Discriminating disorder from difference using dynamic assessment with bilingual children

Natalie Hasson, Bernard Camilleri, Caroline Jones, Jodie Smith and Barbara Dodd

Abstract.

The DAPPLE (Dynamic Assessment of Preschoolers' Proficiency in Learning English) is currently being developed in response to a clinical need. Children exposed to English as an additional language may be referred to SLT because their proficiency in English is not the same as their monolingual peers. Some, but not all, of these children are likely to have a core language learning difficulty. Clinicians need to be able to distinguish disorder from difference due to a child's language learning context. The assessment used a test-teach-test format to examine children's ability to learn vocabulary, sentence structure and phonology. The assessment, which takes less than an hour to administer, was given to 26 children who were bilingual, 12 currently on an SLT caseload and 14 children matched for age and socio-economic status who had never been referred to SLT. The DAPPLE data clearly discriminated the two groups. The caseload group required a greater amount of prompting to identify targeted words in the receptive vocabulary assessment and performed less well in the post-teaching expressive component. For sentence structure, the caseload group required more cues to acquire the targeted clause elements in the teaching phase. The caseload group made more phoneme errors at the initial and final assessments than the controls and the type of errors made differed. Teaching resulted in greater positive change in percent phonemes correct for the caseload participants. Qualitative analyses of individual children's performance on the DAPPLE suggested that it has the potential to discriminate core language deficits from difference due to a bilingual language learning context. Future directions for development of the test are considered.

I Introduction

1 Cultural and Linguistic Diversity

With the UK and other countries becoming increasingly multicultural and linguistically diverse, speech and language therapists (SLTs) are facing increasing challenges to accurately assess and diagnose children at risk of language impairment. Children in London are reported to speak over 300 different languages, and approximately 30% of them use English as an additional language (EAL) (Cortazzi and Jin 2004; Baker and Eversley 2000, cited by De Lamo White and Jin, 2011) According to Law et al (2000) the general prevalence of long-standing speech and language difficulties is around 6-10% and there is little reason why the prevalence of language difficulties in the bilingual population should differ from that of the monolingual population (Stow and Dodd 2003). The Royal College of Speech and Language Therapists (RCSLT) acknowledges the need to provide culturally and linguistically appropriate services (RCSLT, 2003) yet Mennen and Stansfield (2006) point out that many SLTs lack knowledge of cultures and languages other than their own. .

It is widely acknowledged that reliable identification of communication difficulties can be challenging (Law, Boyle, Harris, Harkness and Nye, 1998). This is particularly true when working with children from diverse economic, linguistic and cultural environments (Stow and Dodd , 2003; Mennen and Stansfield 2006; De Lamo White and Jin 2011). It is usually regarded as best practice to assess multi-lingual children in all languages they are exposed to, however, in these difficult fiscal times, and with the diversity of languages encountered, access to interpreters and timely assessments can often be particularly difficult to achieve. As a result, it is often difficult for SLTs to accurately discern if children from diverse linguistic and cultural backgrounds are presenting with a language delay relative to their monolingual English speaking peers, or if they have a core language difficulty.

Different approaches to assessment of bilingual children used by practising SLTs have been reviewed by De Lamo White and Jin (2011). Standardised norm referenced tests are used most frequently, but are usually not standardised on bilingual

populations and are used by SLTs with a disclaimer that the norms may not be applicable to a particular individual for whom English may be a second or additional language. The findings are used qualitatively or descriptively, without recognition, however, that the elicitation of the data is dependent on cultural familiarity with the materials, and that instructions may not be perfectly understood or may require the child to carry out an unfamiliar task (De Lamo White and Jin, 2011). In particular, vocabulary and narrative skills are linked to a child's prior experiences leading to test scores reflecting life experiences and socio-economic status rather than language ability (Lidz and Peña, 2009).

Research has also found that bilingual children can be both under and over represented on SLT caseloads (Winter, 1999; Stow and Dodd, 2003; Mennen and Stansfield, 2006). Reasons for misdiagnosis may include the difficulty that the referring professional has in interpreting and identifying the difference between a child learning English as an additional language or having a core language deficit in all of the languages they are exposed to. This issue of *over*-identification of children with speech and language difficulties can unnecessarily deplete SLT resources. Conversely, there is a potential impact on the identification of children who have latent difficulties with language learning which are masked by learning English as an additional language thus causing them to be *under*-identified on SLT caseloads. Any given SLT caseload may therefore include children who have been inaccurately identified ('false positives') as well as children who truly have a language learning difficulties.

2. Dynamic Assessment

Dynamic assessment (DA) has been proposed as an alternative or complementary format of assessment that is thought to reduce the inherent cultural and linguistic bias attached to static standardised tests (De Lamo White and Jin, 2011; Peña, Iglesias & Lidz, 2001). Static tests, whether standardised or informal procedures, are dependent on a single 'snapshot' assessment of an individual's ability at a particular point in time. The value of static assessment has been disputed for many years with Vygotsky (1978) being one of the earliest educationalists to propose the assessment of learning potential as an alternative. Vygotsky rationalised that static assessment reflects only on past and present functioning rather than an individual's potential to learn in a

situation in which his performance can be supported, referred to as 'modifiability' of the individual. He termed the potential for individuals to learn through the guidance of a more experienced peer as the 'zone of proximal development' (ZPD). He believed the assessment of the ZPD to be a better indicator of a person's functioning and of his future prognosis, than a test of his crystallized ability, such as an IQ test which is held to be an indicator of future achievement. Assessment of the ZPD has been operationalised in various ways to assess an individual's potential to learn, or his/her need for prompting in order to achieve a task.

The most well known proponent of the application of DA to clinical populations is Professor Reuven Feuerstein. Feuerstein's approach to assessment includes evaluation of the individual's response to a specific kind of intervention, the mediated learning experience (MLE) which is incorporated into the 'teach' phase of a test-teach-retest model of DA. In a mediated learning situation, the mediator shapes the experience of the learner, by interposing himself between the stimulus, or the experience, and the individual. He is thus able to help the mediatee attend selectively to relevant stimuli, focus on important aspects, process appropriately using comparisons and links to past experiences, and generalize the experience to new situations (Haywood 1993). In order for an interaction to be characterised as mediational it must contain essential components that also identify the process as metacognitive, focussing on awareness in the individual of the process of change, and the ability to transfer learnt strategies to other applications. The battery of assessments compiled by Feuerstein, the Learning Propensity Assessment Device (LPAD, Feuerstein et al 2002), also, however, contains assessment instruments that do no employ test-teach-retest formats and mediation, but make use of other forms of prompting such as repetition of the stimulus, or specific prompts.

Other researchers have used different models of DA, most commonly the notion of Graduated Prompting (Campione and Brown 1987) in which a succession of increasingly directive cues are provided to assist an individual in solving a problem. The amount of prompting required is a measure of an individual's ZPD. The reader is referred to Grigorenko and Sternberg (1998), Campione (1989), or Haywood and Lidz (2007) for further detail about the diverse procedures that are included in the 'umbrella term' of DA.

Research into applications of DA to speech and language (Hasson and Joffe 2007, Lidz and Peña, 2009) has suggested that DA is especially useful for SLTs both to diagnose language impairments and to inform intervention. Intrinsic to the definition of a developmental language impairment is the intransigent nature of the difficulty, which suggests that it is not readily modifiable without an investment of intervention. The language deficit that results from cultural or linguistic difference is more likely to be remediable with shorter term exposure to good language models or teaching. A dynamic assessment may therefore be able to distinguish the two conditions on the basis of modifiability or stimulability. The use of a DA that taps into ability to learn also enables the assessment to be carried out in English, regardless of the first language of the child, as the assessment is of the child's ability to learn English, and not of his already acquired language. The information gained from DA about a child's modifiability and potential to learn could support SLTs in identifying children who are candidates for intervention or support, as well as selecting suitable therapy targets and facilitations.

A series of studies into the use of DA to assess culturally and linguistically different children (hereafter CLD), and differentiate typically developing children from those with language impairment has been published by Elizabeth Peña and colleagues. Peña, Iglesias and Lidz, (2001), for example, examined the performance of preschool CLD children, using a word learning task, with a pre-test-teach-post-test method. The teach phase consisted of mediated strategies for naming, and the children's performance during these sessions was also rated for modifiability. Post-test scores and ratings differentiated the typically developing children from those with low language ability, who were less able to benefit from the short-term mediated learning experience. Typically developing CLD children markedly improved their performance on post-test, and were also able to transfer learning to other areas of language, showing improved scores on other tests of language that did not specifically tap naming abilities. Dynamic assessment methods were more predictive in this differentiation than static pre-test scores, which have been shown to over-diagnose children with CLD as language impaired.

Further studies using DA to identify children with language impairments include a study of receptive vocabulary by Camilleri and Law (2007). A DA of receptive vocabulary was developed in order to compare the performance of monolingual English speakers with children with English as an additional language (EAL), and of typically developing children with those referred to SLT services. The static administration of the British Picture Vocabulary Scales (BPVS, Dunn et al, 1997) was followed by a DA procedure aiming to facilitate learning of vocabulary by strategic use of relevance, discrepancy and mutual exclusivity criteria, rather than actual teaching of a new word. A hierarchy of mediational prompts was employed to lead the child to the words and to use of the strategies. The DA procedure was found to differentiate between children with normally developing language and those referred to SLT services. Of particular interest was the fact that referred children with EAL achieved a similar range of scores on the DA as monolingual referred children although their static scores on the BPVS were significantly lower. This suggests that the static test may not be suitable for children with EAL and risks over-diagnosing them as language impaired. On the other hand, the DA may constitute a more valid measure of lexical ability for use with children with EAL as well as with monolingual children (Camilleri & Law, 2007).

Moving away from vocabulary studies, Peña et al (2006) examined the classification ability of a DA of narrative ability in first and second grade school children. Two wordless story books, found to be parallel, were used as pre- and post- tests, and two sessions of intervention targeting story components, were carried out in the 'teach' phase. Intervention was mediational in nature, and slightly individualised for each child. Ratings of modifiability according to the 3-criterion modifiability scale (Peña, 2000) were also carried out after the second intervention session. In general, all children performed better on the post-test after the two sessions of mediated learning experience, but the typically developing children showed greater gains than those with language impairment. Pre-test measures of narrative did not accurately classify TD and LI children. The best single predictor was the clinician's modifiability rating, which was seen as consistent with the aims of DA which are to assess responsiveness to instruction. Thus it can be seen that the results of the study using narrative parallel the findings of earlier studies using naming tasks, and confirm the advantage of DA

over static tests for classification purposes, as well as the significant usefulness of modifiability ratings.

Dynamic assessments of expressive language have also been used to differentiate the potential of individual children to benefit from intervention. It is only within this context that DA of word combinations and syntax has been used. Olswang, Bain and Johnson (1992) applied Vygotskian theory and Feuerstein's DA methods to gauge the learning potential of young children in the language acquisition process. The authors constructed a hierarchy of prompts and transfer tasks to assess the potential for children at the single word stage of development to progress to combining two words in various semantic relationships. Two children, aged 32 and 35 months, both using single word utterances only and thus exhibiting language delays in comparison to their chronological age norms, were investigated using the DA protocol. Although the children performed similarly on the static assessment, their response to the prompting during the DA differed markedly. Thus the procedure demonstrated the differing potential for immediate improvement in the two children.

Hasson, Dodd and Botting (2012) investigated the skills of 24 children with language impairments on a DA of sentence formulation, and found that the procedure differentiated potential to benefit from intervention within the group. The DA was predictive of outcomes from intervention which were not related to the results of a static standardised test. Dynamic assessments of syntax have not, however, been used to differentiate bilingual children with language differences from those with specific language impairment, and nor have dynamic assessments of phonology. Numerous procedures have measured stimulability in studies of phonology (e.g., Dodd, Holm, Zhu Hua et al, 2003), and several studies have characterised the phonology of bilingual children (e.g., Goldstein & Swasey, 2001; Grech & Dodd, 2009), but dynamic procedures have not been used in the process of differentiating between children requiring intervention, and those who may improve spontaneously.

In this study the term bilingual will be used to refer to individuals who use two or more languages in any modality, speaking, reading or writing (Mackey, 1968). The present study makes use of abbreviated procedures for the DA of phonology, syntax and vocabulary, combined into a staged procedure that aims to differentiate bilingual

preschool children whose language skills are progressing well from those who may have a language impairment. In this sense, the DA is being presented as a pre-diagnostic or screening assessment. Classically, when evaluating a pre-diagnostic or screening assessment the question is whether the tool under consideration can correctly classify children as 'possibly abnormal' or 'possibly normal'. The screening assessment is validated by comparing its results to a reference test or 'gold standard' which typically consists of a standardised clinical diagnostic test of known validity (Law, Boyle, Harris, Harkness and Nye, 1998). Pass/fail criteria are adopted for the screening test and a statistical cut-off is adopted for the 'gold standard' (e.g. 1.5 standard deviations below the mean) and the results of the two are compared. A screen is considered to be appropriate if the number of children incorrectly identified as possibly normal (false negatives) and incorrectly identified as possibly abnormal (false positives) is kept to a minimum (Law et al., 1998). In other words, the proportion of children correctly identified as possibly normal (specificity) and the proportion of children correctly identified as possibly disordered (sensitivity) needs to be as high as possible.

There are a number of reasons why such a validation process involving a 'gold standard' is not currently feasible in the context of carrying out a preliminary evaluation of a dynamic assessment. The first is that there is no reference test for assessing bilingual children which could be used as a 'gold standard'. Indeed the application of DA to bilingual children is driven partly by the fact that there is no such standardised, norm-referenced assessment. It would be possible to classify children according to pass/fail criteria on a screening test and then to compare this to their clinical status as cases and non-cases (i.e. whether they have been referred and are being treated), but there needs to be an acknowledgement that the latter classification is not a definitive one, precisely because it is so difficult to assess bilingual children who present with a delay at a given time. Two bilingual children with similar levels of difficulty with language may be considered as a case and a non-case, purely on the basis of whether there has been an expression of concern.

At this stage in its development, there is one further reason why it would be difficult to validate the DA against a gold standard, and that is the fact that pass/fail criteria on the DA dynamic assessment would need to be established. The Dynamic Assessment

of Preschoolers' Proficiency in Learning English (DAPPLE) is unlikely to lead to straightforward pass./fail classification. Rather it leads to a profile of responses which then needs to be interpreted in terms of the child's language learning abilities. The main question that this study hopes to answer therefore stops short of a validation of the DA as a screening tool. Rather, it is an exploratory question, namely whether bilingual children who have been referred for speech and language intervention and bilingual children who are considered to be typically developing respond differently to the learning opportunities afforded within a DA context. If clear differences were found within the individual components and/or within the assessment as a whole, it would be feasible to further explore whether the DA can reliably be used to identify children at risk of language difficulties, who would benefit from further diagnostic assessment for language impairment. Such an assessment could constitute a valuable decision making tool for Speech and Language Therapists working with bilingual children.

II Method

1. Aims of the study:

i. To pilot the tasks and materials devised in the three areas of phonology, vocabulary and syntax for their accessibility to children aged 3-5 years.

ii. To determine whether the individual subtests, and the battery as a whole, elicits significantly different performances from children referred to SLT services with concerns about their language, when compared to bilingual children who are considered to be typically developing.

2 Participants

Ethical approval for the study was obtained from South West London Research Ethics Committee. Once ethical approval was obtained the SLT researchers identified suitable participants to take part in the study.

Participants were bilingual children aged 3-5 years and belonged to one of two groups.

i. The 'caseload' participants were identified by the SLTs from whom they were already receiving intervention. Their parents/carers were approached, in person or

over the telephone, for their consent for their child to be assessed using the DAPPLE. Bilingual children from diverse linguistic and cultural backgrounds were included with the only pre-requisite being that the children should have had some exposure to English, given that the assessment would be carried out in English. This exposure was assumed if the participants had attended their educational setting for several months. Children were excluded from the study if they had significant difficulties in addition to, or other than language, including autistic spectrum disorders, syndromes and hearing loss.

ii. Educational settings with whom the researchers had a current working relationship were approached to recruit control group participants, and supported the researchers in gaining parental consent. These educational settings included private nurseries and nursery and foundation classes of mainstream schools. As with the group of children referred from the caseload, children from diverse linguistic and cultural background, with English as one of their languages were included.

The non-verbal cognitive ability of participants was not used as a criterion for inclusion/exclusion in either group

The cohort studied included, the caseload group of 12 bilingual children (mean age 50.25 months, SD 5.79, comprising 4 girls and 8 boys) and the control group of 14 bilingual children (mean age 50.43 months, SD 6.90, comprising 9 girls and 5 boys). Children fell into the following socio-economic status categories based on their postcode (ACORN): Urban Prosperity, Moderate Means and Hard Pressed. Children came from a range of different language learning contexts and English proficiencies. Languages they were exposed to (in addition to English) included Bengali, Gujarati, Polish, Yoruba, Portuguese, Twi, Lingala and Turkish. One child was exposed to three languages, those being Spanish, English and French.

3 The Assessment Tasks

Each language subtest was devised with test–train–retest components (Peña et al, 2000, 2006, 2007) to elicit information about the participant's proficiency in learning English. There were both static and dynamic components. The assessment comprised the following tasks presented in this order:

Task 1: The block building task from the British Ability Scales II (BAS, Elliott, 1996).
Task 2: Dynamic assessment of vocabulary consisting of a static pre-test receptive vocabulary picture selection task followed by a vocabulary teaching phase and a post-test of targeted vocabulary
Task 3: Dynamic assessment of expressive language including static pre-test followed by an expressive language teaching phase,
Task 4: Goodenough Draw A Man Test (Goodenough, 1926)
Task 5: Dynamic phonological assessment (Diagnostic Screen taken from the Diagnostic Evaluation of Articulation and Phonology (DEAP, Dodd et al, 2002)
Task 6: The assessment of expressive language (task 3) was repeated in the post-test phase.

Tasks 1 and 4 of the assessment provided a screening measure of the child's non-verbal capacity. The block building task also served as a warm-up activity to develop a relationship between the SLT and the child. The Draw A Man Test utilised the time between the intervention phase and the retest phase for the expressive language task, while assessing the child's ability and providing an apparent play activity for the child as a 'break' from testing.

a) Task 1: The block building test from the British Ability Scales (Elliott, 1996).
In this assessment the child was asked to see if they could copy the design made by the SLT with his/her bricks. Evidence suggests that by 2 years old a child can build a tower of six or seven cubes, when they reach 3 years they can build several three-cube bridges when given a model. At 4 years a child can build three steps from cubes after a demonstration, and by 5 years a child can use blocks to build elaborate models (Sheridan, 2008).

b) Task 4: The Draw a Man Test (Goodenough, 1926)
This task required each child to draw a person on a piece of paper. The following instruction was given: *"I want you to make a picture of a person. Make the very best picture you can"*. The child was given as long as they would like to draw the picture and the SLT then asked to keep the picture at the end. This test is used to make an estimate of a child's cognitive and intellectual abilities that are reflected in the

drawing's quality. The drawings are marked initially for their pencil control and recognisable human representation and then each additional feature is scored a plus or minus. A final composite score is achieved which is correlated with a chronological age.

c) Task 2: Dynamic assessment of vocabulary learning.

The procedure was based on that described by Camilleri and Law (2007). The first part of the DAPPLE's vocabulary assessment provided a measure of the child's receptive knowledge of pictures of nouns without adult intervention or assistance. The starting point for the intervention phase occurred once up to six words (out of 30) had been identified as being unknown to a child and at least 12 words were identified as being known. The intervention phase of the dynamic assessment was carried out in the form of a posting game where the child posted picture cards into a post box. The child was presented with three cards, one of which was a targeted previously unknown item, and the other two were previously known items, used as distractors. Children were then encouraged to use process of elimination strategies to accurately create a new word-referent match. Prompting to achieve picture recognition was according to a standardised hierarchy of cues (from least to most assistive).

If the child was able to select the picture (independent identification) using only contextual mediation, i.e. using elimination of known items as a strategy, they scored three points. At this point, if the child was unable to identify the correct picture the SLT provided feedback to the child by saying ''*No, that's not the ' judge'; that was a hard word; let's try and find the easier ones first''* (context/language mediation). The SLT proceeded to get the child to find the easier items first which they had already identified correctly in the pre-test. Once the child had identified both distractors correctly, they were again asked to find the difficult target word, and scored two points if they were successful.

If the child was still struggling to identify the correct target item a final level of mediation which made an explicit link between the referent and the word (context/language/context mediation) was used by turning the distractor pictures face down after correct identification. This meant that there was only one possible option left to point at for the target word (explicit identification) scoring one point. The

process was repeated for all of the chosen vocabulary items (up to six). Children were assigned an overall score for mediation (1 to 3) based on the mean score across items. For example, if children consistently adopted independent identification they achieved a mean score of three.

Expressive task. Once the child had identified all three of the vocabulary items, irrespective of the level of mediation, the first expressive task was presented. The child was asked which picture they would like to post in the post box first. This demonstrated whether the child was able to show immediate recall for the target word and use it expressively. All three target items named would be posted away. This continued for each set of three words (up to six target words) with the child naming the distractor items and target items before posting them in the post box.

Post-test. After the child named all of the vocabulary items and posted them away, the target pictures were selected from the post box (maximum of six) and presented to the child all at once; without the distractor items. The child was asked to name the items for a second time and post them away. This represented a measure of each child's retention of items on an expressive level. If the child was unable to name some of the items, the child was asked to point to each remaining picture in turn when spoken by the SLT. This procedure served to check whether the child retained those remaining vocabulary items receptively.

d) Task 3: Dynamic assessment of expressive language

In the expressive language assessment, the child was told that '*We are going to look at some pictures and tell some little stories about them'*. The child was initially given a model of a sentence next to a given picture e.g. '*Look, the fireman is squirting water on the fire'*. The static pre-test then required the child to comment on what was happening in two consecutive different pictures. The child's response was recorded verbatim and no prompting was given.

During the intervention phase the child was presented with four different composite pictures. Each picture was intended to elicit a three or four element sentence from the child e.g. *'the boy is eating the banana'* or *'the grandma is sitting on her chair in the garden''*). The child was shown one picture at a time and asked *'What is happening*?' If the child responded spontaneously with a correct sentence containing three or four

elements they were scored three. If the responses were not achieved and the child did not produce the sentence, the sentence did not contain the correct number of elements or the sentence was not sequenced appropriately, the SLT prompted as necessary e.g. '*what is he eating?*' or '*where is she sleeping?'* If these prompts elicited the required response the child scored two. If the child was still unable to produce the target sentence, the sentence was modelled by the SLT for the child to imitate, and if this level of prompting was required, the child scored one. This procedure of successive prompts is a simplified version of that used by Bain and Olswang (1995) to assess potential for children to learn 2-word combinations.

e) Task 5. The phonological assessment (adapted from the Diagnostic Screen taken from the Diagnostic Evaluation of Articulation and Phonology (DEAP), Dodd, 2002)
The phonological screening consisted of ten colourful pictures that the child was required to name twice. The procedure assessed single word production, consistency of sound production in single words and stimulability of phones involving repeated attempts to elicit accurate production. The child was asked to name all ten pictures. Any speech sounds that were not produced accurately by the child were modelled by the SLT and the child was encouraged to copy the sound to check stimulability. The child was then asked to name the 10 pictures again. The child's two productions were checked for consistency and then the inconsistency calculation is worked out with the number of words produced differently divided by the number of words produced twice. If teaching the articulation of individual sounds enhanced children's pronunciation of words, their inconsistency score would be higher than if the teaching did not affect pronunciation.

f) Task 6: Dynamic assessment of expressive language post-test:
The child was presented with two more composite pictures and again asked '*What is happening*?' No prompting was given, and the child's spoken responses were recorded verbatim as with the static pre-test.

4 Procedure
The assessment was carried out by one of two speech and language therapist researchers in a single session lasting a maximum of 40 minutes. Responses were

transcribed during the procedure and scoring carried out immediately after completion of the test.

5. Analysis

Each of the subtests was scored and the caseload and control groups compared in their performance at pre-test, post test and in terms of the amount of prompting employed in the intervention phase of the procedure. One or two –way ANOVAs were used to explore group performance for each component of the test procedure. Qualitative analyses of the grammatical and phonological structures used by the two groups of children were described. Finally correlations between the performances on the different subtests were carried out to ascertain whether measures reflected potential to learn or strengths in specific aspects of language.

III Results

1 Bilingual Caseload and Bilingual Control Comparisons

The two groups were well matched for socio-economic status: seven children (58%) in the caseload group and eight of the controls (57%) belonging to middle-class ACORN categories; and, five of the caseload group (42%) and six of the control group (43%) belonging to the 'hard pressed' category. The descriptive data for the two groups' chronological age and performance on the Draw-A-Man and brick building tasks is shown in Table 1. One-way analyses of variance indicated that the caseload and control groups of bilingual children did not differ in terms of age ($F_{1,25} = .005$, $p = .994$, NS); but that the control group performed better on the Draw-A-Man test than the children attending therapy ($F_{1,25} = 5.283$, $p < .05$). Raw scores were inspected to determine the distribution of scores in each of the groups. A score of 52, approximately the middle of the range of scores for both groups was used to compare the distribution of individuals' scores. There were nine caseload children who scored less than 52 compared to 5 controls; with only three caseload participants scoring more than 52 as compared to nine controls. The two groups were therefore not matched for non-verbal performance on one standardised procedure. Another commonly used non-verbal cognitive assessment, a brick building imitation task, revealed a similar, but non-significant result, with eight of the caseload children scoring below the normal range compared to four of the controls ($\Xi^2 = 3.77$ 9 $p = .06$).

Table 1 about here

2. Vocabulary

Table 2 shows the means and standard deviation for the vocabulary tasks. In the pre-test of receptive vocabulary, a one-way ANOVA indicated that the caseload group performed less well than the control group ($F_{1,24} = 12.76$, $p<.01$). Having identified up to six words that children did not know in the pre-test, children were then given the opportunity to learn each word in a dynamic interaction (see method). There was a significantly greater need for graded assistance required by the caseload group (as indicated by the 'Mediational Score') in comparison to the control group ($F_{1,24}$ 7.031, $p<.02$) in order to identify the target word in the learning interaction. Once children had identified the target words from amongst two distractors, they were asked to name those pictures. There was no significant difference between the proportion of words named by children in the two groups in this first expressive task ($F_{1,24} = 0.007$, $p = .936$, NS). However, when the children had a second chance to name the same target words after completion of the learning interaction, the caseload group performed significantly less well than the control group ($F_{1,24} = 9.280$, $p<.01$), indicating they were unable to retain/access the expressive form of words learned.

Table 2 about here

The data were examined to explore individual differences. There was one child in the caseload group who had a profile of responses that resembled that of the control group. She required minimal assistance to identify the target words and went on to name all of the target words correctly in the second expressive task. A second child in the caseload group was able to name three of six words targeted in the dynamic interaction despite a poor score on the vocabulary pre-test. Conversely, there was one child in the control group who achieved the lowest score in the pre-test assessing receptive vocabulary and went on to perform poorly on expressive measures within the dynamic interaction.

3. Sentence Structure

The descriptive data for the expressive language measures are shown in Table 3. A two-factor, repeated measures, ANOVA compared the groups' pre and post performance measure of the number of clause elements used by caseload and control bilingual groups. There was a significant groups' difference ($F_{1,23}$ =5.985, p<.025) with the control group performing better than the caseload group. There was also a difference between the two assessments ($F_{1,23}$ = 7.309, p< .025), with better performance in the post-dynamic interaction assessment. The interaction between group and assessment time was not significant ($F_{1,23}$ =2.483, p=.104) indicating that both groups performed better in the post dynamic interaction assessment. However, the score achieved by the children which was based on the number of cues required in order to achieve the criterion measure, was greater for the control group children ($F_{1,25}$ =7.06, p<.025), indicating that they required fewer cues.

Table 3 about here

In order to detect qualitative differences, the clausal structures used by the two groups at each time were examined (see Table 4). Both groups used more subject -verb structures in the post test stage than in the pre-test, which was met in the caseload children by a corresponding decrease in verb-object structures. The greater number of subject-verb structures in the control children resulted from many of the children producing more than one conjoined clause, and telling a longer story at the post-test stage. There were markedly fewer subject-verb-object structures and correspondingly more 4 element subject-verb-object-adverbial structures in the control group children at the post-test stage.

Table 4 about here

4 Phonology

The descriptive data for the four phonology measures are shown in Table 5. A two factor, repeated measures, ANOVA (pre- and post-teaching phase measures of words in error for caseload and control groups) indicated that the control group produced more words accurately than the caseload group ($F_{1,24} = 13.264$, $p = .001$). The conditions term was also significant with participants performing better in the post-test condition ($F_{1,24} = 12.986$, $p = .001$). The interaction was also significant ($F_{1,24} = 5.009$, $p = .035$). Inspection of the means in Table 5 indicated that the caseload children made greater positive change than the controls. A one way analyses of variance, however, showed no significant groups' difference in the number of speech sounds that were not stimulable ($F_{1,25} = 1.586$, $p = .220$ NS).

A repeated measures ANOVA (pre and post percent phonemes correct measures for caseload and bilingual groups) indicated that the control group produced more phonemes accurately than the caseload group ($F_{1,25} = 11.633$, $p < .01$). The conditions term was significant with participants performing better in the post-test condition ($F_{1,24} = 22.902$, $p < .001$). The interaction was also significant ($F_{1,24} = 17.814$ $p < .001$). Inspection of the means for percent phonemes correct (PPC) in Table 2 indicated that the caseload children made greater positive change than the controls. A one-way analysis of variance examining inconsistency of errors was significant ($F_{1,25} = 9.283$, $p < .01$). Caseload children were less consistent than controls when the number of words that were produced differently on the two trials was compared.

Table 5 about here

In order to examine change qualitatively, the types of errors made by children in both groups were examined. All except one of the caseload children used phonological error patterns that are atypical of English phonological development. In comparison, fewer than half of the control children did so, although one child in the control group children exhibited four error patterns atypical of English phonological development.

5 Comparison of phonology, sentence structure and vocabulary.

A series of correlations compared post-test performance of the combined groups (see Table 6). Performance was correlated suggesting that children's response to dynamic assessment was not related to a specific aspect of language, but rather to an ability to make use of graded cues.

Table 6 about here

There were a number of children who were identified by their performance as potentially being in the wrong group. Some participants in the control group performed so poorly that it seems they should be referred for assessment. Conversely, there were some children in the caseload group whose performance was similar to that of the majority of the control group. Table 7 presents these children's profiles across the three aspects of language assessed.

Table 7 about here

IV Discussion

In this study we have carried out the first steps in developing and evaluating a dynamic assessment battery for use with bilingual children. Twenty-six bilingual children, with a mean age of 4;8 years, were assessed on the DAPPLE. Twelve were on the caseloads of speech language therapists and 14 were non-caseload controls. This trial version of the DAPPLE assessed children's ability to learn vocabulary, sentence structure and phonology in a test, teach, retest format. Although the groups were matched for age, the caseload group were found to perform less well on a standardised non-verbal assessment. The DAPPLE data clearly discriminated the two groups in different ways for the three aspects of language examined. For vocabulary, children on the caseload required a greater amount of prompting to identify targeted words in the receptive vocabulary assessment and performed less well in the post-teaching expressive component. For sentence structure, the caseload group required more cues to achieve the targeted clause elements in the teaching phase. For phonology, while the groups did not differ in their sound stimulability, the caseload

group made more phoneme errors at the initial and final assessments than the controls. Teaching resulted in greater positive change for the caseload participants, whereas the controls showed no change in percent phonemes correct. Performance on post-teaching reassessment tasks was correlated across the three aspects of language. Qualitative analyses of individual children's performance on the DAPPLE suggested that it has the potential to discriminate core language deficits as opposed to difference due to a bilingual language learning context. These results will now be considered in more detail.

1 Difference on non-verbal performance

The study reported was a preliminary evaluation of an assessment designed to discriminate between language learning disorder and different language performance due to the language learning context. Consequently, the study compared groups of bilingual children already receiving SLT intervention and those not referred for assessment (controls). It is not surprising that an unselected caseload group would perform less well on non-verbal measures than controls. Their poorer ability to perform in class on a range of cognitive tasks, as well as having poorer communication skills, might have alerted their teachers to a more general learning problem that led to their inclusion on the SLT caseload. Inspection of the raw data, however, showed an overlap between the two groups' non-verbal cognitive performance. It would be expected that the clinical population would contain some children whose non-verbal abilities were within the normal range as well as others with co-occurring non-verbal and linguistic deficits.

2 Vocabulary

The caseload group performed less well than the control group on the pre-test of receptive vocabulary and needed significantly more graded assistance to identify the target word in the learning interaction. While there was no group difference between the proportion of words named in this first expressive test, the caseload group performed less well than the controls in the second expressive assessment, indicating poor maintenance of learning. Previous research has reported similar findings (Camilleri & Law, 2007; Camilleri, 2009) that have been interpreted theoretically in terms of fast-mapping ability. Research on fast mapping has suggested that typically developing children as well children with language impairments are able to identify a

novel word within a stream of words and establish a link between word and referent (Dollaghan, 1985, Rice, Buhr & Nemeth, 1990). However, establishing and retaining the phonetic information, particularly for subsequent expressive use, was found to be the most vulnerable aspect of the fast mapping process (Dollaghan, 1985). The current results are consistent with that idea. The caseload children were less able to retain representations for expressive use than controls despite being able to name some of the targets accurately at the end of the teaching phase. An impaired ability to establish/retain fast-mapped phonological representations might account for differences between caseload and controls. The fact that referred children were able to use words expressively initially suggests a difficulty retaining or accessing the stored representations rather than difficulty entering the phonetic information in the first place. For identification purposes, the combination of a low static pre-test score and a reduced ability to establish and retain new word-referent combinations is likely to be more accurate than a low static score alone. Naturally, children with a good knowledge of English vocabulary (i.e. high static scores) are unlikely to have a language impairment. However, children with lower vocabulary knowledge may be differentiated on the basis of their ability to learn and retain new words. This will need to be explored in future research which investigates the sensitivity and specificity of a measure which combines both a static measure of vocabulary knowledge and a dynamic measure of word learning.

3 Sentence Structure

The control group performed better than the caseload group on both the first and second assessments with both groups performing better in the post-dynamic interaction assessment. The caseload children required more cues to produce sentences containing either three or four clauses. Previous research is limited, however Olswang and Bain (1996) found that DA of children with delayed language development predicted their progress in therapy. Children who were more stimulable for word combination made greater improvements in therapy sessions immediately following the assessment. The results were not conclusive however, and Olswang and Bain noted the need to identify what type of therapy and dosage of intervention would be required to facilitate optimum progress. Similar conclusions were reported by Hasson (2011), following DA of sentence formulation in 8-10 year old children diagnosed with specific language impairment. In the present study the teaching

component of the dynamic assessment procedure appeared to enhance the sentence structure of both caseload and control participants, but children on the caseload, as predicted, needed more cues from the examiner in order to achieve the criterion measure. Qualitative assessment revealed different patterns of sentence structure for the two groups, a finding that deserves further exploration. With sentence structure, it may be that a combination of lower initial scores, a greater number of cues required for improvement and a distinctive pattern of sentence structures used may be successfully adopted to distinguish between different groups of bilingual children who initially demonstrate reduced knowledge of sentence structure in English.

4 Phonology

The phonology assessment led to a rather different pattern of results. Quantitative measures indicated that the control group produced more words and phonemes accurately than the caseload group at both assessments. However, while there was no improvement in the control group's percent phonemes correct (PPC) in the post-teaching phase, the caseload group showed a marked improvement even though the teaching phase was limited to teaching single speech sound production. This reflects the lower starting point and therefore the greater room for improvement in the caseload group. It remains to be seen whether within the caseload group, those children showing greater change in PPC in the DA, benefit more from intervention. Qualitative analyses indicated that the two groups evidenced different predominant patterns of errors, perhaps providing the potential for discriminating difference from disorder. Previous research (McIntosh & Dodd, 2011) has demonstrated that the types of error patterns evident at two years predict later phonological development. Subsequent modifications to the DAPPLE should include qualitative analyses to establish whether that may provide a more cost-effective way of identifying disorder.

5 Comparison of aspects of DAPPLE

It is not surprising that there were low to moderately significant correlations for performance across the three language domains. Research on monolingual children with language impairments has documented that difficulty in one domain is often associated with difficulties in others perhaps because vocabulary, grammar and phonology combine to allow communication (Hoffman, Norris & Monjure, 1990). Further exploration of the nature of the inter-relationships between the three domains

might, however, provide insights into sequential bilingual language acquisition. However, the rather weak association between domains reflects clinical experience that most children with specific language impairment have a primary impairment in vocabulary/semantics, grammar or phonology (Dodd & Crosbie, 2010). The DAPPLE was developed in such a way that the assessment could be carried out in its entirety in one session of approximately thirty to forty minutes. Ultimately, it would also be possible to carry out assessment of the three different domains separately or to choose to assess one (or two) domains if these were the areas which were of primary concern with an individual child. While the DAPPLE remains under development, it is envisaged that all three components will be carried out in order to evaluate its utility as a complete battery as well as its utility in the three separate domains.

6 Conclusions and future research directions

The study presented was an initial evaluation of a new assessment designed to discriminate between disorder and difference due to language learning context in children between the ages of 3 and 5 years acquiring English as an additional language. Such a tool is necessary because clinicians and teachers are currently challenged by the number of children speaking one of a wide range of first languages and English. Findings from the new assessment showed that it was possible to obtain useful data on three language domains in English in well under an hour using a test-teach-retest dynamic assessment format. Comparison of the patterns of performance of caseload and control participants showed that the number of cues needed to learn discriminated the groups for vocabulary and sentence structure learning. Types of errors made discriminated the groups for sentence structure and phonology; and lack of retention of learning was a marker for the caseload participants for vocabulary. Although performance on the post-test speech accuracy measure showed greater improvement for the caseload than control group, a measure for speech seems worth retaining at this stage of the assessment's development given the prevalence of speech impairment irrespective of the number of languages learned. With the phonology assessment, it may well be that the pattern of errors rather than the response to the interactive element may be the more diagnostically valid criterion.

One difficulty interpreting the study's data may have been the caseload group's slight, but significantly poorer, performance on non-verbal cognitive measures. However this

was perhaps an inevitable artefact of the inclusive sampling methodology which reflects the real clinical population. Another difficulty was the group design that obscured individual differences in a heterogeneous population. The next planned step is to collect data on a modified DAPPLE using a case study approach. A multiple single case-study design will allow investigation of factors such as age, cognitive performance and English language exposure. The effect of language pair also needs to be investigated in greater depth to determine if dynamic assessment in English neutralises this variable.

Following further exploratory research using multiple case-studies, it should also be possible to prospectively validate the DAPPLE as a screening assessment, and evaluate whether a particular profile of scores/responses on the DAPPLE successfully identifies children as 'different' versus 'disordered'. Rather than adopting a 'gold standard' assessment, it may be necessary to follow children up longitudinally to determine whether children identified as possibly disordered by the dynamic assessment are the ones who continue to experience language and learning difficulties in the longer term. This would ideally involve a study which followed up a cohort of bilingual children, rather than just a clinical population, to ensure that the assessment could be fully validated in terms of its predictive sensitivity and specificity.

Although dynamic assessment has potential as a tool which informs the process of discriminating difference from disorder in bilingual children, the DAPPLE should be considered a pre-diagnostic assessment. Once a child has been identified as being at risk for difficulties learning English, then they need to be further assessed in English as well as their first language as recommended by the RCSLT (Communicating Quality, 2006).

References

ACORN. Available: http://www.caci.co.uk/acorn-classification.aspx [2012, 3/1/2012]

Bain, B.A. and Olswang, L.B., (1995). Examining readiness for learning two-word utterances by children with specific expressive language impairment: Dynamic assessment validation. *American Journal of Speech-Language Pathology*, 4(1), pp. 81-91.

Camilleri, B. & Law, J. (2007) Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *Advances in Speech-Language Pathology*, December 2007: 312-322

Camilleri, B. (2009) *The dynamic assessment of receptive vocabulary in preschool children.* Unpublished doctoral dissertation, City University, London

Campione, J.C., (1989). Assisted assessment: A taxonomy of approaches and an outline of strengths and weaknesses. *Journal of learning disabilities,* 22(3), pp. 151-165.

Campione, J.C. and Brown, A.L., (1987). Linking dynamic assessment with school achievement. In: C.S. Lidz, Ed, *Dynamic Assessment. An Interactional Approach to Evaluating Learning Potential.* New York: Guilford Press, pp. 82-115.

Cortazzi M. & Jin, L., (2004) Reflection on speech-language therapists' talk: implications for clinical practice and education. *International Journal of Language and Communication Disorders 39* (4), 477-480

Crutchley, A. (1999) Bilingual children with SLI attending language units: getting the bigger picture. *Child Language Teaching and Therapy,*; 201-217

De Lamo White, C. & Jin, L., (2011) Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language and Communication Disorders 46* (6), 613-627.

Dodd, B. & Crosbie, S. (2010). Language and cognition: evidence from disordered language. In U. Goswami (Ed.), *Handbook of childhood cognitive development*. 2nd edition.Oxford: Blackwell. pp 604-625,

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

Dodd, B., Holm, A. Hua, Z., and Crosbie, S.,. (2003) Phonological development: a normative study of British English-speaking children. *Clinical Linguistics and Phonetics,* 17 (8), 617-643.

Dollaghan, C. (1985) Child meets word: "Fast mapping" in preschool children. *Journal of Speech and Hearing Research, 28,* 449-454

Dunn, L.M., Dunn, D.M., Whetton, C. and & Burley, J., 1997. *The British Picture Vocabulary Scale II.* 2nd Edition. Berkshire UK: NFER-Nelson.

Elliott, C. D. (1996) British Ability Scales II. Windsor: NFER-Nelson

Feuerstein, R., Feuerstein, R.S., Falik, L.H. and Rand, Y., (2002). *The dynamic assessment of cognitive modifiability: The learning propensity assessment device: Theory, instruments and techniques.* Jerusalem: ICELP Press

Goodenough, F. L. (1926). *Measurements of Intelligence by Drawings.* NY: World Book Co.

Goldstein, B. & Swasey, P. (2001). An initial investigation of phonological patterns in typically developing 4-year-old Spanish-English bilingual children. *Language speech and hearing Services in Schools*, 32, 153-165.

Grech, H. & Dodd, B. (2009) Phonological acquisition in Malta: a bilingual language learning context. *International Journal of Bilingualism*, 12 (3) 155-171.

Grigorenko, E.L. and Sternberg, R.J., (1998). Dynamic assessment. *Psychological bulletin,* 124, pp. 75-111.

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

Hasson, N., Dodd, N., and Botting N. (2012) Dynamic Assessment of Sentence Structure (DASS): Design and evaluation of a novel procedure for assessment of syntax in children with language impairments. *International Journal of Language and Communication Disorders 47(3) 285-299*

Hasson N. (2011) *Dynamic Assessment and Informed Intervention for Children with Language Impairment*. Unpublished doctoral dissertation, City University, London

Haywood, H.C., (1993). A Mediational Teaching Style. *International Journal of Cognitive Education and Mediated Learning,* 3(1), pp. 27-38.

Haywood, H.C. and Lidz, C.S., (2007). *Dynamic Assessment in Practice. Clinical and Educational Applications.* New York: Cambridge University Press.

Hoffman, P.R., Norris J.A.& Monjure, *J. (1990) Comparison of Process Targeting and Whole Language Treatments for Phonologically Delayed Preschool Children. Language, Speech and Hearing Services in Schools, 21, 102-109*

Law, J., Boyle, J, F., Harris, F., Harkness, A., & Nye, C. (2000) Prevalence and natural history of primary speech and language delay: . Findings from a systematic review of the literature. *International Journal of Language and Communication Disorders, 35, 2.* 165-188

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).

Lidz, C, S., &., Peña, E, D., (2009) Response to Intervention and Dynamic Assessment: Do We Just Appear to Be Speaking the Same Language? *Seminars in Speech and Language, Volume 30*, Number 2

McIntosh, B & Dodd, B. (2011) *The Toddler Phonology Test*. London Pearson.

Mackey, W.F. (1968). The description of bilingualism. In J.A. Fishman (Ed.) *Readings in the sociology of language*. The Hague: Mouton, 554-584.

Mennen, I. & Stansfield, J., (2006) Speech and language therapy service delivery for bilingual children: a survey of three cities in Great Britain. *International Journal of Language and Communication Disorders 41* (6), 635-652.

Olswang, L.B, Bain, B. A & Johnson, G.A. (1992) Using dynamic assessment for children with language disorders. In S.F. Warren & J.Reichle (Eds.) *Causes and Effects in Communication and Language Intervention.* London: Paul. H. Brooks

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

Peña, E. D., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. American Journal of Speech-Language Pathology, 10, 138-154)

Peña, E. D., Gillam, R. B., Malek, M., Felter, R., Resendiz, M., & Fiestas, C. (2006). Dynamic assessment of children from culturally diverse backgrounds: Application to narrative assessment. *Journal of Speech, Language and Hearing Research, 49,* 1037-1057.

Peña, E.D, Resendiz, M, & Gillam, R. (2007). The role of clinical judgements of modifiability in the diagnosis of language impairment. *Advances in Speech-Language Pathology,* December 2007; 9 (4): 332-345

Rice, M.L., Buhr, J.C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Research, 55,* 33-42

Royal College of Speech and Language Therapists (2006). *Communicating Quality 3: Professional Standards for Speech and Language Therapists* London: Royal College of Speech and Language Therapists (RCSLT).

Royal College of Speech and Language Therapists (2010). The RCSLT Annual Report 2009-2010. London

Sheridan, M.D., (2008) From birth to five years: Children's developmental progress. London. Routledge.

Stow, C, and Dodd, B. (2003) Providing an equitable service to bilingual children in the UK: a review. *International Journal of Language and Communication Disorders 38* (4), 351-377.

Valencia, R. & Suzuki, L. (2001). *Intelligence testing in minority students: Foundations, performance factors, and assessment issues*. Thousand Oaks, CA: Sage

Vygotsky, L. (1978) Mind in society: the development of higher psychological processes. Cambridge: Harvard University Press.

Winter, K. (1999) Speech and language therapy provision for bilingual children: aspects of the current service. *International Journal of Language and Communication Disorders 34* (1), 85-98.

Winter, K. (2001) Number of bilingual children in speech and language therapy: Theory and practise of measuring their representation in *International Journal of Bilingualism 5 (4),* 465-495

Table 1. Mean (SD), range for caseload and control bilingual children: chronological age, Draw-a-Man Test and block-building performance.

	Bilingual Caseload (N= 12)	Bilingual Control (N=14)
Chronological age (months)	50.3 (5.8) 42-59	50.4(6.9) 39-58
Draw-a-Man Test (mean score)	47.3 (11.7) 39-78	61.1 (17.8) 39-88
Block building performance Age appropriate level Below age appropriate level	 4 8	 10 4

Table 2 about here

Table 2. Mean (SD) and (range) of performance on vocabulary measures for caseload and control bilingual children

	Bilingual Caseload (N= 12)	Bilingual Control (N=13)
Raw score receptive	17.9 (4.0) (10-24)	24.36 (5.0) (14-29)
Mediational score	1. 9 (1.3) (0-3)	2.6 (1.1) (0-3)
First expressive naming trial (percent named)	46.20 (27.4) (0-83)	47.29 (40.6) (0-100)
Second expressive naming trial (percent named)	33.43 (26.5) (0-100)	69.43 (32.8) (0-100)

Table 3. Mean (SD) and (range) of number of clause elements used by caseload and control bilingual children, and cues required to elicit full sentences in teach phase

	Bilingual Caseload (N= 12)	Bilingual Control (N=13)
Pre-test – no of clause elements present	4.0 (0.74) (3-5)	5.85 (2.9) (1-12)
Post –test – no of clause elements present	4.75 (2.4) (2-11)	8.69 (5.4) 0-19)
Score achieved in teach phase	5.91 (1.6) (4-8)	7.92 (2.1) (4-10)

Table 4. Distribution of sentence structure types according to group and assessment

	Bilingual Caseload – Pre-test	Bilingual Caseload – Post-test	Bilingual Control Pre-test	Bilingual Control Post-test
Clause elements present				
One element only	5	4	2	
SV	2	7	3	12
VO or VA	8	3	4	1
SO or SA	4	2	1	2
SVO	3	4	10	5
SVA	2	4	3	5
SVOA			2	9
SVAA			3	3
Other 3 or 4 elements		1		1
Additional conjoined clause		3	2	9
Additional subordinate clause		1	1	3

Key: A=adverbial, O=object, S=subject, V=verb

Table 5. Mean (SD), range for caseload and control bilingual children: words correct pre-test; words correct post-test; and number of sounds not elicited

	Bilingual Caseload (N= 12)	Bilingual Control (N=14)
Pre-test: words correct (/10)	3.4 (1.4) 0-6	7.2 (3.1) 0-10
Post-test: words correct(/10)	4.3 (1.2) 1-7	7.4 (3.0) 0-10
Sounds not stimulable	1.5 (1.7) 0-5	.7 (1.1) 0-4
Pretest: percent phonemes correct (PPC)	79.3 (7.7) 59-88	93.2 (9.5) 71-100
Post-test: percent phonemes correct (PPC)	85.4 (6.9) 73-93	93.6 (9.1) 73-100
Inconsistency	24.2 (14.4) 0-40	8.6 (11.7) 0-30

Table 6. Pearson's correlations for combined groups' post dynamic interaction assessments of phonology, vocabulary and sentence structure

	Correlation	Significance Level
phonology and sentence structure	.478	.016
sentence structure and vocabulary	.544	.005
vocabulary and phonology	.420	.032

Table 7. Examples of performance profiles of children whose group category was problematic.

Measure	Poorly performing Control	Higher performing Caseload
Pretest: PPC	87.5	82.1
Post-test: PPC	89.3	87.5
Inconsistency	30	20
No. of developmental English error patterns	4	5
No. of atypical English error patterns	4	0
Pre-test – no. of clause elements present	1	4
Post–test – no of clause elements present	4	3
No of cues required	4	4
Raw score receptive	14	22
First expressive naming trial (% named)	17	0
2nd expressive naming trial (% named)	33	17
No.of cues required	3	3