

Assessing Deaf Children's Writing in Primary School: Grammar and Story Development

DIANA BURMAN, DEBORAH EVANS, TEREZINHA NUNES and DANIEL BELL
Department of Education, University of Oxford, UK

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. Copyright © 2008 John Wiley & Sons, Ltd.

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

INTRODUCTION

The aim was to describe the development and validation of an analytic instrument designed to assess the writing of severely or profoundly deaf primary school children who were attending schools for the deaf or units for the hearing-impaired in mainstream schools. Although educators (see, for example,

Isaacson, 1996; Powers et al., 1998; Tur-Kaspa and Dromi, 2001) agree that assessment is a crucial element for setting individualized educational plans (IEPs) for students, and in particular for deaf students, reliable and valid assessments for analysing the development of the writing of deaf primary school children 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 may be integrated into mainstream schools. Previous research (Tymms et al., 2003) has shown that children with mild hearing loss do not differ significantly from hearing children in literacy assessments that were developed for hearing children and have been adapted for the deaf only insofar as the instructions are concerned.

An assessment designed for this same group has been described previously (Burman et al., 2007), and this used a holistic approach similar to that used in the English National Curriculum guidelines. In order to develop the holistic assessments, six experienced teachers of the deaf classified a sample of deaf children's stories into five levels of achievement. The characteristics of the stories that had been classified at each of these levels were then analysed and the criteria that were used to produce these classifications were made explicit. The criteria included the following.

- 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 British Sign Language (BSL), such as the verb 'to be'.
- The use of pronouns and of inflected verb forms to indicate tenses.
- 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 being explicitly enunciated by teachers at the outset. Lastly, it was observed that it is possible for teachers to learn to use the characteristics that were 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 which are reliable across judges when holistic assessments, rather than analytic instruments, are used, and holistic scores make it more difficult to identify students' strengths and difficulties in writing (Heefner and Shaw, 1996). So, after designing a holistic assessment procedure, it was decided to move on to design a more analytic assessment, which could be used reliably by teachers with less training and which could help to identify more precisely the strengths and difficulties of deaf primary school children.

A 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 which may be used in the development of the assessment procedure.

From the analysis of the literature, two major dimensions for the classification of analytical levels used in previous research were identified, and these will also be the two dimensions that are used in the analytic score: syntactic criteria and criteria related to the message quality. Each of these will be reviewed in turn.

Analytic assessments based on syntactic criteria

Past attempts to design analytic instruments for the 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 of profoundly deaf primary school children. Thus, the criteria 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.

- Number of sentences and words used in the composition.
- The complexity of syntactic form used, and composition development.
- Analysis and categorization of errors made in composition.
- Quantitative use made of different parts of speech.
- Quantitative analysis made 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 et al., 2007) will reveal the inadequacy of the above criteria. Table 1 presents sample paragraphs from the beginning of stories classified at the top level of performance, which included approximately seven per cent of the stories analysed. The above criteria could be applied (though with difficulty) to the texts presented in Table 1, but it should be quite clear that there will be obvious obstacles

Table 1: Transcriptions of the first paragraph of a story describing a four-picture sequence, by four children.*

1. On morning Daddy got the big bag for clothes inside He have a family went to the new house
2. The boy was pack the suit cases because ready go to holidays
3. The boy was called Sam. Sam was pack the clothes Because we goes outing.
4. The man is called Tom he put his clothes in pack because it was great a hot day

*These productions comprise the top seven per cent of all the 58 stories analysed by Burman et al. (2007).

when we try to apply them to less-structured productions. Although it is possible to count the number of sentences in these productions — the criterion that is the most straightforward — analysis of variation in complexity of syntactic forms will show a floor effect. Even the best stories produced by the children in the sample showed very little variation in syntactic complexity. It is necessary to search for criteria which describe smaller steps in the development of written language in order to obtain an appropriate assessment of deaf children's written productions.

Powers and Wilgus (1983) also developed an analytic schema, based on syntactic criteria, for assessing the writing of deaf children. Their analysis focused on syntactic patterns and included four levels:

- repetitious use of a single pattern
- the use of a variety of simple sentence patterns (such as Subject–Verb–Object (S–V–O), Subject–Verb–Adjective (S–V–A) or Subject–Verb–Prepositional phrase (S–V–P))
- expansions adding an adverbial or gerundial phrase, or use of compound sentences (e.g. with the word 'and')
- complex sentences, including embedded subordinate clauses.

Although this is a clearly described analytical schema, inspection of the examples presented in Table 1 indicates that most deaf children in the 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 the sample, as the productions in Table 1 above exemplify the best writing 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 the writing samples.

However, the first two items, the use of English sentence patterns and the use of S–V–O order, of Powers and Wilgus (1983) allow for the identification of some elements of English grammar in the 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 the literacy development of pre-school hearing children, and included three aspects: *directional principles*; *language level*; and *message quality*.

Directional principles — that is, writing from left to right and top to bottom — seem to be mastered by deaf children in ways that are very similar to the progress of hearing children (Mayer, 2007); no reversals of directional patterns were observed, even in the simplest of the written productions in the sample. Clay (1991, 1993) 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. It was proposed to use this criterion to differentiate

between writing that may be analysed by use of our assessment criteria and those productions which cannot be analysed according to these criteria because there is no word identification in the writing.

With respect to language level, Clay's (1991, 1993) 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 is 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 which uses predominantly simple sentences, 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 and Stiggins, 1990) for deaf children attending a large school for the deaf in the USA. The majority of the children had severe to profound loss: approximately six per cent had a moderate loss. The procedure for use of 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:

- ideas and content development
- organization (mostly presentation on the page)
- voice
- word choice
- fluency
- 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 (Spandel and Stiggins, 1990, p. 155))

3. developing ('writer begins to take control')
4. maturing (more control, confidence)
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 (1996) 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 8–13 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% were 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 USA) 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 Subtest; significant correlations were obtained for all eight sets of scores, with the smallest being equal to 0.52 but all others were equal to or greater than 0.6.

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 may 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, the study by Heefner and Shaw (1996) also included three dimensions that related to message quality: ideas and content, voice, and word choice.

Isaacson (1996) and Welch and Link (1992) have contributed to the analysis of message quality by offering definitions and examples of elements that may be scored positively when analysing ideas and content. According to Isaacson (1996), ideas and content refer to the *organization* 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 *cohesion* (e.g. maintaining clear links across paragraphs, using personal and demonstrative pronouns appropriately and unambiguously). The analysis of ideas and content also includes *originality* (e.g. attempting humour,

presenting a unique point of view). Welch and Link (1992) proposed that each paragraph may 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 work, but it is possible to select some elements from this analytical approach for use in the present 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 (1996) suggest that it may be coded in three ways: voice-as-guide; voice-as-message; and voice-as-perspective. Although the importance of this dimension of assessment is recognized, inspection of the productions presented in Table 1 in the Results suggests that it is unlikely that such judgements will be appropriate for the assessment of the productions of deaf primary school deaf children.

Word choice refers to use of vocabulary which 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 about the analysis of quality of the message in these studies are useful, it was not possible to implement them in this exact manner in the analysis of the writing of deaf primary school children: similar to what was reported for the dimensions of *fluency* and *conventions*, Heefner and Shaw (1996) 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 and Link (1992).

Harris and Graham (1992) offered more specific suggestions for analysing the message quality. These workers suggested that we 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 the present analysis, it was decided to opt for simpler indicators of message quality, starting from those listed by Harris and Graham (1992): 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, and so forth).

In summary, various analytic instruments have been described in the literature for use with deaf students. Some reports do not include data on the performance of deaf children but contain only a description of possible instruments (Isaacson, 1996); others have been used with older students, in secondary school (e.g. Gormley and Sarachan-Deily, 1987); and others have reported the validation of an instrument and included primary school children in the sample

(Heefner and Shaw, 1996). The study by Heefner and Shaw (1996) 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 which seeks to use scales describing smaller changes in progress in writing is justified. To date, the conclusion reached by Powers et al. (1998) approximately a 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 the present study was to take a step to fill this gap in the literature.

METHOD

Participants

The study participants were 167 children recruited from 22 special schools or mainstream schools with a unit for hearing-impaired children. Their ages ranged from 6;2 to 12;9 (mean 9;11, standard deviation (SD) 1;5). The level of hearing loss was as follows.

- Profoundly deaf (67 children).
- Profoundly–severely deaf (16 children).
- Severely deaf (30 children).
- Moderately–severely deaf (15 children).
- Moderately deaf (15 children).

Twenty-four children 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 two months to obtain test–retest correlations; the remaining children were participating in another project and, for practical reasons, were not retested.

Measures

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

Procedure

A sequence of four pictures was presented to the children by means of coloured printed posters. The first picture (see Burman et al., 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 and 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 analysed in our previous study as a reference (Burman et al., 2007) and examples from the literature (specifically, the criteria identified by Clay (1991, 1993); Heefner and Shaw (1996); Harris and Graham (1992) mentioned above), 17 items were identified to be scored by use of ratings that varied from 0 to 4. These items and a brief explanation of the ratings are presented in Table 2.

The children's productions were awarded a score of '0' 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 may be seen from the description that ratings were phrased in order to allow for summing them 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:

Table 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 (rakfleosanchFevkdsormbir)
- 1 Beginning to (evident once, e.g. wmsm amsdmri)
- 2 Sometimes (evident two or three times)
- 3 Mostly (evidence present under most pictures)
- 4 Systematically and correctly (all writing resembles words)

[2] Put words in subject-verb order (e.g. 'mum put' or 'boy go')?

- 0 No evidence
- 1 Beginning to (evident once)
- 2 Sometimes (evident two or three times)
- 3 Mostly (evidence present under most pictures)
- 4 Systematically; may have one or two errors (appropriate subject-verb order)

Table 2: *Continued*

- [3] Form noun and verb phrases (e.g. 'clothes in car' or 'going holiday?')
- 0 No evidence (uses isolated words, not forming noun or verb phrases)
 - 1 Beginning to (evident once)
 - 2 Sometimes (evident two or three times)
 - 3 Often (at least three-quarters of text shows some connection)
 - 4 Systematically; may have one or two isolated words (appropriate noun-verb phrases)
- [4] Include appropriate prepositions (e.g. 'in', 'to' or 'at?')
- 0 No evidence
 - 1 Beginning to (evident once)
 - 2 Sometimes (e.g. 'in bag', 'in car' or 'on sand')
 - 3 Often (include a variety of prepositions)
 - 4 Systematically, but a few allowed (appropriate variety of prepositions)
- [5] Use the articles 'the' and 'a' appropriately?
- 0 No evidence
 - 1 Beginning to (evident once)
 - 2 Sometimes (evident two or three 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 two or three 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 (one or two isolated changes in verb tense)
 - 2 Sometimes (more than two changes in tense)
 - 3 Often when required (a variety of tenses — some used 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)

Table 2: *Continued*

- 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 (e.g. using 'he' throughout)
 2 Sometimes (using two or three different pronouns)
 3 Often (including 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 or people and/or items, places, time)?
 0 No evidence
 1 Beginning to ('Sam' or 'Dad')
 2 Includes two or three 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 or expressions (e.g. 'far away', 'nearly there', 'stuff')?
 0 No evidence
 1 Beginning to (one or two examples)
 2 Sometimes (three or four examples)
 3 Often
 4 Systematically and appropriately
- [17] Include direct speech?
 0 No evidence
 1 Beginning to (one instance, e.g. 'Be careful')
 2 Sometimes (uses direct speech more than once)
 3 Often (uses direct speech more than twice)
 4 Systematically and correctly (e.g. "Be careful!" said Dad ...')

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.

As indicated in previous work (Burman et al., 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 and 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 2–11), one for the quality of the text (items 12–17) and one for the total production (all items, 2–17).

Only items that may be considered obligatory were included in the grammatical score: 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 unnecessary morphemes (e.g. two verb inflections on the same verb) or words (e.g. two articles or prepositions before a noun).

Items that might be seen as optional elements were considered under text quality: 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.

Table 3: Examples of three stories, with different scores, produced by the children.*

Boy (aged 10;6, score 11 (1 SD below the mean), no help for spelling was requested:
 I Bag cloths Pack redj went to car.
 I Bag wakll car Boot redj.
 I car look sand smay.
 I sand carsl.

Girl (aged 11, score 24 (at mean), no help for spelling was requested:
 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.

Girl (aged 10;8, score 35 (1 SD above the mean), the spelling of the underscored words was provided by the teacher upon request:
 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.

*paragraphs were inserted between pages corresponding to the different pictures.

RESULTS

Table 3 displays a sample story that received an average score, one that received a score 1 SD below the mean, and a third production which scored as 1 SD 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.

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 r correlation between the three scores was $r = 0.91$ for the grammatical score, $r = 0.89$ for message quality and $r = 0.94$ for the total score. These correlations are all high, and significant at the 0.001 level, which indicates that the assessors were able to use the ratings reliably.

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). It was 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 of 10 items, from items 2–11. Table 4, which contains only these items, displays the means and SDs for each of the items included in the grammatical scale. The items have been ordered from the highest to the lowest means.

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 which aim to describe variation in a group of participants. In order to test whether it is possible to add these scores, the Cronbach's alpha reliability was calculated for this scale. This statistic indicates the internal consistency of a group of items. Values equal to or greater than 0.7 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 Cronbach's alpha reliability for this grammatical scale was 0.91, which is very high. Such a high reliability level is encouraging in a scale of measurement such as this one, where items are not rephrasing the same content, but focus on different aspects of grammar.

The maximum possible score for the scale was 40 (ten items multiplied by the maximum score of '4' on each item); scores varied from 0 to 37, which

Table 4: Mean (out of four) and SDs by item for the grammatical scale ($n = 167$).	
	Mean (SD)
Absence of unnecessary words or 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)
SD = standard deviation.	

indicates that there was a good range of variation. The mean score 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 effects. An SD of 9.09 indicates that approximately 66% of the scores varied between 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 (skewedness = 0.09; standard error = 0.19; $z = 0.47$; as 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 it is a sensitive measure for this sample.

The message quality scale

This scale was composed of six items (items 12–13 in Table 2 above). Table 4 displays the means (SDs) for the different items included in the scale. Items have been ordered from highest to lowest means.

In Table 4 the item with the highest mean is the 'use of relevant words', which must 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 0.85. Again, such a high reliability level is encouraging in a scale of measurement such 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 (six items multiplied by the maximum score of '4' for each item); scores varied from 0 to 22, a range which 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 6, which marks the first quarter of the possible range 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. SD was 4.42. The distribution was significantly skewed (skewedness = 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 scale should be revisited if possible to check whether it could be made more sensitive by the use of smaller steps in the development of this aspect of deaf children's writing.

Total score scale for assessing the writing of deaf children

The total score scale was based on the 16 items that comprised both scales. It is a matter of empirical test whether the two scales, although theoretically distinct, may be merged to form a single scale. Cronbach's alpha reliability for this scale was 0.94, 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 multiplied by 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 half-way along the range of possible scores, and indicating that the scores are not accumulated at the lower end of the scale, even though the mean is below the half-way mark. The SD was 12.97, which means that approximately 65% of the scores observed fell between 11.5 and 37.5. The distribution was not significantly skewed (skewedness = 0.29; standard error = 0.19; $z = 1.53$), even though there were 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 retested for this purpose. The Pearson r correlation for the grammatical scale was $r = 0.85$, for the message quality scale was $r = 0.56$, and for the total scale was $r = 0.82$. All the correlations were significant at the 0.001 level. The results for the overall assessment of the children's writing and the grammatical scales are good, but the result for the message quality scale suggests the need for improvement if this aspect is to be judged independently.

Validity

In order to validate this measure, the children's scores were correlated 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 r correlation for these assessments was based on 163 cases. The correlation was $r = 0.72$, which is high, and was significant at the 0.001 level. This provides evidence in support of the validity of the 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. Thirty-one children who had cochlear implants were excluded from this analysis because it is difficult to place them on the same scale of hearing loss as the other children who do not have an implant. A partial correlation was calculated, controlling for age, between the severity of the hearing

loss and the children's scores in the writing assessment. This correlation (based on 133 cases) was, as expected, negative (Pearson's $r = -0.26$): the higher the level of the hearing loss, the lower the results of the writing assessment. It was significant at the 0.002 level, even though it was low. An analysis of covariance (ANCOVA) was also run, controlling for age, and the children who had cochlear implants were compared with those who did not. This analysis did not show a significant difference between the two groups.

DISCUSSION AND CONCLUSIONS

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 hypothesized 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.

The present study showed that it is possible to construct such a scale. Our assessment of children's writing was based on two subscales: 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 IEPs for teaching their pupils how to improve the quality of their writing. This is a very significant finding because, at present, deaf children's progress might not be recognized owing to floor effects on available measures, a result that is frustrating both for the children and for their parents and teachers.

ACKNOWLEDGEMENTS

This project was made possible by grant number EDU/00372/G from The Nuffield Foundation. We are very grateful to the Foundation for this support, without which this project could not have been carried out. We are also very grateful to the teachers and children whose collaboration made this project possible.

REFERENCES

- Burman DL, Nunes T, and Evans D. Writing profiles of deaf children taught through British Sign Language. *Deafness and Education International*, 2007; 9 (1): 2–23.
- Clay MM. *Becoming Literate: The Construction of Inner Control*. Portsmouth, NH: Heinemann, 1991.

- Clay MM. *An Observation Survey of Early Literacy Achievement*. Portsmouth, NH: Heinemann, 1993.
- Gormley K and Sarachan-Deily AB. Evaluating deaf students' writing: a practical approach. *Volta Review*, 1987; 89: 157–170.
- Harris KR and Graham S. *Helping Young Writers Master the Craft: Strategy Instruction and Self-regulation in the Writing Process*. Cambridge, MA: Brookline Books, 1992.
- Heefner DL and Shaw PC. 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 students' 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 AR and Wilgus S. Linguistic complexity in the language of hearing impaired children. *Volta Review*, 1983; 85: 201–210.
- Powers S, Gregory S, and Thoutenhoofd ED. *The Educational Achievements of Deaf Children — A Literature Review*. HMSO: Department for Education and Employment, 1998.
- Shanahan T. Nature of the reading–writing relation: an exploratory multivariate analysis. *Journal of Educational Psychology*, 1984; 76: 466–467.
- Shanahan T and Lomax R. An analysis and comparison of theoretical models of the reading–writing relationship. *Journal of Educational Psychology*, 78: 116–123.
- Spandel V and Stiggins R. *Creating Writers: Linking Assessment and Writing Instruction*. New York, NY: Longman, 1990.
- Tur-Kaspa H and 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, and Jones P. Young deaf children and the prediction of reading and mathematics. *Journal of Early Childhood Research*, 2003; 1: 197–202.
- Welch M and Link DP. Informal assessment of paragraph composition. *Intervention in School and Clinic*, 1992; 23: 145–149.
- Yoshinaga-Itano C and Snyder L. Form and meaning in the written language of hearing-impaired children. *Volta Review*, 1985; 87: 75–90.

Address correspondence to: Diana Burman, Department of Education, University of Oxford, Oxford, UK. (E-mail: diana.burman@edstud.ox.ac.uk)