Assessing deaf children's writing in primary school: grammar and story development

Diana Burman, Deborah Evans, Terezinha Nunes, & Daniel Bell

Department of Education, University of Oxford

Abstract

Assessment is essential for designing individualized educational plans (IEPs) for children. In order to contribute effectively to this process, assessments must be appropriate for the group, show neither floor nor ceiling effects, and help teachers formulate specific aims. Different attempts to develop such assessments for the writing of deaf primary school children have shown floor effects. This paper reports the validation of an analytic instrument aimed at assessing deaf primary school children's writing and at contributing to the design of IEPs. Participants (N=167) were deaf children in the age range 6y6m to 13y11m, who were attending special schools or mainstream schools with units for the deaf; orally educated deaf children fully integrated in mainstream schools were not included. The writing samples were elicited by means of a story picture sequence. The dimensions of analysis included 16 items related to aspects of grammar and story development. A single and reliable scale was formed by these items, generating one score with a normal distribution. High inter-judge reliability, high test-retest correlation and a high correlation with reading comprehension were observed. We conclude that the assessment is a useful, reliable and valid instrument for analysis of deaf children's writing.

Key words: writing assessment, deaf primary school children, validation of literacy measure

Assessing deaf children's writing in primary school: grammar and story development

Our aim in this paper is to describe the development and validation of an analytic instrument designed to assess the writing of severely or profoundly deaf primary school children attending schools for the deaf or units for the hearing impaired in main stream schools. Although educators (see , for example, Isaacson, 1996; Powers, Gregory, & Thoutenhoofd, 1998; Tur-Kaspa & Dromi, 2001) agree that assessment is a crucial element for setting individualized educational plans for students, and in particular for deaf students, reliable and valid assessments for analysing the development of deaf primary school children's writing are still missing. Such an assessment would also be of great value for research and general monitoring of deaf children's development.

The assessment was not intended for use with children with mild hearing loss who can be integrated in mainstream schools. Previous research (Tymms, Brien, Merrell, Collins, & Jones, 2003) has shown that children with mild hearing loss do not differ significantly from hearing children in literacy assessments developed for hearing children and adapted for the deaf only in so far as the instructions are concerned.

We have previously described (Burman, Nunes & Evans, 2007) an assessment designed for this same group, which used a holistic approach, similar to that used in the English National Curriculum guidelines. In order to develop these holistic assessments, six experienced teachers of the deaf classified a sample of deaf children's stories into five levels of achievement. We then analysed the characteristics of the stories that had been classified at each of these levels and made explicit the criteria that were used to produce these classifications. The criteria included children's consistent use of units of meaning larger than isolated words, such as verb

and noun phrases, the use of English word order, the inclusion of articles and other words that exist in English but not in BSL(British Sign Language), such as the verb "to be", the use of pronouns and of inflected verb forms to indicate tenses, and the use of a varied vocabulary, including adjectives and adverbs. These criteria were identified through qualitative analyses of the productions classified in different levels of sophistication rather than explicitly enunciated by teachers at the outset. Finally, we observed that it is possible for teachers to learn to use the characteristics that we identified as achievements at the different levels and to produce reliable ratings of deaf children's texts.

Holistic assessments are familiar to teachers, who often mark children's writing in this way. However, it is more difficult to obtain judgements that are reliable across judges when holistic assessments are used than when analytic instruments are used, and holistic scores make it more difficult to identify students' strengths and difficulties in writing (Heefner & Shaw, 1996). So, after designing a holistic assessment procedure, we decided to move on to the design of a more analytic assessment, that could be used by teachers reliably with less training and that could help identify more precisely the strengths and difficulties of deaf primary school children.

Our brief review of past work on analytic approaches to assess deaf children's written productions has two aims: to critically analyse this past work and to identify possible criteria that can be used in the development of our assessment.

From this analysis of the literature, we identified two major dimensions for the classification of the analytical levels used in previous research, which will also be the two dimensions that we use in our analytic score: syntactic criteria and criteria related to the message quality. We review each of these in turn.

<u>Analytic assessments based on syntactic criteria</u>

Past attempts to design analytic instruments for our target group have not been successful because the criteria proposed for assessment are often above what is accomplished in a large proportion of the written productions by profoundly deaf primary school children. Thus they fail to discriminate between different levels of production and place many productions at floor level.

Yoshinaga-Itano and Snyder (1985), for example, proposed the following criteria: (1) number of sentences and words used in the composition; (2) complexity of syntactic form used and composition development; (3) analysis and categorization of errors made in composition; (4) quantitative use of different parts of speech; and (5) quantitative analysis of types of transformational grammar structures. Even a brief inspection of the writing samples produced by 10/11-year-old profoundly deaf students presented in our previous work (Burman, Nunes, & Evans, 2007) will reveal the inadequacy of these criteria. Figure 1 presents sample paragraphs from the beginning of stories classified at our top level of performance, which included approximately 7% of the stories analysed. These criteria could be applied (though with difficulty) to the texts presented in Figure 1, but it should be quite clear that there will be obvious obstacles when one tries to apply them to less structured productions. Although it is possible to count the number of sentences in these productions – the criterion which is the most straightforward from those proposed by Yoshinaga-Itano and Snyder – the analysis of variation in complexity of syntactic forms will show a floor effect. Even the best stories produced by the children in our sample show very little variation in syntactic complexity. It is necessary to search for criteria that describe smaller steps in the development of written language in order to obtain an appropriate assessment of deaf children's written productions.

Figure 1 about here

Powers and Wilgus (1983) also developed an analytic schema based on syntactic criteria for assessing deaf children's writing. Their analysis focused on syntactic patterns and included four levels: (1) repetitious use of a single pattern; (2) use of a variety of simple sentence patterns (e.g., Subject-Verb-Object: S-V-O; Subject-to be-Adjective: S-V-A; Subject-Verb-Prepositional Phrase: S-V-P); (3) expansions adding an adverbial or gerundial phrase or use of compound sentences (e.g. with the word "and"); (4) complex sentences, including embedded subordinate clauses). Although this is a clearly described analytical schema, inspection of the examples presented in Figure 1 indicates that most deaf children in our sample would at best be placed in Level 1, but the vast majority would not show a written production to attain the criteria specified for Level 2. This scale would produce a floor effect if applied to our sample, as the productions in Figure 1 exemplify the best writing we observed. There were very few examples of expansions through the use of compound sentences and no example of the use of embedded subordinate clauses in our writing samples. However, their first two items, the use of English sentence patterns and the use of S-V-O order allow for the identification of some elements of English grammar in written productions of deaf primary school children.

Later, Isaacson (1996) attempted to adapt Clay's (1991; 1993) descriptions of the development of hearing children's literacy to the assessment of deaf children's compositions. Clay's description was created for assessing preschool hearing children's literacy development, and included three aspects: <u>directional principles</u>, <u>language level</u> and message <u>quality</u>.

Directional principles – that is, writing from left to right and top to bottom – seems to be mastered by deaf children in ways that are very similar to hearing

children's progress (Mayer, 2007); we did not observe reversals of directional patterns even in the simplest of the written productions in our sample. Clay also includes under this analysis the use of spaces between words, a criterion that was not met by all the children, as some produced strings of letters without spaces. We propose to use this criterion to differentiate between writing that can be analysed using our assessment criteria and those productions that cannot be analysed according to these criteria, because there is no word identification in the writing.

With respect to language level, Clay's description of productions at levels 3 and 4 seem to cover a transition that takes place over many school terms, if not school years, in the case of profoundly deaf children: level 3 is the ability to copy and level 4 the ability to compose a message with simple sentences, even if repetitive.

Similarly, the progress between two of the consecutive levels in the assessment of message quality for hearing children seems to take place over a long period for profoundly deaf children: the ability to write word groups (level 2) and the ability to write using simple sentences (level 3). Because this one-level transition would not be appropriate to describe the many steps that profoundly deaf children take to reach writing that uses predominantly simple sentences, we feel that a finer analysis is required for our purposes. However, both criteria seem useful and will be employed as part of our analytical assessment. The ability to write word groups will be defined as the use of noun-phrases and verb-phrases; the ability to write using simple sentences will be divided into smaller steps: the ability to use English S-V-O syntax and the ability to include further elements into this structure.

Heefner and Shaw (1996) investigated the use of the Six-Trait Analytical Scale (Spandel & Stiggins, 1990) for deaf children attending a large school for the deaf in the U.S. The majority of the children had severe to profound loss:

approximately 6% had a moderate loss. The procedure for using the scale is to present students with wordless story books by using slides of the pages, discussing the story development with the students, and then asking them to write a story.

Six dimensions of writing are considered: (1) ideas and content development; (2) organization (mostly presentation on the page); (3) voice; (4) word choice; (5) fluency; and (6) conventions. The last two aspects, fluency and use of conventions, are related to syntactic aspects of the written productions.

Raters must be trained to use the system and develop reliability in its use. For each of the dimensions, five levels are used: (1) beginning to use (struggling); (2) emerging ("moments that trigger" production, to use the authors' description; p. 155); (3) developing ("writer begins to take control"); (4) maturing (more control, confidence); and (5) strong (skilled). Raters are trained to use the intervals between these levels to describe intermediary performance and each trait is assessed independently.

Heefner and Shaw's sample of primary school children's production was large (201 written productions; the number of children is not clear because texts were collected from some children more than once) and included work by children in the age range 8 to 13 years. Unfortunately, their scale showed a marked floor effect: when the average overall ratings were considered, 66% of the children's productions were at the emerging level and the remaining 34% at the developing level. The scale on sentence fluency was able to measure progress across one school year for children in the first four years of primary school, but it is not clear whether the differences in the means were significant. However, in the subsequent four years of primary school (middle school, as it is termed in the United States) the scale did not pick up improvements within the year. The conventions scale did not show improvement in

primary school.

The total score for the six dimensions was validated by means of correlations with the Stanford Achievement Test, Reading Comprehension Sub-test: significant correlations were obtained for all eight sets of scores, with the smallest being equal to .52 but all others were equal to or greater than .6.

Thus it is possible to conclude that deaf children make progress in the use of syntactic aspects of written English during primary school but sensitive scales are necessary to measure this progress. The floor effect observed in this measure and its inability to pick up further developments of children's writing beyond grade 4 are cause for concern in the use of this measure. However, the criteria proposed can be used with more sensitive scales, which assume less progress between levels. This is what our different items attempt to accomplish.

Analytic assessments of message quality

As indicated earlier on, Heefner and Shaw's study also included three dimensions which related to message quality: ideas and content, voice, and word choice.

Isaacson (1996) and Welch & Link (1992) have contributed to the analysis of message quality by offering definitions and examples of elements that can be scored positively when analysing ideas and content. According to Isaacson, ideas and content refer to the <u>organisation</u> of the presentation (e.g. a logical sequence of events, a clear ending, ordering of ideas within paragraphs with the introduction of characters or events at the start, before other points are made) and <u>cohesion</u> (e.g. maintaining clear links across paragraphs, using personal and demonstrative pronouns appropriately and unambiguously). The analysis of ideas and content also includes <u>originality</u> (e.g. attempting humour, presenting a unique point of view). Welch & Link (1992)

proposed that each paragraph can be analysed and scored in three parts: topic sentence, supporting sentences, clinching sentence. Some of the criteria suggested by these researchers seem to be too advanced for the productions we observed in our previous study but it is possible to select some elements from this analytical approach for use in our investigation: for example, introducing the character in the story and the appropriate use of pronouns to maintain cohesion.

Voice refers to the taking of a perspective in a narrative. Heefner and Shaw suggest that it can be coded in three ways: (1) voice-as-guide; (2) voice-as-message; and (3) voice-as-perspective. Although we recognise the importance of this dimension of assessment, inspection of the productions presented in Table 1 suggests that it is unlikely that such judgements will be appropriate for the assessment of primary school deaf children's productions.

Word choice refers to use of vocabulary that demonstrates a sound knowledge of subtle differences between words (e.g. words with the same denotative meaning can have different connotations) and avoidance of repetition.

Although the suggestions made regarding the analysis of quality of the message in these studies are useful, we did not find that it was possible to implement them in this exact manner in the analysis of primary school deaf children's writing: similarly to what was reported for the dimensions of fluency and conventions, Heefner and Shaw observed floor effects for these traits. However, one aspect was identified for further use in our analytic criteria: avoiding repetition of the characters' names by use of pronouns, a criterion that was already identified from the work of Welch & Link (1992).

Harris & Graham (1992) offered more specific suggestions for analysing the message quality. They suggested that one should consider whether the following

elements are present: introduction of a main character, description of locale, statement of the time the story takes place, a precipitating or starting event, a goal formulated by the character in response to the starting event, actions taken to reach the goal, ending result, and final reaction of the character to the outcome of the actions.

In our analysis, we decided to opt for simpler indicators of message quality, starting from those listed by Harris & Graham: naming of the character, the use of pronouns to avoid repetition of a name, the use of appropriate and different pronouns, the addition of elements to the story not present in the picture (e.g. the destination of the travel, a time of day or day of the week, relationships between characters etc.).

In summary, various analytic instruments have been described in the literature for use with deaf students. Some reports do not include data on deaf children's performance but contain only a description of possible instruments (Isaacson, 1996); others have been used with older students, in secondary school (e.g. Gormley & Sarachan-Deily, 1987); and others have reported the validation of an instrument and included primary school children in the sample (Heefner & Shaw, 1996). The study by Heefner & Shaw seems to be the most rigorous one to date in the development of a scale to assess the writing of severely and profoundly deaf children in primary school. However, their instrument showed a marked floor effect with this population, and therefore further research that seeks to use scales describing smaller changes in progress in writing is justified. To date, the conclusion reached by Powers, Gregory, and Thoutenhoofd (1998) about one decade ago still seems valid: assessments appropriate for deaf children are scarce and no validated tests are available to assess the writing and literacy skills of severely and profoundly deaf children in primary school. The aim of this paper is to take a step to fill this gap in the literature.

Method

<u>Participants</u>

Participants were 167 children recruited from 22 special schools or mainstream schools with a unit for hearing impaired children. Their ages ranged from 6y2m to 12y9m, with a mean of 9y11m and a standard deviation of 1y5m. The level of hearing loss was: 67 profoundly deaf; 16 profoundly-severely; 30 severely deaf; 15 moderately-severely deaf; and 15 moderately deaf; 24 had cochlear implants so they are not included in this classification. The children (N=90) from 14 schools were retested within the maximum period of 2 months to obtain test-retest correlations; the remaining children were participating in another project and for practical reasons were not retested.

<u>Measures</u>

All the children produced texts elicited according to the procedures described in the subsequent section and also answered a reading comprehension assessment (see Burman, Nunes, & Evans, 2007) for the purposes of validation. Reading comprehension and writing are significantly correlated (Shanahan, 1984; Shanahan & Lomax, 1986; Parodi, 2006). This validation procedure was also used by Heefner and Shaw (1996).

<u>Procedure</u>

A sequence of four pictures was presented to the children by means of coloured printed posters. The first picture (see Burman, Nunes, & Evans, 2007) shows a man packing a suitcase, the second shows him putting things into the boot of a car, the third shows people in a car on the road, the fourth shows two children playing on a beach. Using a procedure similar to that described by Heefner & Shaw (1996), the teacher shows the posters and discusses with the children what might be happening in this story. The children are given booklets with the pictures printed on each page and

asked to write their story. The children are told that they can ask for the spelling of words that they might want to write and the teacher either fingerspells the word for the child, or provides the written version on a small card, which is placed orthogonally to the writing orientation on a page so that the child must orient the card before copying the word. The cards are attached to the child's written production so that it is possible to know later on that the words were provided.

Scoring

Using the samples we analysed in our previous study as a reference (Burman, Nunes, & Evans, 2007) and examples from the literature (specifically, the criteria identified by Clay (1991; 1993), Heefner & Shaw (1996) and Harris & Graham (1992) mentioned in the previous section), we identified 17 items to be scored, using ratings that varied from 0 to 4. These items and a brief explanation of the ratings are presented in Figure 2. Productions were awarded a zero score if there was no evidence of the use of a syntactic resource or of the optional resources that add to the coherence and quality of the message. They were attributed a 4 if the message showed systematic and correct use of the criterion, allowing for some errors.

It can be seen from their description that ratings were phrased in order to allow for summing the ratings across items: all ratings have 0 as the lowest level of performance and 4 as the highest level. Item 1 is not included in the score: its aim is to help the assessor decide whether or not it is worth carrying out the analysis by scoring the child's production with this instrument. Considering the nature of the criteria proposed, if a production does not show systematic use of writing that resembles words, the criteria will not be appropriate for its analysis.

Figure 2 about here

As indicated in our previous work (Burman, Nunes, & Evans, 2007), it is

necessary to interpret the children's writing in order to decide where the sentence boundaries are, in the absence of punctuation, and how an incomplete sentence might be rephrased. During the initial discussion of how scoring would be carried out, it became apparent that it was not possible to distinguish substitutions from omissions unambiguously, because the same chunk of words could be rephrased in different ways. Although omissions and substitutions have been used as distinct criteria in the literature (e.g. Tur-Kaspa & Dromi, 2001), it was decided that both substitutions and omissions signalled that the child had used incomplete or not perfectly grammatical sentences, and these mistakes would then be scored together. The use of the different criteria for ratings was refined by including descriptive words under the items and examples of written productions observed. A brief scoring manual was developed with explanations for how each item is scored. It is expected that assessors would receive training in order to achieve reliable scoring but that this training does not need to be extensive and could be carried out in one day.

The first item on the scoring sheet simply indicates whether the production can be usefully analysed by means of scores in the different items. If the written production does not systematically use groups of alphabetical letters separated by spaces, the items we developed would not be appropriate for its description. The remaining items form two groups, which can produce three scores: one for the use of grammar (items 2 through 11), one for the quality of the text (items 12 through 17), and one for the total production (all items, 2 through 17).

We included in the grammatical score only items that can be considered obligatory: sentences should contain noun and verb phrases; these should appear in appropriate order; articles and verb tenses should be used appropriately; prepositions and connectives should be used to link phrases and sentences; there should be no

substitutions or omissions; there should be no un-necessary morphemes (e.g. two verb inflections on the same verb) or words (e.g. two articles or prepositions before a noun).

We considered under text quality items that might be seen as optional elements: the use of words clearly relevant to the story; the use of pronouns instead of continuous repetition of the nouns; the inclusion of creative elements (such as naming the character, the place, the time, relationships between the characters, describing feelings and attempting humour); the use of colloquialisms and of direct speech.

These two dimensions have been consistently distinguished in the literature. In our analysis, we test whether they are truly independent by examining whether, when combined, they form an internally consistent scale.

Results

Figure 3 displays a sample story which received an average score, one which received a score one standard deviation below the mean and a third production scored as one standard deviation above the mean. These examples are included in order to provide a sense for the meaning of the scores but they will not be discussed in detail.

Insert Figure 3 about here

Reliability of the scoring procedure

Scoring was carried out on a sample of 60 written productions by two independent assessors (the first and second authors). The Pearson correlation between the three scores was r=.91 for the grammatical score, r=. 89 for the message quality and .94 for the total score. These correlations are all high and significant at .001 level, which indicates that the assessors were able to use the ratings reliably.

The scales of measurement and their internal consistency

The first item rated the children's production on the use of alphabetical letters,

with groups of letters separated by spaces, and with some of these groups forming identifiable words (even if misspelled). We considered that our items would not be useful for the assessment of children who could not attain this criterion. A total of 153 (92%) children were rated as attaining this criterion systematically and thus our assessment would be appropriate for them. This item was not included in the analyses described here but all the participants are included so that it is possible to consider whether there were ceiling or floor effects.

The grammatical scale

This scale was composed by 10 items, from item 2 through to 11. Table 1, which contains only these items, displays the means and standard deviations for each of the items included in the grammatical scale. The items have been ordered from the highest to the lowest means.

Table 1 about here

The items show a gradation in order of difficulty, as indicated by means that vary from low to high scores. This is a desirable feature of scales that aim at describing variation in a group of participants. In order to test whether it is possible to add these scores, we calculated the Cronbach's alpha reliability for this scale. This statistic indicates the internal consistency of a group of items. Values equal to or greater than .7 suggest that the items, as a group, measure the same dimension. As this is a statistic based on the overall correlations between the items and the total score, the maximum score of 1.0 is rarely observed. The alpha reliability for this grammatical scale was .91, which is very high. Such high reliability level is encouraging in a scale of measurement as this one, where items are not rephrasing the same content but do focus on different aspects of grammar.

The maximum possible score for this scale was 40 (10 items times the

maximum score of 4 on each item); scores varied from 0 to 37, which indicates that there was a good range of variation. The mean of 17.16 is close to 20, which is the half-way point in the scale, confirming that the distribution does not show floor or ceiling effect. The a standard deviation of 9.09 indicates that about 66% of the scores varied between about 7 and 24, once again showing that the scores of neither group were at the top nor at the bottom of the range of possible scores . The distribution was not significantly skewed (skewness = 0.09; standard error = 0.19; z = 0.47; as with an absolute value of 2 or more indicates a skewed measure, either approaching ceiling or floor effects). Thus there was no floor or ceiling effect for this scale, which suggests that this scale is a sensitive measure for this sample.

The message quality scale

This scale was composed by 6 items (items 12 through 13 in Table 4). Table 2 displays the means and standard deviations for the different items included in this scale. The items have been ordered from highest to lowest means.

Table 2 about here

The item with the highest mean is the use of relevant words, which must in fact be the easiest one to attain. Again, there is variation in level of difficulty, which is a desirable trait of a sensitive measure for a particular group.

Cronbach's alpha reliability for this scale was .85. Again, such high reliability level is encouraging in a scale of measurement as this one, where items are not rephrasing the same content but focus on different aspects of the quality of a text.

The maximum possible score for this scale was 24 (6 items times the maximum score of 4 for each item); scores varied from 0 to 22, a range that shows that there were productions at all the different levels of the scale. The mean was 7.05, which is close to the value of six, which marks the first quarter of the possible range

16

of scores; this suggests that the scores in this scale are accumulated towards the bottom of the scale, even though some productions show quite high scores. The standard deviation was 4.42. The distribution was significantly skewed (skewness = 0.86; standard error = 0.19; $z = 4.5$), as the scores were accumulated at the weaker end of the scale (i.e., more than 65% of the scores fell between 0 and 12, which is half of the maximum possible score). Thus the analysis of message quality should be revisited, if possible, to check whether it could be more sensitive by the use of smaller steps in the development of this aspect of deaf children's writing.

The total scale for assessing deaf children's writing

The total score was based on the 16 items that composed both scales. It is a matter of empirical test whether the two scales, though theoretically distinct, can be merged to form a single scale. Cronbach's alpha reliability for this scale was .94, which is very high. This high reliability indicates that it is appropriate to add the scores for the two separate scales in order to obtain an overall assessment of the children's writing.

The maximum score for this scale was 64 (16 items times the maximum score of 4 on each item). The scores for the children's overall performance varied from 2 to 59: thus they were observed along almost the whole possible range of scores. The mean was 24.66, which is between the bottom third and halfway along the range of possible scores; this indicates that the scores are not accumulated at the lower end of the scale, even though the mean is below the half way mark. The standard deviation was 12.97; this means that about 65% of the scores observed fell between about 11.5 and 37.5. The distribution was not significantly skewed (skewness = 0.29; standard error = 0.19; $z = 1.53$), even though there are fewer productions at the top than at the bottom end of the scale. Thus the overall scale for the assessment of the children's

written productions was sensitive and showed neither floor nor ceiling effects.

Test-retest reliability

The test-retest correlations were calculated for each specific scale and for the total scale. These correlations were based on the 90 children who were re-tested for this purpose. The Pearson correlation for the grammatical scale was r=.85, for the message quality scale was r=.56, and for the total scale was r=.82. All the correlations were significant at the .001 level; the results for the overall assessment of the children's writing and the grammatical scale are good but the result for the message quality suggests the need for improvement if this aspect is to be judged independently.

Validity

In order to validate this measure, we correlated the children's scores on the overall assessment with a measure of children's reading comprehension. Because four children could not be given the reading comprehension test, the Pearson correlation for these assessments was based on 163 cases. The correlation was r=.72, which is high and was significant at the .001 level. This provides evidence in support of the validity of this assessment.

Further analyses

There was sufficient variation in hearing loss across children to examine the correlation between their performance in the writing assessment and the level of hearing loss. We excluded from this analysis 31 children who had cochlear implants because it is difficult to place them on the same scale of hearing loss as the other children, who do not have an implant. We calculated a partial correlation, controlling for age, between the severity of the loss and the children's scores in the writing assessment. This correlation (based on 133 cases) was, as expected, negative (r = -0.26): the higher the level of the hearing loss, the lower the results of the writing

assessment. It was significant at the .002 level, even though it was low. We also ran an analysis of covariance, controlling for age, and compared the children who had cochlear implants with those who did not. This analysis did not show a significant difference between the two groups.

<p align="center">Discussion and Conclusions</p>

Previous research has shown that it is very difficult to design reliable and valid assessments of deaf children's writing when the children's level of hearing loss is such that they cannot be fully integrated in mainstream classes. There is often a floor effect because the steps between the levels used in the assessments of hearing children are too large for deaf children. It was hypothesised that a more sensitive scale could be developed, which would show a normal distribution and help teachers identify the smaller steps that deaf children need to take in order to achieve competent writing.

This study showed that it is in fact possible to construct such a scale. Our assessment of children's writing was based on two sub-scales, grammatical attainment and message quality. The items were reliably judged by two independent judges, showed high internal consistency, and high test-retest reliability. The overall assessment of the children's writing was validated by means of a correlation with a reading comprehension test, which was high and significant.

We believe that this scale can be a useful tool for teachers in designing individualized educational plans for teaching their pupils how to improve the quality of their writing. This is a very significant finding because presently deaf children's progress might not be recognised due to floor effects on available measures, a result that is frustrating both for the children themselves and for their parents and teachers.

References

Burman D.L., Nunes T. & Evans D. Writing profiles of deaf children taught through British Sign Language. Deafness and Education International. John Wiley & Sons, Ltd. 2007. 9:1, 2-23

Clay, M. M. Becoming literate: the construction of inner control. Portsmouth (New Hampshire): Heinemann, 1991.

Clay, M. M. An observation survey of early literacy achievement. Portsmouth (New Hampshire): Heinemann, 1993.

Gormley, K., & Sarachan-Deily, A. B. Evaluating deaf students' writing: a practical approach. The Volta Review. 1987. 89, 157-170.

Harris, K. R., & Graham, S. Helping young writers master the craft: Strategy instruction and self-regulation in the writing process. Cambridge (MA): Brookline Books. 1992.

Heefner D.L. & Shaw P.C. "Assessing the written narratives of deaf students using the Six-Trait Analytical Scale." Volta Review 1996. 98(1): 147-168.

Isaacson, S. "Simple ways to assess deaf or hard-of-hearing student' writing skills." Volta Review. 1996. 98(1): 183-199.

Mayer, C. "What Really Matters in the Early Literacy Development of Deaf Children". Journal of Deaf Studies and Deaf Education, 2007. 12 (2): 411 - 431.

Parodi G. Reading – Writing connections: Discourse-oriented research. Reading and Writing. 2007. 20: 225-250.

Powers, A. R., & Wilgus, S. Linguistic complexity in the language of hearing impaired children. The Volta Review 1983. 85: 201-210.

Powers S., Gregory S., & Thoutenhoofd, E. D The Educational Achievements of Deaf

Children – A Literature Review. 1998. Department for Education and Employment, Her Majesty's Stationery Office. Norwich.

Shanahan T. Nature of the reading-writing relation: An exploratory multivariate analysis. Journal of Educational Psychology 1984. 76: 466-467.

Shanahan T. & Lomax R. An analysis and comparison of theoretical models of the reading-writing relationship. Journal of Educational Psychology 1986. 78: 116-123.

Spandel, V., & Stiggins, R. Creating writers: linking assessment and writing instruction. New York: Longman. 1990.

Tur-Kaspa H. & Dromi E. Grammatical deviations in the spoken and written language of Hebrew-speaking children with hearing impairments. Language, Speech and Hearing Services in Schools 2001. 32: 79-89.

Tymms, P., Brien, D., Merrell, C., Collins, J., & Jones, P. Young deaf children and the prediction of reading and mathematics. Journal of Early Childhood Research 2003. 1: 197-202.

Welch, M., & Link, D. P. Informal assessment of paragraph composition. Intervention in School and Clinic 1992. 23: 145-149.

Yoshinaga-Itano, C., & Snyder, L. Form and meaning in the written language of hearing-impaired children. The Volta Review 1985. 87: 75-90.

Figure 1

Transcription of four children's first paragraph on a story describing a four-picture sequence.

1. One morning Daddy got a big bag for clothes inside He have a family went to the new house

2. The boy was pack the suit case because ready go to holidays

3. The boy was called Sam. Sam was pack the clothes because we goes outing.

4. The man was called Tom he put his clothes in pack because it was great a hot day

These productions comprise the top 7% of all the 58 stories analysed by Burman, Nunes, and Evans, 2007.

Figure 2

Items used in analysing the children's written productions

Does the child …				
1. Include spaces between groups of alphabetical letters to resemble words?				
0. No evidence (rakfleosanchF evkdmsormbir)	1. Beginning to (Evident once e.g. wmsm amsdmri)	2. Sometimes (Evident 2 or 3 times)	3. Mostly (Evidence present under most pictures)	4. Systematically and correctly (All writing resembles words)
2. Put words in subject-verb word order, e.g. 'mum put'/'boy go'?				
0.No evidence	1. Beginning to (Evident once)	2. Sometimes (Evident 2 or 3 times)	3.Mostly (Evidence present under most pictures)	4. Systematically; may have 1 or 2 errors (appropriate subject-verb order)
3. Form noun and verb phrases, e.g. 'clothes in car'/ 'going holiday'?				
0. No evidence (Uses isolated words, not forming noun or verb phrases)	1. Beginning to (Evident once)	2. Sometimes (Evident 2 or 3 times)	3. Often (At least ¾ of text shows some connection)	4. Systematically; may have 1 or 2 isolated words (appropriate noun-verb phrases)
4. Include appropriate prepositions, e.g. 'in'/ 'to'/ 'at' ?				

0. No evidence	1. Beginning to (e.g. in bag)	2. Sometimes (e.g. in bag/ in car/ on sand)	3. Often (include a variety of prepositions)	4. Systematically but a few errors allowed (appropriate variety of prepositions)

5. Use the articles 'the' and 'a' appropriately?

0. No evidence	1. Beginning to (Evident once)	2. Sometimes (Evident 2 or 3 times, not always appropriately)	3. Often (Often and mostly appropriately)	4. Systematically but a few errors allowed (appropriate use; few omissions)

6. Use connectives such as 'and', 'then'/ 'next'/ 'so'/ 'after'/ 'now'/ 'because'?

0. No evidence	1. Beginning to (Evident once)	2. Sometimes (Evident 2 or 3 times)	3. Mostly (include a variety of connectives)	4. Systematically but a few errors allowed (appropriate variety of connectives)

7. Use full-stops and capital letters correctly?

0. No evidence	1. Beginning to	2. Sometimes	3. Often when required	4. Systematically but a few errors allowed (e.g. Names and starting sentences)

8. Use verb tenses, e.g. 'go'/ 'went'/ 'saw'/ 'opened'/ 'was packing'?

0. No evidence	1. Beginning to (1 or 2 isolated changes in verb tense)	2. Sometimes (more than 2 changes in tense)	3. Often when required (a variety of tenses - some correctly)	4. Systematically but a few errors allowed (appropriate use of a variety of tenses)

9. Use punctuation (**" " , ! ?**) beyond full stops?

0. No evidence	1. Beginning to	2. Sometimes	3. Often when required	4. Systematically but a few errors allowed

10. Include substitutions or omissions, e.g. 'they are so happy to the beach'/ 'he went down and next to the door'?

0. Constantly (this includes single word writing)	1. Often (most sentences are missing words)	2. Sometimes (at most half the time)	3. Rarely (at most a quarter of the time)	4. No evidence

11. Include unnecessary words or morphemes, e.g. '**is** everything **is** locked'/"the a"/ "paided"?

0. Constantly	1. Mostly	2. Sometimes	3. Rarely	4. No evidence

12. Use words relevant to the illustrations?

0. No evidence dnejiri	1. Beginning to man	2. Sometimes Man bag car	3. Mostly Man bag boy door boot	4. Systematically and correctly Many appropriate words

13. Include appropriate pronouns, e.g. 'he'/ 'she'/ 'they'/ 'his'/ 'it'/ 'hers'/ 'their' ?

0. No evidence	1. Beginning to (using e.g. 'he' throughout)	2. Sometimes (using 2 or 3 different pronouns)	3. Often (include a variety of pronouns)	4. Systematically but a few errors allowed (appropriate variety of pronouns)

14. Include information beyond what is depicted, e.g. names (people and/or items), places, time?

0. No evidence	1. Beginning to ('Sam' or 'Dad')	2. Includes 2 or 3 examples	3. Includes many examples	4. Includes sufficient information to create a story

15. Include information on characters, feelings, intent, humour?

0. No evidence	1. Beginning to	2. Sometimes	3. Often	4. Includes sufficient information to create a story

16. Include colloquial language/ expressions e.g. 'far away'/ 'nearly there'/ 'stuff'

0. No evidence	1. Beginning to (1 or 2 examples)	2. Sometimes (3 or 4 examples)	3. Often	4. Systematically and appropriately

17. Include direct speech?

0. No evidence	1. Beginning to (1 instance; e.g. Be careful)	2. Sometimes (uses direct speech more than once)	3. Often (more than 2 times)	4. Systematically and correctly (e.g. "Be careful!" said Dad….)

Figure 3

Examples of three stories produced by children with different scores. Paragraphs were inserted between pages corresponding to the different pictures.

_____

Boy, age 10y 6m, score 11 (1 SD below the mean)

I Bag cloths Pack redj went to car.

I Bag wakll car Boot redj.

I car look sand smay.

I sand carsl.

No help for spelling was requested.

_____

Girl, age 11, score 24 (at mean)

The man is ph bag for hoilday. Then man put chothes in bag.

The man put bag in car. And go now.

They is there now. But is far far away.

The children make sandcathe. They is happy day.

No help for spelling was requested.

_____

Girl, age 10y8m, score 35 (1 SD above the mean)

one day I get my clothes in the case and Im going holiday today and Im going to seen

my Dad and my Mum I got lots of things to do to got watch and my tooth bouch and

my shoes my sunglass lots things in the case that a enough thing.

in boot I got enough case I got one case I close the boot I get in the car.

and I drive my old car and it was suning shieing and parked my old car I saw sea.

and I play with my sandcastle with Mum and may bairt. and play with my ball and

that ead of the story.

The spelling of the underlined words was provided by the teacher upon request.

_____

Table 1

Means (out of 4) and standard deviations by item for the grammatical scale (n=167)

	Mean	Standard Deviation
Absence of unnecessary words/morphemes	3.65	0.80
Use of subject-verb order	2.48	1.54
Use of noun and verb phrases	2.22	1.56
Use of different verb tenses	1.76	1.05
Use of full-stops and capital letters	1.64	1.36
Use of prepositions	1.56	1.22
Use of articles	1.46	1.28
Absence of substitutions or omissions	1.14	1.27
Use of connectives	0.96	0.97

| Use of varied punctuation | 0.28 | 0.78 |

Table 2

Mean (out of 4) and standard deviation by item for the message quality scale (N=167)

	Mean	Standard Deviation
Use of relevant words	3.50	0.91
Use of pronouns	1.19	1.13
Use of information beyond pictures	1.08	1.15
Description of feelings, intent, humour	0.76	0.91
Use of colloquialisms	0.59	0.82
Use of direct speech	0.37	0.87