Applied Psycholinguistics (2020), **41**, 237–258
doi:10.1017/S0142716419000493

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Evaluating Part V of the German version of the Token Test as a screening of specific language impairment in preschoolers

Michaela Schmoeger*, Matthias Deckert, Brigitte Eisenwort, Benjamin Loader, Annemarie Hofmair, Eduard Auff and Ulrike Willinger

Medical University of Vienna
*Corresponding author. Email: michaela.schmoeger@meduniwien.ac.at

(Received 29 August 2018; revised 08 October 2019; accepted 10 October 2019; first published online 03 December 2019)

Abstract

Specific language impairment (SLI) is a very common childhood disorder that is characterized by impairments in expressive and/or receptive language regarding different modalities. Part V of the German version of the Token Test was evaluated as a potential screening tool for the early detection of SLI. Forty-five male and 16 female monolingual native German-speaking preschoolers with SLI (4–6 years) and 61 age- and gender-matched typically developing controls were examined with a German version of the Token Test and an established intelligence measure. Token Test performance was significantly worse in preschoolers with SLI including greater group differences at age 4 than at ages 5 and 6. Analyses showed a detection rate of 77% for Part V of the Token Test in the whole sample as well as 85.1% at age 4, 80.6% at age 5, and a nonsignificant detection at age 6. Correctly detected preschoolers with SLI showed significantly worse performance than typically developing controls regarding nonverbal and verbal intelligence, numeracy, problem solving, working memory, visual attention, and memory. Children with SLI show worse Token Test performance, whereas at ages 4 and 5, Part V of the Token Test could potentially serve as a screening tool for the detection of SLI.

Keywords: cognitive performance; intelligence; language disorder; preschoolers; screening; specific language impairment; Token Test

Specific language impairment (SLI) is a very common childhood disorder with a prevalence rate of approximately 7% in early childhood (Siu, 2015; Tomblin et al., 1997). According to the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-V; American Psychiatric Association, 2013), it is classified as a neurodevelopmental disease that is characterized by impairments in expressive or/and receptive language regarding different modalities.[1] These impairments concern the use and/or comprehension of speech, written language, or symbols (American Speech-Language-Hearing Association, 2016) and in further consequence the linguistic dimensions phonology, morphology, semantic, syntax, and pragmatics (e.g., Kamhi & Clark, 2013).

Besides these characteristic symptoms, SLI is increasingly associated with lower performance in different cognitive domains such as sustained selective attention, attentional shifting, working memory, inhibitory control, cognitive flexibility, problem solving, and planning (e.g., Aljahlan & Spaulding, 2019; Hughes, Turkstra, & Wulfeck, 2009; Kapa, Plante, & Doubleday, 2017; Pauls & Archibald, 2016; Roello, Ferretti, Colonnello, & Levi, 2015; Spaulding, 2010; Willinger et al., 2017). In this context, it was shown that these deficits are at least partly independent of language impairment severity or linguistic demand of used tasks (e.g., Pauls & Archibald, 2016). In this way, SLI is increasingly connected to general processing limitations and difficulties in the acquisition of automatic skills (e.g., Kamhi & Clark, 2013; Nicolson & Fawcett, 2007).

Language problems in SLI should traditionally not be better explained by, inter alia, apparent sensory or neurological deficits (e.g., Kamhi & Clark, 2013). However, recent studies increasingly indicate alterations in brain morphology and functionality in children with SLI. In this context, changes were shown, for example, in language-specific regions such as Broca's area or Wernicke's area, in regions associated with visual perception as well as in white matter tracts connecting these brain regions (e.g., Girbau-Massana, Garcia-Marti, Marti-Bonmati, & Schwartz, 2014; Kamhi & Clark, 2013; Morgan, Bonthrone, & Liégeois, 2016; Van der Lely & Pinker, 2014).

SLI appears early in childhood and often persists into adolescence and even adulthood (for summaries, see Gillam & Kamhi, 2010; Kamhi & Clark, 2013), whereas SLI subtypes and associated symptoms are seemingly not stable over time. In this context, a cross-sectional study on SLI subtypes showed that in kindergarten children with SLI, 35% were classified as the expressive type, 28% as the receptive type, and 35% as an expressive/receptive type (Tomblin, Records, & Zhang, 1996). Nevertheless, longitudinal studies showed that children frequently change from one SLI subtype to another and can do so in a short time (e.g., Conti-Ramsden & Botting, 1999). With respect to SLI subtypes, however, others argue that pure expressive subtypes of SLI do not exist and that expressive problems are associated with impairments in language knowledge and problems in language input processing (Leonard, 2009).

In the course of typical language development, until the age of (approximately) 3, children mainly achieve bottom-up processes of language processing such as phonological word form detection, morphosyntactic categorization, lexical–sematic categorization, lexical access and retrieval, phrase structure and reconstruction as well as aspects of prosodic processing (for a review, see Skeide & Friederici, 2016). Shortly before as well as at the beginning of preschool age (approximately 3–4 years of age), children increasingly develop top-down processes such as the analysis of semantic and syntactic relations (Skeide & Friederici, 2016). In this context, it was shown that between ages 5 and 8 typically developing children but also children with SLI each show relatively stable language development trajectories (Norbury et al., 2017), whereas a greater variability regarding language performance in preschoolers was seen compared to school-aged children (Schmitt et al., 2017). In preschool age, besides language processing, further cognitive abilities that heavily rely on prefrontal brain regions develop increasingly, such as working memory, and cognitive flexibility in terms of shifting ability, inhibition, attentional control, and planning (e.g., Best & Miller, 2010; Garon, Bryson, & Smith, 2008; Jurado & Rosselli, 2007), all of which were shown to be less well developed in SLI (e.g., Aljahlan & Spaulding, 2019; Kapa et al., 2017; Pauls & Archibald, 2016).

Given the multiple problems in SLI as well as its prolonged course, the necessity for an early detection of language impairments is clearly given and has already been strongly recommended by associations like the American Speech-Language-Hearing Association or the American Academy of Pediatrics (Hagan, Shaw, & Duncan, 2008; Siu, 2015). Despite these recommendations, at least in the United States, 30%–55% of the children receive no screening of language impairments by their health service provider (American Speech-Language-Hearing Association, n.d.; Siu, 2015). Furthermore, nonverbal intelligence tests that are necessary for the diagnosis of SLI are often not administered by professionals because nonverbal intelligence performance often does not influence the access to services and such an assessment seemingly falls outside their scope of practice (see, e.g., Kamhi & Clark, 2013). The current situation regarding screening for language impairments is even more critical if everyday consequences of (undetected) SLI are considered as ongoing SLI is associated with, inter alia, poorer psychosocial outcome (Snowling, Bishop, Stothard, Chipchase, & Kaplan, 2006), a higher disposition for social–emotional and behavioral difficulties (Levickis et al., 2018; Willinger et al., 2003), lower independent functioning (Conti-Ramsden & Durkin, 2008), as well as with higher costs for families and states by higher health service utilization (Le et al., 2017).

Although research on tasks for the early and fast detection of language impairments shows promising developments (e.g., Eisenberg & Guo, 2013; Guo & Eisenberg, 2014; Lugo-Neris, Peña, Bedore, & Gillam, 2015; Sim et al., 2015), it has to be noted that most of these screening procedures show disadvantageous properties or missing information with respect to, for example, duration or modality of administration and/or scoring. In this context, a recent review indicated that all openly available language assessments in English show limitations with respect to the evidence of their psychometric quality (Denman et al., 2017). Furthermore, with respect to the daily use of language assessments, it was shown that, unfortunately, psychometric properties of language tests do not influence how frequently they are used (Betz, Eickhoff, & Sullivan, 2013). With respect to the German-speaking area, it was shown that no adequate screening protocols for SLI are currently available (Kasper et al., 2011). Given the multiple and concomitant problems of SLI, the negative consequences of ongoing (undetected) SLI as well as the current situation regarding assessment tools, the necessity for the evaluation of short, effective, and feasible tools for the screening of SLI in preschoolers becomes apparent.

The "Token Test" was originally designed to measure receptive language impairments in patients with aphasia (De Renzi & Vignolo, 1962) but was also shown to be effective in the detection of language problems in children and even preschool children (e.g., Cole & Fewell, 1983, Geyer, Niebergall, Remschmidt, & Merschmann 1978). In this task, the examinees are shown plastic objects (tokens) lying on a table The tokens differ in size, form, and color. Across the different parts of the Token Test, the examinees are given increasingly difficult commands like, for example, "Touch the red circle" (in Part I, the most easiest part), or "Before touching the green circle, pick up the white rectangle" (in Part V, the most difficult part) and have to execute these demands (see, e.g., De Renzi & Vignolo, 1962; Strauss, Sherman, & Spreen, 2006). The Token Test measures language comprehension in the form of understanding nonredundant commands (e.g., Strauss et al., 2006) and was shown to be strongly associated with the performance in language reception but also production tasks (see, e.g., Cole & Fewell, 1983; Gutbrod, Mager, Meier, & Cohen, 1985).

As Token Test performance further relies on cognitive abilities like short-term memory/working memory and inhibition (see, e.g., Lezak, Howieson, Bigler, & Tranel, 2012; Strauss et al., 2006), and is considered to indicate problems in more complex verbal comprehension or even general cognitive processing (e.g., DiSimoni & Mucha, 1982; Geyer et al., 1978; Kamhi & Clark, 2013; Willinger et al., 2017), the Token Test seems to be very suited for screening SLI in preschool children, especially given the previously mentioned language and cognitive development in this age span.

Willinger et al. (2017) investigated the predictive role of a short German version of the Token Test (Orgass, 1982) as a screening of preschool SLI and showed insufficient detection rates in the screening of SLI (based on the recommendations by Plante and Vance, 1994, regarding the classification rates of preschool language tests). Due to the advantageous properties of the Token Test, such as its sound discriminant validity and reasonable reliability, gamelike character, clear instructions, independence from gender and ethnicity, portability, and its cost-effectiveness (Peña-Casanova et al., 2009; Ripich, Carpenter, & Ziol, 1997; Strauss et al., 2006; Taylor, 1998; Willinger et al., 2017), they nevertheless recommend investigating the predicitve role of variations or single parts of the Token Test. In this context, a study involving adult aphasic patients, Orgass, Poeck, Hartje, and Kerschensteiner (1973) indicated that Part V of the Token Test potentially holds the same diagnostic value as the whole Token Test, whereas this result was opposed by Kelter, Cohen, Engel, List, and Strohner (1976, 1977). Nevertheless, Part V could be predicitive of SLI as its increased requirements to process syntax, relational concepts, and subtle changes in symbol formulation as well as its greater syntactic and semantic variability (Orgass, 1982; Lezak et al., 2012; Strauss et al., 2006) in terms of complex speech comprehension is potentially sensitive with regard to the unstable development of syntactic and semantic processing in preschool age (e.g., Schmitt et al., 2017; Skeide & Friederici, 2016). According to the authors of the current study, it can be argued that Part V shows increased requirements not only regarding linguistic aspects but also regarding cognitive abilities like short-term memory/working memory, and inhibition (abilities depicted for the whole Token Test; see Lezak et al., 2012; Strauss et al., 2006) as well as regarding cognitive flexibility in terms of set shifting, attentional control, attentional shifting, and aspects of planning (abilities not previously linked to Part V in test descriptions). Part V with its increasingly longer and difficult commands would, in the notion of the authors, put greater requirements on the child to hold the target objects in mind and to manipulate them (short-term/working memory), to maintain and shift the attentional focus between objects (attentional shifting and attentional control), to withhold touching objects that should be excluded from actions (inhibition), to change between including or excluding objects from actions (cognitive flexibility), and to analyze all information of a command in order to respond adequately (planning). Similar to linguistic aspects, these cognitive requirements could be potentially sensitive with regard to executive function impairments in SLI (as previously mentioned).

AIMS

Given the early occurring linguistic impairments, the weaker performance regarding different cognitive abilities, and the number of concomitant problems, the necessity for the evaluation of short, effective, and feasible tools for the screening of SLI in young age becomes apparent. Willinger et al. (2017) previously investigated the diagnostic value of the Token Test in preschool children and showed insufficient detection rates in the screening of SLI. Nevertheless, considering the advantageous properties of the Token Test as well as the previously mentioned theoretical implications, and further following the recommendation of Willinger et al. (2017), the aim of the current study was to analyze the diagnostic value of the five parts of the 50-item German version of the Token Test (Orgass, 1982). This was done to investigate whether Part V has a higher predictive value than the whole Token Test and therefore to see whether Part V can be considered as an adequate screening of SLI in preschool age. The data for the analyses in the present study were provided by Willinger et al. (2017). Whereas until 2011, adequate screening protocols for SLI in German language were scarce (Kasper et al., 2011), in the recent years promising language screening instruments were increasingly published (see, e.g., the ESGRAF 4-8 [*Evozierte Sprachdiagnose grammatischer Fähigkeiten*], Motsch & Rietz, 2019; PDSS [*Patholinguistische Diagnostik bei Sprachentwicklungsstörungen*], Kauschke & Siegmüller, 2009). Nevertheless, investigating the diagnostic value of the Token Test as a short screening for SLI has further beneficial effects as it can be conducted in different languages. Therefore, the results of the current study have implications not only for the German-speaking area. The result of Part V being an adequate, short screening of SLI would further have another positive diagnostic implication as preschoolers with SLI were shown to have less clear preference for spoken answers upon being questioned and by trend use more supporting und substituting gestures (Lavelli & Majorano, 2016). In order to further compare the results of the present study with those of Willinger et al. (2017), the association between intellectual capacities and Part V of the Token Test was investigated. The hypotheses of the current study, given in the Method section, build on the results of Willinger et al. (2017) and are therefore mainly formulated accordingly. Hypotheses 1–3 addressed the question whether preschoolers with SLI and typically developing controls differ regarding performance on the different parts of the Token Test and can be adequately be classified by Part V, in view of diagnostic implications. Regarding Hypotheses 2 and 3, additional, explorative analyses were conducted with respect to explicit age groups. These reanalyses were inspired by the reviews of the manuscript.[2] Hypothesis 4 and 5 tended to offer intellectual profiles of preschoolers with SLI and typically developing children, in view of therapeutic implications.

Materials and method

Subjects

Forty-five boys (74%) and 16 girls (26%; expected gender ratio; see, e.g., Tomblin et al., 1997) aged between 4 and 6 years ($M = 4.9$ years, $SD = 8$ months) were recruited at the Medical University of Vienna. SLI was diagnosed by a psychologist according to DSM-V (2013) whereby diagnosis implied at least average scores in two nonverbal intelligence tests (Coloured Progressive Matrices—Raven, Raven, & Court, 1998; Columbia Mental Maturity Scale—Burgemeister, Blum, & Lorge, 1954) as well as at least one below average score out of five in standardized linguistic tests (Active Vocabulary Test for 3- to 6-year-old children—Kiese & Kozielski, 1996; Heidelberg Evaluation of Language Development—Grimm & Schoeler, 1990; Peabody Picture Vocabulary Test—Dunn, 1965). Furthermore, a control sample of typically developing, age-matched 45 boys and 16 girls was recruited. The control sample was examined using the same tasks as with the children with SLI whereby these children were required to have at least average scores in all tasks. Scores for all tasks can be seen in Table 1.

All preschoolers (SLI and controls) were monolingual native German speakers and all tasks in this study were performed in German. Before participating in this study, the examiners ensured that all children understood the target words used in the Token Test (colors and forms). In this pretest, no child had to be excluded from further participation. The caregivers of each child were informed and had to give their written consent prior to participation. The study protocol obtained approval by the ethics committee of the Medical University of Vienna and conforms to the provisions of the Declaration of Helsinki.

Materials

Token test

Performing a German 50-item version of the Token Test (Orgass, 1982), children are shown plastic objects lying on a table whereas these objects differ regarding three dimensions: size (small/large), form (rectangles/circles), and color (green/white/yellow/red/blue). Within a fixed order, children have to understand commands and perform accordingly like for example in Part I: "Touch the yellow circle." The Token Test is divided into five parts, and each part becomes increasingly more difficult (e.g., Part II: "Touch the small yellow rectangle"; Part III: "Touch the yellow circle and the white rectangle"; Part IV: "Touch the small yellow circle and the big green circle"). In accordance with the handbook, in each part, the objects are arranged differently on the table. Part V of the Token Test was shown to be the most difficult one (Orgass, 1982; e.g., Part V: "Before touching the green circle, pick up the white rectangle"). Incorrect responses are scored and summed for each part separately, yielding error scores for each part. Administration time of the whole Token Test is approximately 10 min and scoring time is approximately 5 min.

Intelligence scores

Intellectual capacities were examined using the German version of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Eggert, 1975), yielding a verbal

https://doi.org/10.1017/S0142716419000493 Published online by Cambridge University Press

Table 1. Test scores for preschoolers with SLI and typically developing children, with scores given for the total sample patient and control group that was used for calculations in the current study as well as explorative for the different age groups (values are presented as means with standard deviations [SD])

	4-year-olds		5-year-olds		6-year-olds		Total sample	
Variables	SLI	TD	SLI	TD	SLI	TD	SLI	TD
N	37	37	18	18	6	6	61	61
Diagnostic scores:								
Language tests								
AVT—Active vocabulary	29.1 (10.7)	56.1 (9.3)	38.9 (12.3)	62.2 (8.2)	41.3 (12.6)	67.0 (11.6)	33.2 (12.3)	58.8 (10.0)
HELD—Active grammar	6.1 (5.5)	11.4 (2.8)	8.4 (6.0)	15.7 (2.2)	9.3 (4.8)	18.0 (1.1)	7.1 (5.7)	13.3 (3.5)
HELD—Passive grammar	5.1 (3.2)	10.8 (4.0)	6.8 (4.4)	10.8 (2.8)	4.8 (3.6)	13.5 (2.8)	5.5 (3.6)	11.1 (3.6)
HELD—Using language	0.7 (2.1)	8.6 (3.8)	0.2 (0.5)	10.7 (4.9)	1.8 (1.9)	14.7 (1.4)	0.7 (1.8)	9.8 (4.4)
PPVT—Passive vocabulary	51.5 (22.5)	67.4 (16.1)	74.9 (17.6)	78.8 (14.2)	74.7 (10.2)	87.3 (12.5)	61.1 (23.0)	72.7 (16.6)
Nonverbal intelligence								
CMM	40.3 (6.1)	51.4 (7.0)	52.7 (15.2)	58.0 (7.0)	59.5 (6.4)	57.3 (8.4)	44.3 (10.7)	53.8 (8.4)
CPM	14.3 (3.7)	14.4 (2.0)	16.1 (3.4)	16.7 (3.0)	16.8 (7.6)	18.7 (3.0)	15.3 (4.2)	15.5 (2.8)
Empirical scores:								
Token Test								
Part 1*	3.4 (3.4)	0.03 (0.2)	0.6 (1.7)	0.1 (0.2)	0.5 (0.8)	0.0 (0.0)	2.3 (3.1)	0.03 (0.2)
Part 2*	5.8 (3.5)	0.3 (0.7)	1.5 (2.5)	0.3 (0.8)	2.3 (2.9)	0.7 (1.6)	4.2 (3.8)	0.3 (0.8)
Part 3*	7.0 (3.5)	1.1 (2.0)	2.7 (2.8)	0.1 (0.3)	2.2 (2.4)	0.7 (1.7)	5.3 (3.8)	0.8 (1.7)

(*Continued*)

https://doi.org/10.1017/S0142716419000493 Published online by Cambridge University Press

Table 1. (*Continued*)

Variables	4-year-olds		5-year-olds		6-year-olds		Total sample	
	SLI	TD	SLI	TD	SLI	TD	SLI	TD
Part 4*	8.5 (2.4)	2.5 (3.4)	4.9 (3.6)	2.0 (2.4)	4.3 (2.5)	1.5 (2.8)	7.0 (3.4)	2.2 (3.1)
Part 5*	8.2 (2.5)	2.8 (2.6)	5.9 (2.6)	1.6 (1.9)	4.3 (2.9)	2.2 (1.5)	7.2 (2.9)	2.4 (2.4)
Cognitive abilities								
WPPSI—Verbal scale	21.1 (10.0)	45.6 (19.6)	38.7 (9.2)	45.0 (11.5)	42.3 (7.6)	50.8 (5.6)	28.2 (13.0)	45.9 (16.7)
WPPSI—Nonverbal scale	13.7 (8.2)	23.9 (8.6)	32.7 (11.1)	30.0 (10.0)	30.5 (14.2)	47.8 (9.4)	21.0 (13.3)	30.4 (12.4)
WPPSI—Arithmetic	3.8 (3.0)	9.7 (2.1)	8.3 (3.4)	11.7 (2.7)	10.3 (3.4)	14.3 (2.6)	5.8 (4.0)	10.8 (2.7)
WPPSI—Animal house	11.2 (5.0)	23.3 (5.3)	22.4 (6.5)	22.0 (9.4)	26.2 (8.3)	26.5 (8.1)	16.0 (8.3)	23.2 (7.0)

Note: SLI, preschoolers with specific language disorder. TD, typically developing children. AVT, Active Vocabulary Test for 3- to 6-year-old children.HELD, Heidelberg Evaluation of Language Development. PPVT, Peabody Picture Vocabulary Test. CMM, Columbia Mental Maturity Scale. CPM, Coloured Progressive Matrices. WPPSI, Wechsler Preschool and Primary Scale of Intelligence.
*Error scores, higher scores mean worse performance

and a nonverbal scale. Furthermore, the WPPSI subtests arithmetic (measuring numeracy, problem solving, and verbal working memory) and animal house (measuring visual memory, visual attention, and fine motor coordination) were performed.

Hypotheses and statistics

Hypothesis 1. Preschoolers with SLI and typically developing controls will differ significantly regarding their performance on each of the five parts of the Token Test. *T* tests will be performed to analyze this hypothesis.

Hypothesis 2. Preschoolers with SLI and typically developing controls will be significantly classified by all five parts of the Token Test, yielding classification rates of 80% or greater (see Plante & Vance, 1994). A direct discriminant analysis will be performed.

Hypothesis 3. Preschoolers with SLI and typically developing controls will be significantly classified by Part V of the Token Test, yielding classification rates of 80% or greater (see Plante & Vance, 1994). A direct discriminant analysis will be performed.

Hypothesis 4. Preschoolers with SLI who were correctly classified and those who were incorrectly classified by Part V of the Token Test will differ significantly regarding all four scores of the WPPSI. Mann–Whitney *U* tests will be performed (criteria for *T* test were not fulfilled).

Hypothesis 5. Typically developing controls who were correctly classified and those who were incorrectly classified by Part V of the Token Test will differ significantly regarding all four scores of the WPPSI. Mann–Whitney *U* tests will be performed (criteria for *T* test were not fulfilled).

Additional, explorative calculations regarding Hypotheses 2 and 3

Separate direct discriminant analyses for the classification of preschoolers with SLI and typically developing controls will be performed separately for age groups 4, 5, and 6, using all five parts of the Token Test (additional calculations regarding Hypothesis 2) as well as Part V of the Token Test (additional calculations regarding Hypothesis 3) as classifying variables. These additional analyses are explorative, especially regarding age group 6 with its small sample size (see Figure 1). Statistical analyses were performed using IBM SPSS 25 whereas all results fulfilling $p < .05$ were seen as significant.

Results

In the following, the results with respect to each hypothesis will be presented.

Hypothesis 1

Preschoolers with SLI and typically developing controls showed significant differences regarding all five parts of the Token Test, Part I: $T\ (120) = 5.6$, $p \leq .0001$; Part II: $T\ (120) = 8.0$, $p \leq .0001$; Part III: $T\ (120) = 9.2$, $p \leq .0001$; Part IV: $T\ (120) = 8.5$, $p \leq .0001$; Part V: $T\ (120) = 11.0$, $p \leq .0001$. In each part, preschoolers with SLI showed significantly more mistakes than the controls (see Table 1).

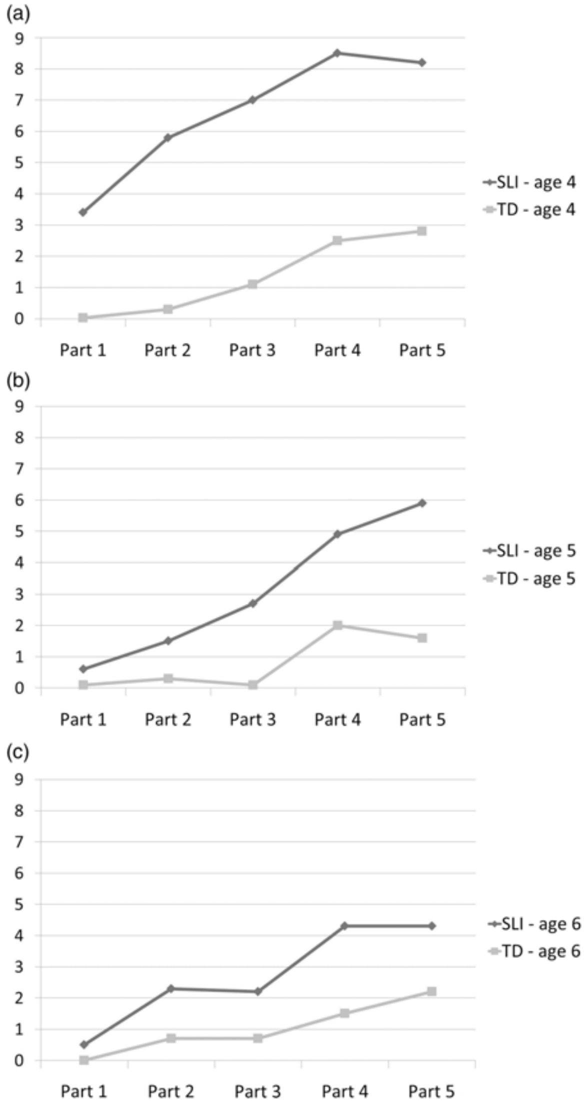

Figure 1. Token Test error scores shown for children with specific language impairment (SLI) and typically developing, age- and gender-matched controls (TD) in the separate age groups (a) 4 years ($n = 72$), (b) 5 years ($n = 36$), and (c) 6 years ($n = 12$). Results indicate great differences in age group 4 and decreasing differences in age groups 5 and 6. Higher scores mean worse performance.

Hypothesis 2

The analysis revealed a significant discriminant function, Canonical Correlation = .7, Wilks' lambda = .5, χ^2 (5, $N = 122$) = 75.9, $p \leq .0001$. In addition, 87% of the typically developing controls and 71% of the preschoolers with SLI were classified correctly by all five parts of the Token Test, showing a total correct classification of 79% (see Table 2).

https://doi.org/10.1017/S0142716419000493 Published online by Cambridge University Press

Table 2. Direct discriminant analyses between preschoolers with SLI and typically developing children with respect to the five parts of the Token Test (error scores)

Variables	Wilks' lambda	*F* value
* Token Test error score—Part I	.793	31.4
* Token Test error score—Part II	.659	62.1
* Token Test error score—Part III	.629	70.6
* Token Test error score—Part IV	.641	67.2
* Token Test error score—Part V	.545	100.4

Note: All results are statistically significant ($p \leq .0001$).

Additional calculations Hypothesis 2

Regarding age group 4, analysis revealed a significant discriminant function, Canonical Correlation = .80, Wilks' lambda = .36, χ^2 (5, $N = 74$) = 70.14, $p \leq .0001$. In addition, 97.3% of the typically developing controls and 81.1% of the preschoolers with SLI were classified correctly by all five parts of the Token Test, Part I: Wilks' lambda = .67, F (1, 72) = 35.44, $p \leq .0001$; Part II: Wilks' lambda = .45, F (1, 72) = 89.31, $p \leq .0001$; Part III: Wilks' lambda = .47, F (1, 72) = 81.2, $p \leq .0001$; Part IV: Wilks' lambda = .49, F (1, 72) = 75.80, $p \leq .0001$; Part V: Wilks' lambda = .46, F (1, 72) = 84.30, $p \leq .0001$, showing a total correct classification of 89.2%.

Regarding age group 5, analysis revealed a significant discriminant function, Canonical Correlation = .73, Wilks' lambda = .46, χ^2 (5, $N = 36$) = 24.46, $p \leq .0001$. In addition, 94.4% of the typically developing controls and 77.8% of the preschoolers with SLI were classified correctly by all five parts of the Token Test, Part I: Wilks' lambda = .95, F (1, 34) = 1.92, $p = .175$; Part II: Wilks' lambda = .91, F (1, 34) = 3.51, $p = .069$; Part III: Wilks' lambda = .69, F (1, 34) = 15.41, $p \leq .0001$; Part IV: Wilks' lambda = .81, F (1, 34) = 7.80, $p = .009$; Part V: Wilks' lambda = .51, F (1, 34) = 32.17, $p \leq .0001$, showing a total correct classification of 86.1%.

Regarding age group 6, analysis revealed a nonsignificant discriminant function, Canonical Correlation = .79, Wilks' lambda = .37, χ^2 (5, $N = 12$) = 7.35, $p = .196$. In addition, 83.3% of the typically developing controls and 83.3% of the preschoolers with SLI were classified correctly by all five parts of the Token Test, showing a total correct classification of 83.3%.

Hypothesis 3

The analysis revealed a significant discriminant function, Canonical Correlation = .7, Wilks' lambda = .5, χ^2 (5, $N = 122$) = 72.6, $p \leq .0001$, by Part V of the Token Test. In addition, 79% of the typically developing controls and 75% of the preschoolers with SLI were classified correctly by Part V of the Token Test, showing a total correct classification of 77% (see Table 2).

https://doi.org/10.1017/S0142716419000493 Published online by Cambridge University Press

Additional calculations Hypothesis 3

Regarding age group 4, analysis revealed a significant discriminant function, Canonical Correlation = .73, Wilks' lambda = .46, χ^2 (1, N = 74) = 55.40, $p \leq .0001$. In addition, 83.8% of the typically developing controls and 86.5% of the preschoolers with SLI were classified correctly by Part V of the Token Test showing a total correct classification of 85.1%.

Regarding age group 5, analysis revealed a significant discriminant function, Canonical Correlation = .67, Wilks' lambda = .51, χ^2 (1, N = 36) = 22.31, $p \leq .0001$. In addition, 77.8% of the typically developing controls and 83.3% of the preschoolers with SLI were classified correctly by Part V of the Token Test showing a total correct classification of 80.6%.

Regarding age group 6, analysis revealed a nonsignificant discriminant function, Canonical Correlation = .45, Wilks' lambda = .79, χ^2 (1, N = 12) = 2.20, p = .138. In addition, 66.7% of the typically developing controls and 33.3% of the preschoolers with SLI were classified correctly by Part V of the Token Test, showing a total correct classification of 50%.

Hypothesis 4

Analysis regarding the univariate differences between correctly and incorrectly classified preschoolers with SLI yielded significant differences regarding all four scores of the WPPSI. Preschoolers with SLI who were correctly classified by Part V of the Token Test showed significantly lower scores on the WPPSI verbal scale, Z = –4.3, $p \leq .0001$, WPPSI nonverbal scale, Z = –3, p = .003, WPPSI subtests arithmetic, Z = –4.3, $p \leq .0001$, and animal house, Z = –3.5, $p \leq .0001$, than those who were classified incorrectly. Details are shown in Table 3.

Hypothesis 5

Analysis regarding the univariate differences between correctly and incorrectly classified typically developing controls showed significant differences regarding two scores of the WPPSI. Controls who were correctly classified by Part V of the Token Test showed significantly higher scores on the WPPSI verbal scale, Z = –2.8, p = .005, and WPPSI subtest arithmetic, Z = –3.3, p = .001, than those who were classified incorrectly. No significant differences were shown for the WPPSI nonverbal scale, Z = –.1, p = .951, and the WPPSI subtest animal house, Z = –8, p = .436. Details are shown in Table 3.

Discussion

SLI emerges early in child development and is characterized by impairments in expressive and/or receptive language (e.g., American Psychiatric Association, 2013; Tomblin et al., 1997). Considering the unfavorable aspects associated with SLI such as its early onset, its prolonged course, its multiple associated impairments, its unfavorable prognosis, the negative everyday life consequences of ongoing (undetected) SLI, the current situation regarding the screening of SLI, as well as the properties of existing assessment tools, the necessity for the evaluation of short,

https://doi.org/10.1017/S0142716419000493 Published online by Cambridge University Press

Table 3. Group differences between correctly and incorrectly classified preschoolers with SLI as well as between correctly and incorrectly classified typically developing children with respect to the four WPPSI scores (classifications are based on the scores of Token Test Part V, and values are presented as mean with standard deviation [SD])

	SLI		TD	
Intellectual abilities	Correctly classified	Incorrectly classified	Correctly classified	Incorrectly classified
N	46	15	48	13
WPPSI—Verbal scale	**24.3 (11.5)**	**41.0 (9.3)**	**48.3 (17.4)**	**37.5 (9.9)**
WPPSI—Nonverbal scale	**17.8 (11.3)**	**30.7 (14.5)**	30.5 (12.8)	30.3 (11.0)
WPPSI—Arithmetic	**4.5 (3.2)**	**9.8 (3.6)**	**11.3 (2.7)**	**8.7 (1.9)**
WPPSI—Animal house	**13.8 (7.2)**	**22.7 (8.1)**	23.6 (7.5)	21.8 (4.6)

Note: WPPSI, Wechsler Preschool and Primary Scale of Intelligence. SLI, preschoolers with specific language impairment. TD, typically developing children. Significant results are bold faced.

effective, and feasible tools for the screening of SLI in early childhood becomes apparent. Recently, Willinger et al. (2017) evaluated the Token Test, a task that was previously shown to detect language problems in preschool children (e.g., Cole & Fewell, 1983) and basically shows good psychometric properties (e.g., Strauss et al., 2006) as well as characteristics that go well with SLI symptoms, even if they change over time (see, e.g., Cole & Fewell, 1983; Conti-Ramsden & Botting, 1999; Lavelli & Majorano, 2016; Leonard, 2009). Although Willinger et al. (2017) showed that the total score of a short German version of the Token Test did not yield sufficient detection rates in the screening of SLI in preschool age (age range 4–6 years), they nevertheless recommend to investigate the predicitive value of variations or single parts of the Token Test, given the advantageuos properties of the task. Following this recommendation as well as theoretical implications, the aim of the study was to investigate whether Part V has a higher predictive value than the whole Token Test and therefore to see whether Part V can be considered as an adequate screening of SLI in preschool age.

The results of the current study showed that preschoolers with SLI exhibit worse results on all five parts of the Token Test, including the easiest (Part I) and well as the most difficult (Part V). When examining these results in detail with regard to explicit age groups (see Table 1 and Figure 1), it can be seen that 4-year-old preschoolers show the worst overall results, including on the easiest part (Part I). Furthermore, it can be seen that after the age of 4, there is seemingly no difference between children with SLI and controls on Part I whereas on Parts II–V children with SLI still show worse results.

These results are presumably explained by excessive demands on language-specific top-down processes such as analysis of semantic and syntactic relations in SLI (for an overview regarding typical language development see, e.g., Skeide & Friederici, 2016). This is depicted by worse results in children with SLI across the whole age range regarding Parts II–V. The overall difference (whole sample)

https://doi.org/10.1017/S0142716419000493 Published online by Cambridge University Press

between groups regarding Part I is seemingly mainly driven by great difficulties of children with SLI at age 4 (see Figure 1). By trend, the greatest differences between groups can be seen in Part V, which would be explained by previously reported greater requirements with regard to semantic and syntactic processing in this part (see, e.g., Orgass, 1982; Strauss et al., 2006). As can be seen in the explorative Figure 1, at age 4 great differences between children with SLI and typically developing children can be seen, whereas the difference between groups on each part shows similar curves at ages 5 and 6. This explorative result has to be treated with caution due to the small sample size in age group 6. Nevertheless, such a development would be in line with studies that indicate that between ages 5 and 8 typically developing children and children with SLI show relatively stable language development trajectories (Norbury et al., 2017) whereas a greater variability regarding language performance can be found in preschoolers (Schmitt et al., 2017).

The results showing an increasing difference between preschoolers with SLI and typically developing children on Token Test performance across the progressively difficult parts could possibly be further explained by aspects on executive functions development and pathology in SLI. Whereas typical development of executive functions in young age is associated with development of the prefrontal cortex (for a review, see Garon et al., 2008), children with SLI were previously shown to exhibit limitations in cognitive abilities that heavily rely on prefrontal brain regions. They show, for example, weaker performance in cognitive aspects such as the ability to shift and maintain your focus (attentional shifting and sustained selective attention); the ability to avoid distraction or refraining from a dominating, incorrect response (inhibition/inhibitory control); the ability to keep information in memory as well as use it for mental operations (short-term/working memory); the ability to stick to rules and adjust to changing rules or task demands (use of rules and cognitive flexibility); or the ability to analyze given information in a way that leads to adequate responses (planning ability; see, for example, Aljahlan & Spaulding, 2019; Kapa et al., 2017; Roello et al., 2015). In the notion of the authors, impairments in one or more of these abilities could lead to worse results on the Token Test, as the task requires, with increasing difficulty across parts, to hold the target objects in mind and to manipulate them, to maintain and shift the attentional focus between objects, to withhold touching objects that should be excluded from actions, to change between including or excluding objects from actions, and to analyze all information of a command in order to respond adequately.

In SLI, the lower performance on the Token Test can be possibly explained by problems in both language processing and cognitive abilities, which were shown above to be associated with Token Test performance, especially regarding Part V. This notion would be supported by literature indicating a bidirectional influence of cognitive abilities and language development such as executive functions influencing vocabulary learning or processing syntax as well as language skills mediating executive functions (for summaries see, e.g., Aljahlan & Spaulding, 2019; Kapa et al., 2017). Further support stems from studies on pathological brain alterations in SLI regarding regions associated with language processing as well as regions associated with higher order cognitive abilities such as executive functions (see, e.g., Agostini et al., 2010; Crinion et al., 2006; Girbau-Massana et al., 2014; Kamhi & Clark, 2013; Krishnan, Watkins, & Bishop, 2016; Van der Lely & Pinker, 2014).

https://doi.org/10.1017/S0142716419000493 Published online by Cambridge University Press

Attentional shifting in terms of shifting attentional resources between targets influences language development and was shown to be impaired in SLI (Aljahlan & Spaulding, 2019). Abnormal electrophysiological brain activity during auditory language perception (Pijnacker et al., 2017), developmental lags and variations in temporal language processing abilities (Choudhury & Benasich, 2011), and alterations in regions associated with visual perception (Girbau-Massana et al., 2014) suggest problems in the integration of auditory and visual information. This would lead to problems in attentional shifting across sensory modalities (see, e.g., Aljahlan & Spaulding, 2019), which is important in the Token Test as verbally given commands have to be combined with visual information regarding the objects. In this context, it was shown that the integration of auditory and visual attentional shifting predicts children's ability to understand verbal directions (Magimairaj & Montgomery, 2013).

With respect to the diagnostic value of the Token Test as a screening tool for SLI, the present study showed that the whole Token Test correctly classified 79% of the preschoolers with SLI and typically developing children whereas Part V of the Token Test correctly classified 77% of the whole sample (age range 4–6 years). Therefore, when taking children at preschool age/early childhood as well as at the end of preschool age/beginning of middle childhood into account (for age groups please see, e.g., Berk, 2014; Feldman, 2017; Steinberg, Vandell, & Bornstein, 2010), the whole Token Test as well as Part V alone shows inacceptable detection rates (lower than 80%), following the guidelines of Plante and Vance (1994), Messick (1989), and the "The Standards for Educational and Psychological Testing" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

Nevertheless, when considering the additional, explorative analyses of the current study, in age groups 4 and 5, the whole Token Test as well as Part V alone yields acceptable identification rates whereas in age group 4 the whole Token Test nearly reaches good levels (for guidelines on detection rates of language screenings, see Plante & Vance, 1994). In this way, in age group 4, groups were correctly classified with a rate of 89.2% by the whole Token Test and 85.1% by Part V, whereas in age group 5, groups were correctly classified with a rate of 86.1% by the whole Token Test and 80.6% by Part V. Viewing these results, it can be seen that the whole Token Test yields higher detection rates than Part V alone whereby Part V also yields acceptable classification rates with respect to age groups 4 and 5. In contrast, age group 6 was not significantly classified by any Token Test measure. It has to be noted that due to varying sample sizes, these results are of rather explorative nature and, therefore, following explanatory approaches, have to be treated with caution.

In SLI, abnormalities in the superior temporal lobe as well as regarding the inferior longitudinal fasciculus were shown (e.g., Girbau-Massana et al., 2014; Van der Lely & Pinker, 2014). These areas are typically associated with language processing in typically developing children, whereas at the age of 6, compared to adults, there is still strong interhemispheric connectivity, mainly within these superior temporal regions (Friederici, Brauer, & Lohmann, 2011). Nevertheless, it can be hypothesized that fast changes in white matter tracts such as the corpus callosum, the inferior longitudinal fasciculus, and the superior longitudinal fasciculus starting around age 6 (Uda et al., 2015) would also apply to children with SLI. Given such changes, at approximately age 6, children with SLI could have reached a language level at which the linguistic demands of the Token Test are not sufficient for correct classification.

In contrast, it can be hypothesized that shown increases in typically developing children at the end of preschool age in executive functions such as attentional control, working memory, inhibition, shifting, and planning (Best & Miller, 2010; Garon et al., 2008; Jurado & Rosselli, 2007) also apply to children with SLI to a certain degree. In this way, it can be further hypothesized that at approximately age 6, children with SLI reach a level of executive functioning at which the cognitive demands of the Token Test are not sufficient for correct classification.

Taking these two explanatory approaches together, it is also conceivable that such increases in language processing and/or cognitive processing lead to compensatory mechanisms, increasing the Token Test performance at age 6 (see Figure 1). Such a hypothesis would also be in line with the previously mentioned bidirectional influence of language processing and executive functioning (see, e.g., Kapa et al., 2017).

The Token Test seemingly does not cover all important language and cognitive aspects to identify a heterogeneous patient group such as SLI across several age groups. Likewise problems were previously discussed by Theodorou, Kambanaros, and Grohmann (2013), as they state that the effectiveness of language tests depends, inter alia, on the multi-facetedness of assessment tools, measuring different aspects of language such as phonology, semantics, morphology, syntax, and sentence repetition. Therefore, the necessity for developing and validating new screening tools for the detection of SLI is still clearly given. In this context, Theodorou, Kambanaros, and Grohmann (2017) showed good results using a sentence repetition task to identify SLI in Cypriot Greek-speaking children.

Summarizing the results on the predictive value of the Token Test, the results of the current study indicate that for the use of the Token Test as well as Part V alone, possible age effects have to be taken into account. The whole Token Test or Part V alone could potentially serve as acceptable tools for the screening of SLI in age groups 4 and 5, whereas regarding age 6, the diagnostic value of the Token Test needs to be further investigated in future studies. Given the current results as well as the advantageous properties of the Token Test, future studies should definitely investigate the diagnostic value of the Token Test as well as its separate parts in the same and even older age groups including greater sample sizes. These studies should further include executive function tasks so as to add empirical data to the previously mentioned hypotheses.

Previous studies suggest that SLI is often associated with weaker cognitive performance (Dyck & Piek, 2010; Henry, Messer, & Nash, 2012; Willinger & Eisenwort, 1999; Willinger et al., 2017; Willinger, Voelkl-Kernstock, Neubauer, & Einenwort, 2001). Another aim of the study was to provide intellectual profiles of children who were previously correctly classified by Part V of the Token Test in comparison with incorrectly classified children. It was shown that correctly classified preschoolers with SLI exhibited significantly worse results with respect to nonverbal and verbal intelligence, numeracy, problem solving, verbal working memory, fine motor coordination, visual attention, and visual memory than incorrectly classified preschoolers with SLI. Furthermore, it was shown that correctly classified typically developing children showed better performances in verbal intelligence and the

https://doi.org/10.1017/S0142716419000493 Published online by Cambridge University Press

WPPSI subtest arithmetic than incorrectly classified typically developing children. On the one hand, these results support the notion of the Token Test being an assessment tool for more general cognitive abilities (see, e.g., DiSimoni & Mucha, 1982; Peña-Casanova et al., 2009; Willinger et al., 2017) as well as the notion of SLI being dependent on general processing limitations (see, e.g., Kamhi & Clark, 2013; Nicolson & Fawcett, 2007). On the other hand, these results as well as the result that preschoolers with SLI showed worse performance regarding each part of the Token Test (even the easiest parts), indicate the urgency of an early detection of SLI as well as early therapy for children with SLI. In this context, recent research shows promising therapeutic approaches like, for example, grammar training (Smith-Lock, Leitano, Lambert, & Nickels, 2013; Smith-Lock, Leitãno, Lambert, et al., 2013), specific read-aloud techniques (Justice, Logan, & Kaderavek, 2017), or metacognitive and metalinguistic training (Schiff, Nuri Ben-Shushan, & Ben-Artzi, 2017). Furthermore, it was shown that higher education of parents as well as more adequate judgment of executive functions in children by teachers increases the chance that children with SLI receive therapy (Wittke & Spaulding, 2018).

Conclusion

Preschoolers with SLI showed significantly worse performance in all five parts of the Token Test than age- and gender-matched typically developing controls including the easiest (Part I) and well as the most difficult (Part V). This result can likely be explained by greater demands on children with SLI regarding language-specific top-down processes typically developing in this age span such as analysis of semantic and syntactic relations. This result could further be explained by greater demands on children with SLI regarding cognitive abilities associated with the Token Test such as short-term memory/working memory, inhibition, attentional control, attentional shifting, set shifting, and planning. Another explanation could probably be SLI-specific problems in the integration of auditory and visual information, which is important in the Token Test as verbally given commands have to be combined with visual information regarding the objects. These explanations would be supported by shown language and executive functioning deficits in SLI and/or corresponding brain pathology. Additional analyses involving explicit age groups showed that across all age groups, apparent differences regarding Parts II–V of the Token Test could be seen. Furthermore, great differences between patient and control group was shown at age 4, whereas afterward this difference seemingly decreases and rather similar curves across all Token Test parts can be seen in age groups 5 and 6. These results are supported by studies that show relatively stable language development trajectories between ages 5 and 8 as well as a greater variability regarding language performance in preschool age. The whole Token Test correctly identified 79% of the preschoolers with SLI and typically developing children, whereas Part V of the Token Test correctly classified 77% when taking the whole sample (age range 4–6) into account. Therefore, at least in a sample comprising children in early as well as middle childhood, the Token Test cannot be seen as an effective tool for the detection of SLI. Nevertheless, additional analyses showed that in age groups 4 and 5, Part V of the Token Test yielded acceptable classification rates (85.1% and

80.6%) whereas age group 6 was not significantly discriminated by any Token Test variable. The Token Test seemingly does not cover all important language and cognitive aspects to identify a heterogeneous patient group such as SLI across several age groups. With respect to the results regarding age group 6, it can by hypothesized that age-specific increases in language processing and executive functioning as well as neurodevelopmental aspects that were to date shown in typically developing children also apply to children with SLI. In this way, increases in language processing and/or cognitive processing could lead to compensatory mechanisms, increasing the Token Test performance at age 6. Given the results on Token Test performance as well as the intellectual profiles provided by the present study, the necessity of an early detection of SLI as well as an early therapy for children with SLI is clearly given.

Conflicts of interest and source of funding. The authors report no conflict of interest. This study did not receive any funding.

Notes

1. In DSM-V (2013), this disorder is termed "language disorder." The authors are aware that the term "specific language impairment" is increasingly discussed controversially and that an increasing number of synonyms exist (e.g., "language disorder," "developmental language disorder," or "primary language impairment"; see, e.g., Westby, 2017). Nevertheless, in this study, the term SLI will be used as it is the most traditional term for that disorder.

2. The authors of the study want to thank the reviewers and the editor for their valuable comments on the manuscript, which inspired these additional, explorative analyses.

References

Agostini, G., Mancini, J., Chabrol, B., Villeneuve, N., Milh, M., George, F., … Girard, N. (2010). Language disorders in children with morphologic abnormalities of the hippocampus. *Archives de Pediatrie*, **17**, 1008–1016.

Aljahlan, Y., & Spaulding, T. J. (2019). The impact of manipulating attentional shifting demands on preschool children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **62**, 324–336.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Speech-Language-Hearing Association. (n.d.). *Definitions of communication disorders and variations*. Retrieved October 26, 2016, from http://www.asha.org/policy/RP1993-00208/

Berk, L. E. (2014). *Development through the lifespan* (6th ed.). Boston: Pearson Education.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, **81**, 1641–1660.

Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools*, **44**, 133–146.

Burgemeister, B. B., Blum, L. H., & Lorge, J. (1954). *The Columbia Mental Maturity Scale*. Hudson, NY: Yonkerson.

Choudhury, N., & Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: Prediction to 3- and 4-year language and cognitive abilities. *Clinical Neurophysiology*, **122**, 320–338.

Cole, K. N., & Fewell, R. R. (1983). A quick language screening test for young children: The Token Test. *Journal of Psychoeducational Assessment*, **1**, 149–153.

Conti-Ramsden, G., & Botting, N. (1999). Classification of children with specific language impairment: Longitudinal considerations. *Journal of Speech, Language, and Hearing Research*, **42**, 1195–1204.

Conti-Ramsden, G., & Durkin, K. (2008). Language and independence in adolescents with and without a history of specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*, **51**, 70–83.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., … Usui, K. (2006). Language control in the bilingual brain. *Science*, **312**, 1537–1540.

Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y. W., & Cordier, R. (2017). Psychometric properties of language assessments for children aged 4–12 years: A systematic review. *Frontiers in Psychology*, **8**, 1515.

De Renzi, A., & Vignolo, L. A. (1962). Token test: A sensitive test to detect receptive disturbances in aphasics. *Brain*, **85**, 665–678.

DiSimoni, F., & Mucha, R. (1982). Use of the Token Test for Children to identify language deficits in preschool age children. *Journal of Auditory Research*, **22**, 265–270.

Dunn, L. (1965). *Peabody Picture Vocabulary Test*. Shoreview MN: AGS.

Dyck, M., & Piek, J. (2010). How to distinguish normal from disordered children with poor language or motor skills. *International Journal of Language & Communication Disorders*, **45**, 336–344.

Eggert, D. (1975). *Der Wechsler Preschool and Primary Scale of Intelligence*. Bern-Stuttgart-Wien: Hans Huber.

Eisenberg, S. L., & Guo, L. Y. (2013). Differentiating children with and without language impairment based on grammaticality. *Language, Speech, and Hearing Services in Schools*, **44**, 20–31.

Feldman, R. S. (2017). *Development across the life span* (8th ed.). Boston: Pearson Education.

Friederici, A. D., Brauer, J., & Lohmann, G. (2011). Maturation of the language network: From inter-to intrahemispheric connectivities. *PLoS One*, **6**, e20726.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, **134**, 31–60.

Geyer, M., Niebergall, G., Remschmidt, H., & Merschmann, W. (1978). Zur diagnostischen Anwendbarkeit des Token Test im Kindes und Jugenalter. *Zeitschrift für Kinder- und Jugendpsychiatrie*, **5**, 203–221.

Gillam, S. L., & Kamhi, A. G. (2010). Specific language impairment. In J. S. Damico, N. Mueller, & M. Ball (Eds.), *The handbook of language and speech disorders* (pp. 210–226). Oxford: Wiley.

Girbau-Massana, D., Garcia-Marti, G., Marti-Bonmati, L., & Schwartz, R. G. (2014). Gray–white matter and cerebrospinal fluid volume differences in children with specific language impairment and/or reading disability. *Neuropsychologia*, **56**, 90–100.

Grimm, H., & Schoeler, H. (1990). *Heidelberger Sprachentwicklungstest*. Goettingen: Hogrefe.

Guo, L. Y., & Eisenberg, S. (2014). The diagnostic accuracy of two tense measures for identifying 3-year-olds with language impairment. *American Journal of Speech-Language Pathology*, **23**, 203–212.

Gutbrod, K., Mager, B., Meier, E., & Cohen, R. (1985). Cognitive processing of tokens and their description in aphasia. *Brain and Language*, **25**, 37–51.

Hagan, J., Jr., Shaw, J., & Duncan, P. (2008). *Bright futures: Guidelines for health supervision of infants, children, and adolescents* (3rd ed.). Elk Grove Village, IL: American Academy of Pediatrics.

Henry, L. A., Messer, D. J., & Nash, G. (2012). Executive functioning in children with specific language impairment. *Journal of Child Psychology and Psychiatry*, **53**, 37–45.

Hughes, D. M., Turkstra, L. S., & Wulfeck, B. B. (2009). Parent and self-ratings of executive function in adolescents with specific language impairment. *International Journal of Language & Communication Disorders*, **44**, 901–916.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review*, **17**, 213–233.

Justice, L. M., Logan, J., & Kaderavek, J. N. (2017). Longitudinal impacts of print-focused read-alouds for children with language impairment. *American Journal of Speech-Language Pathology*, **26**, 383–396.

Kamhi, A. G., & Clark, M. K. (2013). Specific language impairment. In O. Dulac, M. Lassonde, & H. B. Sarnat (Eds.), *Handbook of clinical neurology* (Vol. **111**, pp. 219–227). Amsterdam: Elsevier.

Kapa, L. L., Plante, E., & Doubleday, K. (2017). Applying an integrative framework of executive function to preschoolers with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **60**, 2170–2184.

Kasper, J., Kreis, J., Scheibler, F., Möller, D., Skipka, G., Lange, S., & Von Dem Knesebeck, O. (2011). Population-based screening of children for specific speech and language impairment in Germany: A systematic review. *Folia Phoniatrica et Logopaedica*, **63**, 247–263.

Kauschke, C., & Siegmüller, J. (2009). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen–2. Auflage*. Amsterdam: Elesevier.

Kelter, S., Cohen, R., Engel, D., List, G., & Strohner, H. (1976). Aphasic disorders in matching tasks involving conceptual analysis and covert naming. *Cortex*, **12**, 383–394.

Kelter, S., Cohen, R., Engel, D., List, G., & Strohner, H. (1977). The conceptual structure of aphasic and schizophrenic patients in a nonverbal sorting task. *Journal of Psycholinguistic Research*, **6**, 279–303.

Kiese, C., & Kozielski, P. (1996). *Aktiver Wortschatztest für 3–6 jährige Kinder*. Goettingen: Beltz.

Krishnan, S., Watkins, K. E., & Bishop, D. V. (2016). Neurobiological basis of language learning difficulties. *Trends in Cognitive Sciences*, **20**, 701–714.

Lavelli, M., & Majorano, M. (2016). Spontaneous gesture production and lexical abilities in children with specific language impairment in a naming task. *Journal of Speech, Language, and Hearing Research*, **59**, 784–796.

Le, H. N., Gold, L., Mensah, F., Eadie, P., Bavin, E. L., Bretherton, L., & Reilly, S. (2017). Service utilisation and costs of language impairment in children: The early language in Victoria Australian population-based study. *International Journal of Speech-Language Pathology*, **19**, 360–369.

Leonard, L. B. (2009). Is expressive language disorder an accurate diagnostic category? *American Journal of Speech-Language Pathology*, **18**, 115–123.

Levickis, P., Sciberras, E., McKean, C., Conway, L., Pezic, A., Mensah, F. K., … Reilly, S. (2018). Language and social-emotional and behavioural wellbeing from 4 to 7 years: A community-based study. *European Child & Adolescent Psychiatry*, **27**, 849–859.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York: Oxford University Press.

Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., & Gillam, R. B. (2015). Utility of a language screening measure for predicting risk for language impairment in bilinguals. *American Journal of Speech-Language Pathology*, **24**, 426–437.

Magimairaj, B. M., & Montgomery, J. W. (2013). Examining the relative contribution of memory updating, attention focus switching, and sustained attention to children's verbal working memory span. *Child Development Research*, **2013**, 1–12.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, **18**, 5–11.

Morgan, A., Bonthrone, A., & Liégeois, F. J. (2016). Brain basis of childhood speech and language disorders: Are we closer to clinically meaningful MRI markers? *Current Opinion in Pediatrics*, **28**, 725–730.

Motsch, H.-J., & Rietz, C. (2019). *Evozierte Sprachdiagnose grammatischer Fähigkeiten (EGRAF)–2. Auflage*. Reinhardt: München.

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: Reuniting the developmental disorders? *Trends in Neuroscience*, **30**, 135–141.

Norbury, C. F., Vamvakas, G., Gooch, D., Baird, G., Charman, T., Simonoff, E., & Pickles, A. (2017). Language growth in children with heterogeneous language disorders: A population study. *Journal of Child Psychology and Psychiatry*, **58**, 1092–1105.

Orgass, B. (1982). *Token Test Manual*. Weinheim: Beltz.

Orgass, B., Poeck, K., Hartje, W., & Kerschensteiner, M. (1973). Zum Vorschlag einer Kurzform des Token Tests zur Auslese von Aphasikern. *Nervenarzt*, **44**, 93–95.

Pauls, L. J., & Archibald, L. M. (2016). Executive functions in children with specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, **59**, 1074–1086.

Peña-Casanova, J., Quiñones-Úbeda, S., Gramunt-Fombuena, N., Aguilar, M., Casas, L., Molinuevo, J. L., … Martínez-Parra, C. (2009). Spanish Multicenter Normative Studies (NEURONORMA Project): Norms for Boston Naming Test and Token Test. *Archives of Clinical Neuropsychology*, **24**, 343–354.

Pijnacker, J., Davids, N., van Weerdenburg, M., Verhoeven, L., Knoors, H., & van Alphen, P. (2017). Semantic processing of sentences in preschoolers with specific language impairment: Evidence from the N400 effect. *Journal of Speech, Language, and Hearing Research*, **60**, 627–639.

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, **25**, 15–24.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Sec. 2. The Coloured Progressive Matrices*. San Antonio, TX: Harcourt Assessment.

Ripich, D. N., Carpenter, B., & Ziol, E. (1997). Comparison of African-American and white persons with Alzheimer's disease on language measures. *Neurology*, **48**, 781–783.

Roello, M., Ferretti, M. L., Colonnello, V., & Levi, G. (2015). When words lead to solutions: Executive function deficits in preschool children with specific language impairment. *Research in Developmental Disabilities*, **37**, 216–222.

Schiff, R., Nuri Ben-Shushan, Y., & Ben-Artzi, E. (2017). Metacognitive strategies: A foundation for early word spelling and reading in kindergartners with SLI. *Journal of Learning Disabilities*, **50**, 143–157.

Schmitt, M. B., Logan, J. A., Tambyraja, S. R., Farquharson, K., & Justice, L. M. (2017). Establishing language benchmarks for children with typically developing language and children with language impairment. *Journal of Speech, Language, and Hearing Research*, **60**, 364–378.

Sim, F., Haig, C., O'Dowd, J., Thompson, L., Law, J., McConnachie, A., … Wilson, P. (2015). Development of a triage tool for neurodevelopmental risk in children aged 30 months. *Research in Developmental Disabilities*, **45**, 69–82.

Siu, A. L. (2015). Screening for speech and language delay and disorders in children aged 5 years or younger: US Preventive Services Task Force Recommendation Statement. *Pediatrics*, **136**, e474–e481.

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, **17**, 323–332.

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 & Communication Disorders*, **48**, 265–282.

Smith-Lock, K., Leitão, S., Lambert, L., Prior, P., Dunn, A., Cronje, J., … Nickels, L. (2013). Daily or weekly? The role of treatment frequency in the effectiveness of grammar treatment for children with specific language impairment. *International Journal of Speech-Language Pathology*, **15**, 255–267.

Snowling, M. J., Bishop, D., Stothard, S. E., Chipchase, B., & Kaplan, C. (2006). Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *Journal of Child Psychology and Psychiatry*, **47**, 759–765.

Spaulding, T. J. (2010). Investigating mechanisms of suppression in preschool children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **53**, 725–758.

Steinberg, L., Vandell, D., & Bornstein, M. (2010). *Development: Infancy through adolescence*. Belmont, CA: Wadsworth.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

Taylor, R. (1998). Indices of neuropsychological functioning and decline over time in dementia. *Archives of Gerontology and Geriatrics*, **27**, 165–170.

Theodorou, E., Kambanaros, M., & Grohmann, K. K. (2013). Specific language impairment in Cypriot Greek: Diagnostic issues. *Linguistic Variation*, **13**, 217–236.

Theodorou, E., Kambanaros, M., & Grohmann, K. K. (2017). Sentence repetition as a tool for screening morphosyntactic abilities of bilectal children with SLI. *Frontiers in Psychology*, **8**, 2104.

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, and Hearing Research*, **40**, 1245–1260.

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, **39**, 1284–1294.

Uda, S., Matsui, M., Tanaka, C., Uematsu, A., Miura, K., Kawana, I., & Noguchi, K. (2015). Normal development of human brain white matter from infancy to early adulthood: A diffusion tensor imaging study. *Developmental Neuroscience*, **37**, 182–194.

Van der Lely, H. K., & Pinker, S. (2014). The biological basis of language: Insight from developmental grammatical impairments. *Trends in Cognitive Sciences*, **18**, 586–595.

Westby, C. (2017). Executive Function Deficits in Language Impairment. (Summary of Young and Gray's (2017) Executive function in preschoolers with primary language impairment. (Journal of Speech, Language, and Hearing Research, 60, 379–392.) *Word of Mouth*, **29**, 7–10. doi:10.1177/1048395017749870b

Willinger, U., Brunner, E., Diendorfer-Radner, G., Sams, J., Sirsch, U., & Eisenwort, B. (2003). Behaviour in children with language development disorders. *Canadian Journal of Psychiatry*, **48**, 607–614.

Willinger, U., & Eisenwort, B. (1999). Verbale und nonverbale Intelligenz bei sprachentwickungsgestörten Kindern. *Klinische Pädiatrie*, **211**, 445–449.

Willinger, U., Schmoeger, M., Deckert, M., Eisenwort, B., Loader, B., Hofmair, A., & Auff, E. (2017). Screening for specific language impairment in preschool children: Evaluating a screening procedure including the token test. *Journal of Psycholinguistic Research*, **46**, 1237–1247.

Willinger, U., Voelkl-Kernstock, S., Neubauer, R., & Einenwort, B. (2001). Verbal and non-verbal intellectual capacity in preschool-time and spelling and reading abilites in school-time among children with specific language impairment. In K. Kallus, N. Posthumus, & P. Jimenez (Eds.), *Current psychological research in Austria* (pp. 107–110). Graz: Akademische Druck- und Verlagsanstalt.

Wittke, K., & Spaulding, T. J. (2018). Which preschool children with specific language impairment receive language intervention? *Language, Speech, and Hearing Services in Schools*, **49**, 59–71.

Cite this article: Schmoeger, M., Deckert, M., Eisenwort, B., Loader, B., Hofmair, A., Auff, E., and Willinger, U. (2020). Evaluating Part V of the German version of the Token Test as a screening of specific language impairment in preschoolers. *Applied Psycholinguistics* **41**, 237–258. https://doi.org/10.1017/S0142716419000493