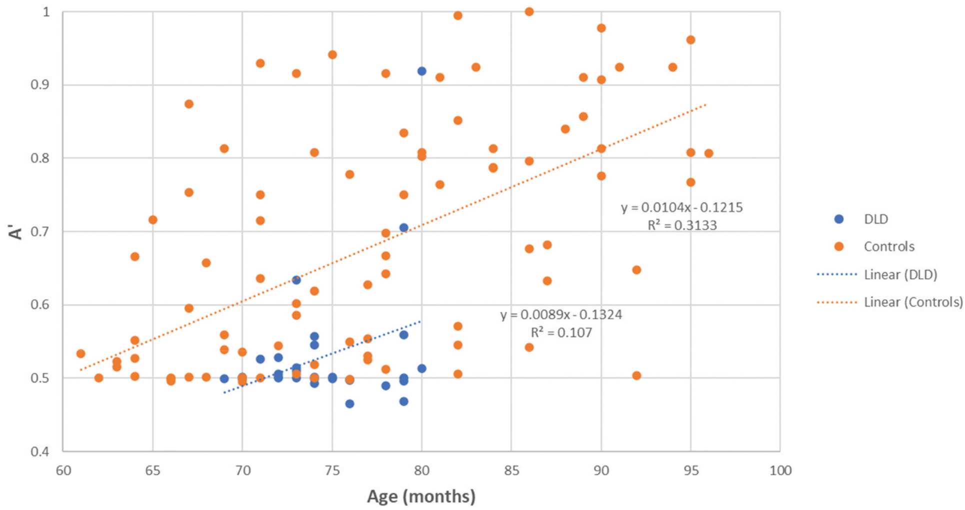

Figure 1. Scatter plot of performance on GJT *A′* total score for DLD and mainstream controls.

Score have been plotted (see Figure 1), demonstrating that while there is an overlap in individual scores between the two groups, the development of DLD participants is below the mainstream controls.

Objective 3: To determine whether the GJT predicts membership to a clinically selected sample of children with DLD

Receiver-operating characteristic (ROC) curve analyses were used to plot the sensitivity against the specificity of the GJT across a range of potential cut-offs to determine the predictive utility of the task where 0.5 (i.e. 50%) indicates no discrimination, 0.7 (i.e. 70%) to 0.8 (i.e. 80%) is considered *acceptable*, 0.8 to 0.9 (i.e. 90%) is considered *excellent*, and above 0.9 is considered *outstanding* (Hosmer & Lemeshow, 2000) (see Figure 2). The area under the curve = 88% indicating the GJT is an *excellent* discriminator for DLD when compared to the mainstream sample. ROC curves are also presented for the TROG-2 (area under the curve = 86%), PPVT-4 (area under the curve = 76%), and the NRT (area under the curve = 79%), suggesting the TROG-2 is also an *excellent* discriminator for DLD, and the PPVT-4 and NRT are *acceptable* discriminators. The choice of cut-off scores for clinical use (and hence sensitivity and specificity) would depend in part on the prevalence of DLD in the sample to which any test was applied to optimise discrimination.

Objective 4: To determine whether measures of receptive vocabulary, receptive grammar, and/or nonword repetition are related to performance on the grammaticality judgement task

Correlation analysis was used to assess the relationship between the GJT and measures of receptive grammar (TROG-2), receptive vocabulary (PPVT-4) and nonword repetition (NRT) for all participants. The results are presented in Table 5 and indicate that there is a moderate to strong relationship between age, all measures, and the GJT.

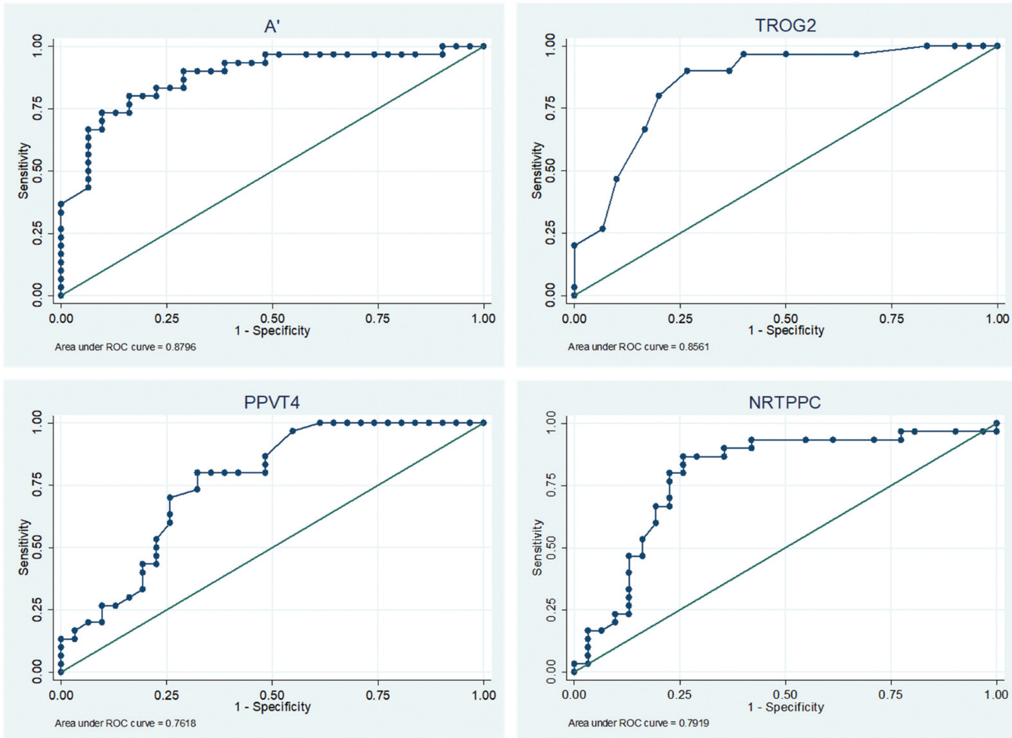

Figure 2. ROC curves for GJT *A′* total score, TROG-2, PPVT-4, and NRT PPC.

Table 5. Correlation matrix between the GJT and other measures for all participants.

	Age	GJT *A′* total score	GJT *A′ 3s*	GJT *A′ ′s*	GJT *A′ ED*	TROG-2	PPVT-4
GJT *A′* total score	0.547*						
GJT *A′ 3s*	0.430*	0.858*					
GJT *A′ ED*	0.501*	0.890*	0.640*				
GJT *A′ ′s*	0.524*	0.933*	0.828*	0.745*			
TROG-2	0.527*	0.599*	0.594*	0.465*	0.605*		
PPVT-4	0.523*	0.490*	0.471*	0.399*	0.478*	0.650*	
NRT PPC	0.309	0.478*	0.478*	0.349*	0.470*	0.439*	0.414*

NRT PPC = Nonword Repetition Test Percent Phonemes Correct (Dollaghan & Campbell, 1998); TROG-2 = Test for Reception of Grammar-2 (D. Bishop, 2003); PPVT-4 = Peabody Vocabulary Picture Test-4 (Dunn & Dunn, 2007).
GJT *A′*= grammaticality judgement task *A′*. *ED* = regular past tense; *3s* = third person singular; *′s* = possessive ′s.

Discussion

The current study aimed to investigate the potential clinical utility of a GJT of inflectional morphology (*ED*, *3s*, *′s*). We investigated whether there were developmental differences in mainstream children's ability to detect grammaticality, and whether a GJT could predict membership to a clinically selected sample of children with DLD. We also investigated whether measures of receptive vocabulary, receptive grammar, and/or nonword repetition were related to performance on a GJT.

We found that the GJT *A'* Total Score was sensitive to developmental differences in mainstream children aged between 61 and 86 months (Objective 1). The age-effect for mainstream children shows that an increase of 12 months corresponds to an increase of 0.12 on the GJT. Items marked with *ED* and *3s* were judged with greater accuracy than those marked with possessive *'s* in all groups. This effect was surprising as previous research has found that performance on GJTs was consistent with expressive language skill and that children with DLD demonstrate persistent difficulty judging the grammaticality of verb morphology (Dale et al., 2018; Rice et al., 1999, 2009). As expressive use of *'s* is typically acquired earlier than finiteness marking (Brown, 1973), we might expect the ability to judge grammaticality of *'s* to similarly precede the ability to judge the grammaticality of finiteness markers. The poorer performance on *'s* could be due to the difference in task stimuli. Audio stimuli in the *ED* and *3s* subtests were accompanied by video, whereas the *'s* audio stimuli were accompanied by still images. The video may have been more engaging and/or felicitous than the still images. Alternatively, utterances were longer in the *'s* condition, which may have increased the demands on phonological memory (e.g. Gottardo et al., 1996; McDonald, 2008). This explanation is supported by the findings of the correlation analyses which found that nonword repetition was related to performance on the GJT Total Score.

Age-adjusted regression analyses indicated that the GJT differentiated Year 1 DLD children (mean age = 74.70 months) from mainstream children (age range = 61–96 months) (Objective 2). Analyses showed that the DLD group performed significantly lower on all variables with the exception of the *'s* allomorphs. However, since the performance on the GJT *'s* subtest overall was significant, the former finding is likely not meaningful. The results from the analyses broadly show that the performance by this sample of Year 1 children with DLD is delayed by one year on all variables, consistent with findings from Smith-Lock (1995). This may suggest that children with DLD experience a delay in the development of language, phonological memory, and ultimately the metalinguistic awareness necessary for detecting grammatical or ungrammatical structures (Gottardo et al., 1996; McDonald, 2008). However, the mainstream children did show differing performance across morphemes, whereas children with DLD in Year 1 performed roughly at chance across all morphemes according to *A'*. Therefore, most children with DLD at Year 1 age may simply not possess grammaticality judgement skills like Pre-Primary age children. Rather, most children with DLD may have performed at chance since they have difficulty applying metalinguistic skills to judge grammaticality.

Our ROC curve analyses indicated that the ability to judge grammaticality was an *excellent* discriminator between children with and without DLD (area under the curve = 88%) (Objective 3) which was similarly reliable to the TROG-2 (area under the curve = 86%), and more reliable than the PPVT-4 (area under the curve = 76%) and the NRT (area under the curve = 79%). Although nonword repetition tasks have been shown to be reliable indicators of DLD (then SLI) in some clinical populations (e.g. Graf Estes et al., 2007), Ebbels et al. (2012) found that adolescents with DLD (then SLI) demonstrated bimodal distribution on a nonword repetition task, which suggests that not all language difficulties associated with DLD are related to factors underlying nonword repetition, such as phonological memory.

Similar to previous findings, receptive grammar and nonword repetition were correlated with GJT *A'* Total Scores, suggesting these are related to grammaticality judgement skills (Ellis Weismer et al., 1999; Gottardo et al., 1996; McDonald, 2008) (Objective 4). Of note, the PPVT-4 as a measure of static receptive vocabulary was also related to performance on the GJT; however, the strength of the relationship between receptive vocabulary was less than the relationship between receptive grammar and nonword repetition.

Limitations

One limitation to this study is the risk of ascertainment bias. Although mainstream participants were recruited with reference to minimal eligibility criteria and hence relatively unselected, all participants attended the same catholic primary school. Therefore, this sample may not represent the wider mainstream public-school population. The DLD sample was clinically selected from a public specialised school and therefore may also increase the risk of ascertainment bias, as these children may not represent all children with DLD, especially those who may be unidentified in mainstream schools.

Further, the use of area under the curve from the ROC curve analysis of the GJT to identify children with DLD should be considered. The cut-off to be used (hence the sensitivity and specificity values) is dependent on the population the test is to be used on. Given the prevalence of DLD is 7% (Calder, Brennan-Jones, et al., 2022; Norbury et al., 2016; Tomblin et al., 1997), using the GJT to screen for children would likely result in a high number of false positives if it were applied to the general population. The purpose of the ROC curve analysis in the current study was to evaluate the potential utility of the GJT as a method of discriminating children who had previously been identified to have DLD from those in mainstream settings, which may indicate better potential for diagnosis than screening.

The GJT in the current study targeted two structures of verbal inflection and one of the nominal inflection, which may have resulted in an incomplete profile of grammaticality judgement of inflectional morphology. Future studies should explore judgement of more grammatical structures, including plural structures to further represent nominal morphology, as well as syntactic structures. Future research might also benefit from considering children's expressive morphosyntactic skill. The evaluation of both grammaticality judgement and production of the same structures would allow exploration of the relationship between these potentially reciprocal skills (e.g. Rice et al., 1999).

Lastly, only one age group of children of DLD was represented in the current study compared to three in the mainstream group. Although this allowed for evaluation of age-related differences in grammaticality judgement of mainstream children, age-matched groups across all age ranges would allow for a richer comparison between children with DLD and mainstream children.

Clinical implications and future research

Given that the performance on the GJT significantly predicted membership to a clinically selected sample of children with DLD, future research should evaluate the use of grammaticality judgement tasks as screeners for DLD at school age. Future research may explore the results of ROC curve analyses in further detail to determine their utility for diagnostic purposes. Specifically, the GJT investigated in the current study could be assessed in combination with other language measures, such as the TROG-2, to discriminate children with DLD from the mainstream population. The GJT may also be assessed as a short form reduced from the total 90 items to improve time efficiency and compared to other assessment tools to determine what it might add to accurate assessment and diagnosis of DLD.

The use of interventions that involve grammaticality judgement should also be investigated. The cueing hierarchy found to be effective in grammar intervention (Smith-Lock et al., 2013, 2015) included requests of the child to evaluate their production and to choose between a grammatical and ungrammatical production after an error. For example, if the child omitted *'s* and said, 'Jack hat' instead of 'Jack's hat', one step in the feedback provided asked, 'Jack hat?'. If that cue was unsuccessful, the child was provided with a forced choice of, 'is it Jack's hat or Jack hat?'. These cues involve implicit grammaticality judgement. Perhaps this type of feedback is more useful for children with good grammaticality judgement skills and is potentially amenable to intervention. Given the relationship between grammaticality judgement, receptive grammar, and nonword repetition, viable interventions may include those that aim to reduce demands on phonological memory and syntactic complexity through visual supports and metalanguage. Calder et al. (2020, 2021) tested such an intervention and found that briefly focusing upon grammaticality judgement did not result in improvements on a GJT. Nevertheless, increasing dose intensity may result in an observable effect. Improving the ability to judge grammaticality might improve self-monitoring skills and the usefulness of therapeutic cues. This could assist children to move from occasionally correct to consistently correct production, as well as to use morphemes as cues to interpret syntax correctly.

Conclusion

To explore the clinical potential of grammaticality judgement tasks, we tested children with DLD and children in mainstream classes on a Grammaticality Judgment Task as well exploring the relationships between measures of receptive grammar, receptive vocabulary, and nonword repetition. Performance on the GJT increased with age in the mainstream children, and the GJT was the most reliable discriminator for membership to a clinically selected group of children with DLD. Therefore, grammaticality judgement shows promise as a clinical tool to identify children with DLD. Further research should directly evaluate its use in the assessment and diagnosis of DLD and its use in grammar intervention.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Australian Research Training Programme.

ORCID

Samuel D. Calder http://orcid.org/0000-0001-6064-5837
Denis Visentin http://orcid.org/0000-0001-9961-4384
Mary Claessen http://orcid.org/0000-0002-1087-5041
Susan Ebbels http://orcid.org/0000-0003-0402-6227

References

Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, *41*(6), 675–693. https://doi.org/10.1080/13682820500442602

Biedermann, B., Fielder, N., & Smith-Lock, K. (2021). Nominal number and language pathologies. In P. C. Hofherr & J. Doetjes (Eds.), *The Oxford handbook of number in language*. Oxford University Press.

Bishop, D. (2003). *Test for reception of grammar-2*. The Psychological Corporation.

Bishop, D. V., Adams, C. V., & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. *Genes, Brain, and Behavior*, *5*(2), 158–169. https://doi.org/10.1111/j.1601-183X.2005.00148.x

Bishop, D. V. M., & Hayiou-Thomas, M. E. (2008). Heritability of specific language impairment depends on diagnostic criteria. *Genes, Brain, and Behavior*, *7*(3), 365–372. https://doi.org/10.1111/j.1601-183X.2007.00360.x

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Schiller, N. O. (2016). CATALISE: A multinational and multidisciplinary delphi consensus study. Identifying language impairments in children. *PLOS ONE*, *11*(7), e0158753. https://doi.org/10.1371/journal.pone.0158753

Brown, R. (1973). *A first language: The early stages*. Harvard University Press. https://doi.org/10.4159/harvard.9780674732469

Calder, S. D., Brennan-Jones, C. G., Robinson, M., Whitehouse, A., & Hill, E. (2022). The prevalence of and potential risk factors for Developmental Language Disorder at 10 years in the Raine Study. *Journal of Paediatrics and Child Health*, *58*(11), 2044–2050. https://doi.org/10.1111/jpc.16149

Calder, S. D., Claessen, M., Ebbels, S., & Leitão, S. (2020). Explicit grammar intervention in young school-aged children with Developmental Language Disorder: An efficacy study using single case experimental design. *Language, Speech, and Hearing Services in Schools*, *51*(2), 298–316. https://doi.org/10.1044/2019_LSHSS-19-00060

Calder, S. D., Claessen, M., Ebbels, S., & Leitão, S. (2021). The efficacy of an explicit intervention approach to improve past tense marking for early school-aged children with Developmental Language Disorder. *Journal of Speech, Language, & Hearing Research*, *64*(1), 91–104. https://doi.org/10.1044/2020_JSLHR-20-00132

Calder, S. D., Claessen, M., Leitão, S., & Ebbels, S. (2022). A profile of expressive inflectional morphology in early school-aged children with Developmental Language Disorder. *Clinical Linguistics & Phonetics*, 1–18. https://doi.org/10.1080/02699206.2021.1931454

Dale, P. S., Rice, M. L., Rimfeld, K., & Hayiou-Thomas, M. E. (2018). Grammar clinical marker yields substantial heritability for language impairments in 16-year-old twins. *Journal of Speech, Language, & Hearing Research*, *61*(1), 66–78. https://doi.org/10.1044/2017_JSLHR-L-16-0364

Davies, B., Rattanasone, N. X., & Demuth, K. (2017). Two-year-olds' sensitivity to inflectional plural morphology: Allomorphic effects. *Language Learning and Development*, *13*(1), 38–53. https://doi.org/10.1080/15475441.2016.1219257

Davies, B., Rattanasone, N. X., Schembri, T., & Demuth, K. (2019). Preschoolers' developing comprehension of the plural: The effects of number and allomorphic variation. *Journal of Experimental Child Psychology*, *185*, 95–108. https://doi.org/10.1016/j.jecp.2019.04.015

Davies, B., Xu Rattanasone, N., & Demuth, K. (2020). Acquiring the last plural: Morphophonological effects on the comprehension of /-əz/. *Language Learning and Development*, *16*(2), 161–179. https://doi.org/10.1080/15475441.2020.1717956

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, & Hearing Research*, *41*(5), 1136–1146. https://doi.org/10.1044/jslhr.4105.1136

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Ebbels, S. H., Dockrell, J. E., & van der Lely, H. K. (2012). Non-word repetition in adolescents with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, *47*(3), 257–273. https://doi.org/10.1111/j.1460-6984.2011.00099.x

Ellis Weismer, S., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, & Hearing Research*, *42*(5), 1249–1260. https://doi.org/10.1044/jslhr.4205.1249

Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, & Hearing Research*, *50*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, *23*(1), 83–94. https://doi.org/10.3758/BF03210559

Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *Journal of Experimental Child Psychology*, *63*(3), 563–582. https://doi.org/10.1006/jecp.1996.0062

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley and Sons. https://doi.org/10.1002/0471722146

Leonard, L. B. (2014). *Children with specific language impairment*. The MIT Press. https://doi.org/10.7551/mitpress/9152.001.0001

Leonard, L. B., Kunnari, S., Savinainen-Makkonen, T., Tolonen, A.-K., Mäkinen, L., Luotonen, M., & Leinonen, E. (2015). Noun case suffix use by children with specific language impairment: An examination of Finnish. *Applied Psycholinguistics*, *35*(4), 833–854. https://doi.org/10.1017/S0142716412000598

Marshall, J. C., & Morton, J. (1978). On the mechanics of EMMA. In A. Sinclair, R. J. Jarvella, & W. Levelt (Eds.), *The child's conception of language* (pp. 225–239). Springer-Verlag.

McDonald, J. L. (2008). Grammaticality judgments in children: The role of age, working memory and phonological ability. *Journal of Child Language*, *35*(2), 247–268. https://doi.org/10.1017/S0305000907008367

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Rice, M. L. (2012). Toward epigenetic and gene regulation models of specific language impairment: Looking for links among growth, genes, and impairments. *Journal of Neurodevelopmental Disorders*, *4*(1), 1–14. https://doi.org/10.1186/1866-1955-4-27

Rice, M. L., Hoffman, L., & Wexler, K. (2009). Judgments of omitted BE and DO in questions as extended finiteness clinical markers of specific language impairment (SLI) to 15 years: A study of growth and asymptote. *Journal of Speech, Language, & Hearing Research*, *52*(6), 1417–1433. https://doi.org/10.1044/1092-4388(2009/08-0171)

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, & Hearing Research*, *39*(6), 1239–1257. https://doi.org/10.1044/jshr.3906.1239

Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech, Language, & Hearing Research*, *38*(4), 850–863. https://doi.org/10.1044/jshr.3804.850

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, & Hearing Research*, *41*(6), 1412–1431. https://doi.org/10.1044/jslhr.4106.1412

Rice, M. L., Wexler, K., & Redmond, S. M. (1999). Grammaticality judgements of an extended optional infinitive grammar: Evidence from English-speaking children with specific language impairment. *Journal of Speech, Language, & Hearing Research*, *42*(4), 943–961. https://doi.org/10.1044/jslhr.4204.943

Smith-Lock, K. M. (1995). Morphological usage and awareness in children with and without specific language impairment. *Annals of Dyslexia*, *45*(1), 163–185. https://doi.org/10.1007/BF02648217

Smith-Lock, K. M., Leitao, S., Lambert, L., & Nickels, L. (2013). Effective intervention for expressive grammar in children with specific language impairment. *International Journal of Language and Communication Disorders*, *48*(3), 265–282. https://doi.org/10.1111/1460-6984.12003

Smith-Lock, K. M., Leitão, S., Prior, P., & Nickels, L. (2015). The effectiveness of two grammar treatment procedures for children with SLI: A randomized clinical trial. *Language, Speech & Hearing Services in Schools*, *46*(4), 312–324. https://doi.org/10.1044/2015_LSHSS-14-0041

Tomas, E., Demuth, K., Smith-Lock, K. M., & Petocz, P. (2015). Phonological and morphophonological effects on grammatical development in children with specific language impairment. *International Journal of Language and Communication Disorders*, *50*(4), 516–528. https://doi.org/10.1111/1460-6984.12152

Tunmer, W. E., & Herriman, M. L. (1984). The development of metalinguistic awareness: A conceptual overview. *Metalinguistic Awareness in Children: Theory, Research, and Implications*, 12–35.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, & Hearing Research*, *40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245

Wulfeck, B., Bates, E., Krupa-Kwiatkowski, M., & Saltzman, D. (2004). Grammaticality sensitivity in children with early focal brain injury and children with specific language impairment. *Brain and Language*, *88*(2), 215–228. https://doi.org/10.1016/S0093-934X(03)00100-7