Functional Listening Evaluation (FLE): Speech Material Effects in Children With Normal Hearing

Jeanne Dodd-Murphy, Ph.D.
Baylor University

Michaela J. Ritter, Ed.D.
Baylor University

Walter Murphy, Ph.D.
Texas A&M University-Central Texas

Functional Listening Evaluation (FLE) performance was determined for 31 typically-developing children with normal hearing (7 to 10 years of age). Investigators sought to provide comparison data to help audiologists using the FLE to justify recommendations of hearing assistance technology and/or other accommodations for special school-age populations with normal hearing. The effect of speech materials (including live-voice versus recorded presentation mode) and scoring strategy was evaluated. Each child was tested in the auditory-only conditions of the FLE (Close/Quiet, Close/Noise, Far/Quiet, Far/Noise) using three different sets of speech stimuli: Recorded FLE using [HINT-C] Sentences (RS), HINT-C sentences presented via monitored live voice (LS), and Children's Nonsense Phrases presented via monitored live voice (LNP). Mean word-level scores collapsed across listening conditions were above 97 percent for all three speech materials. LS yielded significantly higher mean performance than either RS or LNP, with no significant difference between RS and LNP means. Sentence- or phrase-level scores showed greater variability. Variability of individual scores was highest in the Far/Noise condition of the FLE. RS scores showed the highest variability among the three speech materials. Word-level scoring is recommended when conducting the FLE using any of these speech materials. In light of the high word-level scores overall for this sample, even relatively small reductions in scores could be clinically significant for 7- to 10-year-olds with normal hearing and special listening needs.

Introduction

The Functional Listening Evaluation (FLE; Johnson, 2013) is a measure of a child's ability to understand speech in a typical classroom. It was originally designed by educational audiologists to determine the effects of noise and distance on speech recognition for children with hearing impairment under conditions simulating each child's customary school listening environment. The results of the FLE, as part of a comprehensive evaluation of classroom listening needs, can be used to justify the recommendation of hearing assistance technology (HAT) for a particular child (AAA, 2011; Johnson, 2012a). In recent years, the use of classroom HAT has expanded to include children with typical hearing who may need a more favorable listening environment to learn (e.g., children with language/learning disabilities or attention deficits, dyslexia, those learning English as a second language) (see Schafer et al., 2014 for a review). Little research is available on the performance of children with normal hearing (with or without risk factors) on assessments such as the FLE which are commonly used with the hearing-impaired population; this information is needed to establish what FLE results would identify children who are likely to benefit from classroom HAT.

The extent to which school-age children exhibit reduced speech recognition in the classroom varies depending on numerous factors, including level of extraneous classroom noise relative to the teacher's voice (the signal-to-noise ratio, SNR), location of the child relative to the teacher, amount of reverberation (measured in reverberation time, RT), and difficulty of the speech task. Lower signal-to-noise ratios, greater distance between the teacher and child, longer reverberation times, and listening to speech with reduced syntactic or semantic cues would all be associated with poorer classroom speech recognition. Even typically-developing children with normal hearing have been shown to have difficulty under adverse listening conditions in either actual or simulated classroom environments (Bradley & Sato, 2008; Iglehart, 2016; Lewis, Hoover, Choi, & Stelmachowicz, 2010; Neuman, Wróblewski, Hajicek, & Rubinstein, 2010; Ruscetta, Arjmand, & Pratt, 2005; Valente, Plevinsky, Franco, Heinrichs-Graham, & Lewis, 2012; Wolfe et al., 2013; Wróblewski, Lewis, Valente, & Stelmachowicz, 2012).

The FLE is a flexible clinical protocol that guides professionals in the systematic evaluation of a child's speech recognition abilities across differing listening conditions by varying the presence of noise (Quiet versus Noise), speaker-to-listener distance (Close versus Far), and access to visual speech cues (Auditory-Visual versus Auditory only). The unaided FLE can be used as a pre-intervention measure to evaluate educational needs for children with listening difficulties; the FLE can also be administered with HAT to demonstrate benefit. Though a selection of speech materials is recommended in the FLE guidelines, the choice of speech stimulus is left to the examiner based on the age, developmental level, and other abilities of the child. The summary form includes a scorebox into which scores are entered, then automatically placed into an interpretation matrix where averaged scores from particular conditions can be compared to estimate the impact that noise, distance, and/or lack of access to visual cues have on speech recognition in the classroom (Johnson, 2013).

The FLE is administered in the child's classroom or a comparable environment, providing a more authentic representation of the child's daily receptive communication abilities than speech recognition testing performed in a sound booth.

Though originally developed for children with hearing loss, Dodd-Murphy and Ritter (2012) suggested the FLE could be useful in determining classroom listening needs of children at risk for academic delays due to factors other than hearing loss; these researchers administered the FLE to normal-hearing children with language and reading impairments, recommending the use of nonsense phrases to increase sensitivity. Normative data for the FLE would be invaluable to educational speech-language pathologists, audiologists, and to other professionals who assess classroom listening performance to provide evidence for educational need of HAT. The FLE provides quantifiable behavioral data that may carry greater weight when meeting eligibility standards or requesting special service provision from a school district.

Besides providing comparison data for evaluating children with normal hearing, the authors were interested in exploring the FLE performance of typically-developing children on multiple speech materials. Multi-word materials are more similar to the running speech that students listen to in the classroom, and each item is long enough in duration to evaluate the effects of reverberation, which is important for determining the need for HAT. The Recorded FLE Using Sentences (Johnson & Anderson, 2013) was recently made available online. The original version of the FLE specified the presentation of materials by monitored live voice, and instructions for live-voice presentation remain in the latest version. Monitored live voice presentation has been shown to increase both mean performance (Uhler, Biever, & Gifford, 2016) and the variability of scores in speech recognition tasks (Brandy, 1966). The use of live voice presentation in audiological speech recognition assessment has been criticized for decreasing its reliability and complicating both the intra- and inter-individual comparison of recognition scores (Hillock-Dunn, 2015); however, educational audiologists may continue to use live-voice presentation as part of the FLE protocol because of ease of administration and/or a sense that live speech has ecological validity in the school setting. Therefore, the current study compared FLE results for the Recorded FLE with live-voice presentation from the same set of sentence lists.

In addition, simple meaningful sentences had been found to be relatively easy for elementary-school-aged children with reading impairments to identify even in noise and with distance; nonsense phrases were suggested as an alternative because they were considered to offer a more difficult task because of the reduced linguistic content (Dodd-Murphy & Ritter, 2012, 2013). Other researchers have shown larger differences between the perception in noise of sentence and nonsense materials in children within the age group of interest. Ruscetta et al. (2005) found that children with and without unilateral hearing loss had significantly lower scores on the Nonsense Syllable Test (Edgerton & Danhauer, 1979) than on the HINT-C sentences (Nilsson, Soli, & Sullivan, 1994) while listening in the sound field in the presence of competing multi-talker babble. Lewis and colleagues (2010) showed a similar trend for 7-year-olds with normal hearing on a recording of nonsense syllables compared to the Bamford-Kowal-Bench sentences (BKB; Bench, Kowal, & Bamford, 1979) across a range of SNRs. Stelmachowicz et al. (2000) showed a small but consistent effect of semantic context in simple sentence recognition for both adults and children with normal hearing that was most evident under the poorest acoustic conditions. Because the Children's Nonsense Phrases (Johnson, 2012b) had been recommended specifically for use with the FLE as more challenging than meaningful sentences for children with minimal or unilateral losses (Johnson & Anderson, 2013), the current study compared FLE results for meaningful sentences and nonsense phrases.

Finally, Dodd-Murphy & Ritter (2012) showed that sentence-level scoring increased the variability and the sensitivity of the FLE to classroom listening difficulties exhibited by typically-hearing children with language and reading impairments. The current study therefore explored the effects of two scoring strategies: word-level and sentence- level.

The current study is the first that documents the FLE performance of children with normal hearing who are typically developing. The study sought to evaluate the following hypotheses: 1) children with typical hearing and development will show near-ideal speech recognition performance on the FLE using simple, meaningful sentences; 2) children with typical hearing and development will produce higher scores and increased variability on live-voice presentation of the FLE when compared to the recorded FLE; 3) children with typical hearing and development will show greater difficulty on the FLE using Children's Nonsense Phrases than on the FLE using meaningful sentences; and 4) sentence-level scoring will generate lower scores and increased variability when compared to word-level scoring on the FLE.

Methods

Participants

Upon Institutional Review Board approval, participants were recruited from a local elementary school. Recruitment activities included an explanation of study objectives and procedures to the principal of the school and the distribution of flyers stating the general purpose of the study, the participation criteria, and contact information.

In order to participate in this study, the children were required to meet the following criteria: 1) have an age between 7 years and 10 years, 11 months, 2) have English as their first language, 3) have no history of special educational services at school or private therapy, no history of developmental delay, and no history of hearing loss. Informed parental consent was required. Each child also signed to indicate assent at the time of testing and received a payment in cash upon completion of the testing.

Data were collected for 31 children with a mean age of 8 years, 11 months. One child over the age of 11 was also tested, but data from that session were not included in the analyses. The participants included 19 male and 12 female students. All participants passed a pure-tone hearing screening at 20 dB HL at 1000, 2000, and 4000 Hz in each ear using a portable audiometer (Maico MA40).

Materials

Three different sets of speech stimuli, all recommended in the FLE instructions, were used to determine scores under the four auditory-only FLE listening conditions. Auditory-visual conditions were not administered to reduce both test time and the likelihood of participant fatigue. The three speech materials were, in order of presentation: 1) the Recorded FLE Using Sentences [RS] (Johnson & Anderson, 2013); 2) HINT-C sentences presented via monitored live voice [LS]; and 3) Children's Nonsense Phrases presented via monitored live voice [LNP] (Johnson, 2012b). The Recorded FLE consists of a custom recording of a female speaker presenting Hearing in Noise Test for Children sentences (HINT-C; Nilsson et al., 1994). The HINT-C sentence materials, based on the original Bamford-Kowal-Bench sentences (Bench et al., 1979), have eight different but equivalent lists of ten simple sentences; each sentence contains five target words, allowing for the option of word-level or sentence-level scoring. For consistency, the examiners also read from the HINT-C sentences in the live voice presentations of meaningful sentences, using the four lists that were not used for the Recorded FLE. In addition, the mp3 file of ten minutes of continuous classroom noise included with the Recorded FLE was used for all live-voice conditions presented with noise. The Children's Nonsense Phrases, available with the FLE protocol, have eight lists of twenty phrases each and can be scored at either the word or phrase level (Johnson, 2013). The first four lists of the Children's Nonsense Phrases were used in the current study.

Procedure

All testing was conducted in an unoccupied room on site at the elementary school from which participants were recruited. Three undergraduate researchers (senior Communication Sciences & Disorders majors) administered the FLE. The first two authors trained the student examiners and periodically supervised the testing. Children for whom parental permission was received were tested individually during scheduled school days, with two examiners working together at one time. During each session, one examiner served as the speaker for the live-voice presentations, while the other examiner marked and scored the child's responses. The examiners alternated roles for each successive child they tested on a particular day.

The most recent version of the FLE was used; the set-up and test process are described in detail in a document available at this link: http://adevantage.com/uploads/FLE_2013v2a-saveable_autocalculable.pdf. Each participant was asked to repeat sentences or phrases under four different listening conditions presented in the following sequence: Close/Quiet, Close/Noise, Far/Noise, Far/Quiet for each of the three speech materials (see Materials above for descriptions). All of the live-voice conditions used were 'auditory only'; that is, the view of the examiner's face was prevented using a dark screen (loudspeaker cover material held in place by an embroidery hoop), so that visual cues were not available to the child, but undistorted auditory information was available. Each child sat in a desk and wore a lapel microphone connected to a digital recorder; a sound file of the entire test session was saved for each participant. Instructions were given before the beginning of testing with each speech material. Test items were only presented once, and children were instructed to repeat the entire sentence or phrase exactly as it was spoken.

All stimuli for the recorded FLE were played on a laptop computer set on top of a table located with the speaker three feet from the child (Close conditions). The laptop loudspeaker volume was adjusted while playing practice sentences without noise until speech was measured at 65 dB SPL at the child's near ear using a sound level meter application on an ipad or iphone (used SPL Meter for ipad (designer Adam Smith); 711RA RMS SPL Meter, A weighting, slow setting). This volume was then held constant for all four listening conditions. In the Far conditions, the computer was moved to a cabinet located at a distance of 15 feet from the child. During the conditions with noise, the designated Recorded FLE files played sentences mixed with classroom noise at a signal-to-noise ratio of + 5 dB.

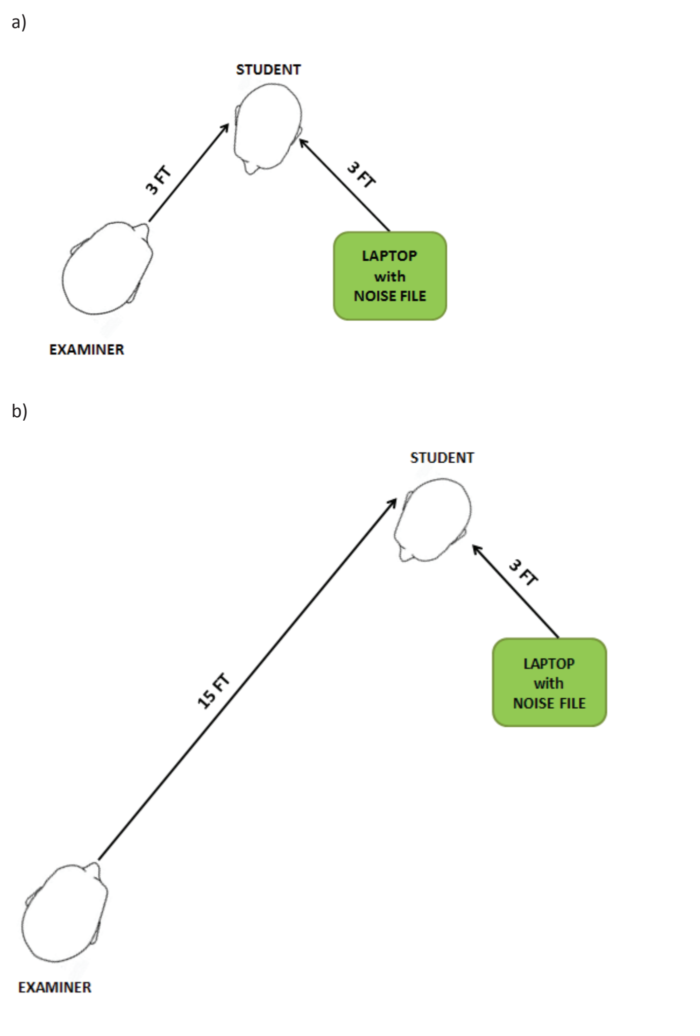

Figure 1. FLE set up for live-voice conditions: a) close condition; b) far condition

For all live-voice conditions with noise, a continuous digital recording of classroom noise was played on the laptop computer on a table set at approximately a 45 degree angle at three feet away from the child's desk. Prior to presentation of the live-voice HINT-C sentences and Children's Nonsense Phrases, the level of the computer speaker was readjusted so that the classroom noise was measured at 60 dB SPL at the child's near ear. The examiner presenting the sentences would then stand three feet away from the child and adjust his or her voice till the level of the practice sentences averaged 65 dB SPL at the child's near ear (see Figure 1 for the FLE set-up for the live-voice testing). At the same time, the research partner would use a second ipad or iphone with the same sound level meter application to determine the approximate dB SPL of the examiner's voice at a distance of one foot from his or her mouth. Then, for all subsequent conditions, the examiner presenting the sentences or nonsense phrases kept his or her voice level as constant as possible using an ipad or iphone located one foot from his or her mouth. The sound level meter applications on the two ipads and one iphone used in the study had been verified to measure dB SPL within one dB of each other. One ipad with the sound level meter device had previously been verified with a type I sound level meter to have accurate dBA SPL measurement above 30 dBA SPL. As with the recorded FLE, the Far conditions were presented (this time by the examiner) from a distance of 15 feet away from the child, shown in Figure 1b.

Analysis

Both the key word level and the sentence/phrase level were analyzed. Scores were computed for each participant based on the percentage of target words and on the percentage of whole sentences or phrases that were correctly identified for each condition, generating a total of 24 scores for the twelve lists. All participants were clearly intelligible. One child had a consistent articulatory problem with /r/; this child's articulation errors were treated so as not to influence the scoring. For example, if the child said /ʃɑp/ for 'sharp', the word was counted correct. If any child repeated the words out of order, the phrase or sentence was counted wrong, but the words were counted as correct. For the HINT-C sentences, the scoring forms indicated that certain words were interchangeable; when scoring sentences, use of either word would be counted as correct. For example, if the recorded voice said 'the' where the form listed 'a/the', the scorer would count the sentence as correct if the child repeated all other words exactly and said 'a' instead of 'the'. Phrases or sentences were considered incorrect if a child inserted a word that wasn't present in the original but otherwise said each word correctly (in that case, words would have been counted as correct). Finally, expanding a contraction (e.g., saying 'she is' instead of 'she's') rendered a sentence or phrase incorrect. When calculating percentage scores, any decimals were rounded to the tenths place.

Mean speech recognition scores for the sample were determined for each FLE listening condition and mean scores overall using the three speech materials were compared statistically. In every comparison involving the Children's Nonsense Phrases (both word and sentence/phrase level score), arcsine transformations of all scores were compared due to the variations in the number of items between the HINT-C sentence lists and the nonsense phrase lists.

The FLE includes an interpretation matrix which analyzes the effects of noise and distance on the child's speech recognition ability. Individual noise and distance effects were determined by calculating the difference between each child's average scores for quiet versus noise conditions and for close versus far conditions, respectively. Mean noise and distance effects were also determined for the sample for each stimulus type. Inter-rater reliability of scoring was also measured by having an experienced graduate student in speech-language pathology listen to approximately half of the recorded sessions and assign both word and sentence level scores for each condition. These scores were then compared to the scores of the original examiners for the same children.

Results

Inter-rater Reliability of Scoring

The speech recognition scores determined by an independent rater were highly correlated with the scores computed by the original examiners. Spearman correlations were similar for both word and sentence level scoring ($r = .775$, $p < .01$, and $r = .771$, $p < .01$, respectively). All data in the current report represent the original scoring.

Percent Correct Word Recognition Scores

Participants showed high word-level scores across speech materials and listening conditions. Mean word recognition scores for the four listening conditions by the three speech materials are shown in Table 1. Mean key-word scores collapsed across listening conditions were above 97 percent for all three speech materials, as indicated in Table 2. Individual scores ranged from 86 to 100 percent. There were only six scores (from six different participants) below 90% on the recorded FLE (RS) in one of the Far conditions; all other scores (366 of 372) were at or above 90%. Because scores were high and similar across listening conditions, group means showed little to no noise or distance effect. Mean noise effects ranged from 0.45% (LS) to 1.3% (RS), while distance effects ranged from 0.1% (LS) to 2.3% (RS).

Table 1

Means/Standard Deviations for Percentages Correct for Listening Condition by Speech Material for Word-Level and Sentence-Level Scoring Strategies

	Listening condition			
	Close/Quiet	Close/Noise	Far/Quiet	Far/Noise
Word level				
Speech material				
LS	99.16/1.34	99.15/0.85	99.87/0.50	98.65/1.82
RS	98.84/1.13	98.71/2.22	97.74/3.45	95.16/4.58
LNP	99.20/1.29	98.28/1.86	97.21/2.50	96.84/2.66
Sentence level				
Speech material				
LS	95.81/6.72	97.74/4.25	99.35/2.50	93.55/8.39
RS	94.84/5.08	95.16/6.26	93.87/8.82	83.87/13.83
LNP	97.26/4.05	92.58/6.69	89.68/8.65	89.35/8.54

Note. LS = Live Voice Sentences; RS = Recorded Sentences; LNP = Live Voice Nonsense Phrases.

Table 2

Means/Standard Deviations for Percentages Correct by Scoring Strategy and Speech Material

	Word level	Sentence level
Speech material		
LS	99.31/1.3	96.61/6.23
RS	97.61/3.43	91.64/10.18
LNP	97.88/2.32	92.22/7.82
Overall	98.27/2.61	93.59/8.49

Note. LS = Live Voice Sentences; RS = Recorded Sentences; LNP = Live Voice Nonsense Phrases.

A repeated measures ANOVA showed a significant main effect of the speech material used on the mean word recognition score for all listening conditions, $F(2,246) = 27.88$, $p<.01$. Post hoc pairwise comparisons using the Bonferroni correction revealed that the mean score for the sentences presented by live voice (LS) was significantly higher than the mean score for the recorded FLE (RS) ($p < .01$) and that the LS mean was also significantly higher than the mean for the live-voice nonsense phrases (LNP) ($p < .01$). There was no significant difference between means for RS and LNP ($p >.05$).

Variability

Overall, scores at the word level showed much less variability than scores at the sentence level. Though mean scores for sentence or phrase level scoring were above 90% (i.e., less than 10% reduction relative to key word scoring) for all three speech materials, individual scores ranged from 50 to 100 percent. Each of the speech materials yielded individual scores less than 90% (a total of 54 scores from 24 different participants).

For both word level and sentence level analyses, variability in scores was greatest for recorded sentences (RS) and least for live-voice sentences (LS), indicated by the standard deviations in Table 1. For all three speech materials, the highest variability of both word and sentence level scores was demonstrated in the Far/Noise listening condition. Figure 2 displays scatterplots of individual scores in the Far/Noise condition by age for each of the three speech materials. These graphs illustrate the much higher variability for sentence or phrase level scoring than for key word scoring and the tighter distribution of scores for the live-voice sentences (LS) when compared with either the recorded sentences (RS) or the live-voice nonsense phrases (LNP).

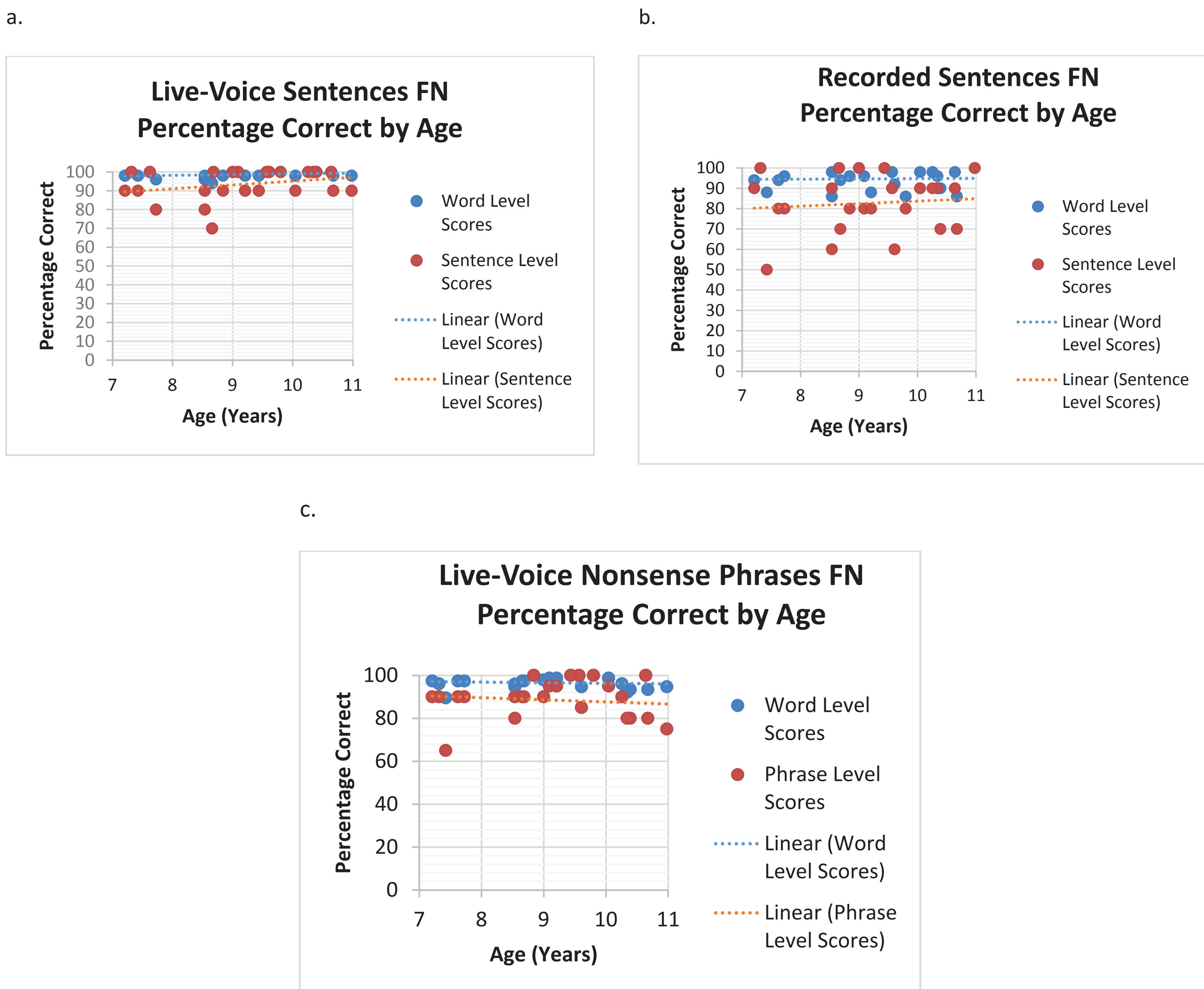

Figure 2. Individual scores by age for the FLE Far/Noise condition for the three speech stimulus types. Dashed lines indicate a line of best fit for each set of scores.

Discussion

The current study is the first that documents the FLE performance of children with normal hearing who are typically developing. We sought to provide comparison data to help educational audiologists using the FLE to justify recommendations of classroom HAT and/or other accommodations for special school-age populations with normal hearing. Establishing criteria that indicate reduced access to speech for auditory learning is particularly important when evaluating children with normal hearing sensitivity because they usually are not expected to need auditory-based interventions. We also intended to demonstrate how choice of speech material may affect FLE performance. To that end, we compared FLE scores in auditory-only conditions using three types of materials in a group of children between 7 and 10 years of age. Though monosyllabic word lists can be used for the FLE, phrase or sentence level materials are more similar to the speech children listen for in classroom settings, and their longer duration may allow a more valid measure of classroom reverberation effects on speech recognition.

Consequently, we chose to conduct the FLE with phrase and sentence materials. The Recorded FLE using Sentences (RS condition: Johnson & Anderson, 2013), recently made available online, has been presented as a convenient way to administer the FLE. Our study compared the recorded version to live-voice presentation of the same sentences (LS). In addition, the same group of children were administered the FLE using the Children's Nonsense Phrases presented by live voice (LNP) to assess whether recognition scores would be reduced with less linguistically predictable material.

Word-level scores were high across the materials and listening conditions when conducting the FLE using either HINT-C sentences (RS or LS) or Children's Nonsense Phrases (LNP), as shown in Tables 1 and 2. Overall mean scores for each type of speech stimulus were above 97 percent. Only six participants scored below 90% in any condition; in all six cases, scores of 86 or 88% were only observed for the Recorded FLE in one of the Far conditions. Otherwise, all scores were at or above 90%. This high level of word recognition performance is consistent with results from Dodd-Murphy and Ritter (2012), who investigated the FLE in elementary school age children with language and reading impairments and typical hearing. Using the BKB-SAE sentences (Bench et al., 1979; Kenworthy, Klee, & Tharpe, 1990) presented via monitored live voice with recorded multi-talker babble as competing noise, we found that means ranged from 96.3 to 98.1 percent in the auditory-only conditions for a sub-group of the sample who were rated by parents to have no significant auditory problems.

The live voice FLE (LS and LNP) presentations in this study were set up to approximate a +5 dB SNR in the Close/Noise conditions and as low as -5 dB SNR in the Far/Noise conditions. Studies from researchers associated with Boys Town National Research Hospital used fixed level SNRs under headphones or in the sound field to study speech recognition of children using recorded versions of the BKB sentences. Lewis and co-workers (Lewis et al., 2010) showed mean scores above 90 percent in a group of 7-year-olds on the BKB-SAEs mixed with speech shaped noise at a +5 dB signal-to-noise ratio under headphones even when scoring sentences correctly only if all three key words were accurate. Another group investigated reverberation effects on key word recognition masked by speech spectrum-shaped noise that was either stationary (i.e., constant in amplitude) or amplitude-modulated in children aged 7 to 14 years old across SNRs from -10 to +10 dB (Wróblewski et al., 2012). The 65 dB SPL speech signal was mixed with noise and with simulated reverberation effects and presented through earphones. Ceiling effects were demonstrated for all participants at +5 and +10 dB SNRs in all conditions, consistent with the results of the current study in the Close/Noise conditions. In the two-meter reverberant condition at -5 dB SNR with the modulated masker (conditions closest to the Far/Noise conditions in the FLE), mean speech recognition score for 7- to 8-year-olds dropped below 80 % and the mean for 9- to 10-year-olds decreased below 90%, though means for 9- to 10-year-olds were not significantly different from that of adults in the same condition (Wróblewski et al., 2012). In comparison, participants in the current study continued to score above 95% on the average; live voice conditions allowed for spatial separation of the speech and the noise, whereas the Wróblewski et al. investigation used a less advantageous spatial orientation of co-located speech and noise. This, however, does not explain why scores on the Recorded FLE remained high in the current group of children. The characteristics of the classroom noise for the Recorded FLE were not specified; however, perceptually, the intensity level of the noise (real talkers interacting in a classroom) varied frequently and to a significant extent, which likely allowed the children opportunities to receive speech cues in the gaps, resulting in improved recognition performance (Griffin, 2015; Stuart, 2005, 2008). Reverberation in the room used in this study may also have been less pronounced, though reverberation time was not measured. The acoustic treatments in Wróblewski et al. simulated reverberation effects at 2 meters (about 6.5 feet) or at 6 meters (almost 20 feet), both distances longer than used in this study.

Wolfe et al. (2013) used key word scoring of recorded HINT-C sentences to evaluate children with normal hearing in a classroom at a variety of fixed SNRs using recorded 4-classroom noise. As expected, children showed near-ideal word recognition in quiet. Mean word recognition dropped slightly below 90% at +4 dB SNR (condition most similar to FLE Close/Noise), with relatively high variability. As SNR dropped to -1 and -6 dB, normal-hearing scores decreased to about 60% and below 20%, respectively. The better performance at negative SNRs shown by the children in the present study is likely related to both the differences in the acoustic properties of the noise and the number of noise sources. Wolfe et al. used four loudspeakers in the corners of the room to present the noise, while the FLE uses only one noise source.

It is more difficult to compare the current study results with those of previous investigations using adaptive or quasi-adaptive procedures; however, in two reports, authors estimated performance functions by SNR or calculated the SNR that would be associated with 95% performance for participant groups, enabling us to evaluate similarities in our findings (Iglehart, 2016; Neuman et al., 2010). Speech recognition scores in noise from the current study were generally higher than predicted by Neuman and colleagues, particularly for the least advantageous FLE condition (Far/Noise, approximating -5 dB SNR). Neuman et al. (2010) used the BKB-SIN with multi-talker babble serving as competition to determine the combined effects of noise and reverberation on speech recognition in 6 to 12-year-old children and adults. As with Wróblewski et al. (2012), speech and noise were co-located and presented binaurally under headphones at several different RTs. The stimuli were designed to simulate the acoustic experience for a child sitting in the back of the classroom at 5.5 m (over 18 feet) from a teacher who is producing a speech level of 70 dB SPL. SNR thresholds for BKB sentences associated with both 50% and 95% performance were calculated. In a graph estimating the performance by SNR function for selected participant groups (Figure 3, p. 342), 9-year-olds' performance was predicted to be within the 65 to 75% range for a SNR of +6 (most similar to conditions for the FLE Close/Noise conditions) while at -6 dB SNR, 9-year-olds' performance was predicted to show floor effects even in the lowest reverberation condition. Eight-year-olds required +11 to +12 dB SNR for 95% performance, while our sample of 7- to-10-year-olds showed mean performance greater than 95% even in a negative SNR.

Performance of children at +5 dB SNR in the current study was more similar to results from Iglehart (2016), who measured speech recognition in actual school classrooms under a variety of SNR and RT conditions using an adapted BKB-SIN procedure for a group of 20-23 children with typical hearing ranging from 5.2 to 16.6 years of age (M=11.1). Iglehart reported mean performance of nearly 95% at +6 dB SNR under both 0.3 and 0.6 RT conditions, comparable to mean scores of 98 to 99% of the present participants in the FLE Close/Noise condition (see Table 1). Iglehart found mean scores for the -6 dB SNR conditions below 30%, though the variability was quite high; while the least advantageous condition of the FLE (Far/Noise) in this study yielded mean scores above 95% for all three speech materials, with relatively low variability. Children in the Iglehart study listened in the sound field, facing the speech signal, with four loudspeakers in the corners of the room generating the four-talker babble from the recorded BKB-SIN. Differences in the acoustic properties of the noise and the number of noise sources probably explain the better performance by children in the current study.

The Recorded FLE Using Sentences (Johnson & Anderson, 2013) is presented as a convenient and standardized alternative to the commonly used live-voice FLE. In this study, live-voice presentation of HINT-C sentences (LS) yielded scores that were slightly but significantly higher than those for the Recorded FLE (RS). Surprisingly, the live-voice presentation of HINT-C sentences produced much lower variability than the Recorded FLE, regardless of whether word or sentence/phrase level scoring was employed. Even with three different speakers in this study, two female and one male, the FLE (LS) using live-voice meaningful sentences showed the highest means and lowest standard deviations of the three types of speech materials, as shown in Table 1. Uhler et al. (2016) reported word/sentence recognition scores with lower standard deviation values for live-voice than for recorded presentation in the sound field in children with hearing impairment in the best aided condition for each child. We chose to present the Recorded FLE using the digitized sound files with the speech and noise pre-mixed because it was the most expedient arrangement and did not require the use of a separate device to present the noise (all sentence lists are also provided in quiet so the examiner may separate the speech and noise sources spatially). Having the speech and noise coming from the same source may have increased the difficulty of the Recorded FLE task for the children based on reported spatial release from masking advantages in children (Cameron, Dillon, & Newall, 2006; Griffin, 2015; Johnstone & Litovsky, 2006). Additionally, the examiners commented that the 5-second interval between sentences on the recording may have introduced variability because some children may have needed more time to respond. The examiner could have monitored the timing, pausing when necessary; however, this reduces the 'press and play' convenience of the recorded test. In this case, then, the greater variability generated by the recorded test is likely related to the difficulty of the task.

The Children's Nonsense Phrases lists are available online and recommended for use with the FLE protocol, particularly for young children with mild/unilateral loss or children with normal hearing who may require classroom listening assessment. There are eight lists, an advantage in the FLE protocol; however, there is little published information about their development, particularly the equivalency of the lists in intelligibility. Children in this study produced significantly lower mean scores on the FLE conditions on Children's Nonsense Phrases (LNP) than on HINT-C sentences (LS) when both were presented via live voice. Though the mean score for nonsense phrases was statistically lower than the mean for simple sentences, the effect size was small, consistent with the findings of Stelmachowicz et al. (2000), who reported average context effects of less than ten percent for 8- and 10-year-old children. Children's Nonsense Phrases items consist of short word sequences without syntactic context, but children would still be able to take advantage of phonotactic probability cues using their experience of phoneme combinations that occur in English (McCreery & Stelmachowicz, 2011). A nonsense syllable task might be more difficult and provide greater contrasts between listening conditions or individuals.

While sentence-level scoring is considered more rigorous, in the current study, using a sentence/phrase scoring strategy yielded scores that were too variable to provide normative reference data. Variability was highest in the Far/Noise condition regardless of the speech material. Figure 2 consists of scatterplots of individual speech recognition scores by age in the Far/Noise condition with scoring strategy as the parameter for each speech material. The recorded version of the FLE (RS) generated higher variability than either of the live voice conditions (LS, LNP) regardless of scoring strategy. The figure illustrates the reduced variability in scores when word-level scoring is used. One advantage of word-level scoring for sentence or phrase materials is that it increases the number of items, which decreases variability. Thus, when administering the FLE using the materials tested in this study, word-level scoring is recommended over sentence-level scoring.

The FLE was designed to compare speech recognition across a variety of conditions simulating realistic classroom listening demands for an individual child, rather than to measure the child's performance relative to normative data. School age children with special listening needs who have normal or near normal hearing, though, may show relatively subtle listening deficits even when they may benefit from HAT or classroom accommodations. This population offers a particular challenge to audiologists for clinical decision-making and providing evidence of educational need. Based on the current results, relatively small differences in scores could be clinically significant for children with typical hearing, indicating difficulties with noise or distance that are outside of normal limits. This pattern was most evident using key word scoring of the live-voice HINT-C sentences (LS), where the noise effect (.45%) and the distance effect (.1%) were practically non-existent when comparing mean data. Therefore, a child showing a noise or distance effect of greater than 5% on the FLE potentially could be at risk for listening difficulties that would reduce her access to spoken language in the classroom.

Certainly, the FLE would be used as only one part of a comprehensive evaluation of classroom listening. Schafer et al. (2014) proposed components and a process for assessing the need for remote-microphone HAT for children with typical hearing who show atypical auditory processing relative to peers. The FLE can provide information useful for the classroom acoustics and observation components of the recommended process. Its interpretation matrix is a useful visual aid in making decisions about classroom placements or communication strategies, counseling children and families, and educating teachers (Gustafson, Hicks, & Lau, 2016). Speech recognition in noise measures are also an important part of this process. Though Schafer et al. (2014) favor the use of the BKB-SIN for this purpose for school age children (see also Schafer, 2010 for a detailed review of specific tests), it is unclear how predictive the SNR threshold it generates is of supra-threshold sentence recognition performance in children across the range of SNRs in typical classroom settings, and it is not appropriate for children younger than elementary school age. Many educational audiologists have had extensive experience with the FLE and value its flexibility. For those professionals who regularly use the FLE as part of their practice, this study has documented that for elementary-school-aged children with normal hearing, word-level scoring generated less variability for sentence recognition, and that the recorded FLE using the current parameters yielded slightly reduced scores when compared to a live voice presentation of sentences from the same set of lists. Knowing that children with normal hearing and typical development have uniformly high scores on the FLE helps strengthen rationales for the provision of HAT and/or other accommodations for 7- to 10-year olds with normal hearing and relatively poor auditory function in the classroom.

References

American Academy of Audiology (2011, April). *Clinical practice guidelines: Remote microphone hearing assistance technologies for children and youth from birth to 21 years (Includes Supplement A).* Retrieved from http://audiology-web.s3.amazonaws.com/migrated/HAT_Guidelines_Supplement_A.pdf_53996ef7758497.54419000.pdf

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology, 13*(3), 108-112.

Bradley, J. S. & Sato, H. (2008). The intelligibility of speech in elementary school classrooms. *Journal of the Acoustical Society of America, 123*(4), 2078-2086.

Brandy, W. T. (1966). Reliability of voice tests of speech discrimination. *Journal of Speech, Language, and Hearing Research, 9*(3), 461-465.

Cameron, S., Dillon, H., & Newall, P. (2006). The listening in spatialized noise test: Normative data for children. *International Journal of Audiology, 45*, 99-108.

Dodd-Murphy, J. & Ritter, M. (2012). Functional Listening Evaluation performance in children with reading difficulties. *Journal of Educational Audiology, 18*, 61-73.

Dodd-Murphy, J. & Ritter, M. (2013). Interpretation of Functional Listening Evaluation results of normal-hearing children with reading difficulties. *Journal of Educational Audiology, 19*, 80-85.

Edgerton, B. J., & Danhauer, J. L. (1979). *Clinical implications of speech discrimination testing using nonsense stimuli*. Baltimore, MD: University Park Press.

Griffin, A. M. (2015). Recognition and comprehension of speech in noise in school-aged children with unilateral hearing loss. Retrieved from http://scholarworks.umass.edu/dissertations_2 (Paper 358)

Gustafson, T. J. S., Hicks, C. B., & Lau, M. (2016). The Functional Listening Evaluation: Exploring how audiologists implement the FLE for classroom enhancement. Retrieved from http://speech-language-pathology-audiology.advanceweb.com/Features/Articles/The-Functional-Listening-Evaluation.aspx

Hillock-Dunn, A. (2015, September). 20Q: Pediatric speech recognition measures – What's now and what's next. Retrieved from http://www.audiologyonline.com/articles/20q-pediatric-speech-recognition-measures-15296

Iglehart, F. (2016). Speech perception in classroom acoustics by children with cochlear implants and with typical hearing. *American Journal of Audiology*, 1–10.

Johnson, C. D. (2012a). Classroom listening assessment: Strategies for speech-language pathologists. *Seminars in Speech and Language, 33*(4), 322-339.

Johnson, C. D. (2012b). Common Children's Phrases and Nonsense Phrases. Available from www.ADEvantage.com

Johnson, C. D. (2013). Functional Listening Evaluation. (Revised, based on C. D. Johnson & P. Von Almen, 1993). Retrieved from http://adevantage.com/uploads/FLE_2013v2a-saveable_autocalculable.pdf

Johnson, C. D. & Anderson, K. L. (2013). The Recorded FLE with Sentences. Retrieved from http://successforkidswithhearingloss.com/fle-recorded

Johnstone, P.M., & Litovsky, R.Y. (2006). Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults. *Journal of the Acoustical Society of America, 120*(4), 2177-2189.

Kenworthy, O.T., Klee, T., & Tharpe, A.M. (1990). Speech recognition ability of children with unilateral sensorineural hearing loss as a function of amplification, speech stimuli and listening condition. *Ear and Hearing, 11*(4), 264-270.

Lewis, D., Hoover, B., Choi, S., & Stelmachowicz, P. (2010). Relationship between speech perception in noise and phonological awareness skills for children with normal hearing. *Ear and Hearing, 31*(6), 761-768.

McCreery, R. W., & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *The Journal of the Acoustical Society of America, 130*(6), 4070-4081.

Neuman, A. C., Wróblewski, M., Hajicek, J., & Rubinstein, A. (2010). Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear and Hearing, 31*(3), 336-344.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95*(2), 1085-1099.

Ruscetta, M. N., Arjmand, E. M., & Pratt, S. R. (2005). Speech recognition abilities in noise for children with severe-to-profound unilateral hearing impairment. *International Journal of Pediatric Otorhinolaryngology, 69*(6), 771-779.

Schafer, E. (2010). Speech perception in noise measures for children: A critical review and case studies. *Journal of Educational Audiology, 16*, 4-15.

Schafer, E. C., Anderson, C., Dyson, J., Wright, S., Sanders, K., & Bryant, D. (2014). A critical review of remote-microphone technology for children with normal hearing and auditory differences. *Journal of Educational Audiology, 20*, 3-13.

Stelmachowicz, P. G., Hoover, B. M., Lewis, D. E., Kortekaas, R. W. L., & Pittman, A. L. (2000). The relation between stimulus context, speech audibility, and perception for normal-hearing and hearing-impaired children. *Journal of Speech, Language, and Hearing Research, 43*, 902-914.

Stuart, A. (2005). Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. *Ear and Hearing, 26*, 78-88.

Stuart, A. (2008). Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children. *Journal of the American Academy of Audiology, 19*, 135-146.

Uhler, K., Biever, A., & Gifford, R. H. (2016). Method of speech stimulus presentation impacts pediatric speech recognition: Monitored live voice versus recorded speech. *Otology & Neurotology, 37*(2), e70-e74.

Valente, D. L., Plevinsky, H. M., Franco, J. M., Heinrichs-Graham, E. C., & Lewis, D. E. (2012). Experimental investigation of the effects of the acoustical conditions in a simulated classroom on speech recognition and learning in children. *The Journal of the Acoustical Society of America, 131*(1), 232-246.

Wolfe, J., Morais, M., Neumann, S., Schafer, E., Mülder, H. E., Wells, N., … Hudson, M. (2013). Evaluation of speech recognition with personal FM and classroom audio distribution systems. *Journal of Educational Audiology, 19*, 65-79.

Wróblewski, M., Lewis, D. E., Valente, D. L., & Stelmachowicz, P. G. (2012). Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults. *Ear and Hearing, 33*(6), 731.

Acknowledgments: This study was funded through a Baylor Undergraduate Research and Scholarly Achievement (URSA) Award. The authors wish to thank the participants and their parents and school staff, as well as the three Baylor student researchers who assisted in data collection: Lori Owens, Alycia Toomey, and Andrew (Zeb) White. We appreciate the efforts of Rachel Hobaugh, who scored half of the recorded sessions for analysis of inter-rater agreement.