Is it necessary to assess fluent symptoms, duration of dysfluent events and physical concomitants when identifying children who are at risk of speech difficulties?

Avin Mirawdeli and Peter Howell

Department of Experimental Psychology,
Division of Psychology and Language Sciences,

University College London

Address correspondence to: Peter Howell, Division of Psychology and Language Sciences, University College London, Gower St., London WC1E 6BT, UK. E-mail: p.howell@ucl.ac.uk.

Phone number: +44 (0)207 679 7566

Fax: +44 (0) 207 436 4276

Abstract Riley's (1994) Stuttering Severity Instrument version three (SSI-3) has three components: a symptom frequency measure (%SS), average duration of the three longest stutters and a physical concomitant (PC) score. An assessment of whether it was necessary to use all of these when using SSI-3 to identify which children are at risk of speech difficulty was performed. Participants were 879 reception class children aged 4-6 years from UK schools. The distributions of the separate components of SSI-3 were examined. Departures from normality were noted for each component. The features seen in the distribution of the individual components were also apparent in the distribution of overall scores (this was not normal and had multiple modes). These findings undermine the usefulness of the overall measure for identifying children at risk of speech difficulty. Prior work used a fixed SSI-3 threshold to identify at risk children. Classification of children as fluent or at risk based on this threshold was compared with classifications based on thresholds applied to the individual components. Classifications were comparable for %SS, but less satisfactory for duration and PC. These findings suggests that %SS performs similarly to overall SSI-3 scores when used to identify at risk children. Riley (1994) conducted correlation analyses to justify the inclusion of all components in SSI-3. This involved part (individual component) scores that were correlated with whole (overall SSI-3) scores. These results were replicated. However, correlations are spuriously inflated if this procedure is employed. Additional correlation analyses showed that part-'whole' correlations were low when the component used as the part was excluded from the 'whole'. Thus Riley's justification for using all components is questionable. Physical concomitants measured on five-point scales (as Riley specified) were no more sensitive than when the scale was collapsed to three or to two points. Since judgments were not affected when the scale was decimated, judges did not appear to be able to use the original scale. Procedures for identifying at risk children in schools need to be short and easy to administer. Thus, since there is no justification for including all components of SSI-3 and duration and physical concomitants are not sensitive measures of fluency, a procedure based on the frequency measure alone is appropriate for use in schools.

Keywords: Stuttering severity instrument, speech difficulty, fluency, assessment of reception class children, screening

Is it necessary to assess fluent symptoms, duration of dysfluent events and physical concomitants when identifying children who have speech difficulties?

Introduction

Different versions of Riley's Stuttering Severity Instrument (SSI) have been widely used for clinical evaluation of stuttering (Riley, 1972; 1994; 2009). Riley (1994) indicated that SSI has other applications including differentiating fluent children from those with fluency difficulties. This suggests that there is a threshold SSI score that can separate fluent children from the rest. However, Riley (1972; 1994; 2009) did not specify what this threshold should be. Table 3 in Riley (1994) shows that 6 is the lowest SSI score that has an associated severity estimate (a label of 'very mild') even though estimates scores down to zero are possible . These observations may indicate that a child who scores below 6 should be considered fluent (Todd, Mirawdeli, Costelloe, Cavenagh, Davis & Howell, 2014). Subsequent estimates were made by Howell and Davis (2011), Howell (2013) and Mirawdeli (2015), who proposed scores of 8, 13 and 16 respectively as the threshold SSI value for the fluent cutoff. These thresholds are higher than that deduced from Riley (1994), probably because sensitive computer-based procedures were used to obtain SSI in these studies. More nonfluent events are detected when computer-based methods are used (Jani, Huckvale & Howell, 2013).

Howell (2013) performed fluency analyses on samples of children's spontaneous speech. He showed that samples that included fluent children and clinical cases of children who stuttered could be separated using Riley's procedure by applying a threshold to the SSI scores that were obtained. Subsequently, Mirawdeli (2015) reported that this approach worked with children in reception classes of mainstream UK schools. In the future, teachers may want to adapt this procedure so that they can administer it in schools. However, the procedure needs to be short, efficient and not be over-technical as schools are only indirectly interested in speech assessment (teachers' main concern being to educate the children). If the procedure was short, it could also be used to assess all children again at later stages in school in order to identify fluency problems that arise post-entry. In some respects, SSI has the efficiency requirement built in (it is short), which commends its use as a basis for identifying children with fluency difficulty. For example, the minimum length of speech sample that is required is 200 syllables (Riley, 1972; 1994). SSI allows users to base the severity score on analysis of a spontaneous speech sample alone. This makes it appropriate for children starting school who cannot read (Howell, 2013). Todd et al. (2014) confirmed that the minimum sample length of 200 syllables gives SSI estimates that correspond to those obtained on longer samples, and that estimates made solely on a spontaneous speech samples do not differ from those obtained when both spontaneous and read materials are used.

Three components are required to obtain an SSI score. The first is the frequency of non-fluent events. Counts of non-fluencies are expressed relative to the number of syllables in a passage and converted to a percentage (percentage of stuttered syllables, %SS, also called the frequency measure). Duration of the three longest non-fluent events in the sample is the second component. The final component measures physical movements of the face and body that occur at the time the speech is collected (termed physical concomitants, PC). When identifying children with fluency difficulties in schools, time would be reduced and the assessment made simpler if one or more of these components of SSI could be discarded. PC is most likely to be redundant as concern has been expressed about how objectively it can be scored (Bakhtiar, Seifpanahi, Ansari, Ghanadzade, & Packman, 2010; Todd et al., 2014). Also, the procedures that Riley (1972; 1994; 2009) outlined for assessing PC are ambiguous (Todd et al., 2014) and PC has lowest reliability of the three components of SSI (Bakhtiar et al., 2010; Lewis, 1995). In addition, it may be possible in the near future to save assessment time by automating %SS and possibly the duration component as well, but this looks to be unlikely in the case of PC.

The main goals of this article were to see whether any of the components could be dropped from the SSI procedure for use in schools. Previous literature is reviewed to obtain indications whether all components are necessary. Then the research questions investigated are presented. It should be noted that the efficiency considerations called for when SSI-based procedures are used in schools may not apply when SSI is used as part of clinical diagnosis. In the latter cases, longer, more comprehensive samples, and measurements on every component, may be desirable.

Evaluation of previous work on Riley's Stuttering Severity Instrument

Riley's Stuttering Severity Instrument measure has been through several revisions, the latest of which is version four (Riley, 2009). However, all versions are the same test (the same standards are used and no new data were collected for the evaluation of later versions). Also, although some changes were recommended in the fourth revision (Riley, 2009), these should not be adopted since the published standards only apply providing that the original procedure is adhered to strictly (Todd et al., 2014).

Assessment of components

All the components for obtaining an SSI score (%SS, duration and PC) are obtained at the time that samples of speech, either one or two, are collected. The raw PC score is added to the transformed %SS and duration scores to obtain the total overall SSI score. The SSI scores can be reported as: the raw numerical values (used throughout this article); a percentile; or a severity descriptor (very mild to very severe).

In clinical samples, although assessment of fluency on multiple dimensions of behaviour is desirable because it provides a picture over and above that obtained from speech *per se*, a single composite score that summarizes severity prevents users from identifying which specific components lead to a speaker's score (Lewis, 1995). Consequently, speakers with the same SSI score can weight differently on the three components. The ambiguity about what a particular SSI score means is a further reason why the contribution that individual SSI components make to overall scores needs examining. Information about separate components is also useful for deciding whether one or more of them can be dropped, thus reducing ambiguity in the interpretation of SSI scores.

Frequency measure (percentage of syllables stuttered, %SS)

A measure of %SS is an essential component when stuttering severity or fluency in general is assessed. The procedure for obtaining %SS in SSI is precise. Procedures to count syllables, and stutters are described and clear and unambiguous definitions of what events are considered as stutters are given in the SSI manuals (Riley, 1994; Riley 2009). The ratio of stutters out of all syllables is then converted to a percentage to give %SS. Transformation of raw %SS is simple (a conversion table is provided in the manual). The way this is subsequently used in computation of the total score is straightforward.

Campbell (2014) examined whether Riley's symptom set needed to be extended for dealing with children in schools where other common types of paediatric fluency problems as well as stuttering (referred to jointly as speech difficulties) are encountered. Less common speech difficulties, such as childhood apraxia, were not considered. The broad classes of speech difficulties that needed to be considered were identified initially based on reviews of the literature. For example, phonological disorder and phonological delay were derived from Dodd, Hua, Crosbie, Holm and Ozanne's (2002) work. The symptoms that were exclusive to each speech difficulty class that were mentioned in the studies reviewed were specified (those for stuttering, phonological disorder, phonological delay). Riley's symptom set explicitly excludes whole-word repetitions as symptoms of stuttering whereas some authors consider them to be core features of stuttering. A separate class of other possible stuttering symptoms was defined that comprised whole-word repetition, revision and phrase repetition). The four sets of symptoms (core stuttering, other stuttering, phonological delay and phonological disorder) were then employed in a preliminary investigation where separate transcriptions were made of the symptoms from each set for 30 participants (pre-analysis diagnosis was fluent for 15 and not fluent for the remaining 15 who had a range of speech difficulties). Based on this analysis, other non-fluent symptoms were identified that were not mentioned as characteristic of symptoms in any of the four initial sets. These were designated as 'other symptoms' and consisted of nasalization, liquid confusion, vowel distortion, voiced alveolar lateral fricative (the replacement of a target alveolar approximant with a labiodental approximant), hyper-, and hypo-nasality. All symptoms in each symptom set were used in hierarchical cluster analyses, and a median split was made as a statistical criterion for separating children who were fluent from those who had speech difficulties. These results were compared with those of the pre-analysis diagnosis to establish how well that symptom set designated children as fluent/having speech difficulties. Next, symptoms were selected that correlated significantly with the overall score for the respective symptom set. The 'pruned' sets identified children's pre-analysis diagnosis better for all of the original sets except the 'other symptoms' set (73% and 67% correct classification for full and pruned symptoms respectively). The most notable feature that is relevant for present concerns was that the core stuttering symptom set (i.e. based on Riley's stuttering symptoms) identified children best overall for both the full (73%) and pruned (87%) symptoms (the symptom pruned from the stuttering symptom set was part-word repetition). Campbell's (2014) findings show that Riley's symptom set is appropriate for use in schools where diverse forms of speech difficulties are encountered when determining whether or not a child has a fluency problem.

Duration

Bloodstein and Ratner (2008) questioned whether duration was useful when fluency is assessed. Moreover, the procedure for estimating raw duration scores in SSI is dated. A manual method involving a stopwatch was used. Making the measurement live requires pre-emptive judgements about what words will be stuttered. Anticipating when a stutter is coming up would impact on the reliability of the assessor's duration measurement because they have to: 1) concentrate on establishing whether the next syllable is stuttered; and 2) perform various multitasking activities such as using the stopwatch (it takes time to start and stop), monitoring the child's speech and movements (for PC), and preparing prompt questions when spontaneous speech samples are collected. All of these factors risk duration measurements being inaccurate. These problems would be particularly acute when the instrument is used in schools (teachers are unlikely to work from recordings).

Duration can be measured objectively from digitally-captured oscillograms of speech (Jani et al., 2013). However, as noted, computers were not used to measure duration when the standards were developed. Therefore, the more accurate computer-based methods should not be used if users want to reference results to the published standards. If duration measurements cannot be made accurately even though procedures to achieve this are widely available, it should be considered whether duration should be omitted when SSI scores are obtained.

PC (physical concomitants)

The reasons why procedures for making a PC score lacked objectivity were discussed by Todd et al. (2014). Breathing noises, jaw movements, eye movements and other bodily movements are each assessed by the user. A short description is given indicating which of the six scale points (0-5) is appropriate. The estimates for each body area are summed to give an overall PC score. No transformation is applied to the PC score. Todd et al. (2014) noted that the guidance for points one and two are difficult to distinguish ("not noticeable unless looking for it" versus "barely noticeable to casual observer"). Similarly, points three, four and five also seem to overlap ("distracting", "very distracting" and "severe and painful looking"). These difficulties probably explain why Bakhtiar et al. (2010) reported that PC measurement was not reliable. Also, PC cannot be reassessed when audio recordings are used nor when a clinician makes the measurements whilst a client is observed live, which are the formats that were used when the standards were derived. In both cases, no check on PCs can be made unless there is a supplementary video recording, which is not always possible. The inclusion of PC in severity scores because of their reported poor reliability and validity and because of the ambiguous descriptors is a major weakness in the SSI assessment procedure.

Combining the three components: Checks of component distributions and transformations of scores

Usually, when tests are developed, checks are made as to whether the data are normally distributed. If not, transformations can be applied to make them normal. When more than one component is incorporated into an overall score, each component may need to be scaled so that they all have the same means and variances. The scores can then be combined and a normal distribution should arise. Riley may have needed to transform %SS and duration but not PC to achieve this. Information about whether checks were made to ensure component distributions were normal and whether any scaling was required are lacking. However, the conversion tables provide some information on what these transformations achieve.

Transformations

For %SS, the raw scores (X-axis) versus transformed scores (Y–axis) using the table on page 10 of Riley (1994) are plotted in Figure 1. Riley's (1994) tables give a single or a range of %SS that converts into one task score value. For example, the non-reader table indicates 6-7 %SS has a task score of 12. The average raw %SS is plotted on the X axis against the corresponding task score on the Y-axis in Figure 1. The task scores are the tabulated values from the non-reader conversion table (X is 6.5% and Y is 12 in the preceding example). This shows that up to 7 %SS, an increase of 1-2 %SS leads to a two point increase in the task score. At values above 8 %SS, an increase of 4 % or more is needed to increase the task score by two points. Figure 1 also shows that the functions are monotonic, but not linear.

Figure 1 about here

The non-linearity changes the sensitivity of the severity measure over different ranges of %SS. The rate of increase of the task scores levels off as %SS increases. Also, the task scores reach asymptote (do not increase further) above 18 %SS. Both these features give low scores more weight relative to high scores. Sensitivity is lost for high task scores (high %SS) at the expense of giving more sensitivity to low task scores. The transformation may have been deliberately designed to achieve the different sensitivities in different ranges or to give the data a normal distribution.

Similar observations apply to the transformation appropriate for duration. The duration measures are also transformed non-linearly, as shown in Figure 2. Again a change in a unit of time at short durations increases the overall task score more than the same change at longer durations. An additional concern when duration is measured is that the length of the speech sample should have been fixed in terms of number of syllables, otherwise the duration scores (specified in absolute time units) would be problematic. This is because the duration score could vary with the length of the sample. For instance, if SSI scores are obtained on 200 and 400 syllable-long samples, the chance of including longer dysfluent syllables is greater in the 400 syllable-long sample.

Figure 2 about here

Use of raw physical concomitant scores

Raw PC scores are added to transformed %SS and duration task scores. Using raw PC scores leads to them having an additive effect across the range of values this measure can take (linear weighting). Taken in conjunction with the levelling off of transformed %SS and duration scores at higher values, PC scores effectively make a bigger contribution to high total overall SSI scores than do %SS and duration scores.

Part-whole correlation analysis as a justification for including all components

Riley (1994) justified the inclusion of all components based on correlation analyses. Riley (1994) correlated each component score with each of the remaining ones and also with the total overall score. The correlation matrix showed that coefficients were lower when individual component scores were compared (e.g. %SS and duration had a correlation coefficient of 0.37) than when any component score was compared with the total score (e.g. %SS and total overall score had a correlation coefficient of 0.83). The set of correlation coefficients for pairwise comparison of individual components is given in Table 1 and the set of correlation coefficients for individual components with total scores is given in Table 2. Based on these analyses, Riley (1994) argued that all components should be included because “None of the parameters used alone will produce the same severity indications as the combination of all three” (Riley, 1994, p.18). The argument justifying inclusion of all components that Riley offered is fallacious since correlating a measure with another quantity that contains that measure inflates the correlation coefficient (Barry, 1983; Snedecor, 1956). One possible alternative way of addressing this issue would be to use partial correlation. However the results to be reported later show that the component variables in SSI depart markedly from normality that preclude use of this analysis procedure.

Tables 1 and 2 about here

Research questions

A streamlined procedure is required when large numbers of participants are assessed for fluency (Howell, 2013). Riley’s (1972; 1994; 2009) SSI instrument has several procedural features that commend its use in such applications. This includes its brevity (Todd et al., 2014) and allowance for assessments to be made with minimal equipment. As seen in the above analysis of SSI, it is not clear why some components were included, whether their inclusion serves a purpose in obtaining an SSI score and why the transformations were made. The following research questions address whether all components are needed for applications where children with speech difficulties are identified in schools.

Research question 1: Distribution and transformation of individual components
The distribution of component scores (raw PC, transformed %SS and duration) should be normally distributed with similar means and variances to allow them to be combined to give normally distributed total overall SSI scores. Whether the distribution of components meet these requirements is examined on a large unselected sample of UK children entering education. For PC, the distribution of raw scores was obtained (Riley does not given a transformation for PC). The distributions of all components and overall SSI scores were assessed for normality and to obtain their mean and higher moment statistics. Then the distributions of the individual components were compared with the distribution of the overall SSI scores. Together, these results provide an indication of whether the data were processed appropriately and indicate any problems introduced when the components were combined.
Research question 2: Scatter plots to ascertain how cutoffs based on individual components compared to cutoffs based on overall SSI scores
A fixed threshold of 16 was applied to overall SSI scores to separate the children who were fluent from those who were deemed to have speech difficulties (Mirawdeli, 2015). This assumes that SSI gives unambiguous case classification. To see how individual components compared, the selected component was plotted on the X axis, and overall SSI scores on the Y axis. These plots show how well that component classifies children with respect to fluency relative to the overall SSI score standard (i.e. how well the participant groups were separated by the individual component compared to the overall SSI score). Contingency tables (classification based on SSI score versus classification based on the individual threshold) were obtained for all children, including statistics to establish classification performance and to compare classifications when different components were used.
Research question 3: Part-whole correlation, and related analyses
Riley's (1972; 1994; 2009) part-whole correlation analysis was conducted on the present data. Although these analyses also showed that correlations between individual component scores were lower than those between each individual component and the overall score, as pointed out above, correlating a score with something that includes that score is fallacious. To address this problem, correlations of individual components with the remaining components were computed.
Research question 4: Evaluation of performance using different numbers of scale values for the PC component
The previous literature indicates that PC is probably not a useful component of SSI severity estimates. To check this, analyses were made to ascertain whether the scale descriptors Riley gives can be used consistently. The PC measured using Riley's five-point scale were compared with PC obtained when the scale was collapsed in different ways. The PC scale was collapsed into two or three categories that were suggested by which scale points seemed to overlap, as reported by Todd et al. (2014). The questions are whether or not the two- and three-point scales give the same results as Riley's (1994) five-point scale. If the five point scale provides a better estimate than two or three point scales, higher correlation coefficients should occur between PC measured this way and the other components of SSI than those found when the two or three point scales are used.

The research questions provide indications as to whether components of SSI can be dropped. As indicated above, PC and duration could potentially be dropped. The research questions were examined in an unselected sample of reception class children.

Method

Participants

Children from the reception classes of 11 London and Ipswich schools were assessed on SSI in the academic years 2012-2013 and 2013-2014. There were 879 children in total (448 males and 431 females); 315 were aged four years, 562 were aged five years and two were aged 6 years. Approximately 57% of the children spoke English as their first language (native language), 9.2% did not disclose language information and 34.2% had English as an Additional Language (EAL). The language data are presented in Figure 3.

Figure 3 and Table 3 about here

Recordings

Picture stimuli from the SSI manuals were used to elicit speech. Topics that children could readily talk about were also employed. Examples of the latter included favorite TV programs, hobbies, days out and school activities. All speech samples were recorded using a Zoom H4N recorder with an internal microphone. The samples were 10 to 15 minutes long and were in English irrespective of the child's first language. PC were scored by the first author at the time the recording was made according to the procedure given in Riley (1994).

Scoring %SS and duration

All audio recordings were uploaded and annotated using Speech Filing System (SFS) software (Huckvale, 2013). SFS (which is available as a free download from http://www.phon.ucl.ac.uk/resource/sfs/) allowed extracts to be selected and played whilst judgments were made. The number of syllables spoken in an extract were counted both for those that were fluent and for those that were not according to the guidelines in Riley (1994). Duration of the three longest stutters was measured from the oscillographic display of the speech sample using SFS.

Scoring PC using different numbers of scale points

The PC data were pre-processed so that they were scored on three different scales. 1) five point scale (Riley, 1972; 1994; 2009); 2) two point scale where points one, two and three on Riley's (1972; 1994; 2009) scale were assigned to new scale point one, and points four and five on Riley's (1972; 1994; 2009) scale were assigned to new scale point two; 3) three point scale where points one and two on Riley's (1972; 1994; 2009) scale were assigned to new scale point one, points three and four on Riley's (1972; 1994; 2009) scale were assigned to new scale point two and point five on Riley's (1972; 1994; 2009) scale was assigned to new scale point three. The collapsing for the new scales (2 and 3) was based on the ambiguous descriptions concerning scale points discussed above (Todd et al., 2014).

The %SS, duration and PC components were used to calculate SSI scores according to the non-reader procedure given in Riley (1994; 2009).

Reliability

Two trained transcribers independently assessed the speech materials from eight speakers chosen at random. Inter-judge reliability for nonfluencies was 96% and for syllable counts was 93%. The associated kappa coefficients were 0.92 and 0.89, which represents agreement well above chance (Fleiss, 1971). The durations of the three longest nonfluencies were identical to within 5 ms for the two judges. This was because the SFS has a calibrated time display on the x-axis of the oscillograms. The average durations of the three longest stutters for the eight speech samples did not differ significantly by *t* test across the judges. For one of the components, PC, inter-judge reliability estimates on these data was not possible because the first author alone was present when the recordings were made and video recordings were not taken (Todd et al., 2014). However, in other work, her inter-judge reliability has been found to be excellent (kappa coefficients above 0.9).

Results

Research question 1: Distribution characteristics of each component (%SS, duration and PC) and their appearance in the distribution of overall SSI scores

The %SS after conversion using Riley's (1994) tables and conversion to *z* scores, are plotted in Figure 4 (standardized scores on the X axis and %SS counts on the Y axis). Figure 4 shows that the %SS scores are positively skewed and that the transformed scores peak approximately at the mean. The duration scores after conversion are plotted in a similar way in Figure 5. This shows that the scores have a positive skew and the transformed scores have a peak below the mean. The raw PC scores were plotted in a similar way to %SS and duration scores in Figure 6. This shows that the scores are not normally distributed. Low scores dominated and there is a long tail through to high values (positively skewed).

The distribution of the overall scores is given in Figure 7 and shows that they are not normally distributed. The distribution has the separate features noted for each of the components. Thus, there is a peak approximately at the mean that reflects that seen in the distribution of transformed %SS scores (Figure 4); There is a second peak below the mean that occurred in the transformed duration scores (Figure 5); There is a long positive tail as observed for the raw physical concomitant scores (Figure 6). Combination of the components according to Riley's (1994) procedure does not yield a distribution that could be used to provide standard severity estimates.

Figures 4 to 7 about here

Skewness and kurtosis statistics were obtained to quantify the deviation of the distributions from normality. Skewness reflects asymmetry about the mean, kurtosis the peakiness of the distributions. After *z* score conversion, a set of scores that are exactly normally distributed would have skewness and kurtosis values of zero. These statistics show that none of distributions of either the individual components or the overall scores was an acceptable fit to a normal distribution. Table 4 also shows that the distributions have different means, which precludes them from having a normal distribution when they are combined.

Table 4 about here

Research question 2: Scatter plots to ascertain how cutoffs based on individual components compare to cutoffs based on overall SSI scores

It is known that an SSI threshold provides reasonable results for identifying children with speech difficulties (Mirawdeli, 2015). The following analyses evaluated how well each raw individual SSI component performed relative to the corresponding overall SSI score. Mirawdeli's (2015) cutoff value of 16, which was used to separate children with speech difficulties from those who were fluent, was applied here to classify children as fluent or not. There were occasional exceptions where children were deemed as having speech difficulties even though their SSI scores were below 16. These cases were classified by the schools as having speech difficulties (these were usually cases where children spoke very little). Scatter plots are given in Figures 8-

10 each with one raw SSI component (%SS, duration and PC respectively) given on the X axis and overall SSI score on the Y axis. Individual children's scores are shown as circles (children with speech difficulties) or squares (fluent). The built-in cutoff criterion ensured overall SSI scores (Y axis) separated the two groups almost perfectly (except for the cases mentioned). Examination of the distribution along the X axis, indicates how well groups were separated when individual raw component scores were used.

To quantify how well each of the components classified children in comparison with a threshold value of 16 applied to overall scores, classifications were compared with the fixed SSI score setting and each component for a range of values on the individual components. The results are summarized in Table 5. Number of cases for overall SSI classification and threshold classification by component are given in columns three to six and sensitivity and specificity in columns seven and eight. The cutoff value of 3% is often used on percentage speech-like dysfluencies (Yairi & Ambrose, 2015). However, the threshold value of 5%SS was superior (100% sensitivity and 89% specificity). Sensitivity never reached 80% for the duration threshold (80% sensitivity and specificity are conventional values for satisfactory classification of medical cases). A PC cutoff of 4 reached above 80% for sensitivity and specificity.

Figures 8 to 10 and Table 5 about here

Research question 3: Part-whole and correlations and related analyses

Riley (1994) justified inclusion of all components based on the part-whole correlation analyses reviewed in the introduction. It emerged here, when the data distributions were examined, that the overall score deviated from normality. Strictly speaking, Pearson product moment correlations require normally-distributed data as it is a parametric test. However, Riley's results are replicated for completeness, followed by ancillary analyses.

Replication of Riley's (1994) part-whole correlation analysis

Table 6 gives the correlations between total overall scores and %SS, duration and physical concomitant scores. The values ranged from .57 to .93 (all were significant by parametric and non-parametric tests and for raw and transformed data, N= 879, df = 877 $p < .001$ in all cases. Table 7 gives the correlation matrix for the three components of SSI. These analyses replicate the results in Riley's (1994) report (i.e. the coefficients in Table 6 are higher than those in Table 7).

The correlations between each individual component with each of the two remaining components (non-overlapping components) are given in Table 8 for the three age groups. Table 8 shows that the correlations between pairs of component and overall SSI scores, where overlap occurs, were high (column five). Correlation coefficients along the ascending diagonal in columns two to four were lower than those for the other cells in these columns. These low values represent cases where components were not shared between the components that were correlated. The correlation coefficients in the cells that are off the ascending diagonal all involve correlations where a component was shared. This should increase the coefficients. This was confirmed as the correlation coefficients in the ascending diagonal are always lower than in the off-diagonal cells. The lower correlations when individual components are compared with the remaining components is not subject to this problem, implying that not all components may need to be included. It is also noteworthy that PC showed the lowest correlations (the next section investigated whether the procedure for measuring this has any objectivity).

Tables 6, 7 and 8 about here

Research questions.4: Evaluation of performance using different scale values for the PC component.

Parametric and non-parametric correlations were performed between PC scores obtained according to the three scales (five, two and three point) and the remaining two SSI components (%SS and duration). The results are given in Table 9. Since the distribution of all components deviated significantly from normality, the parametric coefficients (in brackets) are not appropriate but are given for completeness. All correlation coefficients were significant as N was large (879). The principal issue was whether the five point scale provides useful information, bearing in mind the confusion in the scale-point descriptions. Hence, the pattern of correlations across the three forms of scale and between PC versus %SS or duration were important. Correlations between PC and %SS were better than those between PC and duration whatever scale was used. This underlines, once more, the status of %SS in characterizing severity. The main feature to note in the comparisons is that there was little difference across the different PC scale formats. The non-parametric coefficients for the correlations between PC and %SS ranged from .386 to .391 across scale formats. The non-parametric coefficients for the correlations between PC and duration SS ranged from .315 to .316 across scale formats. Thus it appears that the correlation patterns were not sensitive to the changes in the imposed scale values. Therefore, in terms of the prediction for this research question, the five scale-points did not provide a better estimate than two or three point scales.

Table 9 about here

Discussion

The main goal of this article was to evaluate whether all SSI components are necessary when Riley's instrument is used to identify children starting schools who have speech difficulties. The implications that the findings have for this process are considered. The process of identifying children with speech difficulties and the role a speech assessment has in it are discussed. Finally, implications that the current assessment of SSI has for other situations where the instrument is employed are considered.

Should all SSI components be included when Riley's procedure is used to identify children starting in schools who have speech difficulties?

%SS, duration and PC components of SSI were not normally distributed. %SS peaked around zero after *z* transformation and had a positive skew. The transformed duration scores after *z* transformation peaked below zero and again had a positive skew. The raw PC scores were not normally distributed. Most PC scores were zero and extended to positive values and there was a long positive tail on the distribution. The deviations from normality and positioning of some of the modal values away from zero after *z* transformation implies that combining the components would not lead to a composite that was normally distributed. This was confirmed as the distribution of overall scores after *z* transformation had the features noted in the distribution of the separate components.

These observations on the distributions question the appropriateness of combining the components of SSI to yield unambiguous interpretations of an individual's fluency level (Lewis, 1995). One solution would be to drop any component(s) that showed marked departures from normality. The component with the closest match to a normal distribution that also had a peak near zero after *z* transformation was %SS. However, these observations alone are not sufficient grounds for only using that component in identifying children with speech difficulties as %SS may not achieve satisfactory classification of children. To address whether any of the components on their own provided a basis for classifying children as compared to overall SSI scores, scatter plots were made between scores on each component and the overall SSI scores. The overall SSI score has been successfully applied to identifying children with speech difficulties elsewhere (Howell, 2013; Mirawdeli, 2015). Classifications based on overall SSI scores were evaluated by establishing the degree of correspondence with children about whom schools had expressed concern. Although SSI classifications correspond well with schools' judgments about children, they do not provide a 'gold standard'. Thus, it is possible that any of the components alone could provide better case-classification in some instances than overall scores (although this is unlikely given the close correspondence with schools' judgments). Case classifications were evaluated by applying a threshold SSI score of 16 (Mirawdeli, 2015) to overall scores, and a range of thresholds was applied for each component. A 3% threshold (Yairi & Ambrose, 2005) applied to %SS produced classification with good sensitivity and specificity. However, a 5% SS threshold was even better (100% sensitivity and 89% specificity). When similar analyses were applied to duration and PC, they did not achieve good classifications relative to the score of 16 applied to overall SSI (sensitivity and specificity were lower than for %SS). Thus overall, %SS not only had the most appropriate distribution of scores, but a threshold value applied to this component alone reached satisfactory levels in classifying children as fluent or not (relative to that achieved by an SSI score of 16).

The classification analysis suggests that %SS alone contributed to the success of overall scores for identifying the fluent children. The implication is that duration and PC, could be dropped. Research questions three and four also addressed whether each of the components could be dropped, but in different ways. Research question three replicated and extended Riley's (1994) part-whole correlation analyses. The implication that the part-whole procedure led to spuriously high correlations was confirmed (Barry, 1983; Snedecor, 1956). Therefore, Riley's justification for including all three components was not appropriate. PCs are the most problematic of the components (Bakhtiar et al., 2010; Lewis, 1995; Todd et al., 2014) and were singled out for evaluation in research question four. PC scores did not change their pattern of correlations with either of the remaining components when the scales were collapsed to two or three points. This lack of sensitivity confirmed the concerns expressed about how PC are measured in SSI. On this basis and in the light of other reports (Bakhtiar et al., 2010; Lewis, 1995; Todd et al., 2014), PC should be omitted or their measurement improved in future revisions of SSI.

Although it seems desirable to drop PC in school work, some caveats are appropriate. First, it is possible that PC may play a greater role in, for the present test context, children with EAL and this should be investigated. Second, if the inclusion of PC seemed advisable when dealing with cohorts where there are many children with EAL, it might be appropriate to include a simplified PC measure (e.g. with just 2 values) in a screening tool. Third, the current analyses were based on school, not clinical, samples. Therefore, it is possible that assessment of PC plays a role in children referred to clinic. Further questions about the interplay between the SSI components specifically in clinical samples are also worth considering. For instance, do clinical samples have some children who have relatively infrequent instances of SS but these are of greater duration and/or have very noticeable PCs and, if so, should they, or should they not, be identified as having speech difficulties? If they should be considered children with speech difficulties, duration and PC might contribute uniquely to sensitivity in clinic whereas dropping duration and PC may not add anything as far as identifying fluent children is concerned.

Comments about identifying children with speech difficulties in general

Teachers are in contact with children from the time they enter school and are sensitive to their needs and problems including any speech difficulties. They want a systematic method that can be used to assess children's speech (Dockrell & Howell, 2015). Schools also report that any proposed tool has to be applicable for use with children with EAL (Mirawdeli, Dockrell & Howell, in press).

SSI was chosen as the preferred contender as a starting point for such a tool (Howell, 2013; Mirawdeli, 2015) as it has a short, efficient and clear procedure. In addition basing assessments on audio samples maintains children's confidentiality and it does not require children being able to read (Todd et al., 2014). Any further simplifications of the procedure would be welcomed by schools. In the light of this, the present work confirmed that PC, and probably also duration, can be dropped without affecting performance of a revamped procedure for identifying children with speech difficulties.

Riley's (1994) symptoms focus on breaks, repetition of parts of words or prolongation of the initial part of words, but excluded repetitions of whole words (Howell, 2013). Two features to highlight are: 1) that the fragmentary dysfluencies Riley selected are appropriate for identifying stuttering and other paediatric speech difficulties (Campbell, 2014; Howell, 2013; Mirawdeli, 2015); 2) that the exclusion of whole-word repetitions allows them to be used to identify other forms of difficulty, namely word-finding difficulty. When children cannot find a word, they use hesitation phenomena such as repeating whole words (Fehringer & Fry, 2007). Word-finding difficulty is a particular problem in the speech of children with EAL (Bada, 2010). Riley's procedure effectively does not confuse children with EAL who have word-finding difficulty but are otherwise fluent with those who have speech difficulties - a desirable feature for schools.

Schools would use a procedure that is based on SSI to identify a range of paediatric speech difficulties, not stuttering alone as they are not primarily interested in what form of disorder a child has (which is a matter for Speech and Language Therapists). Campbell's (2014) work that was reviewed in the introduction showed that Riley's symptoms do not miss children with forms of speech difficulties other than stuttering. This suggests that estimating these symptoms would be sufficient for schools' needs. That said, other applications may call for an instrument that includes a set of symptoms that is appropriate for identifying all forms of speech difficulties. An example would be where it is necessary to estimate incidence statistics for different types of speech difficulties. Broomfield and Dodd (2004) reported such estimates for phonological delay and disorder, but not for stuttering. Incidence statistics on all common forms of speech difficulties would be useful given clinicians interest in comorbidity between types of speech difficulties and the etiological significance these may have for patterns of recovery. Extending assessment to different types of speech difficulties, as well as calling for an extension of the symptom set, would require redesign of SSI. When SSI is redesigned, it is important that whole-word repetitions are left out from the speech difficulties set, allowing them to be used to identify word-finding difficulty separately. Identification using an objective speech-based procedure is desirable to schools, but they take other factors into account when deciding whether to refer a child for intervention (the main factor being educational attainment). It is often not recognized that these other factors schools take into consideration bias the samples of children that present to Speech and Language Therapists. Work is taking place on improving coordination which will make both sides aware of features like this so that they can be incorporated into feedback between schools and Speech and Language Therapists (Bercow, 2008, Mirawdeli, Dockrell & Howell, in press).

Relevance of results for wider applications of SSI

A problem with the above proposal on the next stage in development of a suitable instrument to identify children with speech difficulties in schools is that if the suggestions to use audio format, employ a short sample of spontaneous speech and to estimate %SS alone were adopted, they would distance the procedures teachers would want from those which Speech and Language Therapists would want to implement in any revised version of SSI. For instance Speech and Language Therapists would want to upgrade the recording format from audio to audio-video and take longer and more representative samples (Mirawdeli et al., in press).

There are some common considerations that apply to teachers and Speech and Language Therapists. For instance, if PC and duration were dropped in new forms of the instrument, this would call for restandardization. Whilst restandardization is taking place, some attention should be given to examining distributions, tabular conversions and what they achieve, and combination of components as was done here. Also, consideration of ways in which teachers' and Speech and Language Therapists' assessments could maintain some compatibility is essential. The syllable and dysfluency count procedures used in SSI are appropriate and should be retained. The exclusion of whole-word repetitions also seems desirable, although these should be counted separately (Howell & Lu, in press) for the information they provide about word-finding difficulties. If a measure of PC is considered essential, a better option might be to use one based on the premonitory urge for tics scale (Woods, Piacentini, Himle & Chang, 2005).

If it is desirable for teachers and Speech and Language Therapists to coordinate their activities (Bercow, 2008; Mirawdeli et al., in press) it would help if they used scores that were translatable. One possibility would be to embed %SS counted as indicated above into all forms. Speech and Language Therapists could transform the school results to interpret them in their own terms and vice versa. For instance, transforming of %SS and duration in the current form of SSI squashes the scale at the end where teachers score most children, the majority of whom are fluent (least sensitivity where it is most needed). A different transform might be used in schools from the one used in clinics. Software could easily be written that converts scores from one form to another if the scores only differ in respect of the transform that was used. Similar adaptations could be made for versions with and without additional components. If a child is referred to clinicians, the appropriate %SS or SSI score could then be recalculated.

Limitations

The application to identification of children with speech difficulties is an important, but not major application of SSI. SSI needs an equivalent set of performance evaluations in clinical settings to those conducted here. This should examine distributions of scores, the roles of all components etc. Decisions as to whether any of the components can be discarded for clinical application could then be made.

Cases of speech difficulties were identified by teachers. At minimum, this is not as thorough as Speech and Language Therapists' methods of assessing a child. On the other hand, Speech and Language Therapists do not take educational attainment into account in deciding on interventions at present, but teachers do so when identifying children for intervention. The impact of educational attainment on assessments needs further examination.

The children in schools had a range of speech problems. Other speech features may need to be taken into account although Campbell's (2014) work suggests otherwise. If the issue of revising symptom is addressed, new goals of identifying children with speech difficulties should be set and appropriate symptom sets defined depending on these goals. For example, is the procedure intended to identify stuttering alone or a range of paediatric speech difficulties? The issue of using whole-word repetitions to identify word-finding difficulty in children with EAL need further attention.

Conclusions

SSI has many features that make it suitable for use in schools, and a procedure for assessing speech difficulties is desirable to schools. It can be simplified by using the %SS component alone. These alterations would mean that the instrument with %SS alone has to be restandardized. Schools would want many features present in the SSI procedure to be retained. Ironically, Speech and Language Therapist s would want to change some of these. For example, schools would want to stick with the audio format to protect pupils privacy whereas Speech and Language Therapists would want to use audio-video recordings to have permanent records of PC. Therefore, it is possible that any SSI-based procedure for schools would diverge from clinical forms of SSI. This would undermine coordination between school and Speech and Language Therapy services. Alternatively, the importance of schools as a source of referral (Mirawdeli et al., 2015) may require retention by Speech and Language Therapists of at least some of those features that schools would like to keep. If this is the case, some way of converting scores between schools' and Speech and Language Therapists' results would be desirable.

Acknowledgement

This work were supported by a grant from the Dominic Barker Trust, Ipswich, UK, to Howell and their support is gratefully acknowledged. We thank the participants for their help.

References

Bada, E. (2010). Repetitions as vocalized fillers and self-repairs in English and French interlanguages. *Journal of Pragmatics, 42*(6), 1680-1688.

Bakhtiar, M., Seifpanahi, S., Ansari, H., Ghanadzade, M., & Packman, A. (2010). Investigation of the reliability of the SSI-3 for preschool Persian-speaking children who stutter. *Journal of Fluency Disorders, 35*, 87–91.

Barry, W. J. (1983). Some problems of interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology, 9*, 826-828.

Bercow, J. (2008). *The Bercow report: A review of services for children and young people (0–19) with speech, language and communication needs*. Nottingham: DCSF Publications

Bloodstein, O., Ratner, N.B., (2008). A Handbook on Stuttering, 6th ed. Thomson Delmar Learning, Clifton Park, N.Y.

Braun, A.R., Varga, M., Stager, S.,Brocklehurst, P.H., (2013). Stuttering prevalence, incidence and recovery rates. *Journal of Fluency Disorders, 38*, 290-293.

Broomfield, J. & Dodd, B. (2004). Children with speech and language disability: caseload characteristics. *International Journal of Language and Communication Disorders, 39***,** 1-22.

Campbell, A. (2014). *The separation of spoken communication disorders based on symptoms in speech of four and five year old children.* (Unpublished master's thesis). University College London, United Kingdom.

Conture, E. G., & Kelly, E. M. (1991). Young stutterers' nonspeech behaviors during stuttering. *Journal of Speech and Hearing Research, 34*, 1041-1056.

Dockrell, J. & Howell, P. (2015). Meeting the needs of children with Speech Language and Communication Difficulties: Challenges and opportunities. *British Journal of Special Education, 42*, 411-428.

Dodd, B., Hua, Z., Crosbie, S., Holm, A. & Ozanne, A. (2002) *Diagnostic Evaluation of Articulation and Phonology*. London: Pearson Assessment.

Fehringer, C. & Fry, C. (2007). Hesitations phenomena in the language production of bilingual speakers: The role of working memory. *Folia Linguistica, 41*, 37-72.

Fleiss J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378–382.

Howell, P. (2013). Screening school-aged children for risk of stuttering. *Journal of Fluency Disorders, 38*, 102–123.

Howell, P., & Davis, S. (2011). Predicting persistence of and recovery from stuttering at teenage based on information gathered at age eight. *Journal of Developmental and Behavioral Pediatrics, 32*, 196-205.

Howell, P., & Lu, C. (in press). Facts and theories about stuttering. In P. van Lieshout, B. Maassen, & H. Terband (Eds.), *Speech motor control in normal and disordered speech: Future developments in theory and methodology* (pp. XXX–XXX). Rockville, MD: ASHA Press.

Huckvale, M. (2013) *Speech Filing System Windows version 1.75* [computer software]. University College London

Jani, L., Huckvale, M., & Howell, P. (2013). Procedures used for assessment of stuttering frequency and stuttering duration. *Clinical Linguistics & Phonetics*, *27*, 853–861.

Lewis, K. E. (1995). Do SSI-3 scores adequately reflect observations of stuttering behaviors? *American Journal of Speech-Language Pathology, 4*, 46-59.

Mirawdeli, A. (2015). Identifying children who stutter or have other difficulties in speech production in school reception classes. *Procedia - Social and Behavioral Sciences*, *193*, 192-201.

Mirawdeli, A, Dockrell, J. & Howell, P. (in press). A survey of UK Teachers' and Speech and Language Therapists' joint roles with respect to reception class children's Expressive Speech Difficulties. *International Journal of Language & Communication Disorders*.

Riley, G.D. (1972). A stuttering severity instrument for children and adults. *Journal of Speech and Hearing Disorders, 37, 314-322.*

Riley, G. (1994). *The Stuttering Severity Instrument for Adults and Children (SSI-3)* (3rd ed.). Austin, TX: PRO-ED.

Riley, G. (2009). *The Stuttering Severity Instrument for Adults and Children (SSI-4)* (4th ed.). Austin, TX: PRO-ED.

Snedecor, G. W. (1956). Statistical methods, fifth edition. Iowa State University Press, Ames, Iowa.

Todd, H., Mirawdeli, A., Costelloe, C., Cavenagh, P., Davis, S.. & Howell, P. (2014). Scores on Riley's Stuttering Severity Instrument versions three and four for samples of different length and for different types of speech material. *Clinical Linguistics and Phonetics, 28,* 912-926.

Woods, D. W., Piacentini, J., Himle, M.B. & Chang S. (2005). Premonitory urge for tics scale (PUTS): Initial psychometric results and examination of the premonitory urge phenomenon in youths with tic disorders. *Journal of Developmental and Behavioral Pediatrics*, *26*, 397–403.

Yairi, E., & Ambrose, N. (2005). *Early childhood stuttering : for clinicians by clinicians*. Austin, TX: PRO-ED.

Figure captions

Figure 1. X-Y plot for the raw percentage of stuttered syllables and the associated task score of the SSI.

Figure 2. X-Y plot for the raw stuttered syllable length in seconds (average of three longest stuttering events) and its associated task score for the duration component of the SSI.

Figure 3. The distribution of English, undisclosed and children with EAL in the sample (pie chart at left). The frequency of different first languages of the children with EAL is given on the right.

Figure 4. Distribution of %SS scores after conversion using Riley's (1994) table. Standard scores are plotted on the X axis and counts are given on the Y axis.

Figure 5. Distribution of duration scores after conversion using Riley's (1994) duration table. Standard scores are plotted on the X axis and counts are given on the Y axis.

Figure 6. Distribution of raw physical concomitant scores according to Riley (1994). Standard scores are plotted on the X axis and counts are given on the Y axis.

Figure 7. Distribution of overall SSI scores using Riley's (1994) procedure. Standard scores are plotted on the X axis and counts are given on the Y axis.

Figure 8. Scatter plots of raw %SS (frequency) scores (X axis) against overall SSI scores (Y axis). Circle symbols were used for plotting cases where children with speech difficulties (SSI score of > 16) and squares for children deemed fluent. As mentioned in the text, there are cases where children were deemed to have speech difficulties, although their SSI scores were below 16 which were based on schools' judgments.

Figure 9. Scatter plots of raw duration scores (X axis) against overall SSI scores (Y axis). Circle symbols were used for plotting cases where children were designated as having speech difficulties (SSI score of > 16) and squares for children deemed fluent.

Figure 10. Scatter plots of raw physical concomitant scores (X axis) against overall SSI scores (Y axis). Circle symbols were used for plotting cases where children were designated as having speech difficulties (SSI score of > 16) and squares for children deemed fluent.

Table 1. Details of the teachers (column two) and SLTs (column three) who participated. Information sought is listed in the first column. For forced choice responses, the options available are given in column four.

	Teachers	**SLTs**	**Response options allowed (where appropriate)**
Number.	35	35	-
% female.	83%	80%	-
% indicating modal option for time in post in year.	10-20 years, 36%.	Less than 5 years, 56%.	0-5, 5+-10, 10+-20, or more than 20 years).
% of SLT's indicating modal option for time in contact with children with EAL.	-	0-20% of time, 49%.	0-20%, 20+-40%, 40+-60%, 60+-80%, 80+-100%.
% who worked in the modal type of school for participants who worked in schools.	Mainstream, 97%.	Mainstream, 91%.	Mainstream, special, private.

Table 2. Details of the reception classes of schools in which respondents worked. Information reported on is given in column one. Responses are given separately for teachers (column two) and SLTs (column three). Column four gives the available response options for each row.

	Teachers	**SLTs**	**Response options allowed**
% respondents giving the modal option	20+-30, 54%	20+-30, 75%.	10-20, 20+-30 or more than 30.
Main social class of children	Lower class, 76%	Lower class, 52%.	Lower, middle, upper class.
% respondents and modal answer for referrals to SLTs	Up to 5%, 73%.	5+-10%, 50%.	0-5%, 5+-10%, 10+-15%, 15+-20%, more than 20%.
% children who have EAL with modal answer (only for SLTs)	-	0-20%, 53%.	0-20%, 20+-40%, 40+-60%, 60+-80% and 80+ to 100%.
% children with EAL in reception classes with modal answer (only for teachers)	60+-80% , 50%.	Not asked.	0-20%, 20+-40%, 40+-60%, 60+-80% and 80+ to 100%.

Table 3. Responses to questions concerning feedback between teachers and SLTs. the questions asked are given in column one, teachers and SLTs responses are given in columns two and three and response options for the question in row four, where appropriate.

	Teachers	**SLTs**	**Response options allowed (where appropriate)**
Is feedback provided by SLTs post intervention? % yes.	77%	94%	-
Is feedback provided by teachers post-intervention? % yes.	37%	41%	-
Type of feedback sent (teachers) or received (SLTs) from teachers.	Improvements in speech, 80%.	Improvements in speech, 57%.	1) Anti-social behaviour; 2) Improvements in speech; 3) Changes in affective reactions; 4) Educational attainment; 5) All of these.
Type of feedback received (teachers) or sent (SLTs) from SLTs.	Improvements in speech, 63%.	All of these, 79%.	1) Anti-social behaviour; 2) Improvements in speech; 3) Changes in affective reactions; 4) All of these.
Does any feedback you receive post-intervention address issues about children with EAL? % no.	86%	63%	-
Does any feedback you give post-intervention address issues about children with EAL? % yes.	40%	85%	-

Table 4. Questions concerning general language and literacy resources and whether these are suitable for children with EAL.

	Teachers	**SLTs**
Resources for language and literacy? % yes.	91%	77%
If so, are they suitable for children with EAL? % no.	63%	63%
Any specific resources for children with EAL? % yes.	83%	43%

Table 5. Responses to questions concerning resources for identifying speech difficulty. The question and valence of the response are given in row one and options in row four, where appropriate.

	Teachers	**SLTs**	**Response options allowed (where appropriate)**
Assess speech separate from language? % "yes".	91%	80%	-
Anything for assessing speech? % "no".	77%	74%	-
More needed for assessing speech? % "yes".	100%	100%	-
How do children get seen by SLTs? The predominant response is indicated with its %age.	Teachers identify, 83%.	Teachers identify, 56%.	1) SLTs assess all; 2) Teachers identify children, based on personal experience; 3) SLTs advise teachers; 4) Other.
How much time do assessments take? The most frequent choice with its %age is given.	10-15 minutes, 40%.	More than 15 minutes, 64%.	1) 1-2 minutes; 2) 3-5 minutes; 3) 5-10 minutes; 4) 10-15 minutes; 5) More than 15 minutes.
Does assessment time vary? % "yes".	89%	100%	-
Why does assessment time vary (if indicated it does)? The most frequent choice with its %age is given.	Children with EAL take more time, 38%.	The range of difficulties a child has, 83%.	1) Children with EAL take more time; 2) Depends on severity; 3) Academic ability of the child; 4) Whether a child has speech difficulty or not; 5) The range of difficulties a child has.
Would a short systematic procedure for assessing speech be useful? % "yes".	100%	71%	-

Table 6. Rank order (left-hand column) of six factors for their importance when making a referral by teachers (column 2) and SLTs (column 3). The numbers in parentheses are the mean ranks for the response option indicated.

	Teachers	**Speech and Language Therapists**
1	Speech is affected (1.70).	Educational attainment (2.71).
2	Educational attainment (2.70).	Speech is affected (3.14).
3	Child anxious (3.10).	Child bullied (3.29).
4	Antisocial behaviour (3.70).	Antisocial behaviour (3.43).
5	Child bullied (4.00).	Child anxious (3.3).
6	Child has EAL (5.80).	Child has EAL (5).

Table 7. Rank order (left-hand column) of six factors concerning the importance of a short speech test. The rankings for teachers (column two) and SLTs (column three) are given separately. Average ranks are given in parentheses.

	Teachers	**SLTs**
1	External validity (2.22).	Short test can be repeated (3.21).
2	All children can do (2.39).	All children can do (3.43).
3	Short test can be repeated (2.67).	Measure of intervention (3.45).
4	Standardized tests are available (4.11).	Provides an objective measure (3.56).
5	Provides an objective measure (4.83).	Standardized tests are available (3.57).
6	Measure of intervention (4.89).	External validity (4.01).

Table 8. Responses to questions concerning administration of a test for identifying speech difficulty. The question and valence of the response are given in row one, teachers and SLTs responses in rows two and three and response options in row four.

	Teachers	**SLTs**	**Response options allowed (where appropriate)**
Optimum frequency of repeat testing.	Once a year, 40%.	Once a year; more often than once in six months tied at 26%.	1) Once, when a child starts school; 2) Once, towards the end of the first year in school; 3) Once every six months; 4) Once a year; 5) More often than once in six months; 6) Less often than once a year.
Reasons given about how often a test should be repeated.	No clear preference.	Option 3, repeat periodically to identify when problems start, 74%.	1) Not too often to avoid disruption to school activities; 2) Just once, when children enter school, to ensure that children who start school with a speech problem are identified early; 3) Repeat testing periodically to identify when children develop a speech problem.
Specialist who should assess speech difficulty.	Teachers, 80%.	SLTs, 86%.	-
Specialist who should deal with phonological delay.	Teachers, 91%.	SLTs, 86%.	-
Reason for indicating which specialist should do the assessment for speech difficulty.	Knows all the children, 82%.	Has the professional training for identifying speech difficulties, 100%	Teachers: 1) Knows all the children; 2) Has the necessary skills to communicate with children; 3) Appreciates the difficulties associated with hildren with EAL; 4) Has the professional training for identifying speech difficulties. For SLTs, option 1 replaced by: Knows in general about children’s speech difficulties. (i.e. has been educated about a range of speech difficulties)/.

Table 9. Questions addressing practical details of an identification procedure (first four rows), symptoms to include (rows five-six) and reasons for including these symptoms (rows seven and eight). Choices are indicated for teachers (column two) and SLTs (column three) and percentage giving these responses is indicated in parentheses where appropriate. Options available for multiple choice questions are indicated in column four (as appropriate).

	Teachers	**SLTs**	**Response options allowed (where appropriate)**
Length of speech sample.	A short sample, 77%.	A long sample, 74%.	1) As short as possible to avoid disruption in class; 2) A sample sufficiently long to provide a stable measure.
Type of speech sample.	Spontaneous sample alone, 77%.	Representative range of speaking situations, 100%.	1) Spontaneous sample alone; 2) Representative range of speaking situations.
Recording mode	Audio alone, 67%.	Audio-visual, 80%.	1) Audio; 2) Audio-visual.
Type of language sample from children with EAL.	English, 80%.	Alternative language, 46%; English and alternative language, 40%.	1) English; 2) Alternative language; 3) English and alternative language.
Include WWR as indications of speech difficulty?	Would not include WWR, 80%.	Would include WWR, 80%.	-
Include pauses as indications of speech difficulty?	Would not include pauses, 77%.	Would include pauses, 51%..	-
Reason why WWR should be excluded.	Option 3, All children show them, 80%.	Option 3, All children show them, 66%.	1) WWR are indications of word-finding difficulty which teachers should deal with; 2) Children with EAL show a lot of WWR as they do not speak English; and 3) all children show WWR, so they are not unusual.
Reason why pauses should be excluded.	Option 3, all children show them, 77%.	Option 3, all children show them, 85%.	As above with ‘pauses’ substituted for ‘WWR’).

Table 10. Sufficiency of current SLT provision and future needs. The question and valence of the response are given in row one and responses are indicated for teachers (column two) and SLTs (column three). Response options (where appropriate) are given in column four.

	Teachers	**SLTs**	**Response options allowed (where appropriate)**
Is SLT provision for language adequate for all children? % no.	94%.	97%.	-
Is there provision for SLT intervention for children with EAL? % yes.	26%.	29%.	-
What could be provided for children with EAL who have speech difficulty?	One-to-one guidance, 100%.	Interpreter; 66%.	1) Interpreter; 2) One-to-one guidance; 3) Culturally-appropriate material.
What provision should be available for children with EAL for *speech intervention*	One-to-one guidance, 74%.	Interpreter; 63%.	1) Interpreter; 2) One-to-one guidance; 3) Culturally-appropriate material; 4) Linguistically-appropriate material; 5) An SLT who speaks the alternative language

Table 11. Rank order of ten factors for importance (top section) and feasibility for collecting (bottom section) separately for teachers (column two) and SLTs (column three). The numbers in parentheses are the average ranks across participants.

	Teachers	**SLTs**
Importance		
1	Speech (2.79)	Speech (3.57)
2	Intellectual ability (3.61)	Gender (3.93)
3	Other languages (3.94)	Family history (4.14)
4	Gender (4.39)	Comorbid difficulties (5.29)
5	Family history (4.63)	Other languages (5.71)
6	Handedness (5.44)	Intellectual ability (5.86)
7	Closed head injury (7.28)	Closed head injury (6.29)
8	Birth difficulty (7.32)	Extraneous movements (6.36)
9	Comorbid difficulties (7.39)	Handedness (6.64)
10	Extraneous movements (7.72)	Birth difficulty (7.64)
Feasibility		
1	Gender (2.33)	Intellectual ability (2.00)
2	Handedness (3.56)	Gender (4.00)
3	Intellectual ability (3.78)	Family history (4.33)
4	Speech (4.11)	Speech (5.17)
5	Other languages (4.50)	Other languages (5.14)
6	Comorbid difficulties (6.25)	Handedness (5.50)
7	Extraneous movements (6.25)	Extraneous movements (6.17)
8	Closed head injury (7.50)	Comorbid difficulties (6.33)
9	Family history (7.89)	Closed head injury (8.10)
10	Birth difficulty (8.00)	Birth difficulty (8.20)

Figure 1. X-Y plot for the raw percentage of stuttered syllables and the associated task score of the SSI.

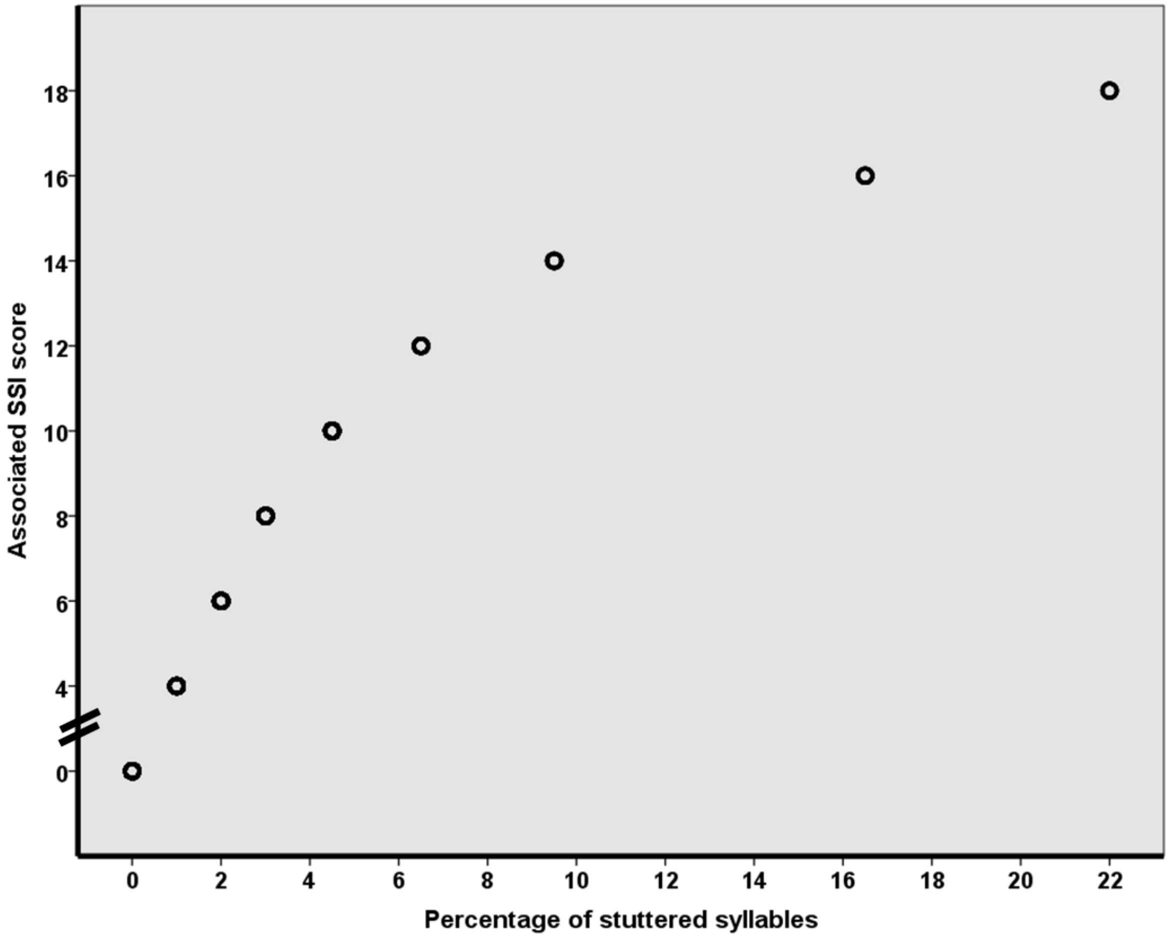

Figure 2. X-Y plot for the raw stuttered syllable length in seconds (average of three longest stuttering events) and its associated task score for the duration component of the SSI.

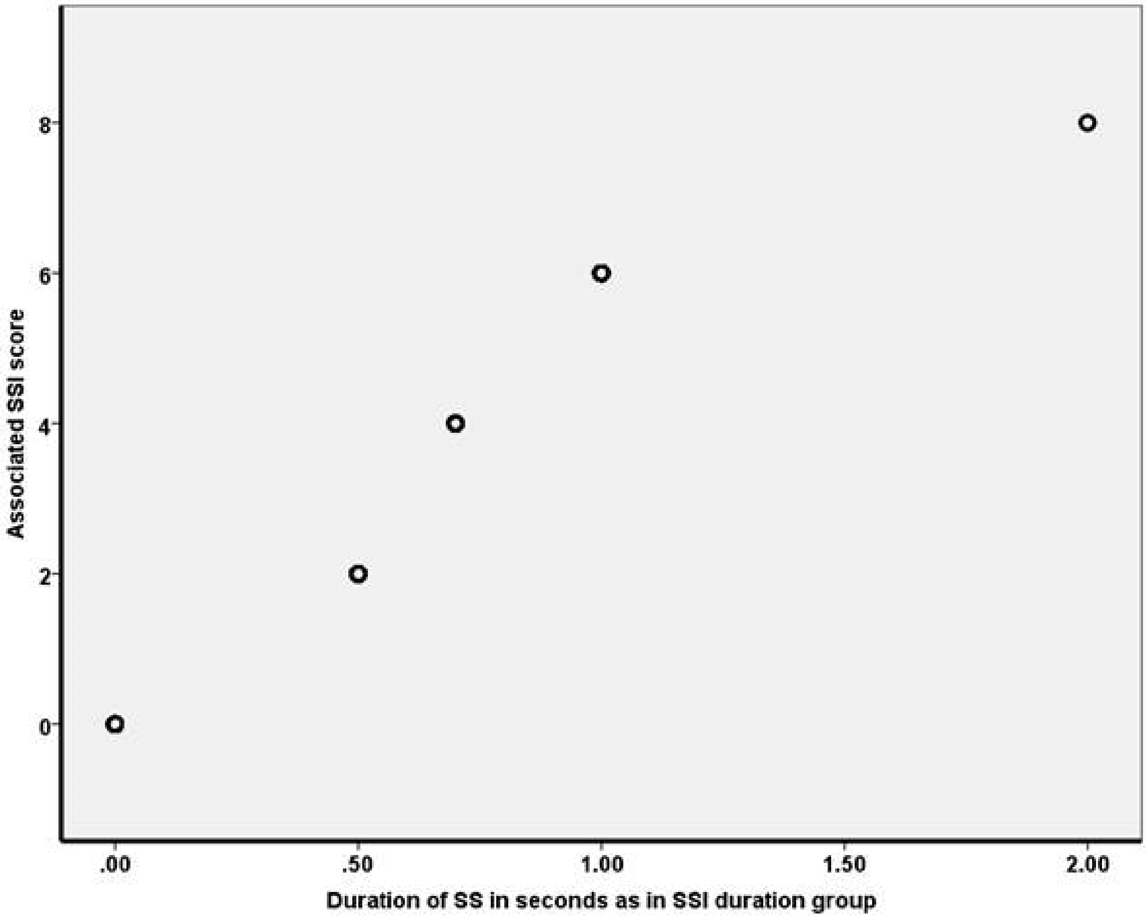

Figure 3. The distribution of English, undisclosed and children with EAL in the sample (pie chart at left). The frequency of different first languages of the children with EAL is given on the right.

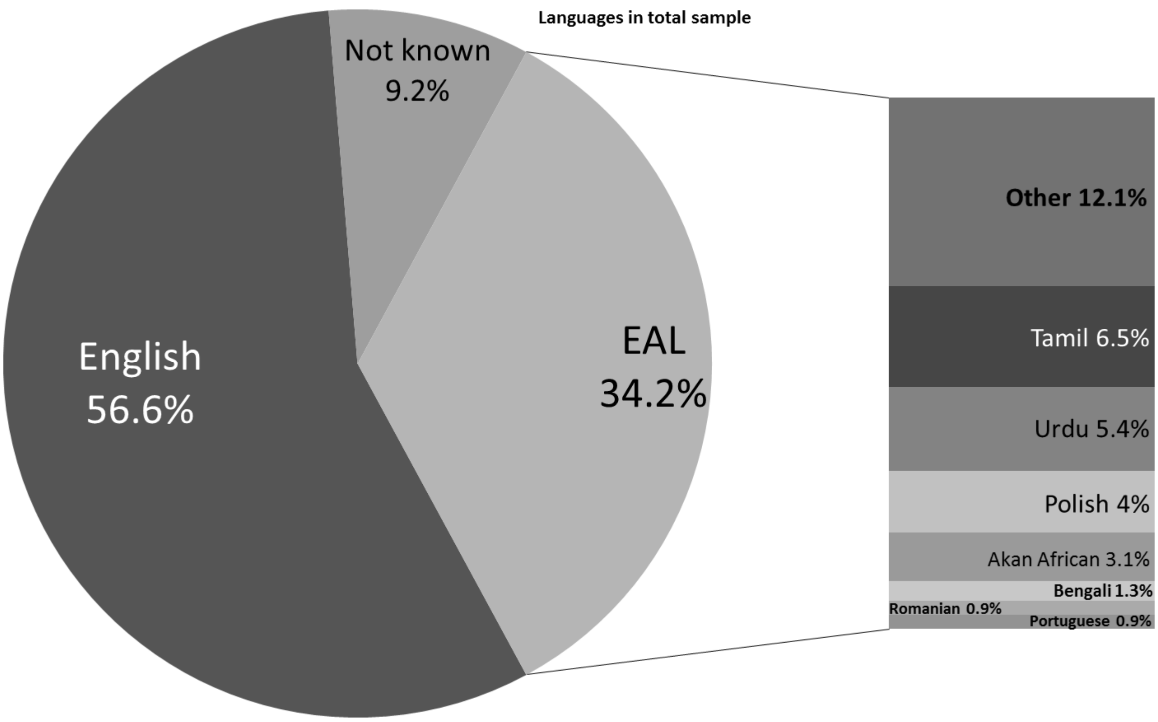

Figure 4. Distribution of %SS scores after conversion using Riley's (1994) table. Standard scores are plotted on the X axis and counts are given on the Y axis.

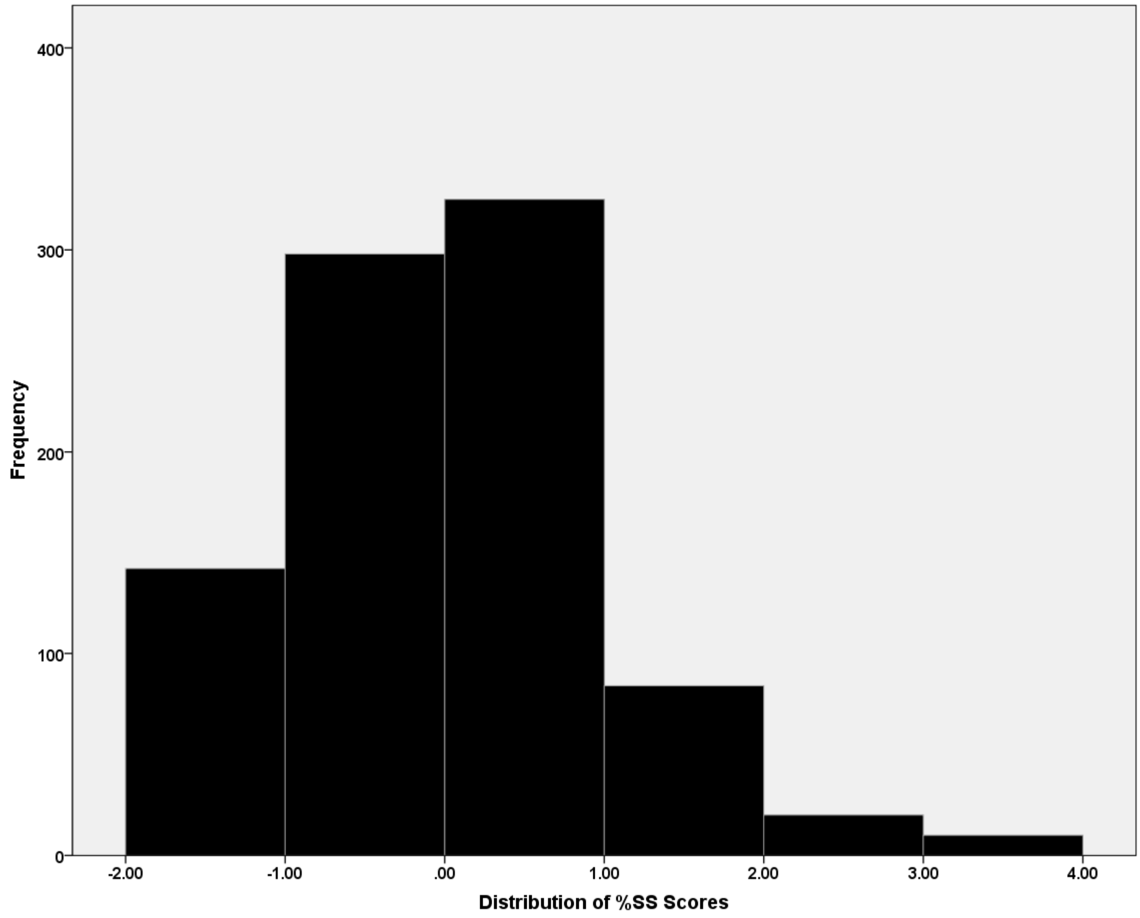

Figure 5. Distribution of duration scores after conversion using Riley's (1994) duration table. Standard scores are plotted on the X axis and counts are given on the Y axis.

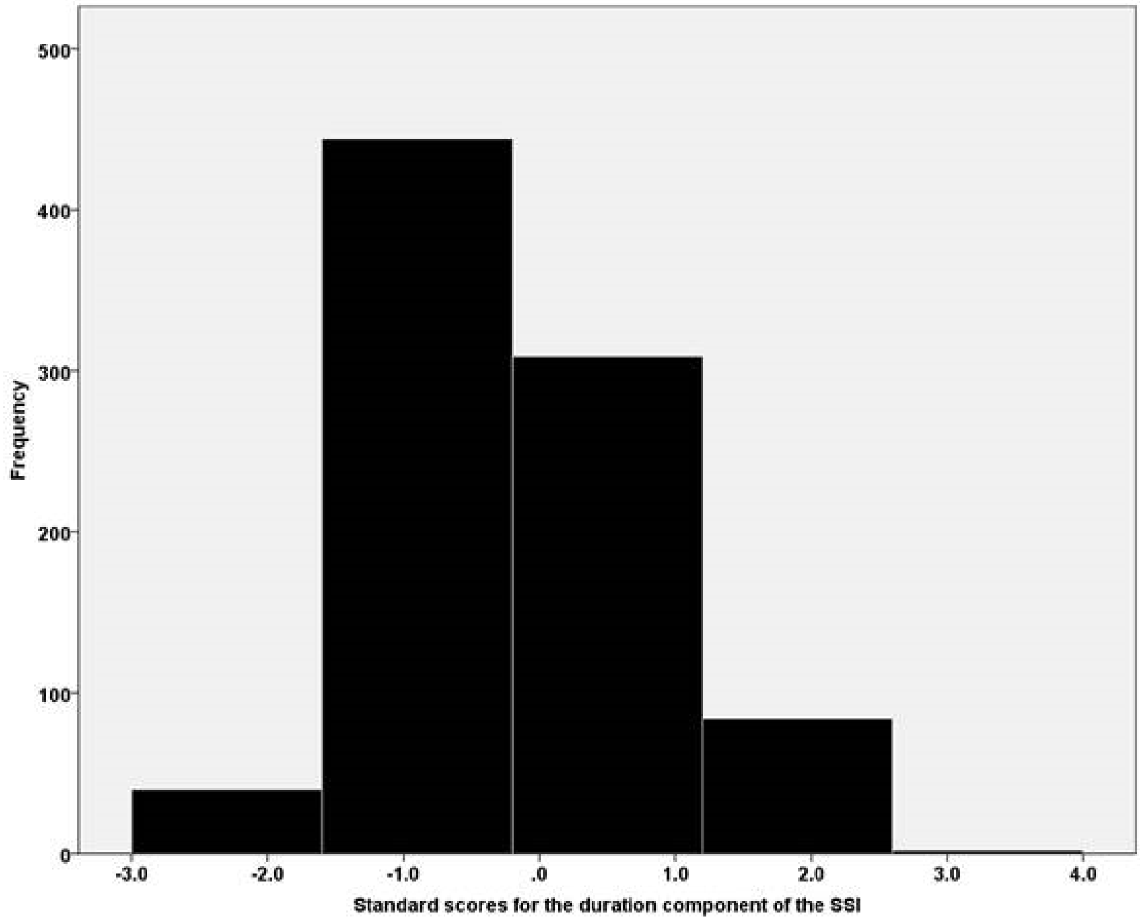

Figure 6. Distribution of raw physical concomitant scores according to Riley (1994). Standard scores are plotted on the X axis and counts are given on the Y axis.

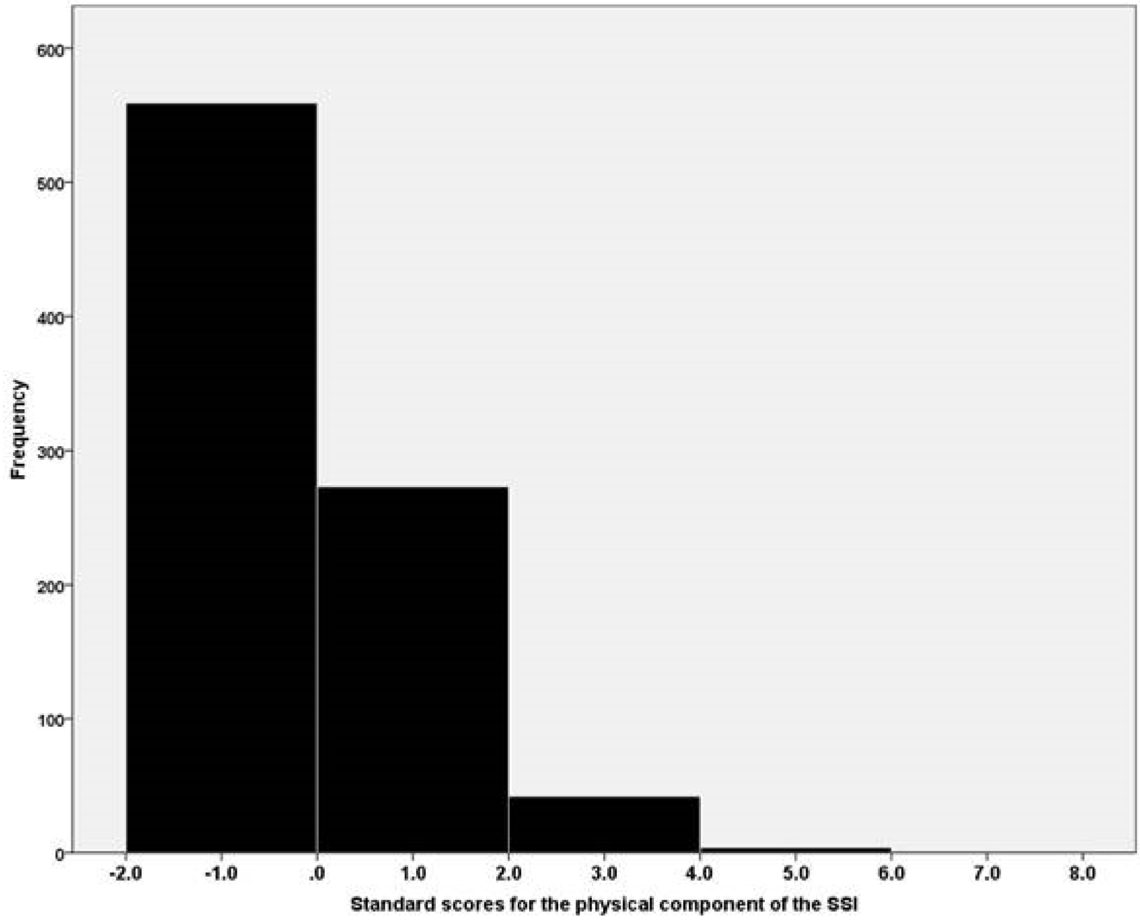

Figure 7. Distribution of overall SSI scores using Riley's (1994) procedure. Standard scores are plotted on the X axis and counts are given on the Y axis.

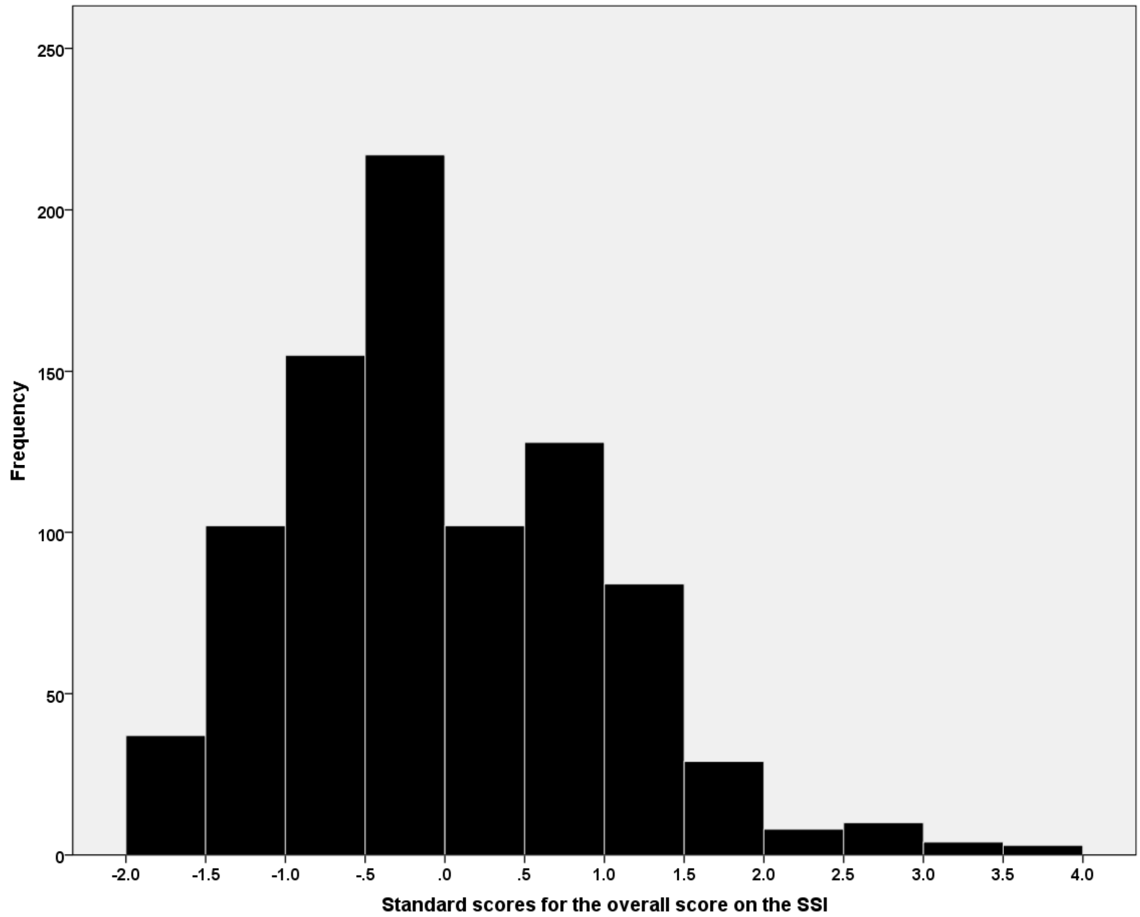

Figure 8. Scatter plots of raw %SS (frequency) scores (X axis) against overall SSI scores (Y axis). Circle symbols were used for plotting cases where children with speech difficulties (SSI score of > 16) and squares for children deemed fluent. As mentioned in the text, there are cases where children were deemed to have speech difficulties, although their SSI scores were below 16 which were based on schools' judgments.

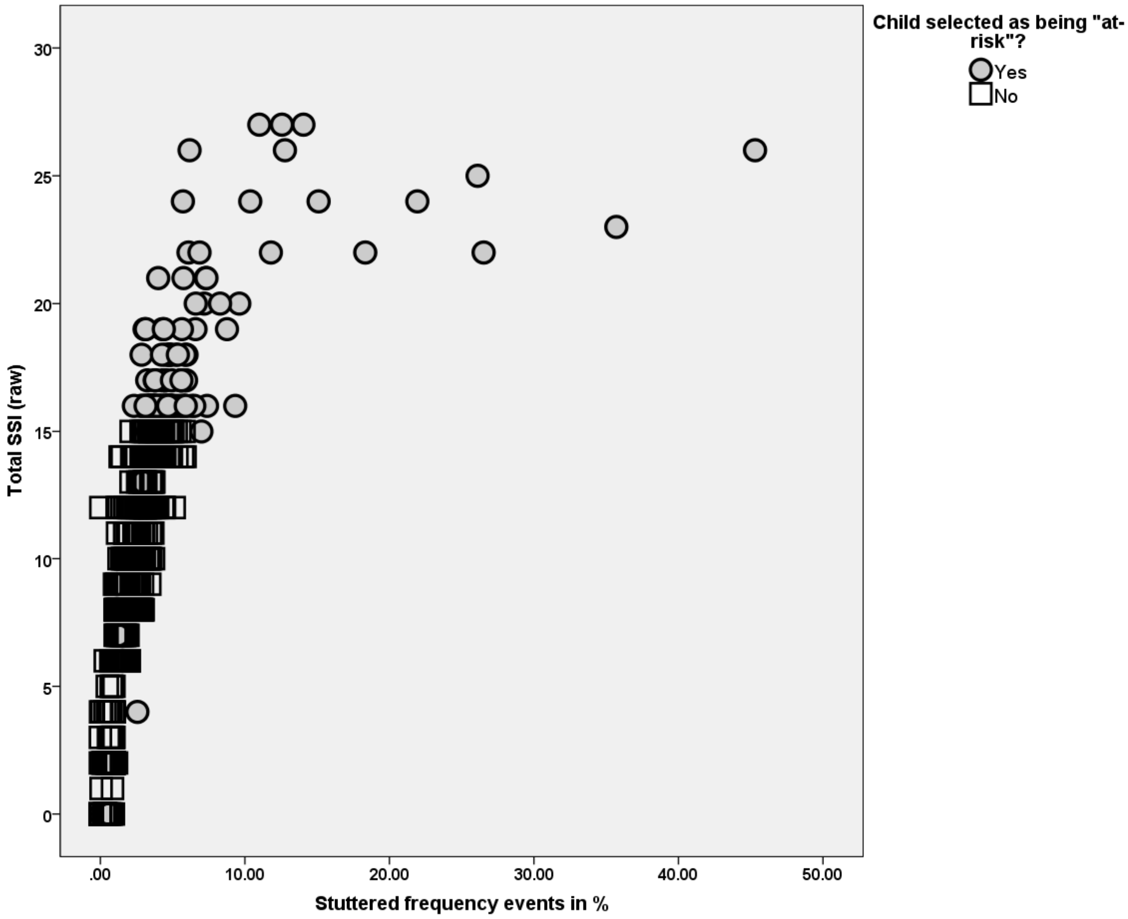

Figure 9. Scatter plots of raw duration scores (X axis) against overall SSI scores (Y axis). Circle symbols were used for plotting cases where children were designated as having speech difficulties (SSI score of > 16) and squares for children deemed fluent.

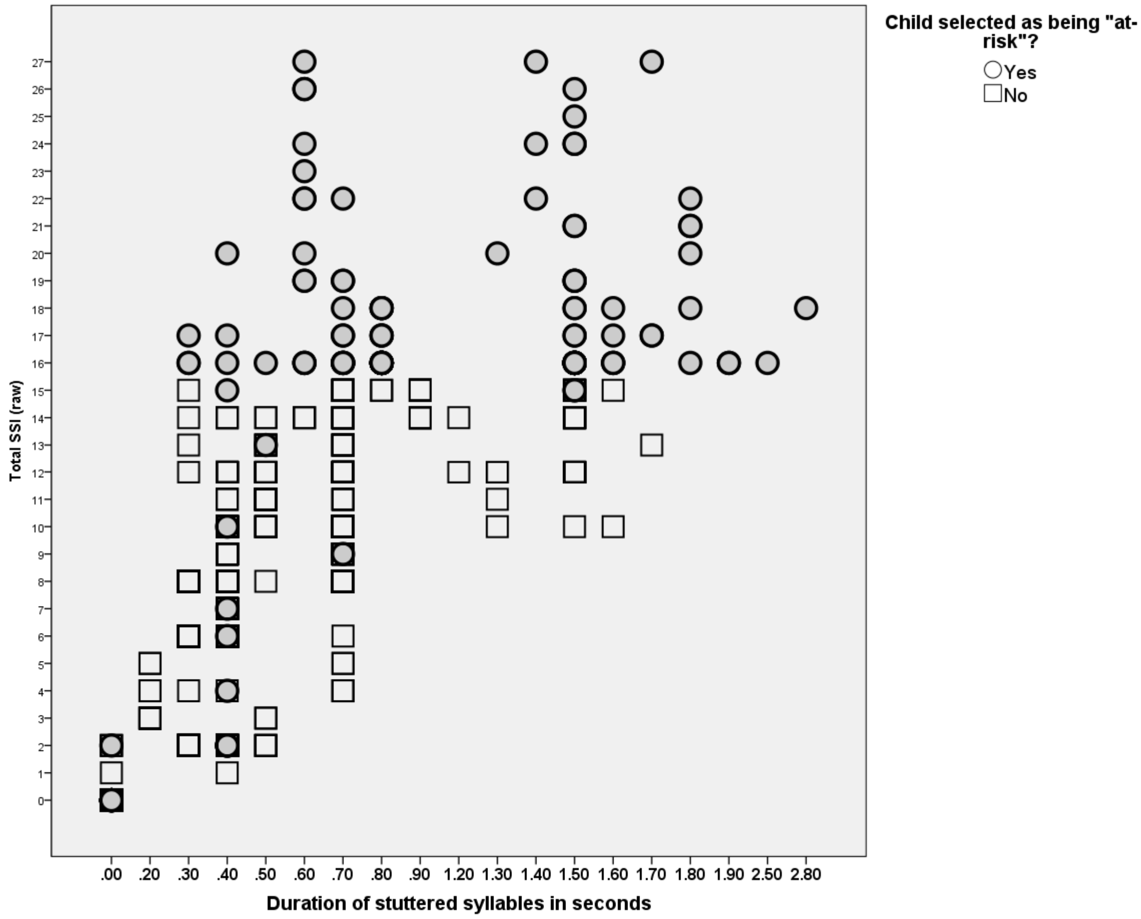

Figure 10. Scatter plots of raw physical concomitant scores (X axis) against overall SSI scores (Y axis). Circle symbols were used for plotting cases where children were designated as having speech difficulties (SSI score of > 16) and squares for children deemed fluent.