City Research Online

City, University of London Institutional Repository

Title:

Dynamic Assessment of Sentence Structure (DASS): Design and evaluation of a novel procedure for assessment of syntax in children with language impairments.

Authors:

Natalie Hasson, Barbara Dodd and Nicola Botting

City University, London

Abstract

This paper describes the development of a Dynamic Assessment task requiring implicit knowledge of syntactic structure. Sentence construction and syntactic organization are known to be poor in children with Specific Language Impairments, but little is known about the way in which children with SLI approach language tasks, and static standardised tests contribute little to the differentiation of skills within the population of children with LI. The Dynamic Assessment of Sentence Structure (DASS) was employed on 24 children aged 8-10, with an identified language impairment, who were tested 4 times, 4 months apart. A range of scores was elicited with no limiting ceiling or floor effects, and the test showed high internal reliability of $\alpha=.833$. Inter-rater reliability of the scoring was highly significant at 0.9. Concurrent validity was demonstrated by significant correlation with scores obtained on the CELF-3(UK) and predictive validity of the measure was also found to exceed that of the standardised test measure. Information about the ability of the children to use strategies and less directive prompts, and to transfer learning between items was elicited, and the information was thought to be useful by Speech and Language Therapists involved in their management.

Introduction

The Assessment of Language Impairment

The primary means of evaluating children's language abilities for the purposes of diagnostic identification of language impairment (LI) and for obtaining information to describe the impairment continues to be through the use of batteries of standardised, norm-referenced tests (Caesar and Kohler 2009). There are specific drawbacks to the use of standardised assessments with children with language impairments, for example that their test performance is frequently not reliable and representative, (Hasson and Joffe 2007; Hasson and Botting 2010, Law et al 1998), that performance may be influenced by processing limitations and working memory (Montgomery 2002), and that test procedures are not culturally fair or standardised on children for whom English is not their first language (Pena 2000). Nevertheless standardised norm-referenced tests are devised for the purpose of diagnosis and have been demonstrated to reliably classify children as language impaired. Assessment for clinical management and the planning of intervention, however, is considered to be a slightly different process, and one for which standardised tests are not specifically designed.

No language tests or measure is sufficient on its own to fully diagnose and describe LI in a child, and most do not attempt to formulate recommendations for intervention (Dockrell, 2001). Rutter (2008) highlighted the need for more dimensional assessments to inform the nature of the treatment that may be offered. Indeed, building a composite profile of abilities from standardised tests, case history data, observations and questionnaires would provide the best baseline from which to make management decisions, and despite the apparent reliance on standardised tests, clinicians in practice do incorporate case history data and their own observations into their assessments (Caesar and Kohler 2009). More holistic assessments

tend to address developmental progress up to the present, gauge risk factors, and evaluate functional and informal skills such as play, attention, non-verbal and interpersonal communication. They also contribute information about contextual understanding of language, for example the following of directions, and the use of language for various functions in conversation. However they do not usually address specific skills such as receptive and expressive vocabulary, semantics and syntax, and for these formal linguistic aspects, tests are the main source of information.

Alternative models of assessment have been proposed to take a broader view of the multiple risk factors implicated by Rutter (2008) in language learning and linguistic ability. Among these, Dynamic Assessment (DA) has emerged as an assessment that is also thought to be more culturally fair to those from different linguistic or cultural backgrounds, and may be more sensitive for measuring incremental change in those with language impairments (Hasson & Botting, 2010). A Dynamic Assessment is one in which the assessor actively intervenes in order to elicit an improved performance from the testee, that is a measure of his potential for learning (Haywood and Lidz 2007). In contrast to a static assessment, that measures the crystallized *product* of learning, the focus in DA is on the *process* of learning, and the responsiveness of the individual to instruction and learning (Haywood and Lidz 2007).

Dynamic Assessment

The theoretical basis for DA is attributed to the work of Vygotsky, and in particular, Vygotsky's notion of the Zone of Proximal Development (ZPD) (Vygotsky 1986). Although he did not provide detail of a DA format, Vygotsky described asking a child to imitate the solution to a problem; to complete an unfinished solution; to work with another more accomplished peer on a problem; and he also explained to the child the problem and the

solution for him to carry it out. Vygotsky also described scenarios in which he established the differing ZPDs in children who scored the same 'intelligence age' on Binet's test, and described the assessment of ZPD as more predictive of level of achievement than their current levels of development. He did not, however, proceed to test this assumption empirically, but laid the groundwork for development of DA procedures by later researchers.

Researchers in several parts of the world indeed pursued the operationalised assessment of the ZPD, which resulted in the development of a number of key DA methodologies. Prominent among these is the Graduated Prompting procedure proposed by Campione and Brown (1987). In summary, the procedure of Graduated Prompting consists of giving a child one item of a problem, (comparable to an item from a standard IQ test), to solve independently. If the child fails, the assessor systematically adds cues, consisting of successively more directive and specific hints, until the child achieves a solution to the task. A count of how many cues, or how much help the child needs to solve the problem is a measure of the child's ZPD. The procedure does not look at how many problems or items a child manages to complete, but instead learning and transfer are calculated by seeing how many hints the child needs to solve an item and subsequent similar items. Testing the ZPD like this involves detailed task analysis of a suitable set of tasks (Campione, Brown and Ferrara 1982). This methodology has been extended and adapted to assess learning of specific subject areas, and to examine strategies for problem solving and is central to many current European studies in DA (e.g., Resing 1997, Resing et al 2009, Hessels, Berger and Bosson 2008).

Independently and contemporaneously with the publication of Vygotsky's writings translated into English, Feuerstein was developing the theory of Structural Cognitive Modifiability

(Feuerstein, Rand and Hoffman 1980), based on his experiences teaching displaced children. Feuerstein believes that Mediated Learning Experience (MLE) wherein learning is intentionally guided and facilitated by a more experienced mediator, is the key to effective learning. Just as every individual has the opportunity to experience MLE from a parent, carer, older sibling or teacher, there is also the possibility that an individual may have experienced inadequate mediation of an aspect of learning, and has the potential to benefit from further appropriately targeted MLE. Thus the potential for modifiability is present to a greater or lesser extent in every individual, and the function of assessment is to identify what may be done to overcome 'cognitive deficiencies' (Feuerstein's term), or to redress the effect of inadequate MLE in the individual's past. The task of assessment is to determine the parameters of learning in the individual, as he relates to a specific task and in the context of MLE from a skilled and motivated assessor.

A further methodology relevant to the current study is the 'Testing the Limits' approach of Carlson and Wiedl (1992), that aimed to demonstrate the improved performance of children with learning disabilities when tests were administered under different conditions. The conditions most significantly altered were those permitting the assessor to give detailed and specific feedback to the testee during the test, in order to facilitate improved performance on subsequent test items. This principle was adapted into the 'clinical interview' procedure, in which children were encouraged to explain and reflect on their own responses during a test in order to better demonstrate their knowledge, and was used by Peña (2001) to assess vocabulary in children with English as an Additional Language (EAL).

The various methods have been adapted and combined in various ways to formulate procedures best suited to specific assessment tasks. For example Vye et al (1987) described a

'continuum of assessment' model in which static testing was followed by graduated prompting and finally by mediational intervention in order to both diagnose children requiring intervention, and assess in more detail their ability to learn from different interventions varying in instructional intensity.

DA of Language

Key studies into the use of DA to assess culturally and linguistically different children (CLD), and differentiate typically developing children from those with LI have been published by Elizabeth Peña and colleagues. Gutierrez-Clellen and Peña (2001) described the tendency of children from diverse cultural and linguistic backgrounds to under-perform on standardised tests, resulting in over-diagnosis of language impairment. DA was found to be a more culturally fair means of assessing the responses of such children to learning experiences, providing opportunities to familiarize them with test expectations and probe their responses. (Peña 2000). Similarly addressing receptive vocabulary, Camilleri and Law (2007) found a DA procedure to differentiate between children with normally developing language and those referred to SLT services, but to equate monolingual children and children with EAL, whose static scores on the British Picture Vocabulary Scales (BPVS) differed. In a follow up study, Camilleri and Law (in preparation) reassessed children from the original study after six months, and found the DA to be predictive of the change in receptive vocabulary, in the group of children identified as Low Scorers on the static BPVS. Peña et al (2006) also showed that a DA of narrative ability had advantages over static tests.

More recently, studies that have sought to use DA to inform further intervention, and make prognoses for improvement, include investigations of phonology (Glaspey and Stoel-Gammon 2007) phonemic awareness, (Spector 1992) receptive vocabulary, and syntax

(Olswang and Bain 1996; Gummersall and Strong 1999). These latter studies found that in some children, both typically developing (TD) and with language impairments (LI), even a simple manipulation elicited performances superior to that of the control TD children. However, no LI control children were used in the study by Gummersall and Strong (1999), so there is no means of estimating the ability of language impaired children to benefit from the intervention, and the study remains inconclusive. Finally, Peña and Gillam (2000) published a series of case studies illustrating the clinical information to be gained from DA and MLE in vocabulary, narrative and explanatory discourse. Procedures for assessment, mediation and scoring of responses were described, based on previous research by the authors into the areas studied, and findings were used to demonstrate how to use DA to determine needs and useful strategies for intervention programmes.

The current study

A hybrid methodology for Dynamic Assessment of the syntactic abilities of children with language impairments was devised. The method, piloted by Hasson and Botting (2010), adopted the use of the Sentence Assembly task from the CELF-3, and developed a novel administration procedure for the assessment of sentence formulation in a population of children with previously identified language impairments. The aim of the research was:

To formulate a valid and reliable procedure for the Dynamic Assessment of sentence formulation that may subsequently yield useful information for planning intervention for children with Language Impairments.

Specifically:

i)   To demonstrate a reasonable range of scores with no key floor and ceiling effects.

ii)   To establish good internal consistency and equivalent parallel forms to ensure change over time represents generalised learning.

iii)   To demonstrate good inter-rater reliability

iv)   To establish concurrent validity and predictive validity compared to a static criterion test

v)   To produce an assessment that is replicable and teachable, in order that any demonstrated utility could then be adopted by practising SLTs in the field.

Method

Construction of the Dynamic Assessment of Sentence Structure (DASS)

The dynamic procedure designed for the current study was a hybrid of graduated prompt and mediational procedures, with the training prompts and mediation directly interposed into the graduated prompt procedure, and without a static post-test measure. Pilot testing of the procedure used a test- train- retest format (see Hasson and Botting 2010), and revealed that the static post- test was not useful for some language impaired children who performed less well in the standardised test situation, than in the intervening training. It was therefore decided to rate performance in the training and not by means of post-test.

The structure of the test was fundamentally a Graduated Prompt Procedure (adapted from Resing 1997) that enabled quantification of the number of prompts required by the individual in order to solve the given language task. It was thought that techniques based on counting prompts or hints, as in the Campione and Brown model of Graduated Prompts, or on applying the identical standardised intervention in the training phase of the DA would improve the precision with which testers could establish *relative* learning ability of children (Vye et al

8

1987). This was considered important in the current study to demonstrate the variability within the population of children labelled as 'Language Impaired' (LI) who are heterogeneous in their ability, but score uniformly low on standardised tests. Furthermore, use of a Graduated Prompt procedure was shown to facilitate identification of strategy use by children from different backgrounds (Resing et al 2009), and identification of strategy use might contribute to future intervention planning.

While the essential number and nature of the prompts was standardised, administration was flexible within the prompt hierarchy, and cues were delivered in an individualized mediational style, enabling feedback to be given and responses to be probed. Mediated learning, as advocated by Feuerstein, is recommended for inclusion into the test procedure to facilitate the gathering of information which may help to formulate individualised interventions. In addition, restructuring the test situation to facilitate improved performance as demonstrated by Carlson and Wiedl (1992) would certainly elucidate the benefits of verbalization and feedback for the individual, and although this may not benefit all language impaired individuals, investigation of those who might benefit is advantageous. This combination of feedback adapted from Carlson and Wiedl's 'Testing the Limits' procedure (1992) and mediation (as recommended by Feuerstein) was intended to facilitate maximum transfer between items and times of testing, and also to highlight metalinguistic knowledge and strategy use by the participants.

Four parallel versions of the test were developed so that simple rote practice effects were controlled for.

The DASS language task -

The task was based on the Sentence Assembly subtest of the CELF-3 (UK) (Semel, Wiig and Secord 1987). This task was chosen for the present study as it enabled sampling of a number of underlying componential skills and processes, found in a pilot test (Hasson and Botting 2010) to be accessible through probing of responses as permitted by a dynamic style of assessment. Kahn and King (1997) similarly used the CELF Sentence Assembly task, giving no reason for its selection, but demonstrating its utility for accessing and assessing cognitive functions.

The materials utilized the same format as that in the CELF-3, in which words are presented visually, printed on a single card, (i.e. not separately in order that they could not be moved about manually) in random order, and the instructions require the child to arrange each set of words into two different sentences. Unlike the CELF-3(UK) in the present study, the grammatical structure of the possible sentences was controlled, requiring different linguistic constructions to be extracted, and presenting items in order of increasing difficulty, and/or increasing length/number of items in the sentence, for each grammatical structure (see Appendix I).

The DASS test items

The specific linguistic structures that were tested by the developed materials were selected in part to represent structures known to be impaired in the language of children with SLI, although it was not possible to accommodate some structures within the format of the test task. The order of presentation of structures was related to grammatical complexity and sequence of age of acquisition, within the constraints of the test format. This degree of manipulation differs from the original CELF-3 subtest, and was designed in order to facilitate training through the presentation and practice of the test items themselves.

Processes and strategies employed to construct the first sentence from the given words, in some examples facilitated construction of the second sentence, by simple reversal of semantically equivalent elements (e.g. the Mum is../ the Dad is..). Semantically reversible sentences have been shown to be of particular difficulty for children with SLI (Bishop 1997). Alternatively, the only possible second sentence may have been a question, requiring auxiliary or copula verb inversion, also an area known to be difficult for children with LI (Rice and Wexler 1996, van der Lely and Battell 2003).

The items were arranged in pairs with a common structure in each pair of items, but with the second increasing the level of difficulty in some way. This was either by increasing grammatical complexity by modifying the phrase structure, for example modifying the present continuous tense verb (*is painting*) to a future tense verb group (*is going to score)*, or by increasing the overall length and number of elements to be manipulated by the child, by adding arguments (e.g. *mum is eating*, to *mum is picking the flowers*). Strategies used to formulate two sentences from the words in one item should facilitate the solving of the subsequent item in the pair, i.e. if a question form was required in Item 1, Item 2 could also be solved by forming a question as one of the sentences.

Administration of the test

All 12 training items were presented, as they cover a range of grammatical structures that are in themselves a source of information about the child's linguistic knowledge. The tasks were presented in a given sequence, due to the increasing level of grammatical complexity of the items. In addition, Feuerstein (1991) points out that grading of items with regard to complexity enables bridging from concrete to abstract levels of functioning.

Items were presented to the child for him to solve independently, and cues were provided only when required to help the child solve the problem item. There were five levels of help available (see Table 1), based on the method employed by Resing (1993) with cues being prescribed in as far as the general type of information and nature of assistance is described, but mediated in a flexible and individualised way, dependent on the responses given by the child. For example, the level of help defined as Level 3, 'Finding Strategies' was mediated by a range of prompts such as 'Which word can we start with?' or 'Can you make a question?' or indeed a number of hierarchically arranged cues, each dependant on the responses of the child. Testing was adaptive, in that cues were applied only if and when required (Resing 1993).

The cues were graded from general metacognitive direction, or no specific prompt (level 1) to more specific strategy based suggestions (levels 2-3), breaking down the task into components and using specific feedback (level 4) and finally to item specific feedback and instruction (level 5; See Table 1). The examiner began with the least directive cue, and added successive, more directive cues only if needed by a child continuing to struggle to achieve the task. The examiner selected which questions to use based on the attempts made by the child, and scaffolded using further questions according to the responses given. The precise number of cues was not recorded, only the highest level of cuing used, according to the descriptors of the levels (see Table 1).

[Table 1 about here]

When a correct solution was presented, the examiner prompted the child for a reflective response, primarily a judgement of grammaticality and sense, as well as an evaluation by the child of the strategies he used, and the level of difficulty he experienced. This procedure was also described by Gutierrez-Clellen and Peña (2001) as a variation on the 'Testing the Limits' procedure, and found to enable children to better demonstrate their knowledge, and the

examiner to better understand the child's thinking and approach to problem solving. The procedure of prompting the child for a reflective response enabled the investigator to informally assess the child's metacognitive awareness, and use this information to make recommendations for the planning of intervention programmes.

Scoring of the procedure

The scoring was based on the level of cues required by the child for him to produce a correctly sequenced sentence. As there were two responses required for each item, (and 12 training items) the total number of cues required was based on a total of 24 responses.

The procedure was scored in several ways.

Quantitative measures enabled comparison of the relative needs for support of the individual children, and could be used to predict future learning.

i)   Total number of cues (max 24 items x 5 cue levels = 120).

ii)   Number of each level of cue required.

Qualitative information derived from the procedure was used to compile a report of the assessment that could subsequently be used by SLTs to inform intervention for the individual.

   i)   Identification of grammatical structures which caused the child the greatest difficulty

   ii)   The effect of amount of content (sentence length) and nature of semantic content on the child's construction of linguistic structures

   iii)   The child's ability to transfer, or generalise learning or strategies i.e. item - to item transfer, as well as which items benefited from transfer effects

iv)   The child's *metalinguistic* knowledge, ability to label, explain and manipulate linguistic concepts

v)   The child's *metacognitive* ability i.e. awareness of the processes and strategies that are used to solve the given task

The number of cues presented was recorded on a scoresheet as the procedure was carried out. The number of each level of cue could be totalled immediately after completion of the assessment session.

Evaluation of the Dynamic Assessment of Sentence Structure

Participants

Managers and SLTs from three trusts and an independent school agreed to take part in the evaluation of the new measure and referred children from their caseloads.

SLTs were asked to identify children who met the following criteria:

- Aged 8-10 years old, and in Years 4 and 5 of school.

- Known to Speech and Language therapy services or Language units/resource bases, or special school.

- Language disorders identified as the primary disorder, likely to score <1SD on a standardised language test.

- English as a first language.

In addition, SLTs were asked to exclude children with:

- Moderate or severe learning difficulties, (IQ<70).

- Hearing impairments.

- Attention deficit disorders.

In total, 31 consent forms were returned to the researcher via the schools, or the self addressed envelopes provided with forms. Permission to conduct the research at schools was obtained from the headteachers of each of the seven schools involved in the study. Testing of potential participants to confirm eligibility for inclusion in the study began at the start of the school term in September.

Measures to confirm eligibility

Although criteria for referral were fairly specific, children were screened to confirm eligibility for inclusion in the study. The following measures were employed:

1. The Coloured Progressive Matrices (CPM) (Raven, Court and Raven 1990) was used as a measure of non-verbal intellectual ability. Children were considered eligible for inclusion in the current study if they scored above the 25th percentile, signifying 'intellectually average' (Raven, Court and Raven 1990 p.CPM30). This ensured that the language difficulties of participants in the study were not secondary to global learning difficulties. Although those with global learning difficulties can be effectively assessed by DA methods, the current study aimed to assess the language abilities rather than the cognitive abilities of children with specific impairments of language.

2. The Strengths and Difficulties Questionnaire (SDQ) (Goodman 1997) was used to detect features of Attention Deficit Hyperactivity Disorder (ADHD). Questionnaires were distributed to class teachers or SLTs who were familiar with the children. Children scoring more than 7 on the items pertaining to hyperactivity were excluded on the basis that attention deficit altered the mediational needs of the children. Pilot testing showed

children with ADHD requiring a higher proportion of mediations addressing behavioural control (Hasson and Botting 2010).

3. CELF-3 (UK) (Semel, Wiig and Secord 1987) to confirm the presence of LI, by a score of 1.25 SD below the norm for CA.

Four children were excluded from the cohort on the basis of the screening tests. A further child was excluded as injury resulted in absence from school during the testing period, and another after the first phase of testing as prolonged illness resulted in intermittent school attendance. One further child was thought to be eligible on the basis of the screening tests, but was subsequently excluded on account of behavioural difficulties that prevented experimental tasks being completed.

The final sample consisted of 24 children, aged between 8;2 and 10;9 at the time of the first test. The mean age of the whole group was 9;3. There were 21 boys, with an age range of 8;2 to 10;9, and 3 girls with ages ranging from 8;2 to 10;7.

Six children were drawn from a special school for children with speech, language or communication needs, in South London. The remainder were referred from six language units or resource bases, two of which were from an Inner London borough, accounting for six children, and the remaining thirteen were from across East and West Hertfordshire.

The Dynamic Assessment of Sentence Structure was administered to the group of children with previously identified language impairments in order to establish sensitivity, reliability and validity of the test. For clinical purposes, only one administration of the test would be

required to inform intervention, although the same or alternative versions of the test may be used later to measure change in performance. Parallel versions of the test were developed so that the test may be used without concerns of simple rote learning and these were administered in randomised order to each child. Thus at each time point a number of children completed each of the parallel versions. Two further trials were used to investigate predictive validity. In this study validity was assessed using correlations with a criterion test (CELF-3). Whilst we acknowledge that this is not the only test of validity possible, given the tension between the dynamic nature of the test and the need for psychometric strength, this was the most appropriate method (Pring 2005 p.178).

Procedure

Approval was gained from Redbridge and Waltham Forest Local Research Ethics Committee and informed written consent was given by all parents of children in the study.

Appointments were made for all children to be assessed over two sessions, usually one week apart. In the first, the Raven's Coloured Progressive Matrices was carried out, followed by subtests of the CELF- 3, for the purposes of confirming eligibility as well as baseline performance. In most of the children, all six subtests of the CELF were not completed, and the remaining subtests were carried out in the second session. In the intervening week, participating Speech and Language Therapists or the teachers of the children were asked to complete the SDQ. The DASS procedure was carried out in the second testing session, after the CELF-3 was completed, and this component was videotaped. The DA took between 25 and 40 minutes, and in all children was completed in one session.

The nature of the DASS test procedure required that all the DA testing was carried out by the investigator. In order to control for experimenter bias, the assessments were videotaped, and

scored by an independent assessor. The scores from standardised tests, and tests for eligibility for the study (Time 1) were made available to the SLTs working with the children in their schools, as well as to the parents of the children. These results could be used by the SLTs to contribute to their planning of intervention. The results of the DASS were not made available to SLTs at this stage, but were released later as part of an additional study investigating the use of the information derived from the DA in informing intervention. All of the children continued in regular ongoing intervention planned and administered by the SLT in their school, for a period of one school term, before the tests were repeated. The DASS and the CELF-3(UK) were then repeated, (T2) and the results compared to the baseline results in order to gauge the validity of the DASS, the reliability of the test, and sensitivity of the tests to change over time. Subsequently, retests were carried out after two further school terms (T3 and T4) in order to ascertain predictive validity of the initial DA. The results of the DA were made available to SLTs in the form of a written report containing the information described above as 'qualitative'. The SLTs were then asked via questionnaire whether they found the information helpful.

An independent rater was given approximately one hour's explanation of the cue rating system used in the test, and practice using examples taken from the recording of one child. Criteria used for cue levels were explained and discussed. She was then asked to rate a sample of the tapes for practice and to measure inter-rater correspondence with the tester. A sample of 25% of the tapes at Time 1 were used to establish consistency of definition and to identify any ambiguities in the scoring criteria. Following Time 3, the independent assessor scored the DAs of all of the participants (100%).

Results

i) <u>Spread of scores at baseline</u>

The scores of the whole cohort at T1 were widely distributed, with a mean score of 61.83 and standard deviation of 20.72. The distribution is shown in Figure 1. Although the distribution is skewed to the lower end somewhat, this is important given the language impaired nature of the sample. No limiting floor or ceiling effects were found.

[Figure 1 about here]

ii) <u>Internal consistency of the test and parallel versions</u>

Internal consistency of the test

Internal consistency was recommended as a measure of reliability by Swanson (1995) for procedures in which test-retest reliability is variable due to the expected change in retest score, such as in DA. Internal reliability of the DASS was tested using responses of all participants to the first item in each pair, i.e. the number of cues required to arrange one sentence from the given words. This was in order to assess the consistency of difficulty of the grammatical structure of items. The second sentence produced was more dependent on metalinguistic knowledge and strategic problem solving. The responses of 24 participants at T1 to 12 sentence items produced an internal reliability Cronbach's α=.833

Equivalence of parallel versions of the test

For the purposes of evaluating reliability of the DASS over time, and monitoring improvements in the children over time four parallel versions of the test were created. Although it was considered that all versions were equivalent, mean scores obtained by all

children taking each version of the test at Time 1, were compared. The results of univariate ANOVA with one between factor (version of test, A, B, C or D) indicated no statistically significant effect of versions of the test, (F (3,20) = .81, p = .503), confirming equivalence between the 4 versions administered.

iii) Inter-rater Reliability

An independent rater was supplied with 6 randomly selected videos of the DAs from the cohort at Time 1. Each item in the test was rated on a scale of 1-5 pertaining to the level of assistance given (as described above). A significant correlation was found between the total scores for each participant (n=6) for each rater (rs = .886, p = .019).

Scores for all participants at Time 3 were rated independently by the same rater. Significant correlation was found between ratings by the experimenter and independent rater (rs = .902, p< .001). Correlation between ratings by the experimenter and the independent rater are shown in Figure 2.

[Figure 2 about here]

iv) Concurrent and predictive validity

DA at T1 in comparison to CELF

A significant correlation was found between participants' scores on the DASS and total raw scores on the CELF-3 (UK) at Time 1 (rs = -.481, p = .017) suggesting a relationship between performance on a standardised test of language and the test of learning potential. The correlation is in a negative direction as favourable performance on the CELF obtains a greater score, while stronger performance on the DASS is shown by need for fewer prompt cues and

hence a lower score. The moderate correlation observed would be expected as the CELF-3 links only with the language knowledge component of the DA.

[Figure 3 about here]

DA scores at Time 2

Table 2 shows the shift in scores from Time 1 to Time 2 in the DASS task. The sample mean has decreased from 61 to 52, reflecting the overall improvement in performance of the group, with similar variance.

[Table 2 about here]

A Wilcoxon test shows that DA scores at T1 and T2 are significantly different, ($Z = -2.786$, $p = .005$), but a highly significant correlation was found between the scores at T1 and T2 (rs $= .706$, $p < .001$), reflecting reliability of the procedure and sensitivity to change over time.

[Figure 4 about here]

DA at T2 in comparison to CELF

A significant correlation was found between the scores obtained by participants on the DASS at T2 and the raw scores on the CELF-3 (UK) at Time 2, (rs $= -.705$, $p = < .001$) suggesting a mapping of gains on the standardised test of language with those on the test of learning potential.

DASS as a predictor of gains in therapy over time

Dynamic Assessments have been used to increase the predictive validity of assessments by identifying the potential of an individual to benefit from intervention. In the current paradigm, the number of prompt cues required in the DASS procedure is regarded as a measure of the individual's ZPD, or their learning potential. Given that there was opportunity for all

participants to continue in prescribed and individualised intervention for the duration of the study, the DA could be related to gains made during that intervention. Progress in therapy was measured by gains on the raw scores of the CELF-3. This measure was used as the standardised administration permitted comparison of scores at each time. Practice effects were presumably present, but according to the principles of DA, ability of a child to learn from and improve on account of previous exposure to stimuli suggests good learning potential, and is a feature to be included, rather than avoided. Raw scores on the CELF were used as they were a more sensitive measure of change in children who gained a few points in a subtest, that was insufficient to alter the standard score for the subtest, and even less likely to be reflected in a total standard score or percentile for the whole test.

A significant correlation was found between the DA at T1 scores and the change in total raw score on the CELF-3 from T1-T4 (rs = .534, p = .033), but not between the CELF-3(UK) at T1 scores with the T1-T4 change scores (rs = .105, p = .7). It was found that the correlations of DA1 scores with CELF change scores at T2 and T3 were not significant, but the longer term outcome was predicted. Furthermore, eleven children improved sufficiently to alter their percentile ranking on the CELF, by Time 3 or Time 4 of the study period. All of these children had scores in the lower range, i.e. below 60 on the DA at Time 1. No children who had higher range DA scores, i.e. poorer learning potential, improved from their lowest percentile ranking. This suggests that the greater learning potential identified by the DA at T1 is predictive of capacity to improve in intervention.

v)   <u>Utility in practice - Feedback from SLTs</u>

A brief questionnaire was used to find out whether the participating SLTs found the information contained in the children's reports after the DA, useful. Responses were received

from eight out of ten SLTs, from six of the seven schools involved in the study. Responses to some of the questions are summarised in Table 3.

[Table 3 about here]

Discussion

Evaluation of the DASS

The results of the trial of the DASS were examined to evaluate sensitivity of the test to individual differences and spread of scores. Evidence was sought that the test would identify that the participants had a range of abilities which could be differentiated by the task. The results indicated an appropriate range of difficulty of the procedure for the population described and the scores suggest sensitivity of the test measure to differentiate within participants in the population. In contrast, standard scores on the CELF-3(UK) reflect significant floor effects, with 20 of the 24 participants scoring the lowest possible total standard scores. Furthermore, results suggested that the content of the DASS was appropriate to elicit a wide range of responses from the cohort, and the anagram task was one that was suitably challenging, but not too difficult for the children to achieve any measure of success. The amount of information gleaned about the children's linguistic knowledge was extensive, and in addition, the procedure enabled the examiner to assess reading ability and working memory limitations as well.

Examination of the responses of the whole group to individual linguistic structures revealed that many of the children required specific support to formulate a question, even given the declarative sentence, and the intention to formulate a question, which is consistent with

reports that question formation with auxiliary verb movement is difficult for children with SLI (Rice and Wexler 1996). Alternatively, observing a second sentence to be achieved more easily than the first sentence, in examples with semantically equivalent items suggests transfer of structural organization from the first to second sentences, and this is explained by Bishop (1997), who noted that ability to substitute semantically equivalent words in a sentence suggests knowledge of the thematic roles in the sentence.

Of interest was that few problems were encountered by the children on the items containing the most words, which were in fact syntactically simple conjoined SVO structures. In contrast, the sentences containing dative structures ('*the boy gave the girl a pencil*' etc) were amongst the most difficult, as evidenced by the need of many more children for higher levels of cue support. In summary, the difficulties encountered by the children were more related to grammatical complexity than to sentence length. The presence of several definite and indefinite articles in a sentence confused many of the children, who were inclined to omit articles, and when an item contained both nominative and accusative case pronouns (you and your), pronouns were also vulnerable to error.

In addition, probing of responses by further questions enabled the examiner to determine that several children interpreted the possessive 's suffix in '*girls' room*' as a plural. There was however, some inconsistency between versions of the test, for example children were required to sequence two actions joined by the temporal conjunctions *'before'* or *'after'*. The items were intended to have a 'correct' sequence, for example in version A, it was clear that *'Joe brushes his teeth before he goes to bed'*. However, some examples emerged as reversible to some of the children, and although the intended structure was '*Dan eats his dinner after he washes his hands*' the examiner had to accept the reverse as correct as well. This meant that in

formulating a second sentence, children were able to reverse elements, when the intended structure of the test was the non-canonical '*after Dan washes his hands he eats dinner*'. In this instance, the scoring of the second sentence was adjusted so that the prompt to 'start the sentence with after', was not counted as a level 4 cue, but this is acknowledged as unreliable.

Evaluation of reliability and validity

Inter-rater reliability was identified as essential to establish unequivocal evaluation of the level of prompting required to elicit correct responses in the child, especially as some variation in the content of prompting was permitted. It was important to the experimenter in the design of the DA, that only a short period of explanation and practice was required before an independent rater could score a sample of videos as the need for lengthy training in DA would impede the uptake of DA by SLTs, and it was intended that the test methodology used in the current study would be accessible to practitioners without further training.

Results obtained on the DA were compared to those obtained on other standardised measures, to gauge the criterion validity of some of the aspects assessed by the DA. The correlation was weak however, as anticipated by the assumption that only part of the achievement on the DA measure is related to achievements in language, as also measured by the CELF-3. This would be consistent with previous interpretations that validity of a DA may be demonstrated by only partial correlation with criterion tests, as the achievement criteria are frequently not the targets of the DA. The remaining variance is influenced by the learning potential measures, in this instance the number of cues required for the child to produce the targeted response. Similarly, Hessels, Berger and Bosson (2008) demonstrated low-moderate correlation (.45) between the Hessels Analogical Reasoning Test (HART) and a static test, the Ravens SPM with which

only some of the same dimensions are assessed. The validity of the HART was further supported with other measures of reliability and validity.

Results also confirmed that the DA was sensitive to the amount of learning that had taken place over time and that performance at T2 was related to original baseline performance, which provided a further measure of reliability. The DA served to differentiate *within* the group of 'children with LI', separating those more likely to progress in grammar in the immediate future from those for whom gains would be smaller, or for whom gains may manifest over a longer period of time. Although increased predictive validity has been derived from Dynamic Assessments, it has been cautioned that the prediction is only as effective as the implementation of intervention that follows. Using DA to make predictions for the progress of children in educational settings is not useful unless poor prognoses can be redressed by specific interventions to improve a child's performance. As the children were all engaged in regular ongoing therapy programmes, and the SLTs were available and willing to implement recommendations of the DA, the present study provided an opportunity to identify whether the DA (at T1) was able to predict which participants were able to profit most from SLT intervention. This would concur with studies (eg. Hessels, Berger and Bosson 2008; Vye et al 1987) showing the greater predictive validity of DA post-test scores in comparison to static test scores.

Limitations of the Study

The greatest limitation of the study evaluating the DASS was that all testing was carried out by the one examiner, and there is as yet no evidence for the effectiveness of the procedure when carried out by other clinicians. This is an area for further study.

A further limitation relates to the participants, and primarily to the age of the children, who ranged between 8 and 10 years. In the lower, range some 8 year olds were able to manage the test adequately, while a small number struggled a great deal, and more than any participants aged 9 and above at the start of the study. In spite of this, no participant scored at the maximum level of cues, the floor of the test, so whilst caution should be applied to children for whom the test would be useful, the age of 8 would probably be an appropriate cutoff. The upper age range may be more flexible, as children aged almost 11 by the final tests still did not score at ceiling, but higher functioning children may in fact reach the point at which all items are achieved spontaneously, without prompting. The age range of application of the procedure is therefore restricted to approximately 3 years.

Linked in part to the issue of age, is the use of the standardised CELF-3(UK) as part of the procedure. A static standardised test such as the CELF is thought to be unreliable for the assessment of children with SLI, whose responses to tests are unreliable and not necessarily representative of their maximum ability. The use of such a measure for participant selection is justified by the normative and discriminant properties of the CELF-3, but for monitoring performance, it is theoretically inconsistent with the rationales for DA development.

Future Directions for Research

Further research to address limitations of the current study and to extend and apply the research questions more widely is recommended. In the first instance, in relation to the clinician administering the test, it would be important to ascertain the reliability of the procedure when administered by other SLT clinicians, who have not been specifically trained in mediation. It is recognized that different examiners will pursue different avenues of mediation, and attempt to mediate different strategies, so outcomes will always be

qualitatively different, but research into whether the measure of prompt levels is similar, and equally reliable when rated by an independent observer, remains to be evaluated.

With regard to procedural variables, experimentation with varying the grammatical structures targeted and the age and severity of language need of the children assessed is also recommended to ascertain range of applicability and potentially extend the range of applicability of the basic DA procedure.

Finally, predictive value of the DASS is an avenue for further research. The current procedure aimed to elicit information that would assist in the planning of appropriate intervention for individuals, and investigation of whether the information does in fact benefit intervention planning and indeed the outcomes of intervention, would be a logical follow up. In the opinion of the current author, this can best be evaluated on a case-by-case basis, and a series of case studies would be welcomed. Similarly, case studies would provide the opportunity for clinicians to try the assessment procedure, and evaluate its reliability as well as clinical usefulness.

Summary

Although some minor alterations to the DASS test items might have improved the utility of the procedure, the data presented have demonstrated that the aim to create a reliable and valid DA of sentence level language has been achieved.

In addition, the high correlation between the scoring of the experimenter and an independent rater with minimal training, as well as the positive responses obtained from SLTs suggest that the procedure would be replicable and useful to practising SLTs. The demand in terms of

administration time and scoring simplicity were controlled as the test was completed within a single test session in every instance, and scoring was completed during the session. The availability of equivalent parallel forms will enable clinicians to test over periods of time without the risk of rote learning confounding results.

The application of DA principles to the assessment of previously diagnosed children with language impairment, for the purposes of finding out more detailed information about them, was a novel one. Access to further information about the way in which they approach a language task, problem solve and self evaluate, and their potential to learn from input from the examiner was sought, and found. No attempt was made to differentiate the children from any other population of their peers, the intention was purely to gain information that would be useful to increase understanding of the children, and in this respect the aims of the study were different to previous studies of DA of language. Furthermore, the targeting of explicit syntactic knowledge by DA added to the body of research, previous studies having addressed word knowledge, (Peña 2000, Camilleri and Law 2007) narrative, (Peña et al 2006,), expository discourse (Peña and Gillam 2000) or two-word combinations (Olswang and Bain 1996).

The combination of established DA techniques devised for the purposes of this project was shown to be successful. The feature common to both Graduated Prompts and Mediated Learning that enabled them to be combined in the procedure is that support for the individual starts with more general, metacognitive and strategic hints, and only becomes more directive if it is needed. Intervention is adaptive and individuals are facilitated in both paradigms to discover their own solutions to problems. The common principle enabled the hybrid procedure to work, and the combination of methods yielded the intended results with the

procedure managing to contain individualised mediation while still being reliably identifiable as consisting of graded cues.

It is the same principle of adaptive support which is advocated for language therapy approaches that are metacognitive in nature. If maximum learning and transfer are facilitated by such methods in the assessment procedure, they should surely be effective *for the same children*, in the intervention process. This does not imply that these methods are effective for all children, and there are those that do not exhibit a great deal of learning in the DA process, that must therefore require more directive or intensive intervention in order to learn, and a useful DA must be able to identify these children as well. The Dynamic Assessment of Sentence Structure presented here represents a positive first step in achieving that outcome.

References

BISHOP, D.V.M., 1997. *Uncommon Understanding. Development and Disorders of Language Comprehension in Children.* Hove: Psychology Press.

CAESAR, L.G. and KOHLER, P.D., 2009. Tools Clinicians Use: A Survey of Language Assessment Procedures Used by School-Based Speech-Language Pathologists. *Communication Disorders Quarterly,* **30**(4), pp. 226.

CAMILLERI, B. and LAW, J., 2007. Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *International Journal of Speech-Language Pathology,* **9**(4), pp. 312-322.

CAMPIONE, J.C. and BROWN, A.L., 1987. Linking dynamic assessment with school achievement. pp. 82-115.

CAMPIONE, J.C., BROWN, A.L. and FERRARA, R.A., 1982. Mental retardation and intelligence. pp. 392-490.

CARLSON, J.S. and WIEDL, K.H., 1992. Principles of dynamic assessment: The application of a specific model. *Learning and Individual Differences,* **4**, pp. 153–166.

DOCKRELL, J.E., 2001. Assessing Language Skills in Preschool Children. *Child Psychology and Psychiatry Review,* **6**(2), pp. 74-84.

FEUERSTEIN, R. and FEUERSTEIN, S., 1991. Mediated learning experience: A theoretical review. In: R. FEUERSTEIN, P.S. KLEIN and A.J. TANNENBAUM, eds, *Mediated learning experience (MLE) : theoretical, psychosocial and learning implications.* Freund Publishing House, pp. 3-51.

FEUERSTEIN, R., RAND, Y. and HOFFMAN, M.B., 1980. *Instrumental Enrichment: A intervention programme for cognitive modifiability.* University Park Press.

GLASPEY, A. and STOEL-GAMMON, C., 2007. A dynamic approach to phonological assessment. *International Journal of Speech-Language Pathology,* **9**(4), pp. 286-296.

GOODMAN, R., 1997. The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry,* **38**, pp. 581-586.

GUMMERSALL, D.M. and STRONG, C.J., 1999. Assessment of complex sentence production in a narrative context. *Language, Speech, and Hearing Services in Schools,* **30**(2), pp. 152-164.

GUTIERREZ-CLELLEN, V.F. and PEÑA, E., 2001. Dynamic assessment of diverse children: A tutorial. *Language, Speech, and Hearing Services in Schools,* **32**(4), pp. 212.

HASSON, N. and BOTTING, N., 2010. Dynamic assessment of children with language impairments: A pilot study. *Child Language Teaching and Therapy,* **26**(3), pp. 249-272.

HASSON, N. and JOFFE, V., 2007. The case for dynamic assessment in speech and language therapy. *Child Language Teaching and Therapy,* **23**(1), pp. 9-25.

HAYWOOD, H.C. and LIDZ, C.S., 2007. *Dynamic Assessment in Practice. Clinical and Educational Applications.* New York: Cambridge University Press.

HESSELS, M.G.P., BERGER, J.L. and BOSSON, M., 2008. Group Assessment of Learning Potential of Pupils in Mainstream Primary Education and Special Education Classes. *Journal of Cognitive Education and Psychology,* **7**(1), pp. 43-69.

KAHN, R.J. and KING, S.R., 1997. Dynamic Procedures for Assessing Children's Cognitive and Emotional Strengths and Needs. *Journal of Cognitive Education,* **6**(2), pp. 101-114.

LAW, J., BOYLE, J., HARRIS, F., HARKNESS, A. and NYE, C., 1998. Screening for speech and language delay: a systematic review of the literature. *Health Technology Assessment,* 2(9).

MONTGOMERY, J.W., 2002. Information processing and language comprehension in children with specific language impairment. *Topics in Language Disorders,* 22(3), pp. 62-84.

OLSWANG, L.B. and BAIN, B.A., 1996. Assessment information for predicting upcoming change in language production. *Journal of speech and hearing research,* **39**(2), pp. 414-423.

PEÑA, E.D., 2001. Assessment of semantic knowledge: Use of feedback and clinical interviewing. *Seminars in speech and language,* **22**(1), pp. 51-63.

PEÑA, E.D., 2000. Measurement of modifiability in children from culturally and linguistically diverse backgrounds. *Communication Disorders Quarterly,* **21**(2), pp. 87-97.

PEÑA, E.D., GILLAM, R.B., MALEK, M., RUIZ-FELTER, R., RESENDIZ, M., FIESTAS, C. and SABEL, T., 2006. Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research,* **49**(5), pp. 1037-1056.

PEÑA, E.D. and GILLAM, R.B., 2000. Dynamic Assessment of Children referred for Speech and Language Evaluations. In: C.S. LIDZ and J. ELLIOTT, eds, *Dynamic Assessment: Prevailing Models and Applications.* Elsevier science, pp. 543-575.

PRING, T., 2005. *Research methods in communication disorders.* London: Wiley.

RAVEN, J.C., COURT, J.H. and RAVEN, J., 1990.
Manual for Raven's Progressive Matrices and Vocabulary Scales - Section 2: Coloured Progressive Matrices. Oxford: Oxford Psychologists Press.

RESING, W.C.M., 1997. Learning potential assessment: the alternative for measuring intelligence? *Educational and Child Psychology,* **14**, pp. 68-82.

RESING, W.C.M., 1993. Measuring inductive reasoning skills: The construction of a learning potential test. In: J.H.M. HAMERS, K. SIJTSMA and A.J.J.M. RUIJSSENAARS, eds, *Learning Potential Assessment.* Amsterdam: Swets & Zeitlinger, pp. 219-242.

RESING, W.C.M., DE JONG, F.M., BOSMA, T. and TUNTELER, E., 2009. Learning During Dynamic Testing: Variability in Strategy Use by Indigenous and Ethnic Minority Children. *Journal of Cognitive Education and Psychology,* **8**(1), pp. 22-37.

RICE, M.L. and WEXLER, K., 1996. Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of speech and hearing research,* **39**(6), pp. 1239-1257.

RUTTER, M., 2008. Diagnostic concepts and risk factors. In: C.F. NORBURY, J.B. TOMBLIN and D.V.M. BISHOP, eds, Hove: Psychology Press, pp. 205.

SEMEL, E., WIIG, E.H. and SECORD, W.A., 1987. *Clinical Evaluation of Language Fundamentals.* 3rd Edition(UK) edn. UK: The Psychological Corporation.

SPECTOR, J.E., 1992. Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of educational psychology,* **84**(3), pp. 353-363.

SWANSON, H.L., 1995. Using the Cognitive Processing Test to assess ability: Development of a dynamic assessment measure. *School Psychology Review,* **24**(4), pp. 672-693.

VAN DER LELY, H.K.J. and BATTELL, J., 2003. Wh-movement in children with grammatical SLI: A test of the RDDR hypothesis. *Language,* , pp. 153-181.

VYE, N.J., BURNS, M.S., DELCLOS, V.R. and BRANSFORD, J.D., 1987. A comprehensive approach to assessing intellectually handicapped children. pp. 327-359.

VYGOTSKY, L.S., 1986. *Thought and language.* Cambridge MA:MIT Press.

APPENDIX I Grammatical structure of Test items

Item No	Structure of 1st and 2nd possible sentence	No of words
1	Declarative with auxiliary plus main verb + PossPron NP Interrogative with aux reversal	6
2	Declarative with auxiliary plus main verb + future tense VP Interrogative with aux reversal	6
3	Declarative with coordination SVcSV Reversible content	7
4	Declarative with coordination SVOcSVO Reversible content with some semantic constraints	11
5	SVOdOi Nand N in Subject position Reversible content	10
6	Declarative SVOiOd N and N in Obj position	10
7	Declarative with copula verb and Adj NP Interrogative with verb reversal	5
8	Declarative with copula verb, contracted negative + possessive NP Interrogative with verb reversal	5
9	Declarative with modal auxiliary plus main verb and prep phrase SVPrepP Interrogative with aux reversal	6
10	Declarative with modal auxiliary plus main verb, Object, and prep phrase SVOPrepP Interrogative with aux reversal	8
11	Declarative with (temporal) subordinate clause SVOsSVO (Use of anaphoric pronoun)	9
12	Declarative with (causative) subordinate clause SVsSVO (Semantic constraints)	7

Key to abbreviations: S=Subject; V=verb; O=Object; N=Noun; Adj=Adjective;
PossPronNP=possessive pronoun Noun phrase eg 'my room' his hair'; PrepP=Prepositional phrase;
VP=Verb phrase, i.e. main verb + auxiliary verb(s); Od=direct object; Oi=indirect Object;
 c=co-ordinating conjunction (usually 'and'); s=subordinating conjunction

WHAT THIS PAPER ADDS

What is already known on this subject:

Dynamic Assessment (DA) has been shown to be a more culturally fair means of assessment that reliably distinguishes children with cultural and linguistic differences from those with Language Impairments. It has been shown to have greater predictive validity than static standardised tests.

What this study adds:

This study provides a methodology for DA of language suitable for use with language impaired children that is specific, reliable and accessible for practicing SLTs. It accesses sentence structure knowledge and the strategies used by language impaired children, and highlights their readiness to learn. Recommendations that are useful for the planning of intervention are extracted.

Table 1. Prompt Sheet for administration of the DASS

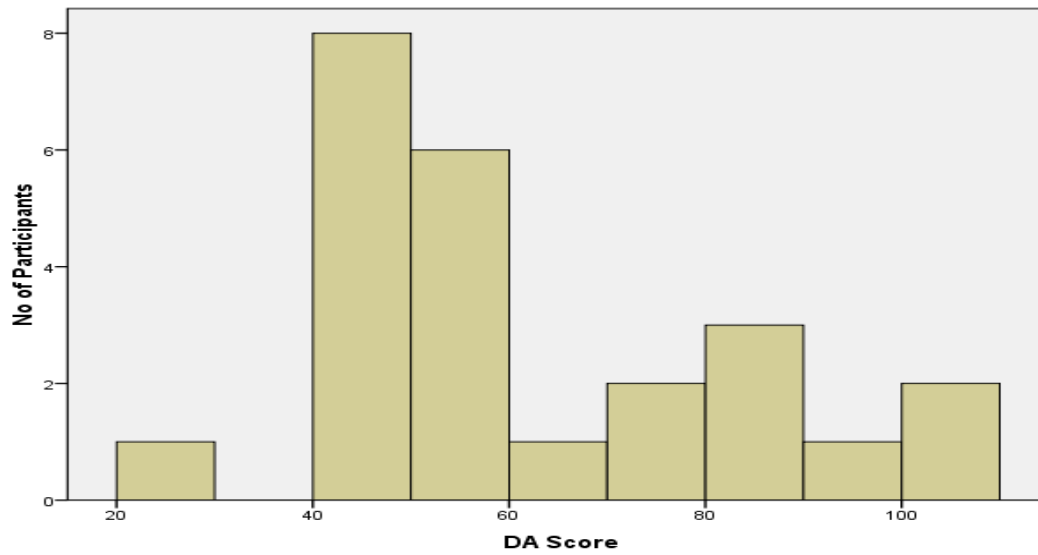| Cue Level	Description of Cue	Example of Phrasing
1	Metacognitive direction Spontaneous response	Do you know what you have to do?
2	Drawing on previous knowledge	How did you do this before? Do you know all the words? Is that right? Can you fix it?
3	Finding strategies	Which one can you start with? You need to start with something different this time Can you make little groups of words? Can you make a question? Can you swop the words around?
	Problem solving	Have you used all the words? What have you left out? Reminder – 'Last time you said..'
4	Breaking down the task.	Which one shall we start with? Which one can you start with to make a question? Start with..X. What comes next?
	Using specific feedback	You have left this one out – where does it go? Repeat part of answer already used Giving part of answer
5	Learning from feedback and instruction	Scaffolding sentence bit-by-bit Presenting cloze task. Explaining. Identifying errors Modelling for imitation
	Reflection - when the answer is correct	Is that the right answer? Why was it not ok? Can you tell me how you did that? How did you know how to do that? Was it easy or hard? Why?

Table 2. DA at Time 1 and Time 2

	No of participants	Mean	Std deviation
DA scores at Time 1	24	61.8	20.7
DA scores at Time 2	24	52.2	19.0

Table 3.   Feedback from SLTs

	Do you think the information supplied to you by the experimenter about the children, after they had had the DA, was useful?	Were you able to make use of this information in your planning of intervention for the children?
No of SLTs (n=8)	Yes - 6 No - 2	Yes- 5 Partly - 1 No - 2

Figure 1. Distribution of DA scores at Time 1.

DA Score. Range Min 24-Max 120

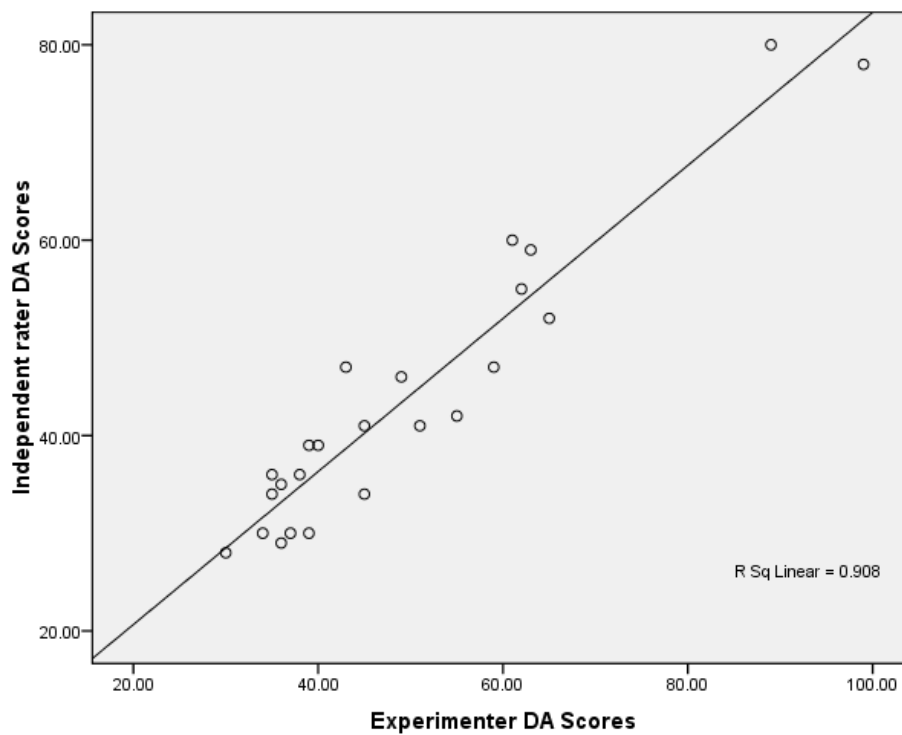Figure 2. Inter-rater agreement for all participants at Time 3.

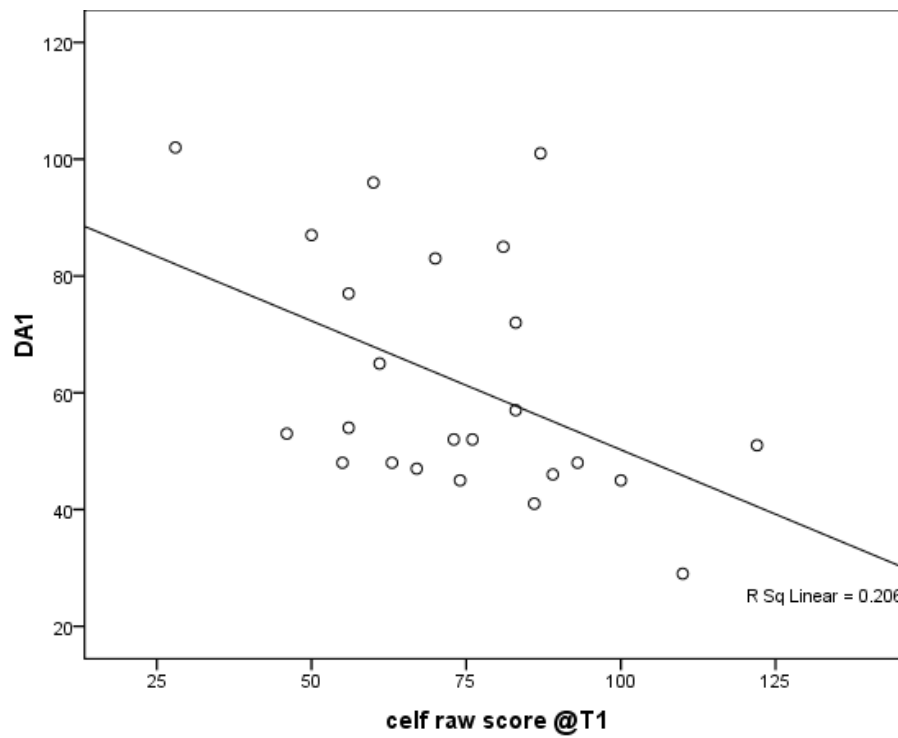Figure 3. Correlation between DASS scores and CELF-3 Total raw scores

at Time 1

Figure 4. Correlation between DA scores at Time 1 and Time 2