International Journal of Speech-Language Pathology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/iasl20

Identifying developmental language disorder (DLD) in multilingual children: A case study tutorial

Saboor Hamdani, Angel Chan, Rachel Kan, Shula Chiat, Natalia Gagarina, Ewa Haman, Magdalena Łuniewska, Kamila Polišenská & Sharon Armon-Lotem

To cite this article: Saboor Hamdani, Angel Chan, Rachel Kan, Shula Chiat, Natalia Gagarina, Ewa Haman, Magdalena Łuniewska, Kamila Polišenská & Sharon Armon-Lotem (20 May 2024): Identifying developmental language disorder (DLD) in multilingual children: A case study tutorial, International Journal of Speech-Language Pathology, DOI: 10.1080/17549507.2024.2326095

To link to this article: https://doi.org/10.1080/17549507.2024.2326095

International Journal of Speech-Language Pathology, 2024; Early Online: 1–15

Identifying developmental language disorder (DLD) in multilingual children: A case study tutorial

SABOOR HAMDANI[1,2], ANGEL CHAN[1,2,3], RACHEL KAN[1], SHULA CHIAT[4], NATALIA GAGARINA[5], EWA HAMAN[6], MAGDALENA ŁUNIEWSKA[6], KAMILA POLIŠENSKÁ[4,7] & SHARON ARMON-LOTEM[8]

[1]*Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong,* [2]*Research Centre for Language, Cognition, and Neuroscience, The Hong Kong Polytechnic University, Hung Hom, Hong Kong,* [3]*Peking University Research Centre in Chinese Linguistics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong,* [4]*City, University of London, London, UK,* [5]*Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin, Germany,* [6]*Faculty of Psychology, University of Warsaw, Warsaw, Poland,* [7]*The University of Manchester, Manchester, UK, and* [8]*English Literature and Linguistics and the Gonda Interdisciplinary Brain Research Center, Bar-Ilan University, Ramat Gan, Israel*

Abstract

Purpose: A long-standing issue in identifying developmental language disorder (DLD) in multilingual children is differentiating between effects of language experience and genuine impairment when clinicians often lack suitable norm-referenced assessments. In this tutorial we demonstrate, via a case study, that it is feasible to identify DLD in a multilingual child using the CATALISE diagnostic criteria, Language Impairment Testing in Multilingual Settings (LITMUS) assessment tools, and telepractice.

Method: This tutorial features a case study of one 6-year-old Urdu-Cantonese multilingual ethnic minority child, and seven age- and grade-matched multilinguals. They were tested via Zoom using Urdu versions of the Multilingual Assessment Instrument for Narratives (LITMUS-MAIN), the Crosslinguistic Lexical Task (LITMUS-CLT), the Crosslinguistic Nonword Repetition Test (LITMUS-CL-NWR), and the Sentence Repetition Task (LITMUS-SRep).

Result: The child scored significantly lower in the LITMUS tests compared to her peers in her best/first language of Urdu. Together with the presence of negative functional impact and poor prognostic features, and absence of associated biomedical conditions, the findings suggest this participant could be identified as having DLD using the CATALISE diagnostic criteria.

Conclusion: The result demonstrates the promise of this approach to collect reference data and identify DLD in multilingual children. The online LITMUS battery has the potential to support identification of multilingual DLD in any target language.

Keywords: *developmental language disorder; multilingual ethnic minority children; Urdu; CATALISE; LITMUS battery; telepractice*

Introduction

At least 7–11% of 5-year-olds worldwide are affected by difficulties in using their first language, which can affect their everyday communication but are not linked to a clear biomedical aetiology (Bishop et al., 2017; Norbury et al., 2016; Tomblin et al., 1997). This condition is identified as developmental language disorder (DLD), which negatively impacts an individual's academic progress, increases the risk of mental health problems in adolescence, and limits career choices in adulthood, if left untreated. A long-standing issue in identifying DLD has been differentiating the effects of

Correspondence: Angel Chan, Speech Therapy Unit, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. E-mail: angel.ws.chan@polyu.edu.hk; Saboor Hamdani, Speech Therapy Unit, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. E-mail: saboor.hamdani@connect.polyu.hk

Supplemental data for this article can be accessed online at https://doi.org/10.1080/17549507.2024.2326095.

ISSN 1754-9507 print/ISSN 1754-9515 online

Published by Taylor & Francis
DOI: 10.1080/17549507.2024.2326095

language experience from impairment in multilingual children. We use the term "multilingual" based on the definition by the International Expert Panel on Multilingual Children's Speech (IEPMCS; 2012):

> People who are multilingual, including children acquiring more than one language, are able to comprehend and/or produce two or more languages in oral, manual, or written form with at least a basic level of functional proficiency or use, regardless of the age at which the languages were learned (IEPMCS, 2012, p. 1).

In this tutorial we demonstrate, via a case study, that it is feasible to identify DLD in a multilingual child using the CATALISE diagnostic criteria, Language Impairment Testing in Multilingual Settings (LITMUS) assessment tools, and telepractice. Before presenting the case study, as background we will first highlight the CATALISE diagnostic criteria, a battery of assessment tools named LITMUS, crosslinguistic evidence for the LITMUS battery in identifying language disorder (LD) or specifically DLD in children, and some current literature on telepractice, in the following sections.

Diagnostic criteria by the CATALISE consortium (Bishop et al., 2017)

The CATALISE consortium provides a binary decision tree that can be followed to help decide whether a multilingual child presenting with speech, language, and communication needs might potentially be identified as having DLD. For a multilingual child who shows difficulty in using language in a way that affects everyday functioning, one needs to consider whether the child is unfamiliar with the majority language, especially when it is also the target language that the child is assessed in. If the child is unfamiliar with the majority language but competent in another language, this may reflect language needs due to insufficient exposure to the majority language. This would not be considered language difficulties that constitute a disorder. However, if a child who is unfamiliar with the majority language is also not competent in another, or any, language, then features of poor prognosis, such as difficulties in multiple language domains, especially receptive ones, may suggest language difficulties that are likely to persist.

For the children identified with language difficulties not associated with a biomedical condition, it is termed DLD under the diagnostic criteria of Bishop et al. (2017). Otherwise, the condition is termed LD associated with that biomedical condition. Both types of disorders can co-occur with disorders in other domains such as attention deficit hyperactivity disorder (ADHD) and dyslexia, or manifest differently among children in terms of the language areas that are more or less impaired. It is important to note that having low non-verbal ability (but not intellectual disability) and/or the presence of biological or environmental risk factors do not exclude a child from being identified as having DLD.

Based on the CATALISE diagnostic criteria (Bishop et al., 2017), multilingual children presenting with the following characteristics could be identified as having DLD: multilingual children (suspected of) having DLD would have demonstrated evidence in terms of negative functional impact, lack of competence even in the best language, presence of poor prognostic features, and absence of associated biomedical conditions.

LITMUS battery

Recall one major consideration in the CATALISE criteria in determining whether a multilingual child has LD/DLD or not is whether "there is evidence that the child does not have age-appropriate skills in any language" (Bishop et al., 2017, p. 1071; Statement 4). Addressing this consideration requires the support of assessment tools to collect objective and informative evidence from multiple language domains to guide the diagnostic decisions. A battery of assessment tools named LITMUS, developed under The European Cooperation in Science and Technology (COST) Action IS0804 (2009–2013), is a good candidate for this purpose. This LITMUS battery offers a set of crosslinguistically and cross-culturally applicable tools that allow us to test multiple language domains. These assessment tools have been repeatedly shown to be able to differentiate children with and without DLD in crosslinguistic studies (e.g. Altman et al., 2016; Armon-Lotem & Meir, 2016; Boerma et al., 2016; Gagarina, Gey, et al., 2019; Kapalková & Slančová, 2017; Saliby et al., 2017; Tsimpli et al., 2016).

These assessment tools have been adapted into numerous languages (see Table 1 in Supplementary Materials). Their crosslinguistic and cross-cultural appeal allows researchers and clinicians to assess children acquiring diverse languages growing up in different parts of the world and assess multilingual children in their multiple languages using assessment tools with parallel designs. Table 1 in our Supplementary Materials displays the number of language versions available for each tool, with information extracted from its respective website and/or provided by the key members of the LITMUS committee (shown in brackets; also authors of this paper), including for the Crosslinguistic Lexical Tasks (LITMUS-CLT: Haman & Luniewska), the Multilingual Assessment Instrument of Narratives (LITMUS-MAIN: Gagarina), the Crosslinguistic Nonword Repetition Test (LITMUS-CL-NWR: Chiat & Polišenská), and the Sentence Repetition Task (LITMUS-SRep: Armon-Lotem).

LITMUS tools: Crosslinguistic evidence of their clinical utility in identifying DLD

The LITMUS battery includes different tools to assess multiple language areas. For example, the LITMUS-MAIN (Gagarina, Klop, et al., 2019) assesses narrative abilities, the LITMUS-CLT (Haman et al., 2015) assesses lexical competence, the LITMUS-CL-NWR (Chiat, 2015) targets nonword repetition abilities, and the LITMUS-SRep (Marinis & Armon-Lotem, 2015) examines morphosyntactic abilities.

LITMUS-MAIN for assessing narrative abilities. The clinical utility of these LITMUS tools in identifying DLD in multilingual children has been repeatedly reported in different crosslinguistic studies. In assessing narrative abilities, the LITMUS-MAIN has been used in multiple studies to examine the narrative competence of children with and without DLD from diverse linguistic backgrounds including multilingual children acquiring Russian and German or Turkish aged 4;5 (years, months; Gagarina, Gey, et al., 2019), multilinguals acquiring Russian and Hebrew aged 5;6–6;6 (Fichman et al., 2017), multilinguals acquiring English and Hebrew aged 5;3–6;5 (Altman et al., 2016), multilinguals acquiring Greek aged 9;0–9;1 (Tsimpli et al., 2016), and multilinguals acquiring Greek and Albanian aged 6–8 (Peristeri et al., 2020). These studies have consistently captured significant typically developing (TD)/DLD group differences, showing that children with DLD scored significantly lower than their TD peers in narrative abilities. Moreover, Boerma et al. (2016) examined the diagnostic accuracy of the Dutch version of the LITMUS-MAIN. In the monolingual group, the LITMUS-MAIN was able to identify 85% of the children with LD (sensitivity) and 79% of the children with typical development (specificity). In the multilingual group, the reported sensitivity and specificity were 79% and 88%, respectively.

LITMUS-CLT for assessing lexical competence. Regarding lexical competence, crosslinguistic studies using the LITMUS-CLT have also reported TD/DLD significant group differences involving multilingual children acquiring Lebanese and English or French aged 5;7 to 7;10 (Saliby et al., 2017), multilinguals acquiring Gaelic and English aged 6–8 (Chondrogianni et al., 2021), and monolinguals acquiring Slovak aged 4;3–5;5 (Kapalková & Slančová, 2017) in expressive and/or receptive lexical abilities. These findings showed that children with DLD scored significantly lower than their age-matched TD peers at both receptive and expressive lexical levels (Chondrogianni et al., 2021; Kapalková & Slančová, 2017; Saliby et al., 2017). Moreover, at an individual level, Eikerling et al. (2023) evaluated the diagnostic accuracy of the verb comprehension subtest in the Italian version of the LITMUS-CLT, reporting a specificity of 78% and a sensitivity of 80% in multilingual Italian-speaking children with and without DLD.

LITMUS-SRep for assessing morphosyntactic abilities. There have also been multiple studies using the LITMUS-SRep to examine the morphosyntactic abilities in children with and without DLD from different language backgrounds, for example, French-speaking children aged 5–8 (de Almeida et al., 2017), French-speaking children aged 5;2–8;9 (Fleckstein et al., 2016), French or German-speaking children aged 5;6–8;11 (Tuller et al., 2018), Russian and German or Turkish children aged 4;5 (Gagarina, Gey, et al., 2019), and Russian and Hebrew children aged 5;5–6;8 (Armon-Lotem & Meir, 2016). These studies reported that the LITMUS-SRep could successfully capture significant TD/DLD group differences. Some of these studies also examined the diagnostic accuracy of the LITMUS-SRep in distinguishing multilingual children with and without DLD at an individual level. de Almeida and colleagues (2017) reported the diagnostic specificity and sensitivity of the French LITMUS-SRep to be around 72% and 76%, respectively. Armon-Lotem and Meir (2016) reported diagnostic accuracy of the LITMUS-SRep to be around 81% in first language (L1) Russian. They also reported the sensitivity and specificity of the LITMUS-SRep to be 100% and 89%, respectively, in second language (L2) Hebrew.

LITMUS-NWR for assessing nonword repetition abilities. Some of the studies cited above (e.g. Armon-Lotem & Meir, 2016; de Almeida et al., 2017; Tuller et al., 2018) and others (e.g. Grimm, 2022: 8–10-year-old monolingual and L2 learners of German) examined nonword repetition abilities using the LITMUS-NWR, and reported TD/DLD significant group differences. Moreover, at an individual level, Armon-Lotem and Meir (2016) reported the sensitivity and specificity of the Russian LITMUS-NWR to be around 70% and 76%, respectively, in L1 Russian multilinguals, and the sensitivity and specificity of the Hebrew LITMUS- NWR to be around 81% and 79%, respectively, in L2 Hebrew multilinguals. de Almeida et al. (2017) reported the diagnostic accuracy of the French LITMUS-NWR to be around 80% in identifying children with DLD.

Telepractice

While the feasibility of providing virtual assessments for families residing in remote and rural areas has been established for some time (Coleman et al., 2015; Sutherland et al., 2016, 2017; Wright, 2018), the outbreak of COVID-19 in late 2019 transfigured this understudied and alternative mode of delivery into a pressing need. Sudden lockdown orders disrupted in-person activities, including face-to-face testing, implying that telepractice was no longer seen as an alternative mode of delivery but the only

available mode for conducting assessments during this unusual period (Nelson & Plante, 2022).

Lately, there are some recent efforts in evaluating the efficacy of telepractice for conducting diagnostic assessments to identify children with and without LD/DLD, and the findings are promising. For example, Nelson and Plante (2022) compared the findings of the Test of Integrated Language and Literacy Skills (TILLS) when administered in person and online. A total of 51 participants with and without language or literacy disorders aged between 6 and 18 years were tested using different subtests of the TILLS. The findings yielded 96% agreement between the two testing modes in identifying children with and without language or literacy disorder. In another study, Eikerling et al. (2023) validated the remote screening of DLD in multilingual Spanish-Italian speaking children using a new web-based application called MuLiMi. Their study recruited 36 Spanish-speaking preschoolers aged 4–6 years old, 16 of whom were already diagnosed with DLD. Language abilities in both L1 and L2 were assessed using some dynamic tasks (novel word learning) and static tasks (verb comprehension, grammaticality judgement, and nonword repetition). They reported significant correlations between the screening task scores and outcome measures from parent, speech-language pathologist (SLP), and teacher questionnaires, and also screening task scores and standardised test scores, providing supporting evidence that the Spanish-Italian MuLiMi application has the diagnostic potential in identifying children at risk for DLD.

This paper

The purpose of this paper is to offer a tutorial, with a case study, to illustrate the feasibility of identifying DLD in multilingual children using the CATALISE criteria in combination with the LITMUS tools via telepractice.

Like many other places of the world, Hong Kong has a growing population of multilingual ethnic minority children, with Pakistanis as a major minority population. The size of the Pakistani population in Hong Kong has increased by 37% from 2006 to 2016 (Census and Statistics Department, Government of the Hong Kong Special Administrative Region, 2017). Many Pakistani children in Hong Kong acquire Urdu as their home language and L1, and Cantonese as a L2 in their school and community. Multilingual Pakistani children have been increasingly encountered by SLPs in recent years. As these children are acquiring both Urdu and Cantonese under reduced input, they may lag behind their monolingual peers in each language and therefore face difficulties in accessing education in schools where they need Cantonese, as well as less input to support development and maintenance of Urdu as their heritage language (e.g. Bosch & Sebastián-Gallés, 2003; Paradis, 2010). However, there are obstacles in providing adequate and appropriate language support, as SLPs and educators in Hong Kong know little about the speech and language development of these children, including the L1 Urdu and its influence on the development of the L2 and community language Cantonese (e.g. Cummins, 2000; Gu & Patkin, 2013). In addition, speech and language assessment tools are not available for Urdu, and tools for Cantonese only have monolingual norms and may not be applicable for this multilingual population. Therefore, language difficulties (i.e. yet to be diagnosed as LD/DLD) cannot be reliably identified in these children—a common issue also faced by researchers and clinicians globally when assessing multilingual ethnic minority children (Armon-Lotem, 2018; Bedore & Peña, 2008). In order to improve the differentiation of Urdu-Cantonese children in Hong Kong with and without DLD, the LITMUS tools have been adapted to Urdu and Cantonese by our research team in order to provide assessment tools for multilingual ethnic minority children.

In this tutorial, we demonstrate how online adaptations of the LITMUS tools can provide informative data when we identify potential DLD in a multilingual child in light of the CATALISE diagnostic criteria, using Urdu-Cantonese multilingual ethnic minority children in Hong Kong as an illustrative example. Importantly, the online versions of these tests have the potential to support the identification of multilingual DLD in any target language in terms of global impact. This article mainly discusses the conceptual and methodological considerations, in the context of a case study, presented as a proof of concept that could pave the way to more wide-ranging research that would have positive societal implications in other parts of the world.

Method

Participants

The participant in this tutorial identified as potentially having DLD (suspected DLD, the S-DLD child), was a female aged 6;8 at the time of testing. She was referred to the in-house speech therapy clinic of the university of the first three authors of this paper by her schoolteacher, who expressed concerns about the child's language abilities. The participant was born in Hong Kong, and her mother reported typical pregnancy and delivery. She had no significant biomedical history, and her mother reported typical development in motor, emotional, and social areas. Although both Urdu and Punjabi were reported as family/home languages, Urdu was reported as being used more frequently and also as the participant's strongest language. The participant also spoke English and Cantonese. Cantonese, the societal and school language, was reported as the weakest language according to parental report (see Table 2 in

Supplementary Materials). She started schooling at the age of 3;6, where the medium of instruction was Cantonese. Her mother expressed concerns related to speech intelligibility, memory when related to academic tasks, and learning ability. The school also expressed concerns about the participant's language abilities that affected everyday social interactions and educational progress. However, her Raven's Progressive Matrices (Raven and Court, 1998) standard score of 91 suggested no intellectual disability.

The other participants included seven Urdu-Cantonese multilingual ethnic minority primary first graders in Hong Kong, considered TD in this tutorial (TD-1 to TD-7), aged between 6;1 and 7;4 (mean age = 6;6, *SD* = 4 months). There were three female and four male participants, and they all achieved a Raven's Progressive Matrices standard score of above 75, indicating no intellectual disability (mean = 92.43, *SD* = 9.13). All participants spoke Urdu as their strongest and home language as ranked by the caretakers (mothers), which is consistent with information on the amount of language exposure also provided in the parental questionnaires (see Table 2 in Supplementary Materials). They were all enrolled in local primary schools with Cantonese as the medium of instruction and were studying at the same grade level (P1). All participant families, including the S-DLD child, had comparable socioeconomic status (SES). Although this sample size is relatively small, this is not uncommon in language acquisition studies on multilingual ethnic minority children (e.g. Farndale et al., 2016; Hu et al., 2014; Serrano-Hidalgo, 2018).

All the caretakers of the participants gave informed written consent to participate in this study. The study was approved by the ethics committee at the university of the first three authors of this paper.

Instruments

The tests used included Urdu versions of the Multilingual Assessment Instrument for Narratives (LITMUS-MAIN; Gagarina, Klop, et al., 2019; Hamdani, Kan, Chan, & Gagarina, 2020; Kan et al., 2020), the Crosslinguistic Lexical Task (LITMUS-CLT; Haman et al., 2015; Hamdani, Kan, & Chan, 2020), the Crosslinguistic Nonword Repetition Test (LITMUS-CL-NWR; Chiat, 2015; Polišenská & Kapalková, 2014; Hamdani, Chan, Kan, Chiat, & Polišenská, 2020), and the Sentence Repetition Task (LITMUS-SRep; Marinis & Armon-Lotem, 2015; Hamdani, Chan, Kan, & Armon-Lotem, 2020).

Multilingual Assessment Instrument for Narratives (LITMUS-MAIN)

The LITMUS-MAIN is used to assess narrative abilities and consists of narrative telling, retelling, and comprehension tasks. The instructions for telling and retelling are delivered through videos that show the hands folding and unfolding the story pictures, with recorded audio narration. The narrator is shown as a cartoon character of a girl, who emphasises that only the participant can see the pictures. For the retelling mode, the whole recorded story script is played to the participants before they are prompted to retell the story. For each story, up to 10 questions for the comprehension mode are presented in live voice, because the wording of the questions depends on the responses given by the participant in previous questions. Specific relevant pictures are highlighted in turn by special effects of a red coloured frame and "pulse" animation in PowerPoint for each comprehension question.

In this study, participants first listen to a prerecorded model story script with picture support and have to retell the story with picture support. Then, they answer the 10 standard comprehension questions about the story. This procedure is repeated for another story, to collect narrative data from two stories. After completing some other language tasks as a "gap," they would be asked to tell the two stories again in turn with the same picture support (but without prior listening to a story script). That is, each child produces a total of four narrative samples, two during retelling and two during telling.

Crosslinguistic Lexical Task (LITMUS-CLT)

The LITMUS-CLT is used to assess lexical abilities and includes production and comprehension modes, with 32 nouns and 32 verbs tested for each mode (including two trial items per mode). For the production mode, participants are asked to orally name the illustration shown on the screen. For the comprehension mode, participants are asked to select the illustration matching the audio stimuli from four options by saying the number of the selected illustration, with the position of the target illustrations varied across trials. The target words are selected from a pool of illustrations standardised for all language versions, with balancing for the semantic type and level of difficulty. The semantic types for noun stimuli include animate natural objects, inanimate natural objects, and artefacts; the verb stimuli include actions performed by humans, by animals, and states/unintentional actions. The level of difficulty is determined by a complexity index, which depends on grammatical and phonological features of the lexical item, and the age of acquisition of the target word, which is determined by ratings from 25 native speakers of Urdu. In terms of the order of presentation, tasks of the same mode are presented together, with the order of mode and order of noun vs. verb counterbalanced across participants. The order of presentation for trials within a subtask is fixed.

Crosslinguistic Nonword Repetition Test (LITMUS-CL-NWR)
The LITMUS-CL-NWR, as the name indicates, is used to assess nonword repetition abilities. Difficulty with repeating nonwords has been put forward as a clinical marker of DLD crosslinguistically (Schwob et al., 2021). The LITMUS-CL-NWR task is presented as a game in which the participants are asked to fix a necklace by repeating the stimuli (Polišenská & Kapalková, 2014). Stimuli are between two and five syllables long, with four items per length. They are selected from a pool of four to six candidates per item that are matched in length (allowing selection of different alternatives for different language versions in case one or more of the candidates is a real word in the language), with the further criterion that the component phonemes should be acquired by the age between 3 and 3;6 years in Urdu-speaking children (Bari & Ajmal, 2016; Zahra, 2016). The stimuli are presented in one randomised order. Two training stimuli are also administered before the task begins.

Sentence Repetition Task (LITMUS-SRep)
The LITMUS-SRep is used to measure morphosyntactic abilities and the task is presented as a treasure-hunting game, where the participants are requested to repeat sentences in order to help a bear find a treasure. There are three blocks of stimuli, corresponding to three levels of syntactic difficulty. The length of the stimuli in each block is shown in Table 3 (Supplementary Materials). A total of 10 syntactic structures are tested, with three in the language specific condition and seven in the language independent condition, which are considered challenging for children with DLD in the target language and across languages, respectively. The target structures were chosen considering the structures tested in the original English SRep task (Marinis & Armon-Lotem, 2015), crosslinguistic literature on DLD regarding vulnerabilities, the typological characteristics of Urdu, and clinical experiences with Urdu-speaking children having language disorders. Each structure (except one) is tested using three sentences, with nominal and verbal inflexional morphology signalling number, gender, and tense controlled within items across structures. One structure (SOV with one auxiliary/modal verb) has eight sentences, as we wanted to assess nominal and verbal morphology (number and gender agreement and case marking), which are vulnerable in children with DLD, more comprehensively in simple structures. The task is presented in three randomised orders. Two training stimuli are also administered before the task begins.

Raven's progressive matrices
A non-verbal IQ test (Raven's Progressive Matrices; Raven & Court, 1998) is also administered to participants to screen out intellectual disability. In Raven's Progressive Matrices, the participants answer 60 multiple choice questions, where they have to identify the missing piece that completes a pattern from six to eight options. A standard score of 70 or above is considered within the normal range.

Parents of bilingual children questionnaire (LITMUS-PaBiQ; Tuller, 2015)
The Urdu version of the LITMUS-PaBiQ was also administered during the recruitment phase before the assessment session(s). The first author, an experienced Urdu-speaking SLP, contacted the caregivers (mothers) of the participants and filled in the questionnaire over a phone call.

The LITMUS-PaBiQ (Tuller, 2015), available for free, collects data on the participant's language environment; language exposure; receptive and expressive competence in each language; frequency of usage of different languages in daily routines; early language and other developmental history; and family history related to speech, language, and hearing issues. This parental questionnaire also gathers information related to the parent/caregiver's education level, occupation, language competence in different languages, and frequency of usage with the child in each language.

The LITMUS-PaBiQ can help obtain important information related to a child's developmental language history, language exposure, language use, and presence or absence of any biomedical conditions. This information could help clinicians determine whether a child's (poor) linguistic competence is associated with certain risk factors related to LD/DLD, factors related to multilingualism, or any significant biomedical condition. This also ensures that a parent/caregiver can share their observations and concerns about a child's speech and language difficulties with the researcher/clinician in a structured way. The LITMUS-PaBiQ has been reported to be informative in differentiating children with and without LD or DLD, or in predicting LD or DLD, in monolingual and multilingual children in several crosslinguistic studies (Boerma & Blom, 2017; Hreich & Messarra, 2013; Tuller et al., 2018).

Online testing procedure

All testing sessions were conducted by an Urdu-speaking SLP. The LITMUS tests used were computerised versions that were adapted for telepractice, which were shown to the participants via the "share screen" function of Zoom. We followed the testing procedures of the original tests designed for face-to-face testing, and only added numbers to each picture in the LITMUS-CLT tasks to facilitate online testing (i.e. instead of pointing at pictures, participants were asked to say the numbers added to the pictures). Participants used a standard set of headphones equipped with a microphone to standardise audio quality, which was mailed to them and offered as a token of appreciation for their participation before

the testing session. The experimenter also used the same microphone for each testing session. In addition, participants were required to use a computer, laptop, or tablet in a quiet environment with stable Wi-Fi and with only an accompanying parent or guardian (testing in public was not recommended), and to switch on their camera during testing. Testing took place at the participants' homes (parents were instructed to keep the ambient noise to minimum level during testing), with most testing completed in one session (around 60–70 min), but if required to suit an individual child's attention, it was divided into two sessions.

The parents/caregivers were requested not to instruct or interrupt the participants during the testing session. They were further advised to use the highest volume level of the headset. Although the volume can also be controlled by the computer, and therefore the exact volume level was not standardised across participants, the volume level was controlled with the following arrangement: At the outset, the participants were asked to indicate whether the speech sounds they were hearing through the headset were loud and clear enough on their side. Moreover, some simple practice trials were conducted before starting the test items in each of the nonword repetition, sentence repetition, and crosslinguistic lexical tasks, to obtain some objective evidence that the participant could hear the auditory stimuli clearly before proceeding to the testing phase.

Scoring

For the comprehension task of the LITMUS-MAIN, participants answer at most 10 questions per story, with one point per question. As participants tell two different stories in Urdu, the maximum score is 20. For the production task, scores for storytelling and retelling are calculated separately. For each mode, participants are given scores on story structure (maximum = 17 per story). Participants are given one score for each of the 16 scoring items, apart from the first one where participants can get a maximum of two points. Scoring items for story structure include the setting for the story (time and place reference), internal state terms (IST) as part of an initiating event[1], and goal (G), attempt (A), outcome (O) and IST as a reaction for each of the three episodes.

Story complexity counts the number of sequences or episodes in terms of AO, GA/GO, GAO, and single G. Since there is currently no gold standard in terms of the scoring approach one should use that best captures the DLD/TD differences in story complexity, this study has chosen three scoring schemes that have been used in crossinguistic studies whose score weighting assignment aligns with the level of story structure complexity based on the binary decision tree by Westby (2005) and represents three levels of granularity of different levels of structural complexity in macrostructure. Among them, the scoring scheme by Gagarina, Gey, et al. (2019) has the finest level of differentiation (six levels), the scoring scheme by Sheng et al. (2020) has the coarsest level of differentiation (two levels), and the scoring scheme by Maviş et al. (2016) is in the middle (four levels). See Table 4 in Supplementary Materials. Furthermore, the scoring schemes by Gagarina, Gey, et al. (2019) and Sheng et al. (2020) have both successfully captured significant DLD/TD group difference in story complexity.

As far as we know, no studies to date have examined and compared how different scoring schemes that vary in the level of granularity of the different levels of story complexity might affect (the degree of) differentiation between children with TD and those with DLD. It is, however, interesting and important to address this methodological issue as it will bear on diagnostic accuracy. Since scoring approaches may vary in their ability to capture the DLD/TD differences, this tutorial therefore chose to use all three scoring schemes for a more informative comparison. See Table 5 in Supplementary Materials for an example of parts of a narrative sample to illustrate the scoring points.

Furthermore, the use of ISTs was also assessed. ISTs refer to terms that describe the internal states of a character, mainly referring to mental states and feelings such as emotions, thoughts, intentions, and reactions (Gagarina, Klop, et al., 2019). Table 6 (Supplementary Materials) presents some English and Urdu examples of ISTs according to the subtypes of ISTs listed in the LITMUS-MAIN manual. This case study scored the use of ISTs in both token and type measures. Token measures count the number of individual words in a narrative sample (i.e. counting also the repetitions of individual words), while type measures count the number of unique word forms (i.e. repetitions of an individual word will only be counted once).

For the LITMUS-CLT, scores are calculated excluding the trial items, with 30 items in total for each subtask and 4 subtasks in total (noun comprehension, verb comprehension, noun production, verb production). Participants are given one score for each correct item.

For the LITMUS-CL-NWR, there are 16 test items in total, excluding the training items. The maximum score is 16 for whole items correct and 56 for segments correct.

In the LITMUS-SRep, there are 35 test items, excluding the trial items. Participants get one score for producing the test sentence verbatim. The maximum score for this task is 35.

Interrater reliability

For the LITMUS-MAIN, CLT, and SRep, all the data were transcribed by a part-time research assistant who is also an Urdu-speaking SLP and were double-checked by the first author (also an Urdu-

speaking SLP). All the scores for these three measures were first completed by one rater and then double-checked by the second rater (both Urdu-speaking SLPs) to make sure that all the utterances and items were scored correctly. Regarding the LITMUS-MAIN, the narrative samples were scored for both narrative comprehension (two stories per child) and narrative macrostructure production (four stories per child). The initial percentage of agreement between two raters for the narrative comprehension questions was 100% (160 over 160 items), and was as follows across different macrostructure measures: story structure = 99.6% (542 over 544 items), story complexity when scored using the scheme by Maviş et al. (2016) = 99.6% (287 over 288 items), story complexity when scored using the scheme by Sheng et al. (2020) = 100% (96 over 96 items), and story complexity when scored using the scheme by Gagarina, Gey, et al. (2019) = 99.8% (575 over 576 items). The initial percentage of agreement between two raters for the LITMUS-CLT was 99.7% (1021 over 1024 items), and for the LITMUS-SRep agreement was 99.6% (279 over 280 items). These disagreements were minor in nature, were due to occasional omissions and typos, and were resolved to 100% agreement after inter-rater checks.

For the LITMUS-CL-NWR, all the data were transcribed independently by two raters. Both raters were Urdu-speaking SLPs and had relevant training in phonology. Out of a total of 128 items (16 target items per participant, eight participants in total), there was disagreement on nine items where both raters perceived and transcribed the phoneme differently and the disagreement was not resolved by discussion. In this case, a third rater who was also an Urdu-speaking SLP and was working in the field of Urdu phonology was invited to transcribe the discrepant items independently. The transcription for the discrepant items was finalised based on the total number of votes (that is, choosing the transcription that had two out of three votes). Overall, the percentage of agreement between the two raters was 93% (119 over 128 items).

Result

This section presents findings for different language domains assessed using all four measures.

Narrative comprehension

For the LITMUS-MAIN comprehension, the S-DLD child was able to answer only four out of 20 questions correctly. On the other hand, her age- and grade-matched peers scored on average 16.71 ($SD = 2.93$) out of 20. The S-DLD child scored 4, which was the lowest score among all participants, while other participants scored between 13 and 20. Table I presents the narrative comprehension scores.

Narrative production

Table II shows the result of narrative production. Overall, scores in the retelling mode (which was also conducted first) were slightly higher than the telling mode. For story structure, the S-DLD child scored 16 in the retelling mode and 4 in the telling mode. Overall, her age- and grade-matched peers scored on average 15.71 ($SD = 2.56$; range = 11–18) out of 34 in the retelling mode, compared to 16.00 out of 34 in the telling mode ($SD = 3.27$; range = 11–21). Therefore, while the S-DLD child scored similarly relative to her peers in retelling, she had a noticeable drop in performance when the mode switched to telling without the support of a prior script, unlike her peers who could maintain a relatively stable performance across story retelling and telling modes (c.f. Sheng et al., 2020).

Scores for story complexity depended on the scoring scheme used. The S-DLD child scored 8 in retelling and 2 in telling according to the scheme by Maviş et al. (2016), 2 in retelling and 1 in telling according to the scheme by Sheng et al. (2020), and 16 in retelling and 4 in telling according to the scheme by Gagarina, Gey, et al. (2019). Overall, her age- and grade-matched peers scored on average 6.43 ($SD = 2.94$; range = 2–12) out of 18 in retelling compared to 6.29 ($SD = 2.75$; range = 3–10) out of 18 in telling according to the scheme by Maviş et al. (2016), on average 2.00 ($SD = 1.00$; range = 1–4) out of 6 in retelling compared to 1.71 ($SD = 1.38$; range = 0–3) out of 6 in telling according to the scheme by Sheng et al. (2020), and on average 14.57 ($SD = 5.09$; range = 8–25) out of 36 in retelling compared to 14.29 ($SD = 4.42$; range = 10–17) out of 36 in telling according to the scheme by Gagarina, Gey, et al. (2019). This shows that the S-DLD child scored significantly lower than other participants in telling mode, especially according to scoring schemes by Maviş et al. (2016) and Gagarina, Gey, et al. (2019). The S-DLD child also produced no complete GAO episodes, similar to the TD children who overall produced very few complete GAO episodes, from 0–3 for retelling and 0–2 for telling. Even though the S-DLD child produced a relatively high number of

Table I. Scores of narrative comprehension tasks assessed via Language Impairment Testing in Multilingual Settings—Multilingual Assessment Instrument for Narrative (LITMUS-MAIN; Gagarina, Klop, et al., 2019; Hamdani, Kan, Chan, & Gagarina, 2020).

Participant	Comprehension scores
S-DLD	4
TD-1	14
TD-2	20
TD-3	15
TD-4	20
TD-5	13
TD-6	16
TD-7	19
Mean for TD participants	16.71
SD for TD participants	2.93

Note. S-DLD = suspected DLD; TD = typically developing.

Table II. Scores of story retelling and telling tasks assessed via Language Impairment Testing in Multilingual Settings—Multilingual Assessment Instrument for Narrative (LITMUS-MAIN; Gagarina, Klop, et al., 2019; Hamdani, Kan, Chan, & Gagarina, 2020).

Participant	Mode	SS	SC-M	SC-Sh	SC-Ga	GAO	IST tokens (type)
	Retelling						
S-DLD		16	8	2	16	0	24 (4)
TD-1		18	12	4	25	3	21 (9)
TD-2		17	7	2	14	0	11 (4)
TD-3		18	6	2	14	0	17 (6)
TD-4		17	6	1	13	0	9 (5)
TD-5		11	2	1	8	0	9 (6)
TD-6		15	6	2	14	1	12 (6)
TD-7		14	6	2	14	1	12 (5)
Mean for TD participants		15.71	6.43	2.00	14.57	0.71	13.00 (5.86)
SD for TD participants		2.56	2.94	1.00	5.09	1.11	4.43 (1.46)
	Telling						
S-DLD		4	2	1	4	0	28 (5)
TD-1		18	8	3	17	0	19 (4)
TD-2		16	4	0	10	0	11 (4)
TD-3		21	10	3	20	2	15 (5)
TD-4		17	4	0	10	0	13 (8)
TD-5		11	3	1	10	0	7 (6)
TD-6		16	6	2	14	0	10 (5)
TD-7		13	9	3	19	2	8 (3)
Mean for TD participants		16.00	6.29	1.71	14.29	0.57	11.86 (5)
SD for TD participants		3.27	2.75	1.38	4.42	0.98	4.18 (1.51)

Note. S-DLD = suspected DLD; TD = typically developing; SS = story structure; SC-M = story complexity according to Maviş et al. (2016); SC-Sh = story complexity according to Sheng et al. (2020); SC-Ga = story complexity according to Gagarina, Gey, et al. (2019); GAO = goal, attempt, outcome; IST = internal state terms.

ISTs (tokens) compared to her peers, 24 tokens in retelling and 28 tokens in telling, the type measures were relatively low (four in retelling, five in telling). For the age- and grade-matched peers, the IST token measures were on average 13 (*SD* = 4.43; range = 9–21) in retelling and on average 11.86 (*SD* = 4.18; range = 7–19) in telling, while the type measures were comparable to the S-DLD child: on average 5.86 (*SD* = 1.46; range = 4–9) in retelling and on average 5 (*SD* = 1.51; range = 3–8) in telling.

Lexical abilities

Table III shows the scores for the LITMUS-CLT. The S-DLD child obtained lower scores in different subtasks compared to her peers. In comprehension, she scored a total of 28, with 16 in nouns and 12 in verbs. For production, she scored a total of 16, with 9 in nouns and 7 in verbs. Overall, her age- and grade-matched peers achieved higher scores for nouns and verbs in the comprehension mode, with an average of 25.14 (*SD* = 2.73; range = 21–27) out of 30 in nouns and 25.29 (*SD* = 2.06; range = 23–28) out of 30 in verbs. In the production mode, scores in nouns (mean = 17.57, *SD* = 6.16; range = 10–26) out of 30 and verbs (mean = 14.71, *SD* = 2.43; range = 11–18) out of 30 were also higher.

Nonword repetition abilities

Table IV presents the LITMUS-CL-NWR scores. One participant (TD-1) had suspected speech sound disorder based on her substitution errors in speech that were not age-appropriate, despite no parental concerns about her language development (mother expressed some concerns about her speech problems affecting her speech intelligibility but reported normal language development) and her performance in the other receptive and expressive language tasks being comparable to the other age- and grade-matched peers. Some of her errors were inconsistent and were marked as incorrect, but the error of substituting retroflex sounds /ʈ/ and /ɖ/ with dental sounds /t̪/ and /d̪/, respectively, was relatively consistent and was not penalised following the standard scoring principle for giving allowance to consistent substitution errors. Overall, the S-DLD child scored the lowest compared to all other participants (including TD-1), with a score of 3 when measured by item and 17 when measured by segment. On average, her age- and grade-matched peers scored 7.00 (*SD* = 1.53; range = 5–9) out of 16 when measured by item and 42.00 (*SD* = 3.46; range = 37–47) out of 56 when measured by segment.

Table III. Scores representing lexical abilities assessed through Language Impairment Testing in Multilingual Settings—Crosslinguistic Lexical Tasks (LITMUS-CLT; Haman et al., 2015; Hamdani, Kan, & Chan, 2020).

Participant	Comprehension			Production		
	Total	Noun	Verb	Total	Noun	Verb
S-DLD	28	16	12	16	9	7
TD-1	46	21	25	28	12	16
TD-2	52	25	27	25	10	15
TD-3	57	29	28	37	22	15
TD-4	47	24	23	24	12	12
TD-5	46	23	23	38	20	18
TD-6	54	27	27	42	26	16
TD-7	51	27	24	32	21	11
Mean for TD participants	50.43	25.14	25.29	32.29	17.57	14.71
SD for TD participants	4.28	2.73	2.06	6.95	6.16	2.43

Note. S-DLD = suspected DLD; TD = typically developing.

Morphosyntactic abilities

Table V shows the scores from the LITMUS-SRep task. The S-DLD child scored the lowest at 1,

Table IV. Scores highlighting nonword repetition abilities assessed via Language Impairment Testing in Multilingual Settings—Crosslinguistic Nonword Repetition (LITMUS-CL-NWR; Chiat, 2015; Hamdani, Chan, Kan, Chiat, & Polišenská, 2020).

Participant	Nonword repetition	
	Whole item	Segment
S-DLD	3	17
TD-1	5	43
TD-2	9	47
TD-3	9	42
TD-4	6	43
TD-5	7	44
TD-6	6	38
TD-7	7	37
Mean for TD participants	7.00	42.00
SD for TD participants	1.53	3.46

Note. S-DLD = suspected DLD; TD = typically developing.

Table V. Scores featuring morphosyntactic abilities assessed using Language Impairment Testing in Multilingual Settings—Sentence Repetition Task (LITMUS-SRep; Marinis & Armon-Lotem, 2015; Hamdani, Chan, Kan, & Armon-Lotem, 2020).

Participant	Score
S-DLD	1
TD-1	22
TD-2	14
TD-3	21
TD-4	22
TD-5	32
TD-6	23
TD-7	16
Mean for TD participants	21.43
SD for TD participants	5.77

Note. S-DLD = suspected DLD; TD = typically developing.

compared to the overall score of 21.43 ($SD = 5.77$; range = 14–32) out of 35 obtained by her age- and grade-matched peers.

Discussion

This tutorial demonstrated the feasibility of identifying DLD in multilingual children using the CATALISE diagnostic criteria as the conceptual framework, with the support of the LITMUS tools to collect objective and informative evidence from multiple language areas to guide the diagnostic decisions.

The binary decision tree proposed by Bishop et al. (2017) was used to guide diagnostic decision-making to aid the identification of DLD in the S-DLD child. The S-DLD child's mother and her school reported that she faced difficulty in using language in a way that affects everyday functioning and educational progress. Since the input she received in the majority language Cantonese is likely reduced compared to other predominantly monolingual L1 Cantonese-speaking children in Hong Kong, following the binary decision tree, the next step to consider is whether the S-DLD child is competent in another language. In this case, Urdu, reported as the S-DLD child's best language, was assessed. Urdu versions of the LITMUS-MAIN (Hamdani, Kan, Chan, & Gagarina, 2020), the LITMUS-CLT (Hamdani, Kan, & Chan, 2020), the LITMUS-CL-NWR (Hamdani, Chan, Kan, Chiat, & Polišenská, 2020), and the LITMUS-SRep (Hamdani, Chan, Kan, & Armon-Lotem, 2020) were used to collect objective and informative evidence from multiple language areas.

Results showed that the S-DLD child obtained particularly low scores in most of the tests and subtests, including comprehension, story structure (telling mode), and story complexity (telling mode) in the LITMUS-MAIN, all the LITMUS-CLT-subtasks, the LITMUS-SRep, and the LITMUS-CL-NWR according to both scoring schemes compared to her age- and grade-matched peers. Moreover, as Table 2 (Supplementary Materials) indicated, the amount of Urdu input to the S-DLD child was not less than that to her other TD peers, but the S-DLD child scored noticeably lower than the other TD peers across multiple language areas in the assessments, and some TD peers scored better even with less reported amounts of Urdu exposure. These data collectively suggest that the S-DLD child is in fact not competent for her age in Urdu. As Urdu is this participant's L1 and best language, the findings suggest language difficulties affecting multiple domains of language. Previous studies using LITMUS tasks have also found that compared to TD children, children with DLD show weaker narrative production and comprehension (e.g. Boerma et al., 2016; Hržica & Kraljević, 2020; Tsimpli et al., 2016), lower nonword and sentence repetition accuracy (e.g. Boerma et al., 2015; dos Santos & Ferré, 2018; Fleckstein et al., 2016; Gagarina, Gey, et al., 2019), and lower picture naming and selection accuracy (e.g. Kapalková & Slančová, 2017; Saliby et al., 2017).

In addition, the parental questionnaire also suggested that the difficulties had been observed for some time and persisted even when the child reached school age, which together with the evidence of lack of competence even in receptive language, suggests poor prognosis according to Bishop et al. (2017). Since there were no associated biomedical conditions reported, the S-DLD child can be identified with DLD.

There are two further observations from our data that could be clinically useful. First, for recall in the narrative assessment, each child was first tested in story retelling, which provides a model story for reference, and then had a break doing some other language tasks and was assessed in telling the same stories without a model story. The S-DLD child had a noticeable drop in story structure and story complexity scores from retelling to telling of narratives. This contrasts with her TD peers, who showed more consistent performance across both story retell and tell tasks. This in turn led to a marked difference in performance on story structure and story complexity between the S-DLD child and children with TDin the story telling task, but not in the model-supported retelling task. This pattern of findings was also reported in Sheng et al. (2020). Their at-risk-for-

DLD group of children exhibited comparable performance on story structure and syntactic complexity as TD peers in story retelling when supported with a prior adult model. However, the story structure and syntactic complexity scores of the at-risk group decreased significantly once the adult model was removed in the story telling task, unlike their TD peers who could maintain a high level of performance across both tasks. Sheng et al. (2020) related this phenomenon to a main assumption of dynamic assessment, in which TD children having intact language learning abilities are expected to show evidence of a stronger learning potential, in response to some support such as training or modelling, than children with DLD having a weaker language learning capacity. In the current context, the benefit of uptaking the sophisticated language and richer story structure modelled in the first retell task resulted in more sustainable performance across retell and tell tasks in the TD children, but this benefit was fragile and diminished quickly for the child with, or at risk for, DLD once the model was removed in the telling task, requiring the child to generate a story from pictures. Although the current paradigm is not equivalent to a dynamic assessment, the differences in performance from the retell task to the tell task between children with TD and (at risk) DLD suggests that it could be clinically informative to evaluate a child's sustainability of modifiability (improvement) upon modelling in assessing multilinguals. To the extent that some children with DLD may show benefits from modelling, we hypothesised that these benefits would be more fragile/transient from story retelling to telling, relative to the TD peers.

The second observation concerns the different schemes for scoring story complexity in macrostructure. Different scoring approaches may vary in their ability to capture the DLD/TD group differences. In general, a scoring scheme with finer levels of granularity (i.e. more levels of score weightings) to differentiate between the different levels of structural complexity should be better able to capture the possible DLD/TD differences if the two groups differ in their story complexity. Recall the three scoring schemes used in the case study: the scoring scheme by Gagarina, Gey, et al. (2019) has the finest level of differentiation (six levels), the scoring scheme by Sheng et al. (2020) has the coarsest level of differentiation (two levels), and the scoring scheme by Maviş et al. (2016) is in the middle (four levels). The current findings showed that the S-DLD child scored noticeably lower than other TD participants in the telling mode, especially according to scoring schemes by Maviş et al. and Gagarina et al. While it is hard to make a strong recommendation based on the case study featuring only one child with DLD and seven TD children, we have also taken note from another ongoing study of our team examining a group of Cantonese-speaking children with DLD ($N = 25$) and TD ($N = 25$) using the Cantonese MAIN and compared these three scoring schemes in story complexity.

The result also suggested that the scoring schemes of Maviş et al. (2016) and Gagarina, Gey, et al. (2019) were better able than the scoring scheme of Sheng et al. (2020) to differentiate between the two groups of children. This result aligns with the idea that since the scoring schemes of Maviş et al. (2016) and Gagarina, Gey, et al. (2019) offer relatively finer levels of granularity (i.e. more levels of score weightings) to differentiate between the different levels of structural complexity, they are better able to differentiate the two groups of children who differ in their story complexity.

Based on these findings, we could offer some preliminary advice that the scoring schemes of Maviş et al. (2016) and Gagarina, Gey, et al. (2019) appear to be better than the scoring scheme of Sheng et al. (2020) in differentiating DLD/TD. However, it is hard to conclude from these findings whether the scoring scheme of Gagarina, Gey, et al. (2019) is better than the scoring scheme of Maviş et al. (2016). Further investigation of this observation including larger sample sizes and examining different age ranges and diagnostic/classification accuracy, such as sensitivity and specificity, is warranted. Moreover, we would also need to consider practicality if we were to make recommendations for clinicians: While a scoring scheme with the finest level of granularity is likely most differentiating, it could also be slower/harder to score. Having more informative data in the future comparing the scoring schemes of Maviş et al. (2016) and Gagarina, Gey, et al. (2019) would allow us to consider both diagnostic accuracy and practicality, to make more concrete recommendations on which scoring system is more ideal or good enough for use by clinicians in speech and language therapy clinics.

Before closing, we would like to point out some novelties and limitations of this tutorial and suggestions for further research. As far as we know, we are likely the first to have adapted the LITMUS tools into Urdu and there is so far no published research documenting the diagnostic potential/accuracy of these Urdu LITMUS tools. This tutorial is likely the first to address the clinical utility of the Urdu LITMUS tools in identifying DLD in Urdu-speaking children in this major, but understudied, world language. This tutorial also presents the first published data on the Urdu materials as far as the LITMUS-Crosslinguistic Lexical Task (LITMUS-CLT), the Crosslinguistic Nonword Repetition (CL-NWR), and the LITMUS-Sentence Repetition (LITMUS-SRep) data are concerned. As for Urdu MAIN, we published some TD data featuring another group of older elementary school children on their production of Urdu narratives using the Urdu MAIN, examining the relationship between narrative macrostructure and microstructure (Chan et al., 2023). Therefore, if

we are referring to the Urdu MAIN data from (suspected) DLD children, and new data addressing the clinical utility of the Urdu LITMUS tools in identifying DLD in Urdu-speaking children, this tutorial also presents the first published data using the Urdu MAIN in this regard. Despite these novelties, the diagnostic accuracy of these new assessment tools in identifying Urdu-speaking multilingual children with and without LD or DLD will still need to be established as we step up to future research with a larger sample size, while the diagnostic accuracy of in person versus telepractice using these LITMUS tools will also need to be examined in future research. In terms of diagnostic accuracy measures for the Urdu-speaking multilingual children suspected of LD or DLD, further studies examining and comparing the diagnostic accuracy of each of our newly-adapted Urdu LITMUS tools are needed. We also acknowledge the following general principles from our current knowledge base in improving the diagnosis of LD or DLD in multilingual children: (a) using a combination of measures for collective evidence to improve the diagnosis of LD or DLD in multilingual children; (b) using measures that would not disadvantage multilingual children with reduced experience to the target language (e.g. measures that are less affected by language-specific vocabulary and morphosyntactic knowledge, such as nonword repetition and dynamic assessment); and (c) using parental questionnaires (in combination with other measures) tapping information on developmental history, and language background and experience. Future studies, such as Li'el et al. (2019), could also be conducted to address this issue.

One merit of this tutorial is that it compares the performance of the S-DLD child with her age- and grade-matched TD peers from similar language backgrounds. Using TD peers with similar language backgrounds as a reference group is necessary to guide the developmental expectations. This is consistent with the principle that multilingual children should be benchmarked against multilingual norms or reference data but not data generated from monolingual norms or samples (Armon-Lotem, 2018), because comparing multilinguals to a monolingual sample would likely disadvantage multilingual children with reduced target language experience, leading to the risk of over-identification of DLD.

Moreover, recall one major consideration in the CATALISE criteria in determining whether a multilingual child has LD or DLD or not is whether "there is evidence that the child does not have age-appropriate skills in any language" (Bishop et al., 2017, p. 1071; Statement 4). We can therefore infer that if a multilingual child has a clear language dominance profile, then evidence showing that the child does not have expected competency even in her best language would constitute adequate evidence that the child would very likely not have expected competency in any language, because she is likely even less proficient in her other weaker/non-dominant languages. This tutorial features a multilingual child who happened to have a clear language dominance profile, with her first language Urdu being the strongest language and Cantonese and other languages being the second/weaker languages. In this regard, therefore, evidence from Urdu would be informative and adequate in addressing this major consideration in the CATALISE criteria, and evidence from Cantonese was not necessary/critical in guiding the diagnostic decisions for this child in this context. Having said that, we are aware that multilingual children are heterogeneous, and therefore for children with a different language dominance profile (for example, children with a more balanced dominance profile, or children whose second language Cantonese is their stronger/best language instead), evidence from Cantonese or languages other than Urdu would become informative in addressing whether "there is evidence that the child does not have age-appropriate skills in any language" (Bishop et al., 2017, p. 1071; Statement 4). Further work could demonstrate how this approach can be applied to multilingual children with different language dominance profiles. When one extends the investigation to multilingual children with different language dominance profiles, there are two more reminders to pay attention to. First, some skills may look "inadequate" when tested only in one language of a multilingual child, even in a child without DLD, e.g. in the case of distributed vocabulary. For these skills, it could be informative to ascertain whether there is a lack of demonstrated and expected competency even when all languages are considered for a child suspected of having DLD. Second, while the L1 of the multilingual children in this tutorial happened to be also the best language of these children, L1 does not always coincide with the best language for all multilingual children and, therefore, one needs to be cautious in interpreting the demonstrated competence of a multilingual child's L1 in the case of language attrition or subtractive multilingualism, if dominance is undergoing a shift.

In closing, this tutorial adds further evidence to the literature that it is possible to identify DLD in multilingual children using telepractice. The LITMUS tools are designed to be appropriate for identifying DLD in multilingual children from different cultures and can be administered online with video conferencing software. These tools can also be used to establish reference data in TD children, so as to guide developmental expectations for specific populations where there are not yet appropriate tools and multilingual norms for assessing multilingual children. Our Urdu adaptations of the LITMUS tools are available for free use by the international clinical and research community, together with instructions for users. They can be used to support telepractice in Urdu in different countries and facilitate language

testing for Urdu-speaking children. Importantly, the result demonstrates the promise of using the CATALISE diagnostic criteria with support from the LITMUS battery tasks adapted to telepractice for identifying DLD and collecting reference data in multilingual children, not only in the multilingual context presented here but also in any target language(s).

Note

1. Initiating event has been commonly considered as a story grammar element in the story grammar framework of analysing narrative macrostructural competence in the literature (e.g. Stein & Glenn, 1979). *Initiating event* refers to an event that triggers/initiates the intentionality of a story character (in the form of a "goal," another story grammar element), which motivates the character to carry out a goal-directed action (manifested as an "attempt," another story grammar element). An initiating event can be expressed with or without the use of an internal state term (IST) but, according to the MAIN scoring criteria, it adopts a more stringent scoring criterion for the story grammar element initiating event. An utterance expressing an initiating event with the use of an IST (e.g., Baby goat was **scared**, or the mother goat **saw** that the baby goat was **scared**, or the mother goat was **worried** about the baby goat in the water) would only be scored one point, and it therefore uses the phrase/term "IST as initiating event" in its scoring manual.

Disclosure statement

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

ORCID

Saboor Hamdani http://orcid.org/0000-0003-4387-5619
Angel Chan http://orcid.org/0000-0002-9547-0210
Rachel Kan http://orcid.org/0000-0003-3893-4549
Shula Chiat http://orcid.org/0000-0002-8981-8153
Natalia Gagarina http://orcid.org/0000-0002-5136-1071
Ewa Haman http://orcid.org/0000-0003-1615-711X
Magdalena Łuniewska http://orcid.org/0000-0001-5504-9766
Kamila Polišenská http://orcid.org/0000-0001-7405-6689
Sharon Armon-Lotem http://orcid.org/0000-0003-3360-9472

References

Altman, C., Armon-Lotem, S., Fichman, S., & Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English-Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics*, *37*(1), 165–193. https://doi.org/10.1017/S0142716415000466

Armon-Lotem, S. (2018). SLI in bilingual development: How do we approach assessment. In A. Bar-On, D. Ravid, & E. Dattner (Eds.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistic perspectives* (pp. 617–641). De Gruyter Mouton.

Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, *51*(6), 715–731. https://doi.org/10.1111/1460-6984.12242

Bari, S., & Ajmal, S. (2016). The first flight: Acquisition of Urdu consonants in native children. *International Journal of English and Education*, *5*(6), 56–71.

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 1–29. https://doi.org/10.2167/beb392.0

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Adams, C. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders*, *66*, 65–76. https://doi.org/10.1016/j.jcomdis.2017.04.001

Boerma, T., Chiat, S., Leseman, P., TImmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research: JSLHR*, *58*(6), 1747–1760. https://doi.org/10.1044/2015_JSLHR-L-15-0058

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *International Journal of Language & Communication Disorders*, *51*(6), 626–638. https://doi.org/10.1111/1460-6984.12234

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, *46*(Pt 2-3), 217–243. https://1 0.1177/00238309030460020801 https://doi.org/10.1177/00238309030460020801

Census and Statistics Department, Government of the Hong Kong Special Administrative Region. (2017). *Hong Kong 2016 Population By-census - Thematic report: Ethnic minorities*. Census and Statistics Department, Government of the Hong Kong Special Administrative Region.

Chan, A., Chen, S., Hamdani, S. Z., Tse, B., & Cheng, K. (2023). Story telling in bilingual Urdu-Cantonese ethnic minority children in Hong Kong: Macrostructure and its relation to microstructural linguistic skills. *Frontiers in Psychology*, *14*, 924056. https://doi.org/10.3389/fpsyg.2023.924056

Chiat, S. (2015). Nonword repetition. In S. Armon-Lotem, J. L. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 123–148). Multilingual Matters.

Chondrogianni, V., Butcher, M., & Garaffa, M. (2021). Supporting children with Developmental Language Disorder in Gaelic-medium primary education. In W. McLeod, A. Gunderloch, & R. Dunbar (Eds.), *Rannsachadh na Gàidhlig 2018 proceedings* (pp. 205–2018). Aberdeen University Press.

Coleman, J. J., Frymark, T., Franceschini, N. M., & Theodoros, D. G. (2015). Assessment and treatment of cognition and communication skills in adults with acquired brain injury via telepractice: A systematic review. *American Journal of Speech-Language Pathology*, *24*(2), 295–315. https://doi.org/10.1044/2015_AJSLP-14-0028

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Multilingual Matters.

de Almeida, L., Ferré, S., Morin, E., Prévost, P., Santos, C., Tuller, L., Zebib, R., & Barthez, M.-A. (2017). Identification of bilingual children with specific language impairment in France. *Linguistic Approaches to Bilingualism*, *7*(3–4), 331–358. https://doi.org/10.1075/lab.15019.alm

dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition*, *25*(1), 58–71. https://doi.org/10.1080/10489223.2016.1243692

Eikerling, M., Andreoletti, M., Secco, M., Luculli, B., Cha, G., Castro, S., Gazzola, S., Sarti, D., Garzotto, F., Guasti, M. T., & Lorusso, M. L. (2023). Remote screening for Developmental Language Disorder in bilingual children: Preliminary validation

in Spanish–Italian speaking preschool children. *Applied Sciences*, *13*(3), 1442. https://doi.org/10.3390/app13031442

Farndale, A., Harris, P., & de Courcy, M. (2016). Social, emotional and linguistic experiences of developing bilingual preschoolers as they learn English as an Additional Language (EAL). *The International Journal of Learning: Annual Review*, *23*(1), 41–58. https://doi.org/10.18848/1447-9494/CGP/41-58

Fichman, S., Altman, C., Voloskovich, A., Armon-Lotem, S., & Walters, J. (2017). Story grammar elements and causal relations in the narratives of Russian-Hebrew bilingual children with SLI and typical language development. *Journal of Communication Disorders*, *69*, 72–93. https://doi.org/10.1016/j.jcomdis.2017.08.001

Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E., & Zebib, R. (2016). How to identify SLI in bilingual children: A study on sentence repetition in French. *Language Acquisition*, *25*(1), 85–101. https://doi.org/10.1080/10489223.2016.1192635

Gagarina, N., Gey, S., & Sürmeli, N. (2019). Identifying early preschool bilinguals at risk of DLD: a composite profile of narrative and sentence repetition skills. *ZAS Papers in Linguistics*, *62*, 168–189. https://doi.org/10.21248/zaspil.62.2019.448

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual assessment instrument for narratives – revised. *ZAS Papers in Linguistics*, *63*(, 20. https://doi.org/10.21248/zaspil.63.2019.516

Grimm, A. (2022). The use of the LITMUS quasi-universal nonword repetition task to identify DLD in monolingual and early second language learners aged 8 to 10. *Languages*, *7*(3), 218. https://doi.org/10.3390/languages7030218

Gu, M. M., & Patkin, J. (2013). Heritage and identity: Ethnic minority students from South Asia in Hong Kong. *Linguistics and Education*, *24*(2), 131–141. https://doi.org/10.1016/j.linged.2012.12.008

Haman, E., Łuniewska, M., & de Pomiechowska, B. (2015). Designing cross-linguistic lexical tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. L. Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 196–240). Multilingual Matters.

Hamdani, S., Chan, A., Kan, R., & Armon-Lotem, S. (2020). *Sentence repetition task: Urdu version* [Unpublished material]. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University. Hong Kong.

Hamdani, S., Chan, A., Kan, R., Chiat, S., & Polišenská, K. (2020). *Urdu crosslinguistic Nonword Repetition test* [Unpublished material]. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University. Hong Kong.

Hamdani, S., Kan, R., & Chan, A. (2020). *Crosslinguistic Lexical Tasks: Urdu version (CLT-UR). A part of LITMUS COST IS0804 Battery* [Unpublished material]. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University. Hong Kong.

Hamdani, S., Kan, R., Chan, A., & Gagarina, N. (2020). The Multilingual Assessment Instrument for Narratives (MAIN): Adding Urdu to MAIN. *ZAS Papers in Linguistics*, *64*, 257–261. https://doi.org/10.21248/zaspil.64.2020.580

Hreich, E. K., & Messarra, C. (2013). *Assessment of L2 children in the Lebanese multilingual context: Using questionnaire for parents of bilingual children (PABIQ) scores as diagnostic indicators for language impairment* [Unpublished Manuscript]. Saint Joseph's University.

Hržica, G., & Kraljević, J. K. (2020). The Croatian adaptation of the Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, *64*(August), 37–44. https://doi.org/10.21248/zaspil.64.2020.555

Hu, J., Torr, J., & Whiteman, P. (2014). Australian Chinese parents' language attitudes and practices relating to their children's bilingual development prior to school. *Journal of Early Childhood Research*, *12*(2), 139–153. https://doi.org/10.1177/1476718X13515429

International Expert Panel on Multilingual Children's Speech (IEPMCS). (2012). *Multilingual children with speech sound disorders: Position paper*. Research Institute for Professional Practice, Learning and Education, Charles Sturt University.

Kan, R. T. Y., Chan, A., & Gagarina, N. (2020). Investigating children's narrative abilities in a Chinese and Multilingual context: Cantonese, Mandarin, Kam and Urdu adaptations of the Multilingual Assessment Instrument for Narratives (MAIN). *Frontiers in Psychology*, *11*, 573780. https://doi.org/10.3389/fpsyg.2020.573780

Kapalková, S., & Slančová, D. (2017). The vocabulary profile of Slovak children with primary language impairment compared to typically developing Slovak children measured by LITMUS-CLT. *Clinical Linguistics & Phonetics*, *31*(11–12), 893–909. https://doi.org/10.1080/02699206.2017.1308016

Li'el, N., Williams, C., & Kane, R. (2019). Identifying developmental language disorder in bilingual children from diverse linguistic backgrounds. *International Journal of Speech-Language Pathology*, *21*(6), 613–622. https://doi.org/10.1080/17549507.2018.1513073

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, J. L. de Jong, & N. Meir (Eds.), *Assessing multilingual children: disentangling bilingualism from language impairment* (pp. 116–143). Multilingual Matters.

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, *37*(1), 69–89. https://doi.org/10.1017/S0142716415000429

Nelson, N. W., & Plante, E. (2022). Evaluating the Equivalence of Telepractice and Traditional Administration of the Test of Integrated Language and Literacy Skills. *Language, Speech, and Hearing Services in Schools*, *53*(2), 376–390. https://doi.org/10.1044/2022_LSHSS-21-00056

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, and Allied Disciplines*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Paradis, J. (2010). Bilingual children's acquisition of English verb morphology: Effects of language exposure, structure complexity, and task type. *Language Learning*, *60*(3), 651–680. https://doi.org/10.1111/j.1467-9922.2010.00567.x

Peristeri, E., Andreou, M., Tsimpli, I. M., & Durrleman, S. (2020). Bilingualism effects in the narrative comprehension of children with Developmental Language Disorder and L2-Greek: Links with language, executive function and theory of mind. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual assessment instrument for narratives* (pp. 297–330). John Benjamins. https://benjamins.com/catalog/sibil.61

Polišenská, K., & Kapalková, S. (2014). Improving child compliance on a computer-administered nonword repetition task. *Journal of Speech, Language, and Hearing Research: JSLHR*, *57*(3), 1060–1068. https://doi.org/10.1044/1092-4388(2013/13-0014)

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales* (Vol. 759). Oxford Psychologists Press.

Saliby, K. A. C., Dos Santos, C., Kouba Hreich, E., & Messarra, C. (2017). Assessing Lebanese bilingual children: The use of cross-linguistic lexical tasks in Lebanese Arabic. *Clinical Linguistics & Phonetics*, *31*(11–12), 874–892. https://10.1080/02699206.2017.1308016 https://doi.org/10.1080/02699206.2017.1308554

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R., & Skoruppa, K. (2021). Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and

meta-analysis. *Journal of Speech, Language, and Hearing Research: JSLHR*, *64*(9), 3578–3593. https://doi.org/10.1044/2021_JSLHR-20-00552

Serrano-Hidalgo, E. (2018). A case study of the experiences and perspectives of Hispanic immigrant parents on heritage language maintenance and bilingual education in the rural community of Brandon, Manitoba. *BU Journal of Graduate Studies in Education*, *10*(2), 4–7.

Sheng, L., Shi, H., Wang, D., Hao, Y., & Zheng, L. (2020). Narrative production in Mandarin-speaking children: Effects of language ability and elicitation method. *Journal of Speech, Language, and Hearing Research: JSLHR*, *63*(3), 774–792. https://doi.org/10.1044/2019_JSLHR-19-00087

Stein, N., & Glenn, C. (1979). An analysis of story comprehension in elementary children. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Ablex Publishing.

Sutherland, R., Hodge, A., Trembath, D., Drevensek, S., & Roberts, J. (2016). Overcoming barriers to using telehealth for standardized language assessments. *Perspectives of the ASHA Special Interest Groups*, *1*(18), 41–50. https://doi.org/10.1044/persp1.SIG18.41

Sutherland, R., Trembath, D., Hodge, A., Drevensek, S., Lee, S., Silove, N., & Roberts, J. (2017). Telehealth language assessments using consumer grade equipment in rural and urban settings: Feasible, reliable and well tolerated. *Journal of Telemedicine and Telecare*, *23*(1), 106–115. https://doi.org/10.1177/1357633X15623921

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: JSLHR*, *40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245

Tsimpli, I., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, *37*(1), 195–216. https://doi.org/10.1017/S0142716415000478

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. L. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 301–330). Multilingual Matters.

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., Dos Santos, C., Abed Ibrahim, L., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders*, *53*(4), 888–904. https://doi.org/10.1111/1460-6984.12397

Westby, C. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (Eds.), *Language and reading disabilities* (2nd ed., pp. 157–232). Allyn & Bacon.

Wright, A. J. (2018). Equivalence of remote, online administration and traditional, face-to-face administration of the Woodcock-Johnson IV cognitive and achievement tests. *Archives of Assessment Psychology*, *8*(1), 23–35.

Zahra, N. (2016). *Acquisition of Urdu Consonants in Typically Developing Children* [Unpublished Masters Thesis]. Centre for Clinical Psychology, University of the Punjab.