Identification of Stuttering in Bilingual Lebanese Children Across Two Presentation Modes

Selma Saad Merouwe[1, 2], Raymond Bertram[2], Sami Richa[5], and Kurt Eggers[2, 3, 4]

[1] Department of Speech-Language Pathology, Saint-Joseph University, Lebanon

[2] Department of Psychology and Speech-Language Pathology, Turku University, Finland

[3] Department of Rehabilitation Sciences, Ghent University, Belgium

[4] Department of Speech-Language Pathology, Thomas More University College, Belgium

[5] Faculty of Medicine, Saint-Joseph University, Lebanon

Author note

Selma Saad Merouwe https://orcid.org/0000-0003-2062-1827

Raymond Bertram https://orcid.org/0000-0002-2709-3644

Sami Richa https://orcid.org/0000-0002-3895-1325

Kurt Eggers https://orcid.org/0000-0003-4221-2063

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Selma Saad Merouwe, Dept. of Speech-Language Pathology, Saint-Joseph University, Damascus Street, B.P. 17-5208 – Mar Mikhaël, Beirut 1104 2020, Lebanon. Email: selma.saad.merouwe@gmail.com

Abstract

The goals of this study were to investigate whether Lebanese speech-language pathologists (SLPs) are accurate at identifying stuttering in bilingual children, to examine whether the use of video-recordings instead of audio-recordings allows for better analyses, and to explore factors that may affect the SLPs' judgments. In phase 1, 32 SLPs listened to narrative samples in Lebanese Arabic of 6 children who do not stutter (CWNS) and 2 who stutter (CWS). They were instructed to label each child as stuttering or not, and to explain what motivated their decisions. Afterwards, they were asked to provide background information by means of a questionnaire. In phase 2, they were asked to judge the same speech samples on the basis of video-recordings, and to explain for each child which speech characteristics they relied on to make their decisions. The results showed that misidentification happens frequently, is significantly more likely to happen with CWNS than with CWS, but also varies within these categories. Moreover, the use of video-recordings does not provide more reliable analyses of speech disfluencies, and speech samples' characteristics and bilingual profile rather than SLPs' characteristics seem to influence the judgments. Qualitative analyses indicate that, in the current study, misidentification may be driven by neglecting or misinterpreting physical concomitants. In general, the findings indicate that identifying and analyzing speech fluency behaviors remain a challenging perceptual task, which underlies the need for developing consistent methods for training students and clinicians in identifying stuttering, especially in a bilingual context.

Keys-words: stuttering-like disfluencies, other disfluencies, bilingualism, typical development, stuttering, identification

Identification of Stuttering in Bilingual Lebanese Children Across two Presentation Modes

1. Introduction

The manifestation of stuttering in bilingual children is an area of relevance to both clinicians and researchers. Despite the fact that the number of bilinguals has considerably increased in the last decades (e.g., Beyea, 2020; Grosjean, 2021; Zeigler & Camarato, 2019), monolinguals are still considered to be the norm in research and clinical practice (Shin, 2017). Therefore, the available data on stuttering in bilinguals are still insufficient. More specifically, there is a lack of empirical data about the development of normal fluency patterns and the manifestation of stuttering in bilingual children (Byrd et al., 2015b).

Different definitions of bilingualism have been proposed throughout the years, but until now, researchers have not settled on a single one. What is agreed upon though is that it is unusual to impeccably master two or more languages (Wei, 2007). In this paper, we adhere to the definition of Chin and Wigglesworth (2007), who suggest defining bilingualism as the competence to communicate in at least two languages. This competence fluctuates in both languages, which leads to various profiles of bilinguals. For the majority of bilinguals, there is an imbalance between the two languages (Kohnert, 2010), such that there is a dominant language (L1), which is the one that is used most often and therefore the one that is best acquired, and a non-dominant language (L2), which is less frequently used and mastered to a lower degree (Kohnert, 2010). Only a minority of bilingual speakers are considered to be balanced bilinguals (Sorge et al., 2016). Stuttering manifests itself in both languages of the bilingual (e.g., Saad Merouwe et al., 2022). A consequence of the demographic evolution (that is the increase in bilingualism) in many countries is that speech-language pathologists (SLPs) are also more frequently faced with the evaluation and diagnosis of stuttering in bilingual children (Beyea, 2020; Winter, 2016).

In general, a stuttering assessment should be multidimensional, based on case history and parental concern in addition to affective, behavioral, and cognitive dimensions (e.g., Shapiro, 2011; Yaruss et al., 2006) as well as the analysis of speech disfluencies and physical concomitants. A frequently used classification of speech disfluencies is the one suggested by Ambrose and Yairi (1999). They distinguished between 'Stuttering-Like Disfluencies' (SLD) including monosyllabic word repetitions, syllable repetitions, sound repetitions, prolongations, blocks and broken words; and 'Other Disfluencies' (OD) including multisyllable word repetitions, phrase repetitions, interjections, unfinished words or sentences and revisions. Ambrose and Yairi (1999) further suggested that children exhibiting '3 or more SLD per 100 syllables should be suspected as exhibiting stuttering' (p. 906). This 3% syllable-metric threshold was used for monolingual English-speaking children, but also appeared in some studies with other languages like Spanish and German (Carlo & Watson, 2003; Natke et al., 2006). Other researchers (e.g., Conture, 2001) have argued to use a 3% SLD word-based metric threshold as an indicator for stuttering, which was applied in a number of studies (Boey et al., 2007; Leclercq et al., 2017; Pellowski & Conture, 2002; Tumanova et al., 2014). The current study follows the latter conjunction in using the word metric counting disfluencies.

Either way, it seems judicious to ask whether a clinical threshold established on monolinguals can be used in bilinguals. Some studies have shown that bilingual CWNS often demonstrate a high rate of disfluencies, including interjections, sound, syllable and monosyllabic word repetitions (e.g., Bedore et al., 2006). These disfluencies, also called mazes, disrupt the forward flow of speech with no contribution to the meaning of the message (Bedore et al., 2006; Byrd et al., 2015a; Carias & Ingram, 2006). It has been suggested that speech fluency in bilinguals is influenced by both speaker-related language (e.g., language dominance and proficiency), and linguistic typology-related factors (e.g., language structure)

(for systematic review, see Chaudhary et al., 2021). Therefore, it stands to reason that among bilinguals, there may be a great amount of variation in disfluencies, given the different language dyads and differences in L1 and L2 language exposure, use and proficiency.

Based on previous studies showing that identifying speech-language disorders in bilingual children is difficult (e.g., De Lamo White & Jin, 2011), resulting in either over- or under-representation of bilingual children in clinicians' caseloads (e.g., Bedore & Peña, 2008; Byrd et al., 2015b), one may ask to what extent clinicians are capable of accurately identifying stuttering in bilingual children and to what extent this is influenced by the frequency and type of disfluencies.

1.1. Speech disfluencies in bilingual children who do not stutter

Models of language development and use in bilinguals underline that the linguistic knowledge of bilingual speakers is not necessarily the same as that of monolinguals (Grosjean, 1989). Bilinguals typically use each language less than their monolingual peers due to which bilingual and monolingual individual's representation of language may differ. For example, it has been shown that bilingual speakers have smaller lexicons than monolingual speakers in either language (Peña et al., 2002). Picture naming studies have shown that bilinguals report more 'tip of tongue' phenomena (Gollan et al., 2005), implying some differences in the strength of connections between semantic and phonological information. In the morphosyntactic domain, differences in language representation may lead to grammatical errors that do not appear in monolinguals (Jacobsen & Schwartz, 2005). Carias and Ingram (2006) suggested that the use of many languages and/or limited language knowledge might overload the children's language processing system, leading to disfluent speech. Bilingual CWNS tend to have more disfluencies in their speech compared to their monolingual peers (Bedore et al., 2006). Usually, these disfluencies can be expected when complex ideas are expressed, for instance including spatial, temporal or causal relationships

(Loban, 1976). Fiestas et al. (2005) report a difference especially in repetitions with their bilingual children using twice the number of repetitions as functionally monolingual children. They suggest that this allows bilinguals a second chance using audition to monitor their own speech output (Fiestas et al., 2005).

Byrd et al. (2015a) described the frequency and types of speech disfluencies produced by bilingual Spanish-English (SE) CWNS. The participants were 18 Hispanic CWNS (5.06 - 6.07 years) of whom 6 were Spanish (S) dominant, 6 English (E) dominant, and 6 balanced bilinguals. Based on narrative speech samples in both languages, the majority of children (14/18) produced a high frequency of disfluencies exceeding the 3% SLD threshold in all their speech samples (ranging from 3% to 22%). The most frequent SLD detected were repetitions of monosyllabic words and sounds (with rhythmic iterations, tension-free), while the most frequent OD were revisions and interjections. This study was recently replicated by Eggers et al. (2019). They included 59 bilingual Yiddish-Dutch (YD) CWNS, divided into two age groups 6.01-7.07 and 9.00-10.04 years. All children were Yiddish dominant, and a high number of participants exceeded 3% SLD in both languages, but more in L2 than L1 (46% in Yiddish and 78% in Dutch). Among the SLD, a predominance of monosyllabic word and sound repetitions were observed. Similar results were found in a study conducted by Saad Merouwe et al. (2019) including 30 bilingual Lebanese-French (LF) CWNS, aged between 5.00 and 7.00 years. The children produced significantly more SLD and OD in French, their L2, and most of them exceeded 3% SLD in both languages (56.6% in Lebanese Arabic and 65% in French), with a high predominance of monosyllabic word repetitions. Given the high frequency of SLD found in SE, YD, and LB CWNS, one may conclude that typical speech fluency of bilinguals differs from monolinguals; therefore, if clinicians were to apply monolingual guidelines to identify stuttering in bilinguals based on speech samples alone, they would often misidentify CWNS as CWS.

1.2. Manifestation of stuttering speech behaviors in bilinguals

Only a few studies have investigated the manifestation of stuttering in bilinguals. The findings across these studies diverge and may depend on several variables like how bilingualism was defined, the language dyad, the age of acquisition and the level of proficiency of each language, how stuttering was diagnosed, the number of participants and their age. Bilingual CWS were found to produce a sizable number of SLD and OD in both spoken languages (e.g., Ambrose & Yairi, 1999; Carias & Ingram, 2006), but some studies found SLD to be more prevalent in the L1 (Shenker et al., 1998; Carias & Ingram, 2006; Lee et al., 2014) while other reported more SLD in the L2 (Ardila et al., 2011; Szmalec, 2013). A clear explanation for more SLD in the L1 compared to the L2 is not readily available (Chaudhary et al., 2021). However, when more SLD were reported in the L2, this has been attributed to factors such as language dominance and L2 proficiency (Schäfer & Robb, 2012). More precisely, a lower L2 proficiency level may come with suboptimal linguistic formulations, which may lead to an increase in disfluencies (Jankelowitz & Bortz, 1996; Maruthy et al., 2015). The differences in frequency and type of disfluencies between bilingual CWS's languages have also been attributed to crosslinguistic differences (e.g., Ardila et al., 2011; Carias et Ingram, 2006; Maruthy et al., 2015; Schäfer & Robb, 2012). These differences pertain to differences in linguistic structure (e.g., Ardila et al., 2011; Maruthy et al., 2015), e.g., phonetic structure (Maruthy et al., 2015; Schäfer % Robb, 2012), leading to difficulties in formulation in one of the two languages (Jankelowitz & Bortz, 1996). For example, Spanish may yield more disfluencies compared to English, as it is a syllable-timed language with complex syntactic structures (Tatliancich-Klinger et al., 2013). Nevertheless, differential stuttering in L1 and L2 as a consequence of differences in linguistic structure has only been investigated superficially and therefore this issue remains unsettled.

Further, some researchers have implied that early exposure to a second language would make a child more vulnerable to the development and persistence of stuttering (e.g., Howell et al., 2009). However, this has been questioned by recent studies showing that bilingual CWNS tend to produce SLD at frequencies comparable to monolingual CWS (Byrd, et al., 2015a; Eggers, et al., 2019, 2020; Saad Merouwe et al., 2019). This could - in addition to the language-related factors possibly affecting the manifestation of disfluencies in bilinguals - lead to a false-positive identification of stuttering among bilinguals, that is, if clinicians were to make diagnostic decisions based on the frequency of speech disfluencies alone.

1.3. Identification of stuttering speech behaviors in bilinguals

The basis for identifying stuttering speech behaviors by clinicians is perceptual judgment (Einarsdóttir & Ingham, 2009). However, reliable identification of stuttering, even in monolinguals, has been shown recurrently to be a challenging perceptual judgment task, because it is vulnerable to idiosyncratic observer characteristics. A few published studies have shown a considerable disagreement in judgment of stuttering despite the evaluators' familiarity with the speakers' language (Cordes & Ingham, 1994). This happens among graduate students who have taken classes in stuttering and fluency disorders (Ingham & Cordes, 1992), but also among experienced clinicians (Brundage et al., 2006). Watson and Kayser (1994) found that SLPs were able to identify atypical speech behaviors in severe stuttering cases with certainty, regardless of language familiarity. Yet, that was not the case when stuttering was mild and an in-depth analysis was needed.

This implies that also in case of bilingual children, severe stuttering may be relatively easy to identify, but the difficulty to distinguish mild stuttering from typical disfluent speech in bilinguals may be exuberated in comparison to monolinguals.

In addition to this, some studies showed that SLPs are more accurate in diagnosis of stuttering in a language familiar to them than in a less known or unknown language (Humphrey, 2004; Van Borsel & Britto Pereira, 2005; Van Borsel et al., 2008). It has also been found that assessments in bilinguals' L1 may be more accurate than assessments in bilinguals' L2 (Einarsdóttir & Ingham, 2009). Consequently, in the current study, we chose to let SLPs judge the L1 speech output of bilingual children, which was at the same time the language SLPs were more familiar with.

1.4. Using audio recordings and video recordings for assessing stuttering in bilinguals

Clinicians often report difficulties identifying whether the disfluencies noticed in their clients' speech are indicative of stuttering or not, especially with bilingual children who tend to experience a higher level of linguistic uncertainty (Byrd, 2018). This might be influenced by the fact that sound, syllable and whole-word repetitions are typical speech characteristics of bilingual children (Bedore et al., 2006; Byrd et al., 2015; Eggers et al., 2019, 2020; Fiestas et al., 2005), but are considered indicators of stuttering when frequently occurring in monolingual children (Ambrose & Yairi, 1999). Some authors have suggested that if a child's disfluencies are not accompanied by noticeable tension such as eye blinks or body movements, then it is likely that the child is not an incipient stutterer (e.g., Watson & Kayser, 1994). That being said, although some published data postulate comparable accuracy in auditory versus visual analysis of stuttering (e.g., Bloodstein & Bernstein Ratner, 2008; Panico et al., 2005), it seems reasonable to assume that an audiovisual speech sample would offer a more representative face to face listening experience, and would enable a listener to make a more accurate judgment of the speaker's fluency compared to audio recordings alone. Given that bilinguals typically have a higher proportion of disfluencies than monolinguals, it seems even more important to assess

whether speech disfluencies are associated with observable features, such as physical tension or other physical concomitants.

Recently, Byrd et al., (2015b) conducted a study in which they asked 14 SE SLPs to listen to audio-recordings of narrative re-tells in Spanish and English that were obtained from two bilingual SE children (a CWS with a confirmed diagnosis of stuttering and a CWNS confirmed as being typically fluent) matched for age, gender and language proficiency. The majority of SLPs (71.4%) were able to correctly identify the CWS as a CWS; however, 85.7% of them falsely labeled the CWNS as a CWS. This false positive identification of stuttering was attributed to the frequent sound, syllable and monosyllabic word repetitions in the speech of the bilingual CWNS. No relationship was found between diagnostic accuracy, and other factors such as years of experience, client caseload, and educational and professional background of the SLPs, although this was not statistically tested due to the small number of cases and SLPs. The majority of SLPs stated that they were confident in their assessment abilities of monolingual and bilingual children. This highlights that they were not aware of the lack of critical knowledge that would compromise their capacity to correctly distinguish bilingual CWS from CWNS. Byrd et al.'s (2015b) study included only audio samples, which - as noted above - do not necessarily allow for reliable analyses in terms of secondary behaviors (e.g., facial tension, eye blinking, movement of the extremities). This limitation was raised by one of the participants who mentioned the need for a visual analysis of stuttering and a thorough background history.

2. Purpose of the present study

Recent studies indicated that bilingual CWNS exhibit a high frequency of SLD, which are comparable to what is indicative of stuttering in monolinguals. The study conducted by Byrd et al. (2015b) showed that if clinicians were to apply a screening test, in which diagnostic decisions are made on limited information based on the frequency and typology of speech

disfluencies alone, bilingual CWNS might be at a risk for false-positive identification of stuttering. However, their study used a relatively small number of SLPs and included one CWNS and one CWS only, which is not representative of the variability seen in bilinguals who do and do not stutter. Furthermore, since only audio samples were used, a visual analysis of the physical concomitants was not possible. In other words, one cannot make too far-fetching conclusions on the basis of the Byrd et al. study. This underlines the need for a study including a larger number of participants, and more varied speech samples in terms of language proficiency, frequency and typology of disfluencies, with an audiovisual presentation mode. The current study meets these needs.

Based on the findings that bilingual CWNS produce a considerably higher number of sound, syllable, and monosyllabic word repetitions than monolinguals, and that they seem to be at risk for false-positive identification as CWS, we postulate that Lebanese SLPs may falsely identify Lebanese bilingual CWNS as CWS. The current study thus aims to replicate and extend the Byrd et al. (2015b) study in the Lebanese context, which is ideally suited for bilingual research because of the particularly high prevalence of bilingual children. Communication in Lebanon is characterized by a dominant spoken language, Lebanese Arabic, and a great variety in the private and public uses of standard Arabic, English and French. The multilingual diversity is also apparent in the education system, where French, English and Standard Arabic are the languages of instruction (Thonhauser, 2000). Consequently, Lebanese children are confronted with many languages from an early age and parents have to opt for an English, French or even a trilingual educational system (Abou, 1962; Shaaban, 1997).

The research questions of the current study are:

(1) Is misidentification more likely to occur with bilingual CWNS compared to bilingual CWS? Our hypothesis is that misidentification occurs more often for bilingual CWNS than for bilingual CWS.

(2) Is misidentification more likely to occur when judgments are made on the basis of audio-recordings compared to video-recordings? Our hypothesis is that video-recordings allow for better analyses of speech disfluencies with secondary behaviors, decreasing the percentage of misidentification.

(3) Is the misidentification frequency approximately equal for each child within a bilingual CWNS-group and within a bilingual CWS-group? Our hypothesis is that misidentification frequency also varies within both groups.

(4) What quantitative variables influence the frequency of successful judgments of SLPs? This hypothesis could not be statistically tested in the study conducted by Byrd et al. (2015b) due to the small number of cases and SLPs. Our initial hypothesis is that both SLPs characteristics (i.e., work experience, number of bilingual CWS treated, self-ratings of confidence in diagnosing stuttering in a bilingual context, and self-ratings of difficulties in distinguishing SLD and OD) and speech samples characteristics (% SLD and % OD) affect the results.

(5) What considerations underlie the correct and incorrect judgments of CWNS and CWS? Byrd et al. (2015b) concluded that “the precise reasons why each individual participant made his or her decision is difficult to determine” (p. 81). They added that “based on the relatively ambiguous nature of the survey questions related to these factors, there is no precise way of knowing how accurate the participants were in their interpretation of the disfluencies they heard” (p. 82). This underlines the need to investigate the factors that influence diagnostic decision making more carefully. Our hypothesis is that for CWNS

correct judgments are based on low SLD frequency and the absence of physical concomitants, whereas incorrect judgments are based on a high frequency of repetitions without considering the absence of physical concomitants. For CWS, we hypothesize that correct judgments are based on high frequency of SLD and physical concomitants, and incorrect judgements on relatively low frequency of SLD and mild physical concomitants.

3. Methods

3.1 Participants

The participants were 32 licensed Lebanese SLPs, between the ages of 21 and 40. All participants were Lebanese Arabic dominant, according to the self-rating they reported. In terms of spoken language-dyads, 12.5% were bilingual Lebanese Arabic-French, and 15.6% were bilingual Lebanese Arabic-English, while 71.8% were multilingual, speaking French, English and sometimes a third language in addition to Lebanese Arabic.

The participants were recruited via three universities that have SLP departments in Lebanon (Saint-Joseph University, Lebanese University and Phoenicia University). The study was approved by Saint-Joseph University's committee of ethics. Emails including an informed consent and a cover letter describing the study were sent to the email addresses obtained from the three SLP departments. Receivers were asked to respond to the email indicating whether they agreed to participate in the study or not. Thirty-two SLPs agreed to participate in the first phase of the study, which involved the completion of a survey, and the analysis of eight audio recordings. Five SLPs withdrew from the second phase of the study, which involved the video analysis of the eight recordings of the same group of children. The major reason to not participate in the second phase was their busy schedule.

The SLP participants' background information was obtained from the participants through a Google form questionnaire. It was available in French and English, the languages used academically in the Lebanese educational system. SLP participants could choose the

language of the survey based on their preference and academical background. The SLP characteristics extracted from the questionnaire that were used in the analyses were work experience, number of bilingual CWS treated, self-ratings of confidence in diagnosing stuttering in a bilingual context, self-ratings of difficulties in distinguishing SLD and OD and general experience. General experience was a sum variable with a minimum score of 5 and a maximum of 10 based on 5 categorical questions with respect to educational level (Ba degree = 1; Ma degree = 2), years of experience (less than 3 = 1; more than 3 = 2), professional experience relevant to stuttering (less than 5 CWS treated = 1; more than 5 = 2), professional experience relevant to bilingualism (no = 1; yes = 2), and specialization programs in stuttering (no = 1; yes = 2). More details on all the variables can be found in Table 1.

3.2 Materials

During the first phase of the study, the Lebanese SLPs analyzed audio recordings of Lebanese-Arabic speech samples of eight children (two with a confirmed diagnosis of stuttering and six confirmed as CWNS – see criteria below in the section) to establish whether they are accurate, via a screening task, in identifying stuttering in bilingual children. During the second phase of the study, four months later, 27 out of the 32 SLPs analyzed the video recordings of the same speech samples. Our goal was to evaluate whether the use of video recordings instead of audio-recordings would improve the accuracy with which stuttering is identified, using the same set of speech samples as in the first phase.

3.2.1 Stimulus materials

The stimulus materials for the first phase were audio recordings of the narrative productions in Lebanese Arabic of 2 bilingual Lebanese CWS, and 6 bilingual Lebanese CWNS matched for age, language dominance (Lebanese Arabic), and elicitation mode (narrative based on the picture book ‘Frog goes to dinner’ by Mercer Mayer, 1974). The samples were not only audio-recorded, but also videotaped. The videotaped samples were

used in the second part of the study. The specific 8 audiotaped and videotaped samples were selected from a larger set of data with about 80 children. Each of these children produced a narrative based on a picture book and a spontaneous speech sample in both the dominant and non-dominant language, amounting to 4 speech samples per child. This larger set of data was collected to analyze speech disfluencies of bilingual Lebanese CWS and CWNS with different linguistic profiles in both their languages, in order to gain insight in the role of bilingualism in general and language dominance in particular in producing speech disfluencies (Saad Merouwe et al., in preparation). The current study selected one speech sample of 8 children, namely the picture book narrative in their dominant language (Lebanese Arabic).

The CWS were recruited through an open call via emails sent to all Lebanese SLPs working in Lebanon. The inclusion criteria were: (a) speaking two languages of which one was Lebanese Arabic; (b) diagnosed with stuttering by their treating SLP based on a comprehensive full assessment; (c) regarded by parents as having a stuttering problem; (d) identified independently as a CWS by the first author and another Lebanese SLP (both certified multilingual SLPs, fluent in Lebanese Arabic, and both fluency disorders specialists with extensive experience in treating stuttering in bilingual children); (e) stuttering severity rated by parents, the first author and another SLP as 2 or higher on an 8-point scale (0 = no stuttering, 1 = borderline, 2 = mild, 3 = mild to moderate, 4 = moderate, 5 = moderate to severe, 6 = severe, 7 = extremely severe). The stuttering severity rating attributed by the first author and the other SLP (on the basis of analyses of all 4 speech samples collected for the larger dataset) was based on type and frequency of disfluencies as well as the presence of physical concomitants and/or arrhythmicity of iterations; (f) age-appropriate speech-language skills based on Parents of Bilingual Children Questionnaire (PaBiQ: see description below)

and the first author's evaluation on the basis of observation; and (f) no reported neurological or learning disorders based on the interview conducted by the first author with the parents.

The identification of stuttering of the two CWS used in the current study was established independently by the first author and a second Lebanese SLP through a comprehensive assessment including video recording analyses of all 4 speech samples per child. Both children had English as a second language. The information of the parents on the child's stuttering was obtained over the phone via an interview conducted by the first author on the basis of a questionnaire. The interview allowed us to obtain more information about the case history of the CWS, the onset of stuttering, its development, the child's reactionary attitudes, parents' concerns and the treatment provided.

CWNS whose speech samples were used as stimulus materials were recruited through an open call sent to schools all over the country. Schools' principals were contacted by emails and telephone calls. After obtaining the principals' consent, letters were sent to parents to also obtain their consent. Children had to fulfill all the following inclusion criteria to be included as a CWNS: (a) speaking 2 languages of which one was Lebanese Arabic; (b) no parental or teacher concern regarding stuttering; (c) age-appropriate speech-language skills based on PaBiQ questionnaire, and the teacher's and the first author's evaluation on the basis of observation; (d) no parental or teacher concern regarding learning abilities; (e) no family history of stuttering; (f) no history of speech-fluency intervention; (g) no physical concomitants and normal rhythmicity of iterations; (h) the first author and the other SLP (stuttering specialists with high proficiency in all languages) evaluated independently the 4 speech samples of each child collected for the larger dataset, and judged the speech as characteristic of typically developing and typically fluent children (tension-free with only regular iterations). Both SLPs rated independently the severity levels of CWNS with the score 0 (normal) for all their speech samples. Among the 6 CWNS included in this study, two had

French as a second language, two had English as a second language, and two were exposed to both English and French in addition to Lebanese Arabic.

The PaBiQ (Tuller, 2015), a standardized parental questionnaire, was used to determine the language profile, including language dominance of each child. This questionnaire asks parents to report information about language input and output in different settings. On the basis of this, a language disorder risk (the No risk index), a language proficiency score (in each language), and a language dominance index can be calculated. The No risk index covers factors that indicate whether a child has a language disorder or not: age at the early stages of lexical and syntactic acquisition, parental concern and the existence of language difficulties in the family. The maximum No risk index is 23 and a score of 19-20 is taken as an indication of typical language development of a child. Language proficiency is estimated by asking parents to rate the current language skills of their child in each spoken language (absolute skills and skills in relation to monolingual children). Lastly, the 'languages used within the family and languages used elsewhere' scores make it possible to compare the child's languages with each other in terms of frequency of use (also other than the family). This information on current language use is combined with information on the quantity and quality of early exposure before the age of 4 to reach an estimation of the language dominance. The maximal possible obtained score is 50 (for a more detailed description, see Tuller et al., 2015). This questionnaire was validated for the Lebanese population, and further analysis showed that the No risk index (tapping into early language skills) highly correlated with the child's current language skills (Hallal et al., 2016). The No risk index and the language dominance index are presented in Table 2.

In this study, 6 bilingual CWNS and 2 CWS were included with a variety of profiles in terms of frequency and type of disfluencies, produced to be judged by 32 SLP participants. To make the samples' analyses feasible in terms of time needed to accomplish the task, we

included only one full speech sample (a narrative elicited via the picture book 'Frog goes to dinner') per child in Lebanese-Arabic, which was the dominant language of 7 out of 8 children. One child (CWNS5) was a balanced bilingual with approximately equal dominance scores in Lebanese Arabic and English. As stated above, Lebanese Arabic was also the dominant language of all SLP participants, which is important, as SLPs are better at identifying people who do and do not stutter in their native language than in a less well-known or foreign language (e.g., Humphrey, 2004; Van Borsel & Britto Pereira, 2005; Van Borsel et al., 2008).

The narrative speech samples of CWNS and CWS included in the current study were similar in terms of number of words (M_{CWNS} = 291.67; SD = 120.99; M_{CWS} = 394; SD = 22.6), $t(6) = -1.13$; $p = .301$, and number of syllables (M_{CWNS} = 448; SD = 181.9; M_{CWS} = 664; SD = 38), $t(6) = -1.5$; $p = .164$. The total sample duration (in seconds) was significantly longer for the CWS (M_{CWNS} = 266.17; SD = 82.76; M_{CWS} = 444.5; SD = 72.83), $t(6) = -2.69$; $p = .036$, which was expected given the higher frequency of disfluencies in their speech. We decided to calculate the percentage of disfluencies on the basis of the whole speech sample in order to obtain the most accurate disfluency estimations possible for each child. We included the full speech samples as they provide a better ecological validity (Olness, 2006), that is, in this way, each child's speech sample covers the narrative of the whole picture book. Had we cut the speech samples to obtain the same number of words, it would have meant that some children's speech samples would have been based on only part of the picture book.

3.2.2 Sample recording

In order to collect the narrative samples, the interviewer showed the wordless book "Frog goes to dinner" (Mayer, 1974) to each of the 8 children explaining that the story is about a boy, a frog and a dog. Before starting to describe the pictures in the book, each child was given time to go through the book and look at each picture to understand the general

theme of the story and storyline. Afterwards, the child was instructed to tell the story while going through the pictures in the book one by one. Each picture required a description to make the story proceed smoothly. The narrative samples of the 8 children were recorded and videotaped using a MacBook with a built-in microphone, which was placed on the table in front of the child for optimal quality. On the basis of these video-recordings, the samples were subsequently transcribed and coded to indicate different types of speech disfluencies.

3.2.3 Sample transcription and coding

The first author and 3 trained undergraduate SLP students transcribed and coded the narratives. Romanized Arabic was used to transcribe the speech samples produced in Lebanese Arabic. To categorize the disfluencies, a similar system to that of Byrd et al. (2015a) was used. Disfluencies were, cf. Ambrose and Yairi (1999), classified as SLD including monosyllabic word repetitions, syllable repetitions, sound repetitions, blocks, broken words and prolongations, and OD including multisyllable word repetitions, phrase repetitions, interjections, unfinished words/sentences and revisions. The revisions were further categorized in lexical, grammatical and phonological revisions (Bedore, et al., 2006).

Before transcribing and coding the speech samples, the first author trained the undergraduate SLP students in transcribing and coding speech samples and determined the inter-judge reliability based on three speech samples not included in the current dataset. Differences among the coders were solved through examination and discussion. In addition, the first author examined all the transcribed samples used for the current study to assure accuracy. The inter-judge reliability, calculated based on the agreement index percentage (point-by-point for location and type cf. Ambrose & Yairi, 1999, and Suen & Ary, 1989), was 0.87.

Mean SLD and OD percentages were calculated based on words as in most previous studies (Boey et al., 2007; Byrd, et al., 2015a, 2015b; Eggers, et al., 2019, 2020; Leclercq et

al., 2017; Pellowski & Conture, 2002; Tumanova et al., 2014). Table 3 lists the percentages for each type of disfluency calculated on the basis of the number of occurrences over the total number of words in the sample. Table 3 also lists the time it took each participant to complete the narrative and the numbers of words and syllables that were produced.

3.3 Procedures

3.3.1 Phase 1: Survey and audio recordings analysis tasks

Before listening to the 8 samples, the SLPs were instructed to complete a questionnaire developed in Google forms through which they answered questions related to personal information, their education and professional experience with respect to stuttering and bilingualism. In addition, they provided self-ratings of confidence in diagnosing stuttering in a bilingual context, and self-ratings of experienced difficulties in distinguishing SLD and OD.

The audio recordings were uploaded to a secure password-protected platform. The participants received an email including a link and a personal code in order to have access to the audio recordings. Before getting started, they had to read the regulations related to the general data protection regulation and give consent for their participation. They were informed that the task would last 50 minutes without interruption (including the speech samples analysis and completing the survey), and that listening again to the audio recordings was not possible. They had to check the sound system of their laptop in advance and were advised to use their earphones. The audio recordings were stored in a way that made them not downloadable. Only a play button was available, so the participants had to listen to the samples in one go without pauses.

After listening to each audio recording, the SLPs were instructed to indicate whether the sample was produced by a CWS or CWNS, which is quite similar to a screening task.

Consequently, they were asked to provide details about the speech characteristics that drove their decisions in general.

3.3.2 Phase 2: Video recordings analysis tasks

For phase 2, which took place 4 months after phase 1, 8 video recordings were uploaded to the same secure password-protected platform. The 8 videos included the same speech output as in phase 2 from the same 8 children. The same procedure was followed as in phase 1. However, different from phase 1, the SLP participants were now asked to explain after each video for each child individually, why they decided to label the audio-visual speech sample as one produced by a child who stutters or not. This task took about 40 minutes.

3.4 Statistical analyses

First, global analyses were conducted by an ANOVA with child category (CWNS versus CWS) and presentation mode (audio versus video) as within (repeated) variables. Second, separate ANOVAs were conducted for CWNS and CWS children independently to assess whether there was individual variation within one category. For the CWNS group, this yielded a 2 x 6 analyses with child (CWNS1 versus CWNS2 versus CWNS3 versus CWNS4 versus CWNS5 versus CWNS6) and presentation mode (audio versus video) as within (repeated) variables. The effect of CWNS was followed up by post-hoc comparisons. For the CWS pair, we performed a 2 x 2 analyses with child (CWS1 versus CWS2) and presentation mode (audio versus video) as within (repeated) variables. This analysis was followed up by two paired t-tests comparing CWS1 and CWS2 for each presentation mode separately. As the data were not distributed normally, we also analyzed the data by non-parametric analyses. As we found the same effects with these analyses as with the ANOVAs, we only report the latter. Second, the relationship between the percentages of correct answers and SLPs' characteristics (e.g., work experience) or samples' characteristics (e.g., % SLD) was explored by means of Pearson's correlation-analyses. These correlations were based on the performance in the audio

part of the experiment, as here more participants (N=32) were included than in the video part (N=27). It also allowed for a more direct comparison with the audio experiment of Byrd et al. (2015b). Subsequent qualitative analyses were done on the answers SLPs gave to motivate their choices with respect to whether a child was a CWS or CWNS. SPSS software (Version 27) was used for all analyses and all multiple comparisons were performed using Bonferroni corrections.

4. Results

4.1 Misidentification of bilingual CWNS compared to CWS in both presentation modes.

For audio presentation, the mean percentage of misidentification was significantly higher for CWNS (45.31%) compared to CWS (12.5%). This was the case for video presentation as well (CWNS 45.65% versus CWS 14.8%). Only one SLP was entirely accurate at identifying all the CWNS and CWS in both presentation modes. Accordingly, the difference in identifying CWS and CWNS yielded a significant main effect for child category accompanied by a large effect size, $F(1, 26) = 11.32$, $p = .02$, $n_p^2 = .30$). There was no effect of presentation mode ($F(1, 26) = 0.16$, $p = .69$, $n_p^2 = .006$) nor an interaction between presentation mode and child category ($F(1, 26) = 0.25$, $p = .62$, $n_p^2 = .010$), and in both cases the effect size was rather small as well. More detailed information is depicted in Figure 1.

4.2 Differences in misidentification between individual children

A significant difference in misidentification emerged between the different CWNS children accompanied by a large effect size, $F(5, 130) = 9.62$, $p < .001$, $n_p^2 = .27$. Post-hoc analyses to further assess this main effect revealed significant differences between CWNS4 versus CWNS1 ($p < .001$), CWNS4 versus CWNS2 ($p = .019$), CWNS4 versus CWNS3 ($p = .001$), CWNS4 versus CWNS5 ($p < .001$), CWNS4 versus CWNS6 ($p = .002$), and CWNS5 versus CWNS2 ($p = .049$). The percentage of misidentification was the lowest for CWNS4

(11%) and the highest for CWNS5 (67%) with the other rates falling in between. There was no effect of presentation mode, $F(1, 26) = 0.12$, $p = .91$, $n_p^2 = .000$, nor an interaction between presentation mode and CWNS child, $F(5, 130) = 1.10$, $p = .36$, $n_p^2 = .041$, and the effect sizes here were minimal or small. More detailed information on the main effect of CWNS is depicted in Figure 2a.

A significant difference in misidentification emerged between the CWS children $F(1, 26) = 9.35$, $p = .005$, $n_p^2 = .27$). There was no significant effect of presentation mode ($F(1, 26) = 0.00$, $p = 1.00$, $n_p^2 = .000$), but there was a tendency for an interaction between presentation mode and CWS child, ($F(1, 26) = 2.74$, $p = .11$, $n_p^2 = .095$). Given that also the effect size was between moderate and large, we followed the interaction up with two paired t-tests. These t-tests showed that, while there was no difference between CWS1 and CWS2 in the audio part of the experiment, $t(31) = -0.81$, $p = .42$, the difference in the video part of the experiment was significant, $t(26) = -3.31$, $p = .003$. This is represented in Figure 2b.

4.3 Variables predictive of a correct identification

In this section, we consider the overall success rate in correctly diagnosing CWNS as CWNS and CWS as CWS in the audio part of the experiment (as here there were more participants than in the video part), and how this may be related to SLP characteristics (e.g., work experience in SLP) and child-related/case-related factors (e.g., number of SLDs). The success rate of SLPs was on average 63% and ranged from 25% to 100%.

4.3.1 SLPs characteristics

Here we determined the relationship between the overall success rate of SLPs and five SLP characteristics, i.e., work experience, number of bilingual CWS treated, self-ratings of confidence in diagnosing stuttering in a bilingual context, self-ratings of difficulties in distinguishing SLD and OD and general experience. None of the variables showed a significant correlation with SLP scores (work experience: $r(30) = -.012$, $p = .948$; number of

bilingual CWS treated: $r(30) = -.187$, $p = .305$; self-ratings of confidence in diagnosing stuttering in a bilingual context: $r(30) = -.112$, $p = .541$; self-ratings of difficulties in differentiating SLD from OD: $r(30) = -.324$, $p = .070$; general experience: $r(30) = -.273$, $p = .130$).

4.3.2 Speech samples characteristics

Potential correlations between the percentage of correct answers given by the SLPs for CWNS and the percentages of SLD and OD were assessed. For CWNS, correlations were not significant for either of the variables and success rate (SLD: $r(4) = -.29$; $p = .57$; OD: $r(4) = .086$; $p = .87$). Figure 3 shows that the lack of correlation may be driven by one outlier, CWNS5, who has relatively little SLD and OD, but was misidentified very often (72%). After removing this outlier, a significant negative correlation was found between the success rate and the %SLD (SLD: $r(3) = -.934$; $p = .02$; OD: $r(3) = -.659$; $p = .226$). For CWS, we refrain from correlational analyses as there are only 2 data points. However, the CWS with a higher number of SLD and OD was more often accurately identified than the CWS with low SLD and OD, so also here there is some hint that there is a relation between the number of disfluencies and successful identification.

4.4 Speech characteristics that SLPs consider when judging the group of CWNS and CWS

Table 5 and table 6 present an overview of the descriptors used by the SLPs to identify CWNS and CWS in the case of correct and incorrect identification. For CWNS, all participants' responses could be clustered under 5 main categories. When CWNS were correctly identified, the most frequently used descriptors included the exhibition of typically fluent speech with normal disfluencies without tension coming along with the disfluencies. When they were misidentified as CWS, the most frequently provided explanations were the exhibition of SLD (including word repetitions, part-word repetitions, prolongations, and

blocks), either without or with physical concomitants. For CWS, the participants' responses were clustered under 4 main categories in the case of correct identification, and 3 main categories in the case of wrong identification. When CWS were correctly identified, the main descriptors included the presence of SLD and/or physical concomitants. When CWS were misidentified as CWNS, the participants mainly attributed it to the low frequency of disfluencies and physical concomitants.

5. Discussion

The major findings of this study in Lebanese bilinguals showed that (1) misidentification was significantly more likely to happen with bilingual CWNS than with CWS, (2) the audio-visual presentation mode did not result in better performance than the audio-only presentation mode, (3) the percentages of misidentification were different within the CWNS group as well as between the two CWS, (4) SLPs' characteristics did not have a clear influence on diagnostic judgments, but there was some indication that an increased number of SLD leads to a higher percentage of misidentification in CWNS and a higher success rate in CWS, and (5) in CWNS, misidentification frequently occurred through misinterpreting physical concomitants, while in CWS, it was related to the low frequency of disfluencies and physical concomitants observed in the child with mild stuttering.

5.1 Identification of CWNS and CWS

The SLPs were significantly more correct in identifying CWS than CWNS. In other words, SLPs tend to misidentify bilingual CWNS more often as CWS than they are misidentifying CWS as CWNS. This main result is in line with the findings by Byrd et al. (2015b). However, the mean percentage of misdiagnosis of the one CWNS in their study was much higher than the average percentage of misidentification of the group of CWNS in the current one (85.71% versus 44.8%; note these percentages are both based on audio recordings). This underlines the usefulness of including CWNS with different disfluency

profiles, as the high percentage of misdiagnosis of CWNS in Byrd et al. (2015b) is most likely related to the CWNS having 17% disfluencies in both languages, whereas the CWNS in our study had between 5 and 14% disfluencies. The current study – including more cases and a larger number of SLP participants – supports thus on one hand the findings of Byrd et al. (2015b), that the chance for not identifying stuttering in bilingual CWS is lower than the chance of false positive identification of CWNS. On the other hand, it also indicates that the misidentification rate for CWNS may be exaggerated in the Byrd et al.'s study, and that in case of CWNS within a more typical range of disfluencies, the percentage of misidentification may be substantially lower.

The present study included 6 CWNS and 2 CWS with different profiles in terms of speech fluency and language proficiency. We opted for this larger variety in order to get more insight into the characteristics driving misinterpretation of disfluencies in bilinguals. The child that was most often correctly identified as a CWNS (i.e., CWNS4) had the lowest percentage of disfluencies (%total disfluencies = 4.84; %SLD = 2.15). However, the child with the second lowest number of disfluencies (%total disfluencies = 6.35; %SLD = 3.47), CWNS5, was most often misidentified. One characteristic that set this child apart from the other 5 CWNS was her language dominance score. More specifically, this child had a substantially lower language dominance score in Lebanese Arabic compared to other children. In fact, this dominance score was more or less at par with her English dominance score (see Table 2). In addition, this child has considerable exposure to French at school as well. In other words, CWNS5 is a trilingual, with rather balanced competencies between Lebanese Arabic and English. Therefore, it may well be that other features of bilingualism shone through (less extensive lexicon, shorter sentences), and that this, in combination with her SLD – consisting mostly of monosyllabic word repetitions – led to a higher percentage of misidentification. As there is only one such case in this sample, this finding requires further investigation, but it

anyway indicates that linguistic profile may impact the perception of stuttering. With regards to CWS, no difference was found between the percentages of misidentification of CWS1 and CWS2 on the basis of audio-only recordings, but the difference was significant on the basis of audio-visual recordings, with a higher percentage of misidentification of CWS2 compared to CWS1. We come back to this discrepancy in the next section where we discuss the impact of presentation mode on assessments.

5.2 Presentation mode

Our main analyses showed that the mode of presentation (audio versus video) does not improve the identification of CWNS and CWS. This result is against our hypothesis and rather intriguing because one would expect better results from an audio-visual experience, as it allows to consider both verbal (speech disfluencies) and non-verbal (presence/absence of physical concomitants) cues. For the CWS there was no difference in misidentification in the audio-only mode, but a significant difference was observed in the audio-visual mode. However, this difference was not consistent across the two CWS. CWS1 (who showed a higher percentage of total disfluencies associated with frequent and noticeable physical concomitants) elicited flawless judgments in the audio-visual but not in the audio-only mode, whereas CWS2 (who had fewer disfluencies and less frequent physical concomitants) elicited better performance in the audio-only than the audio-visual mode. These findings are in line with Watson and Kayser (1994) who showed that SLPs easily observe atypical speech behaviors, or secondary behaviors in severe cases of stuttering, but not when a more fine-grained level of analysis – also with respect to physical concomitants - is required as in case of mild stuttering.

In sum, although further studies might be needed, these findings imply that audio-visual presentation mode does not seem to provide sufficient details to substantially improve identification of stuttering in bilinguals in general. This is not in line with an early study of

Luper (1956), who showed that visual samples led listeners to identify more stuttering behaviors than auditory samples. However, it corroborates with later studies that found that the mode of presentation of speech samples does not influence listener's identification and judgments of stuttering (Bloodstein & Bernstein Ratner, 2008; Martin & Harddson, 1992; Panico et al., 2005). One explanation is that video-recordings do not allow an accurate observation of all physical concomitants (e.g., movements of extremities) as a live experience would allow. Moreover, in the current study, SLPs had to watch the speech samples in one go, without pauses or rewinding, because these options were not available on the platform. Consequently, some SLPs may have missed the physical concomitants of CWS2 (mild stuttering), as they were less frequent than those of CWS1, requiring a greater need for in-depth analysis.

5.3 SLP characteristics that may impact the correct identification of CWNS and CWS

Previous research has shown that experience and training may influence the reliability and accuracy of stuttering judgments (Cordes & Ingham, 1999; Cordes et al., 1992). Therefore, we expected to find some associations between the percentages of correct answers of the SLPs and factors related to their professional experience (i.e., work experience, general experience, number of bilingual CWS treated, self-ratings of confidence in diagnosing stuttering in a bilingual context, and self-ratings of difficulties in distinguishing SLD and OD). Interestingly, no correlation was found between the percentages of correct answers given by the SLPs and all the variables stated above. One possible explanation is the limited variance within our group, especially in terms of age and experience. That is, almost all of the SLPs in this study were relatively young and had less than 10 years' work experience. On the other hand, Byrd et al. (2015b) also found within their group of participants that there was no distinct association between identification accuracy and years of experience or number of clients treated, even though their sample included almost 50% SLPs with 10 years of work

experience or more. Also, the participant that was clearly the most experienced in their study (37 years of experience as SLP) and had assessed speech fluency of more than 50 bilingual clients did not accurately identify the CWNS and the CWS. In contrast, the only participant who correctly identified both CWNS and CWS had limited experience (treated only one bilingual Spanish-English CWS). Moreover, in Byrd et al. (2015b), no apparent relationship between the educational background (professional development courses completed on stuttering and bilingualism) and identification accuracy was found. Byrd et al.'s findings and our findings are in line with the conclusions of Cordes and Ingham (1995) suggesting that high levels of disagreement can occur among graduate students who have taken classes in stuttering and fluency disorders and less experienced clinicians, but also even among highly experienced clinicians. The results show that reaching a high level of proficiency in stuttering identification is not easy and is undoubtedly even more complex in the bilingual context. Together, the results indicate the constant requirement for SLP students and clinicians to be trained in identifying stuttering speech behaviors with reasonable levels of accuracy and reliability, and the crucial need for recent and relevant information specific to the manifestations of stuttering in bilingual children.

5.4 The relation between percentage of disfluencies and success rate of SLPs

When considering all CWNS, there were no correlations between the success rate of SLP participants, and the characteristics of the speech samples SLPs had to analyze (% SLD and %OD). This is partly related to the false-positives elicited by CWNS5 who was frequently misidentified despite the relatively low percentage of SLD found in her speech sample. When this case was removed from the group of CWNS, it was found that the success rate did correlate with the SLD percentage, as it increased when the %SLD in the speech samples decreased. Even though this finding is exploratory, it is in line with the findings of Byrd et al. (2015b).

It should be noted that the SLD in all CWNS included mostly word and part-word repetitions. This is not exclusive for our study, as also Bedore et al. (2006), Byrd et al. (2015b), Eggers et al. (2019) and Fiestas et al. (2005) found that bilingual CWNS exhibit a high level of whole-word, syllable and sound repetitions in their speech. Interestingly, Byrd et al. (2015b) report that sound and syllable repetitions, and whole-word repetitions were the most commonly cited disfluencies that contributed to the false-positive identification of CWNS. What is important is that these disfluencies are specified in assessment protocols as stuttering-like (Ambrose & Yairi, 1999), even though for bilinguals, they often are not indicative of stuttering. This implies that the use of guidelines established for monolinguals to identify stuttering in bilinguals may not be appropriate. This increases the need to look for other clinical features accompanying these types of disfluencies when assessing the speech fluency of bilingual children. For one thing, both Byrd et al. (2015b) and Eggers et al. (2019) found that the high rates of speech disfluencies bilingual children produced were not associated with atypical tension or arrhythmicity. We conclude that the timing of SLD and accompanying tension, along with parental concern, must be considered even more carefully as clinically relevant discriminators of stuttering in the bilingual than in the monolingual population.

As for the two CWS, the success rate of the SLPs was higher for the CWS with a higher percentage of SLD and OD. This result is in line with the conclusions of Watson and Kayser (1994) who state that SLPs easily identify atypical speech behaviors or secondary behaviors in cases of severe stuttering (thus, when the frequency of atypical speech behaviors is higher), but the identification can be more challenging in case of mild stuttering, which requires a more in-depth analysis of the observed speech behavior.

5.5 Descriptors used by the SLPs to label CWNS and CWS as stuttering or not

The diagnostic criteria used to determine whether a child was a CWS or a CWNS were also provided by the SLPs after each case in the video part of the experiment. For the group of CWNS, misidentification of a case was most often motivated by the perception of word and part-word repetitions, prolongations and blocks without consideration of physical concomitants, which – as we argued above – are of great importance to ascertain whether a disfluency is stuttering-like. The second most frequent criterion in case of misidentification was SLD associated with physical concomitants, referring specifically to tensed breathing and absence of eye contact. This explanation is theoretically valid, however, none of the 6 CWNS was judged by the first author and the other SLP (both experienced fluency specialists) as exhibiting physical concomitants (see Table 3). Only one child (CWNS4) occasionally produced a breathy voice, but it was not associated with tension. In addition, the absence of eye contact does not provide a valid rationale given the setting of the data collection, as children were asked to tell the story while looking at the pages and turning them. Hence, it is natural to be focused on the task instead of looking all the time at the camera or at the interviewer. Most importantly perhaps, characteristics that are sometimes qualified or interpreted as physical concomitants, and which occurred in the speech samples (noisy breathing pattern, poor eye contact) were in most cases not occurring simultaneously with SLD and can therefore not be taken as evidence for stuttering. The lack of consideration of physical concomitants and the misinterpretation of SLD may be related to the relatively low levels of experience of the SLPs in the study. Some other explanations in case of misidentification included the exhibition of typical disfluencies associated with word retrieval difficulties or word retrieval difficulties related to bilingualism. Fortunately, these explanations were rather infrequent, as one should be aware that word retrieval difficulties are a common phenomenon in the bilingual context.

When CWNS were correctly identified, it was pointed out that disfluencies were (1) normal disfluencies, or (2) related to word retrieval difficulties that can be attributed to bilingualism or (3) disfluencies without any associated physical concomitant. These correct explanations imply that there is awareness of criteria for stuttering, but the relatively high percentage of misidentification and the invalid explanations for these do also imply that these criteria were not always correctly or consistently applied, at least not by all SLPs. This underlines that the identification of stuttering is a difficult perceptual task –especially with bilingual children who tend to experience a high level of linguistic uncertainty in the first place (Byrd, 2018).

For CWS, the most frequently used explanation when they were correctly identified was the exhibition of SLD with observed physical concomitants, which shows that overall SLPs applied a valid classification criterion (e.g., Ambrose & Yairi, 1999) to judge whether disfluencies were stuttering-like. Misidentification was observed predominantly for CWS2 (the child with mild stuttering), and the most commonly provided explanation (57.2%) referred to the low frequency of disfluencies, sometimes in combination with reference to a limited lexicon (28.5%). This confirms the findings of previous publications (e.g., Watson & Kayser, 1994) which showed that it is more difficult for SLPs to detect cases of mild stuttering with low frequency of SLD and mild physical concomitants requiring a fined-grain level of analysis. It also confirms that the correct analysis of speech behaviors and the identification of stuttering are challenging perceptual but also cognitive tasks, which require extensive abilities of the SLPs, particularly in the bilingual context.

5.6 Additional considerations

In this study, SLPs were provided with only one speech sample per child, in Lebanese Arabic, which was the dominant language for all SLPs. Van Borsel et al. (2008) emphasized the importance of analyzing SLD in both languages in order to diagnose a bilingual child with

stuttering. In this respect, our design may have been suboptimal. However, our rationale for incorporating one speech sample per child from the dominant language only was built on the following. First, we wanted to include a larger number of children (compared to Byrd et al., 2015b), with different profiles in terms of frequency and types of disfluencies as well as language dominance score, in order to better understand which combination of characteristics lead to a higher risk of misidentification. We also wanted to have a larger number of SLP participants (compared to Byrd et al., 2015b) to assure more certainty about the conclusions. Finally, we not only wanted to use audio-recordings, but also wanted to investigate whether video-recordings allow for better judgments given that the observation of secondary behaviors can be important to diagnose a bilingual child as a child who stutters (Byrd et al., 2015b; Watson & Kayser, 1994). This also meant that we had to compromise, e.g., by not presenting a speech sample in the non-dominant language, as there was a maximum time set to accomplish the different tasks of this study (altogether 90 minutes).

Furthermore, we decided to include the dominant-language instead of the non-dominant because different studies have made a case that a more accurate diagnosis can be made on the basis of a child's dominant language (e.g., Watson & Kayser, 1994). Speech disfluencies in the non-dominant language could for a great part reflect difficulties with the acquisition and development of this language, and is therefore less suitable to make a proper assessment. We also wanted to match the dominant language of the children with the dominant language of the participating SLPs to control additional variables related to language familiarity. Lastly, assessment of the dominant language can be argued to be ecological from the clinical point of view, because in clinical practice, SLPs base their decisions relatively often on speech behaviors in the dominant language while diagnosing stuttering. The current study revealed that 26% of the participants assess only the dominant language.

Next it can be argued that – unlike in our experiment – a clinician has the opportunity to listen to speech samples more often before reaching a decision. This is in principle true, but time pressure may not always allow for such a detailed assessment and we can at least say that the current experiment resembles the common practice of screening children in schools where SLPs listen to a child speaking in class and refer some children for a more detailed assessment.

One final aspect to consider is the relatively young age of our participants. The recruitment process was challenging and several experienced SLPs could not participate due to their very busy schedule. On the other hand, speech-language pathology is a relatively young field in Lebanon, so our sample is fairly representative of the Lebanese reality given that clinicians are typically young.

6. Conclusions

The present study shows that it is a real challenge to determine whether a bilingual child is stuttering or not. This challenge is especially apparent for bilingual CWNS, given that misidentification is more likely to occur in CWNS than in CWS, although in case of milder stuttering symptoms, our findings indicate that misidentification of CWS is not uncommon either. Importantly, the differences in misidentification rate within the CWNS group imply that not every bilingual child is at high risk to be misidentified. It appears that bilingual CWNS who exhibit few SLD are less likely to be misidentified, or reversely that wrong judgment is especially triggered by the exhibition of a high frequency of SLD in the dominant language. However, when there is no clear dominant language (e.g., in balanced bilingualism), a relatively low number of SLD may already be enough to make a stuttering-like impression. The exact nature of this phenomenon should be further investigated, but it may be that clinicians take some speech behaviors related to low language proficiency (e.g., avoidance of difficult words, avoidance of long sentences) into account when judging a

CWNS to be a CWS. Lastly, the presentation mode does not seem to improve diagnostic accuracy. The reasons that SLPs gave to qualify a CWNS as a CWS indicate that they often do not consider the absence of physical concomitants or wrongly interpret certain behaviors as physical concomitants. This might be related to the requirements of the task itself (having to listen to the speech sample in one go) and/or to their limited experience. The lack of correlational significance between accuracy rate and any of the SLP-characteristics underscores this notion.

Given the increased likelihood to misidentify bilingual CWNS, it is important that clinicians carefully assess the rhythmicity of iterations and the tension concomitant with SLD. At the same time, it is important to develop standardized methods to educate students and clinicians to distinguish between typical disfluencies and stuttering speech behaviors to reliably assess whether somebody stutters or not, especially in case of bilingual children who have to navigate different linguistic systems, and therefore tend to experience a higher level of linguistic uncertainty.

7. Acknowledgements

The present study was supported by the University of Turku and Saint-Joseph University of Beirut. The authors would like to thank the SLPs, children, and parents who participated in the study. They would also like to thank the students Nadia Sabeh'Ayon, Reem Abi Akar, Laudy Arida, Ghena Antoun, Christel Daaboul, Dana Barakat and Fatmé Khalife for their assistance in collecting and transcribing the data.

8. Reference list

Abou, S. J. (1962). *Le bilinguisme arabe-français au Liban. Essai d'anthropologie culturelle.* Paris: Presses Universitaires de France.

Ambrose, N. G. & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language, and Hearing Research*, *42*(4), 895–909. Doi: https://doi.org/10.1044/jslhr.4204.895

Ardila, A., Ramos, E. & Barrocas, R. (2011). Patterns of stuttering in a Spanish/English bilingual: A case report. *Clinical Linguistics & Phonetics*, *25*(1), 23-36. Doi: https://doi.org/10.3109/02699206.2010.510918

Baker, C. (2001). *Foundations of bilingual education and bilingualism* (3rd ed.). Clevedon, England: Multilingual Matters.

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. Doi: https://doi.org/10.2167/beb392.0

Bedore, L. M., Fiestas, C. E., Peña, E. D., & Nagy, V. J. (2006). Cross-language comparisons of maze use in Spanish and English in functionally monolingual and bilingual children. *Bilingualism: Language and Cognition, 9*(3), 233-247. Doi: https://doi.org/10.1017/s1366728906002604

Beyea, K. (2020). Bilingual Practices in Speech-Language Pathology in Nebraska Schools. *Theses/Capstones/Creative Projects.* 73. https://digitalcommons.unomaha.edu/university_honors_program/73

Bloodstein, O., & Bernstein Ratner, N. (2008). *A handbook of stuttering* (6th ed.). New York, NY: Thomson Delmar Learning.

Boey, R. A., Wuyts, F. L., Van de Heyning, P. H., De Bodt, M. S., & Heylen, L. (2007). Characteristics of stuttering-like disfluencies in Dutch-speaking children. *Journal of Fluency Disorders*, *32*(4), 310–329. Doi: https://doi.org/10.1016/j.jfludis.2007.07.003

Brundage, S. B., Bothe, A. K., Lengeling, A. N., & Evans, J. J. (2006). Comparing judgments of stuttering made by students, clinicians, and highly experienced judges. *Journal of Fluency Disorders, 31*(4), 271 – 283. Doi: https://doi.org/10.1016/j.jfludis.2006.07.002

Byrd, C. (2018). Assessing bilingual children: Are their disfluencies indicative of stuttering or the by-product of navigating two languages? *Seminars in Speech and Language*, *39*(4), 324-332. Doi: https://doi.org/10.1055/s-0038-1667161

Byrd, C. T., Bedore, M. L., & Ramos, D. (2015a). The disfluent speech of bilingual Spanish-English children: Considerations for differential diagnosis of stuttering. *Language, Speech, and Hearing Services in Schools, 46*(1), 30–43. Doi: https://doi.org/10.1044/2014_LSHSS-14-0010

Byrd, C. T., Watson, J., Bedore, L., & Mullis, A. (2015b). Identification of stuttering in bilingual Spanish-English speaking children. *Contemporary Issues in Communication Science and Disorders, 42*, 72-87. Doi: https://doi.org/10.1016/s0094-730x(03)00004-4

Carias, S. & Ingram, D. (2006). Language and disfluency: Four case studies on Spanish–English bilingual children. *Journal of Multilingual Communication Disorder*s, *4*(2), 149–157. Doi: https://doi.org/10.1080/14769670601092663

Carlo, E. & Watson, J. (2003). Disfluencies of 3- and 5-year old Spanish-speaking children. *Journal of Fluency Disorders, 28*(1), 37–53. Doi: https://doi.org/10.1016/s0094-730x(03)00004-4

Chaudhary, C., Marythy, S., Gudattu, V., & Krishnan, G. (2021). A systematic review on the role of language-related factors in the manifestation of stuttering in bilinguals. *Journal of Fluency Disorders*, 68, 1-16. Doi: https://doi.org/10.1016/j.jfludis.2021.105829

Chin, N. B. & Wigglesworth, G. (2007). *Bilingualism: An advanced resource book*. Routledge.

Cordes, A. K., & Ingham, R. J. (1994). The reliability of observational data: II. Issues in the identification and measurements of stuttering events. *Journal of Speech and Hearing Research, 37*(2), 279-294. Doi: https://doi.org/10.1044/jshr.3702.279

De Lamo White, c., & Jin, L. (2011). *Evaluation of speech and language assessment approaches with bilingual children. International Journal of Language and Communication Disorders*, *46*(6), 613-627. Doi: https://doi.org/10.1111/j.1460-6984.2011.00049.x

Eggers, K., Van Eerdenbrugh, S., & Byrd, C. T. (2019). Speech disfluencies in bilingual Yiddish-Dutch speaking children, *Clinical Linguistics & Phonetics, 34*(6), 1-17. Doi: https://doi.org/10.1080/02699206.2019.1678670

Eggers, K., Van Eerdenbrugh, S. & Byrd, C. (2020). Spraakonvloeiendheden bij tweetalig Jiddisch-Nederlands-, Turks-Nederlands- en eentalig Nederlandssprekende controlegroep. [Speech disfluencies in bilingual Yiddish-Dutch, Turkish-Dutch and monolingual Dutch control group]. *Logopedie* [*Logopedics*], mei-juni [May-June], 21-34.

Einarsdóttir, J., & Ingham, R. (2009). Does language influence the accuracy of judgments of stuttering in children. *Journal of Speech, Language, and Hearing Research, 52*(3), 766-779. Doi: https://doi.org/10.1044/1092-4388(2009/07-0248)

Fiestas, C. E., Bedore, L. M., Peña, E. D., & Nagy, V. J. (2005). Use of mazes in the narrative language samples of bilingual and monolingual 4- to 7-year-old children. In J. Cohen, K. T. McAllister, K. Rolstad, & J. MacSwan (dir.), *Proceedings of the 4th international symposium on bilingualism*. 730–740. Cascadilla Press.

Finn, P. & Cordes, A. K. (1997). Multicultural identification and treatment of stuttering: A continuing need for research. *Journal of Fluency Disorders*, *22*(3), 219-236. Doi: https://doi.org/10.1016/S0094-730X(97)00008-9

Gollan, T. H., Montaya, R. I. & Bonanni, M. P. (2005). Proper names get stuck on bilingual and monolingual speakers' tip of the tongue equally often. *Neuropsychology, 19*(3), 278-287.

Grosjean, F. (2021). *Life as a bilingual*. Cambridge: Cambridge University Press. Doi: https://doi.org/10.1017/9781108975490.003

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*(1), 3-36. Doi: https://doi.org/10.1016/0093-934X(89)90048-5

Hallal, F., Kouba Hreich, E., & Messarra, C. M. (2016). *Approche méthodologique pour l'évaluation orthophonique dans le context plurilingue Libanais: Evaluation de l'utilisation du "questionnaire pour parents d'enfants bilingues" (litmus-pabiq) dans une démarche de bilan orthophonique*. [*Methodological approach for speech therapy assessment in the Lebanese multilingual context: Assessment of the use of "the parents*

of bilingual children questionnaire" (litmus-pabiq) in a speech therapy assessment process], [Unpublished thesis]. Saint-Joseph University of Beirut.

Howell, P., Davis, S., & Williams, R. (2009). The effects of bilingualism on stuttering during late childhood. *Archives of Disease in Childhood*, *94*(1), 42–46. Doi: https://doi.org/10.1136/adc.2007.134014

Humphrey, B. D. (2004). Judgments of disfluency in a familiar vs. an unfamiliar language. In A. Packmann, A. Meltzer, & H. F. M. Peters (Eds.), *Theory, research and therapy. Proceedings of the Fourth World Congress on Fluency Disorders*, 424-427, Nijmegen, the Netherlands: Nijmegen University Press.

Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode on repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, *12*(2), 198-208. Doi: https://doi.org/10.1044/1058-0360(2003/066)

Ingham, R. J., & Cordes, A. (1992). Interclinic differences in stuttering events counts. *Journal of Fluency Disorders, 17*(3), 171-176. Doi: https://doi.org/10.1016/0094-730X(92)90009-F

Jacobsen, P. F. & Schwartz, R. G. (2005). English past tense use in bilingual children with language impairment. *American Journal of Speech Language Pathology, 14*(4), 313-323. Doi: https://doi.org/10.1044/1058-0360(2005/030)

Jankelowitz, D. L., & Bortz, M. A. (1996). The interaction of bilingualism and stuttering in adult. *Journal of Communication Disorders, 16*(4), 287-297. Doi: https://doi.org/10.1016/0021-9924(95)00050-X

Karniol, R. (1992). Stuttering out of bilingualism. First Language, 12, 255-283. Doi: https://doi.org/10.1177/014272379201203604

Kohnert K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders, 43*, 456-473. Doi: https://doi.org/10.1016/j.jcomdis.2010.02.002

Leclercq, A.-L., Suaire, P., & Moyse, A. (2017). Beyond stuttering: Speech disfluencies in normally fluent French-speaking children at age 4. *Clinical Linguistics & Phonetics*, *32*(2), 166–179. Doi: https://doi.org/10.1080/02699206.2017.1344878

Lee, A. S., Robb, M. P., Ormond, T., & Blomgren, M. (2014). The role of language familiarity in bilingual stuttering assessment. *Clinical Linguistics & Phonetics*, *28*(10), 723–740. Doi: https://doi.org/10.3109/02699206.2014.892154

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English.

Martin, R. R., & Haroldson, S. K. (1992). Stuttering and speech naturalness: Audio and audiovisual judgments. *Journal of Speech and Hearing Research*, *35*(3), 521-528. Doi: https://doi.org/10.1044/jshr.3503.521

Maruthy, S., Raj, N., Geetha, M. P., & Priya, C. S. (2015). Disfluency characteristics of Kannada-English bilingual adults who stutter. *Journal of Communication Disorders, 56*, 19-28. Doi: https://doi.org/10.1016/j.jcomdis.2015.06.001

Mayer, M. (1974). *Frog goes to dinner*. New York, NY: Dial Books for Young Readers

Natke, U., Sandrieser, P., Pietrowsky, R., & Kalveram, K. (2006). Disfluency data of German preschool children who stutter and comparison children. *Journal of Fluency Disorders*,

31(3), 165–176. Doi: https://doi.org/10.1016/j.jfludis.2006.04.002

Olness, G. (2006, May 29-June 2). *Structural assessment of narratives: Some issues and Illustrative Data.* [Paper presentation]. Clinical Aphasiology Conference, Ghent, Belgium. https://aphasiology.pitt.edu/id/eprint/1750

Panico, J., Healey, C. E., Brouwer, K., & Susca, M. (2005). Listener perceptions of stuttering across two presentation modes: quantitative and qualitative approach. *Journal of Fluency Disorders, 30*(1), 65-85. Doi: https://doi.org/10.1016/j.jfludis.2005.01.003

Pellowski, M. W. & Conture, E. G. (2002). Characteristics of speech disfluency and stuttering behaviors in 3- and 4-year-old children. *Journal of Speech, Language, and Hearing Research, 45*(1), 20–34. Doi: https://doi.org/10.1044/1092-4388(2002/002)

Peña, E. D., Bedore, L. M. & Zlatic-Guinta, R. (2002). Category-generation performance of bilingual children: The influence of condition category, and language. *Journal of Speech, Language and Hearing Research, 45*(5), 938-947. Doi: https://doi.org/10.1044/1092-4388(2002/076)

Saad Merouwe, S., Bertram, R., Richa, S. & Eggers, K. (2022). Bilinguisme et bégaiement: Revue de la littérature et implications sur le diagnostic clinique. *Glossa, 132*, 67-91.

Saad Merouwe, S., Richa, S., Bertram, R., & Eggers, K. (2019, November 21-23). *Speech disfluencies in Lebanese Bilingual Speakers: Preliminary Findings.* [Poster presentation]. American Speech-Language-Hearing Association ASHA, Orlando, USA.

Shapiro, D. (2011). *Stuttering intervention*. (2nd ed.). Pro-ed. Austin, Texas.

Schäfer, M., & Robb, M. P. (2012). Stuttering characteristics of German-English bilingual speakers. *Clinical Linguistics & Phonetics, 26*(7), 597-612. Doi: https://doi.org/10.3109/02699206.2012.689918

Shaaban, K. A. (1997). Bilingual education in Lebanon. In J. Cummins & D. Corson (Eds.), *Encyclopedia of Language and Education, 5,* (pp. 251-259). The Netherlands: Kluwer.

Shenker, R. C., Conte, A., Gingras, A., Courcey, A., & Polomeno, L. (1998). The impact of bilingualism on developing fluency in a preschool child. In E. C. Healey, & H. F. M. Peters (Eds.), *Second world congress on fluency disorders proceedings*, San Francisco, August 18-22 (pp. 200-204). Nijmengen: Nijmegen Univ. Press.

Shin, S. J. (2017). *Bilingualism in schools and society: Language, identity and policy* (2nd ed.). Routledge. Doi: https://doi.org/10.4324/9781315535579

Sorge, B. G., Toplak, M. E., & Bialystok, E. (2016). Interactions between levels of attention ability and levels of bilingualism in children's executive functioning. *Developmental Science, 20*(1), e12408. Doi: https://doi.org/10.1111/desc.12408

Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioral observation data.* Mahwah, NJ: Erlbaum. Doi: https://doi.org/10:4324/9781315801827

Szmalec, A. (2013). Bilinguisme et bégaiement chez l'enfant. *Enfance*, *3*(3), 287-298. Doi: https://doi.org/10.4074/S0013754513003091

Tatliancich-Klinger, C. L., Byrd, C. T., & Bedore, L. M. (2013). The disfluent speech of a Spanish-English bilingual child who stutters. *Clinical Linguistics & Phonetics, 27*(12), 888-904. Doi: https://doi.org/10.3109/02699206.2013.813076

Thonhauser, I. (2000). Multilingual education in Lebanon: "Arabinglizi" and other challenges of multilingualism. *Mediterranean Journal of Educational Studies, 6*(1), 49-2000.

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J.d. Jong, & N. Meir, *Assessing Multilingual Children Disentangling Bilingualism from Language Impairment*, 300-330, Bristol: Multilingual Matters. Doi: https://doi.org/10.21832/9781783093137-013

Tumanova, V., Conture, E. G., Lambert, E. W., & Walden, T. A. (2014). Speech disfluencies of preschool-age children who do and do not stutter. *Journal of Communication Disorders*, *49*, 25–41. Doi: https://doi.org/10.1016/j.jcomdis.2014.01.003

Van Borsel, J., & Britto Pereira, M. M. (2005). Assessment of stuttering in a familiar versus an unfamiliar language. *Journal of Fluency Disorders*, *30*(2), 109-124. Doi: https://doi.org/10.1016/j.jfludis.2005.04.001

Van Borsel, J., Leahy, M. M., & Britto Pereira, M. M. (2008). Judging stuttering in an unfamiliar language: The importance of closeness to the native language. *Clinical Linguistics & Phonetics, 22*(1), 59-67. Doi: https://doi.org/10.1080/02699200701647289

Watson, J. B., & Kayser, H. (1994). Assessment of bilingual/bicultural children and adults who stutter. *Seminars in Speech and Language, 15*(2), 149-164. Doi: https://doi.org/10.1055/s-2008-1064140

Wei, L. (2007). *The Bilingual Reader*. (2nd ed.), Routledge. London.

Winter, K. (2016). Numbers of bilingual children in speech and language therapy: Theory and practice of measuring their representation. *International Journal of Bilingualism, 5*(4). Doi: https://doi.org/10.1177/13670069010050040401

Yaruss, J. S., Coleman, C., & Hammer, D. (2006). Treating preschool children who stutter: Description and preliminary evaluation of a family-focused treatment approach. Language, Speech, and Hearing Services in Schools, *37*, 118 – 136. Doi: https://doi.org/10.1044/0161-1461(2006/014)

Zeigler, K., & Camarota, S. A. (2019). 67.3 Million in the United States Spoke a Foreign Language at Home in 2018. Center for Immigration Studies, Washington, DC.

Table 1

Survey-based description of the SLP participants characteristics

Work experience	*M* = 3.0 years	Range = 1-11
Highest educational degree	Bachelor degree 53.1%	Master degree (obtained or enrollment) 46.9%
Number of bilingual clients who stutter/career	*M* = 8.75	Range = 1-30
Experience with preschoolers who stutter	Yes (46.8%)	No (53.2%)
Courses/workshops about stuttering	Yes (87.5%)	No (12.5%)
Courses/workshops about bilingualism	Yes (59.3%)	No (40.7%)
Specialization program in stuttering	Yes (46.8%)	No (53.2%)
Self-ratings estimating difficulties in distinguishing SLD and OD (scale 1-5; 1 = not at all; 5 = extremely difficult)	*M* = 2.2	Range = 1-5
Self-ratings of confidence in accurately diagnosing stuttering in bilingual children (scale 0-9; 0 = not confident at all; 9 = extremely confident)	*M* = 6.0	Range = 0-9
General experience	*M* = 9.06	Range = 8-12

Table 2

Linguistic profile of children based on Parents of Bilingual Children questionnaire (PaBiQ)

	CWS1	CWS2	CWNS1	CWNS2	CWNS3	CWNS4	CWNS5	CWNS6
Sex	male	male	male	male	female	male	female	male
Age	5;3	5;5	5;9	5;9	5;5	5;4	5;6	5;8
No risk index	23	23	23	21	21	23	23	23
Language dominance Arabic	40	31.25	37.5	35.5	38.5	41	23.5	31.75
Language dominance French	/	/	11.5	16.5	/	/	15	4.5
Language dominance English	19	26.5	/	/	12.25	20	25.5	10.5

Table 3

Description of the speech samples of children who stutter and who do not stutter

Disfluency type[1]	**CWS1**	**CWS2**	**CWNS1**	**CWNS2**	**CWNS3**	**CWNS4**	**CWNS5**	**CWNS6**
MonoWR	2.44	1.85	4.26	3.65	1.93	1.61	2.66	2.63
SndR	4.39	1.59	0.00	0.00	0.55	0.00	0.20	0.33
SylR	8.29	1.06	0.00	0.40	0.28	0.54	0.41	1.32
P	1.95	0.26	3.04	0.00	2.76	0.00	0.00	0.00
B	0.24	1.59	0.00	0.00	0.28	0.00	0.00	0.33
BW	0.00	0.53	0.00	0.00	0.00	0.00	0.20	0.33
MultiWR	1.71	2.38	1.21	0.00	1.38	0.00	0.82	0.00
PhR	0.49	1.06	1.82	1.62	0.55	0.00	0.61	0.00
UW/S	2.20	1.06	0.00	2.43	3.04	1.61	0.61	1.64
I	4.63	0.26	3.65	1.21	4.14	1.08	0.82	3.95
LRev	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00
Grev	0.24	0.26	0.60	0.00	0.00	0.00	0.00	0.99
PhonRev	0.49	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total disfluencies	27.07	12.71	14.58	9.31	14.92	4.84	6.35	11.51
%SLD	17.31	6.88	7.3	4.05	5.8	2.15	3.47	4.94
%OD	9.76	5.28	7.28	5.26	9.11	2.69	2.86	6.58
Ratio SLD/OD (in %)	64/36	56.6/43.4	50/50	43.6/56.4	39/61	44.5/55.5	54.8/45.2	42.9/57.1
Total number of words	410	378	164	246	362	186	488	304
Total number of syllables	691	637	233	320	660	327	650	498
Speech sample length	8’16’’	6’33’’	3’56’’	3’10’’	4’35’’	3’33’’	7’03’’	4’20’’
Stuttering severity rating	5	3	0	0	0	0	0	0

[1] SLD: Monosyllabic word repetitions (MonoWR), syllable repetitions (SylR), sound repetitions (SndR), blocks (B), broken words (BW) and prolongations (P)
OD: Multisyllable word repetitions (MultiWR), phrase repetitions (PhR), interjections (I), unfinished words/sentences (UW/S), lexical (LRev), grammatical revisions (GRev) and phonological revisions (PhonRev)

Physical concomitants	yes	yes	no	no	no	no	no	no

Table 4

Percentages of misidentification per child, on the basis of audio and video recordings

	CWS1	CWS2	CWNS1	CWNS2	CWNS3	CWNS4	CWNS5	CWNS6
%Misidentification audio	9.4	15.6	56.3	40.6	40.6	25	71.9	37.5
%Misidentification video	0	29.6	59.2	40.7	48.1	3.7	70.4	51.8

Table 5

Descriptors used by the participants to label CWNS as stuttering or non-stuttering

Correct identification of CWNS	
Exhibiting typically fluent speech with normal disfluencies	37%
Exhibiting typically fluent speech without physical concomitants (or tension coming along with the speech)	13.33%
Exhibiting typically fluent speech with word retrieval difficulties possibly associated to bilingualism	29.33%
Exhibiting word retrieval difficulties possibly because the child might be more comfortable in his second language	6.67%
Absence of physical concomitants (or tension coming along with the speech)	1.33%
Misidentification of CWNS	
Exhibiting word repetitions, part-word repetitions, prolongations and blocks	51.47%
Exhibiting word repetitions, part-word repetitions, and physical concomitants described as tensed breathing and absence of eye contact.	33.82%
Exhibiting typical disfluencies (repetitions and interjections), with 3 iterations and/or possibly word retrieval difficulties	7.35%
Exhibiting word retrieval difficulties possibly because the child might be more comfortable in his second language	1.47%
Presence of physical concomitants described as tensed breathing or absence of eye contact	5.88%

Table 6

Descriptors used by the participants to label CWS as stuttering or non-stuttering

Correct identification of CWS	
Exhibiting stuttering-like disfluencies with physical concomitants	44.4%
Exhibiting stuttering-like disfluencies (e.g., repetitions, blocks, prolongations)	37.7%
Exhibiting stuttering-like disfluencies and other disfluencies (interjections)	13.3%
Exhibiting physical concomitants (e.g., tension, moving extremities, agitation)	4.44%
Misidentification of CWS	
Exhibiting few disfluencies	57.2%
Exhibiting disfluencies possibly due to poor lexicon	28.5%
Exhibiting other disfluencies (e.g., grammatical revisions)	14.3%

Figure 1. *A depiction of the percentage of misidentified CWNS and CWS on audio-based versus video-based judgments. The error bars indicate the 95% confidence intervals.*

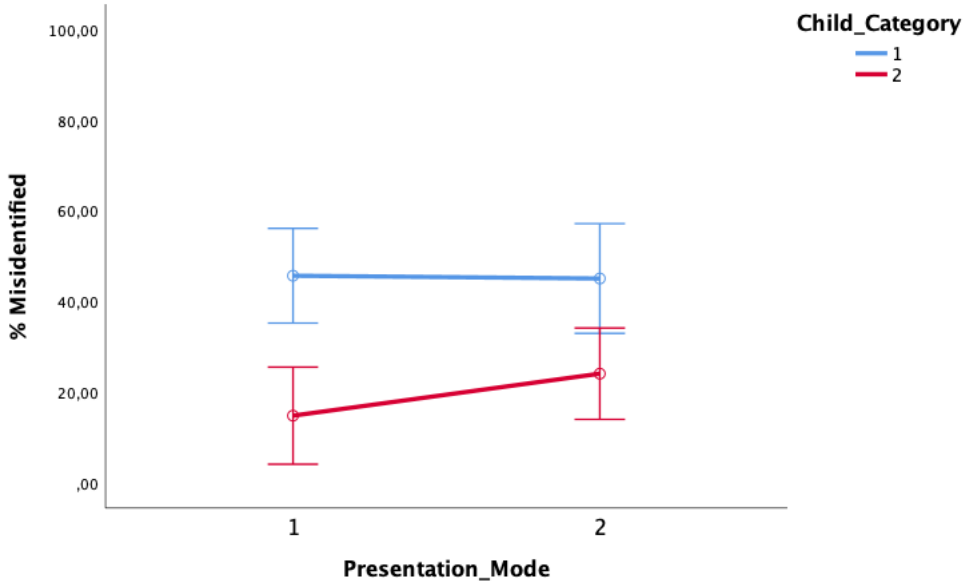

Figure 2a. *A depiction of the percentage of misidentified cases- for each individual CWNS- across audio-based and video-based judgments. The error bars indicate the 95% confidence intervals.*

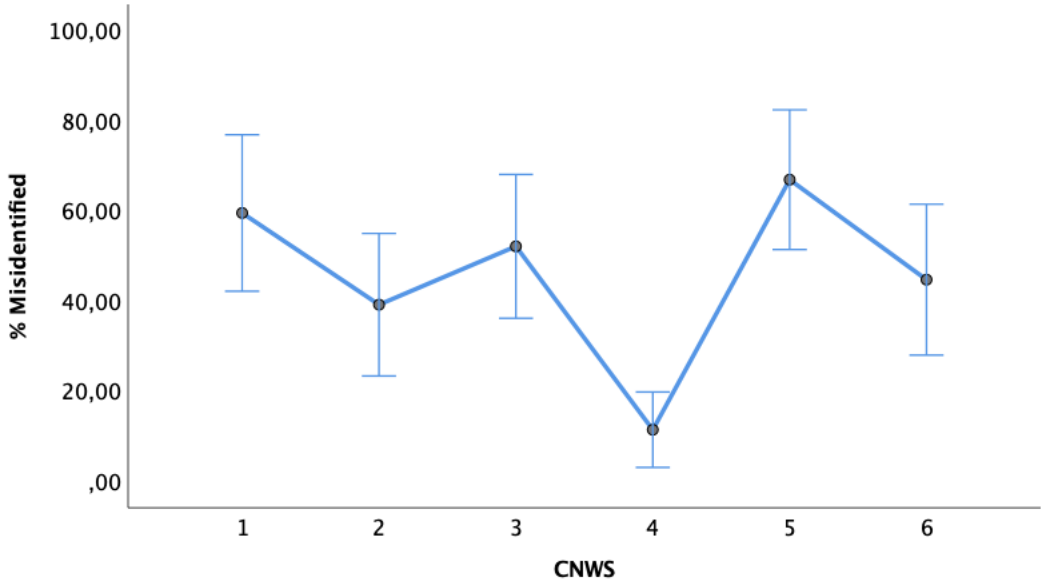

Figure 2b. *A depiction of the percentage of misidentified CWS1 and CWS2 on audio-based versus video-based judgments. The error bars indicate the 95% confidence intervals.*

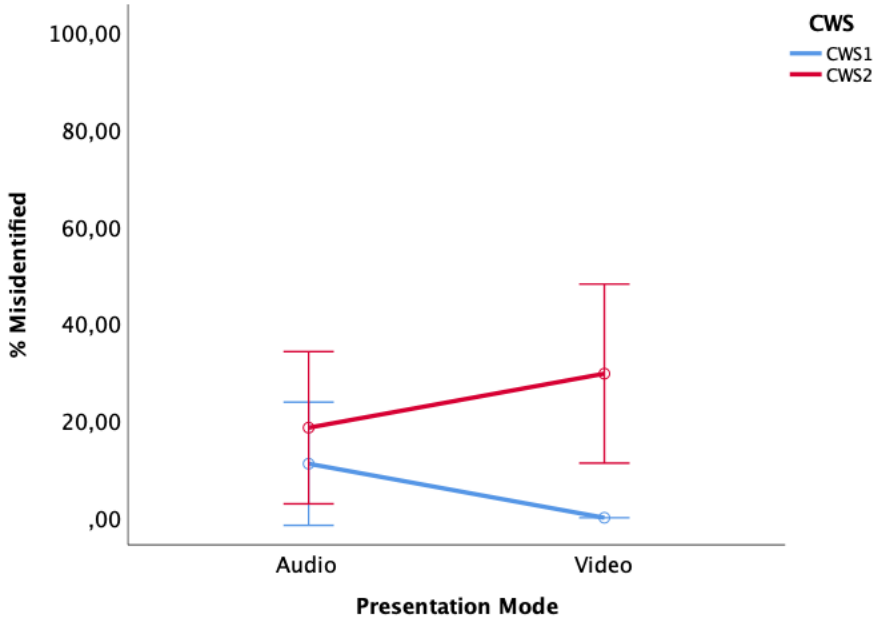

Figure 3

The relation between respectively SLD, OD, and misidentification (%)