Beyond static assessment of children's receptive vocabulary: The dynamic assessment of word learning (DAWL)

Bernard Camilleri, Nicola Botting

Department of Language and Communication Science, City University, London, England

Correspondence: Bernard.camilleri.1@city.ac.uk

Abstract

Background: Children's low scores on vocabulary tests are often erroneously interpreted as reflecting poor cognitive and/or language skills. It may be necessary to incorporate the measurement of word-learning ability in estimating children's lexical abilities (Gray, Plante, Vance, & Henrichsen, 1999).
Aims: This paper explores the reliability and validity of the Dynamic Assessment of Word Learning (DAWL) a new dynamic assessment of receptive vocabulary
Method and procedures: A dynamic assessment (DA) of word learning ability was developed and adopted within a nursery school setting with fifteen children aged between 3;07 and 4;03, ten of whom had been referred to speech and language therapy.
Outcomes and results: A number of quantitative measures were derived from the DA procedure including measures of children's ability to identify the targeted items and to generalise to a second exemplar, as well as measures of children's ability to retain the targeted items. Internal, inter-rater and test-retest reliability of the DAWL was established as well as correlational measures of concurrent and predictive validity. The paper discusses the usefulness of the DAWL and other DA measures in providing information which can be used in making decisions regarding educational and therapeutic provision.

Keywords: vocabulary; dynamic assessment; language; referred

What this paper adds

Dynamic Assessments (DA) of developmental skills are increasingly used in practice and recent studies have shown that good psychometric properties can also be established.

DAs of language ability are only just beginning to be developed. This paper reports psychometric properties of the Dynamic Assessment of Word Learning (DAWL), a new measure of receptive vocabulary.

Introduction

Whether in educational, clinical or research contexts, measures of vocabulary, receptive vocabulary in particular, are frequently adopted as indicators of language ability and of school outcome. They also frequently constitute part of the assessment battery employed in the diagnosis of language impairments/disorders. A number of studies, have adopted a test of receptive vocabulary, such as the British Picture Vocabulary Scale - BPVS (Dunn, Dunn, Whetton, & Burley, 1997) either as a measure for comparison or as a criterion measure for selecting participants. Children's low scores on vocabulary tests are often interpreted as reflecting poor cognitive and/or language skills. However, these tests have been shown to possess

unacceptable reliability and validity for the identification of language impairments in preschool children (Gray et al., 1999).

One of the reasons for this is that vocabulary is an area of language which is particularly susceptible to environmental factors such as socialization practices and parental approaches to directing children's attention (Tomasello & Farrar, 1986). A review of studies related to the effects of the language environment on language development in infants and preschool children concluded that children's early experience had an important role to play in early lexical development (Harris, 1992). With the exception of severe neglect, the language environment may not, in itself, cause language difficulties severe and specific enough to constitute a language impairment (LI) (Bishop, 1997). However the environment's effect on lexical development may be sufficient to lead to a depressed score in lexical assessments and possibly to the mistaken impression of LI. Conversely, children's real language difficulties may be ignored if limited receptive and/or expressive abilities are assumed to be due to limited exposure due to diverse language background or other environmental factors. As Gray et al. (1990) suggest, it may be necessary to incorporate a dynamic measurement of word-learning ability in estimating the lexical abilities of children who achieve low scores on static assessments of vocabulary. The need for assessing children's lexical ability in terms of their ability to learn new words is reinforced by research which indicates that children diagnosed with "specific language impairment" (SLI) have impaired word-learning abilities when compared to both chronologically and language-aged matched normally developing children (Dollaghan, 1987; Rice, Buhr, & Nemeth, 1990; Rice, Oetting, Marquis, Bode, & Pae, 1994).

Dynamic assessment of children

The possibility of developing and adopting assessment procedures which incorporate learning opportunities as an alternative or complementary tool to standardised assessments has been investigated in recent years. In the context of such discussions, standardised assessments are referred to as "static" in that individuals carry out a set of test items with little or no instruction or feedback from the assessor, whereas assessments that incorporate learning opportunities are referred to as "dynamic assessments" (DA). In static assessments, any instruction is usually limited to initial guidance regarding modality of

response. Feedback is perceived as constituting a source of error of measurement which invalidates the procedure and therefore is to be avoided at all costs. When adopting a static assessment, the aim is to obtain a measure of the individual's independent performance on a particular domain at that point in time. In principle, no external variables or sources of bias are introduced. However the extent to which this is possible is disputed (Brown & Ferrara, 1985, Sternberg & Grigorenko, 2002). In direct contrast with static assessments, dynamic assessments are actually defined by the fact that a degree of instruction and feedback is incorporated in the assessment process and the individual's response to this is the main focus of the assessment (Elliott, 2003; Sternberg & Grigorenko, 2002). A construct which is central to dynamic assessment is the "zone of proximal development" (Vygotsky, 1978) which, as a theoretical concept, accounts for the child's transition from one period of development to the next. In traditional, static assessments the child either responds correctly without assistance from the examiner, or fails to respond, therefore failing the item. From the theoretical perspective of the zone of proximal development, a child may be somewhere in between success and failure, being unable to respond independently but able to achieve success with a degree of assistance. Within this framework, one can envisage two children scoring similarly on "static" measures, but responding differently when given assistance (Vygotsky, 1978). As Lidz (1991) and Minick (1987) point out, Vygotsky did not live long enough to realize the notion of the "zone of proximal development" into an actual assessment procedure. However, it remains a powerful theoretical concept for those who are interested in developing interactive/dynamic procedures (Lidz, 1991). It has led to the development of dynamic assessments across a wide range of areas including language. Across a number of populations, the use of static, standardised assessment is particularly problematic. This includes children with English as an additional language as well as children from socially disadvantaged populations (Law, McBean, & Rush, 2011). This is one of the reasons why there has recently been an increase in interest in the development of dynamic assessments of children's language (Hasson & Joffe, 2007; Hasson et al, 2013).

Dynamic Assessment and language ability

As with standardised assessment, dynamic assessment was developed in the field of psychology and has been adopted more recently in the field of language assessment. For example DA of language has been addressed from an educational, rather than a LI perspective to address the assessment of the ability to learn an additional language (Lantolf

& Poehner, 2004; Poehner & Lantolf, 2005). Poehner and Lantolf (2005) present an application of DA of second language learning within the classroom setting. The assessor/tutor provides appropriate feedback and mediation which is "negotiated with each learner with the aim of promoting language development" (p. 243). Poehner and Lantolf (2005) point out that each participant in their studies varied greatly in their abilities in the second language. Therefore English (the students' first language) was used to mediate their performance, to ensure that students understood the assessor's prompts and suggestions. Poehner and Lantolf (2005) argue that DA reduces the risk of erroneous evaluations of an individual's abilities and that their approach involves understanding and promoting the individual's conscious awareness and control of abilities rather than simply providing assistance towards completing a task.

With regards to the assessment of children for whom speech, language or communication is a cause for concern, research on DA has been sporadic. One reason for the limited amount of research may be the inherent issues involved in adopting a methodology which relies heavily on language as the medium for carrying out a defining part of the procedure (i.e. the mediational interaction), when the domain being assessed is itself language and when the individual being assessed has language difficulties. In spite of this inherent issue Peña and colleagues have successfully adopted a test-mediation-retest approach for assessing areas of expressive language including use of labels and narrative skills (Peña, Quinn, & Iglesias, 1992; Peña & Quinn, 1997; Peña & Gillam, 2000; Peña, Iglesias, & Lidz, 2001; Peña et al., 2006; Peña, Resendiz, & Gillam, 2007). The aim of this research was primarily to investigate the use of DA as a way of reducing bias when assessing children from diverse cultural and linguistic backgrounds. Some of these children achieved low scores on static tests in spite of adequate language learning abilities.

Peña and colleagues relate the comparison between pre-teaching and post-teaching performance to Vygotsky's concept of the "zone of proximal development" (Vygotsky, 1978). They also compare their work to Feuerstein's (Feuerstein, Falik, & Feuerstein R., 2003) in terms of his focus on individual's responses within the intervention/mediation process. Ultimately, as they point out (Peña et al., 2006), their application of DA of different aspects of language is most closely based on Lidz's application of cognitive DA (Lidz, 1991; Haywood & Lidz, 2007). Similarly to Lidz's (Haywood & Lidz, 2007) different applications of DA of cognitive functions, their interventions are scripted to different degrees, while encouraging assessors to respond

to the individual needs of children being assessed. Peña and colleagues consistently found that dynamic measures of change from pretest to posttest and/or measures of modifiability (within the mediation phase) worked much better than their static counterparts at distinguishing between typically developing children (with additional languages) and children with reduced language abilities (Peña et al., 1992; Peña et al., 1997; Peña et al., 2000; Peña et al., 2001; Peña et al., 2006; Peña et al., 2007). They also documented children's other areas of strength and weakness as measured by scores on a number of static tests, including assessments of non-verbal and verbal cognitive abilities. Children who failed to make progress from pretest to posttest tended to be the ones with lower language abilities, as measured by some of these additional static measures. Insofar as the DA was intended to differentiate between typically developing children with cultural differences and children with true language disorders, the dynamic assessments succeeded in constituting less biased measures than the equivalent static assessments. The extent to which children's differential responses could have been explained by their varying receptive abilities was however not explored.

Bain and Olswang (1995; 1996), on the other hand, controlled for receptive language by carrying out research with preschool children (aged 30 to 36 months) whose receptive language was at or above average levels but who had a "specific expressive language impairment" (SELI). Their DA protocol adopted a graduated series of prompts, based on similar approaches used in the assessment of cognitive abilities (Brown & Ferrara, 1985; Campione & Brown, 1987). They targeted the immediate potential for children performing at the one-word stage of expressive language development, as assessed through static assessment, to produce two-term utterances.

Children were involved over a 9-week period consisting of three 3-week phases – baseline, treatment and withdrawal. Across studies, the authors found that the dynamic measure, based on the amount of prompts children required to achieve the criterion (two-term utterances) best predicted which children demonstrated the greatest change in their rate of learning language over the duration of the study (Bain & Olswang, 1995; Olswang & Bain, 1996). They concluded that children's potential for immediate change can be assessed through their responsiveness to prompts and cues within a DA context, and that their procedure had successfully identified which children with SELI were ready to produce two-word relations.

Of the other static and discrepancy measures derived from the baseline phase, only the discrepancy between receptive and expressive language ages correlated significantly with children's "proportional change index", a measure of the progress they made from the baseline phase to the end of the withdrawal phase. Bain and Olswang (1996) argue that this indicates that individual static measures are not themselves good predictors for immediate language change but maybe somewhat more useful when used in combination to calculate discrepancy scores (Olswang and Bain, 1996). Interestingly, the more meaningful discrepancy, was the one between receptive and expressive language skills, rather than the one between non-verbal cognitive skills and language skills, which is often used diagnostically.

This finding reinforces the importance of assessing receptive language skills in children with suspected LI. Indeed, it was only recently that DA of receptive vocabulary was specifically addressed, albeit with a different population, when Alony and Kozulin (2007) investigated the assessment of receptive vocabulary in 30 children with Down syndrome aged between 47 and 96 months. Eighteen children were assessed using the DA version, while the remaining 12 children were tested without mediation using the Peabody Picture Vocabulary Test – PPVT (Dunn & Dunn, 1982). With the former group, when a child made an error on the static test, the assessor was able to provide mediation, first in the form of "focussing" (p. 327) and then through verbal mediation, depending on the child's needs (Alony & Kozulin, 2007). One example related to the word "track". Verbal mediation started with a general verbal explanation of the word, in the form of "Tracks are signs left by something that was here before" (p. 327). If further elaboration was required, the assessor might add "Have you ever walked on wet sand? Did you notice the marks that were left on the ground where you were walking earlier?" (p. 327).

In investigating developmental trends in vocabulary, Alony and Kozulin (2007) compared children who received mediation with ones who did not, as well as with normative data. They found that, whereas in the static condition the delay in receptive vocabulary between children with Down syndrome and typically developing children increased with age, under the mediation condition, children with Down syndrome demonstrated a similar developmental trend to normally developing children in the static condition. Alony and Kozulin (2007) demonstrated that a DA of receptive vocabulary could reveal underlying abilities which had been underestimated by the static version of the assessment. It is worth pointing out that the developmental

trend observed by Alony and Kozulin (2007) arose from the fact that older children responded better to the mediation than younger children. This is likely to reflect the fact that their starting receptive language levels were better (than those of younger children) and this in turn had an impact on their ability to benefit from verbal mediation.

The DA approach adopted in the current research incorporates an element of verbal mediation similar to Alony and Kozulin's (2007). However, great care was taken to adopt vocabulary and language structures which would not constitute barriers to children's learning of the targeted vocabulary, even for children with limited language abilities. Additionally a degree of standardisation was also incorporated by adopting elements of a graduated prompting approach – adopting a hierarchy of prompts, from least to most assistive (Bain et al., 1995; Campione et al., 1987) to ensure that children could benefit from the learning interaction, even if the prompts needed to be maximally assistive.

Guthke and Wingenfeld (1992) state that, when considering "the degree of individualization of learning tests, diagnosticians are caught between a rock and a hard place" (p. 83). By their very nature, dynamic assessments (or "learning tests" as referred to by Guthke) involve the provision of learning opportunities in response to the individual's need. Yet if that provision is highly individualised, the administration and scoring becomes more subjective making the analysis and interpretation of results less reliable and valid (Guthke & Wingenfeld, 1992). They argue that when confronted with the choice between the equally important features of individualisation and standardisation, assessors are faced with the option of trying to balance these two features or to lean more heavily towards standardisation, particularly if the ultimate purpose is diagnostic. This tension might be one reason for the low levels of clinical DA use in speech and language therapy. For a DA of any skill, establishing psychometric properties is likely to be a challenge. For example, because the assessment is measuring small amounts of change that occur in interaction with a mediator, a traditional analysis of reliability is difficult to apply (Haywood & Lidz, 2007). Moreover, the fact that DA aims to provide a more sensitive measure, providing information that was not identified using standard existing measures means that an appropriate evaluation of validity is also demanding. However, it is not impossible to apply such standards to a more flexible task such as the one presented here and depends upon test design features which allow both individualised elements

(items, mediation and reflection) as well as objective, quantitative administration and scoring procedures. Hasson, Dodd & Botting (2012) have recently shown that good reliability and validity can be achieved within a dynamic graduated prompts approach to measuring expressive syntax ability. This needs to be further explored in the area of receptive vocabulary and word learning skills.

In addition to the regular issues of psychometrics, dynamic assessments of language abilities need to be very carefully designed when dealing with individuals with language difficulties. The types of cues or prompts that will be most beneficial will depend on the nature and extent of these difficulties (Bain & Olswang, 1995). Assessors therefore need to be fully aware of the child's language abilities, particularly receptive language levels, as well as the language demands of the task. This is crucial in order to avoid the circularity of a situation whereby an individual with language difficulties does not demonstrate improvements through an instructional interaction, not because they lack the potential to learn, but because the learning opportunities or the mediation were inappropriate for them (Law & Camilleri, 2007). While it is impossible to totally eliminate the impact of a child's receptive language abilities on their ability to respond within the dynamic/interactive phase of a DA, it was established that this dynamic assessment of word learning (DAWL) would not primarily involve a metacognitive approach which is dependent on executive verbal control as a means for generalisable development. Rather, it would give children opportunities to demonstrate different aspects of word learning in naturalistic contexts, providing an opportunity to uncover the underlying processes. The current study set out to develop a reliable DA of receptive vocabulary which would assess children's potential to learn new vocabulary items in naturalistic contexts, building on a previous protocol developed by the first author (Camilleri & Law, 2007).

The original protocol consisted of an adapted version of the methodology first used by Dollaghan (1985) to investigate children's "fast mapping" abilities. Within this "fast mapping" research, the novel word "koob" was introduced in an exposure task" which consisted of a hiding game where the child was presented with three objects to hide – a pen, a fork and finally the "koob" (a white, oddly shaped ring). Children's ability to remember the word "koob" for both receptive and expressive purposes was subsequently investigated, with typically developing children demonstrating a remarkably good ability to identify and also to attempt productions of

the word "koob" after only one or two exposures. Camilleri and Law's (2007) procedure was based on these procedures, but adopted real words (that the child had been assessed as not knowing) rather than novel words. Additionally children were asked to identify the harder (unknown) item first, to demonstrate their ability to work out which was the targeted item in the presence of two known distractors (items that child had been assessed as knowing). If children were unable to identify the targeted item, they were given increased assistance and ultimately placed in a "best-case" situation, similar to Dollaghan's "exposure" task – with the distractors being eliminated, prior to identifying the target word. Camilleri and Law's (2007) protocol constituted a simple procedure which assessed children's ability to establish initial word-referent matches in a simple linguistic/contextual presentation. In developing the current protocol – the dynamic assessment of word learning (DAWL) - the aim was to build on the existing protocol by developing a more naturalistic assessment of word learning that would closely approximate children's exposure to words in their everyday school context. Rather than assessing their fast mapping in a simplified presentation or "best case" scenario, children would initially be required to pick out a targeted word within a longer stream of speech while being presented with the extralinguistic information (e.g. a pictorial scene). Similar word-learning opportunities might occur, for example, in a classroom context when a teacher describes pictures in a story book to a group of children. This particular type of fast mapping is referred to as "quick incidental learning" (QUIL) and involves children's ability to make sense of a scene and of the accompanying language, without the need to negotiate joint reference (Rice et al., 1994). The word itself is embedded within a simple syntactic context allowing the child to combine fast mapping with "syntactic bootstrapping" (Brown, 1957) in order to establish an initial representation for the word. In typically developing children, this ability is established by the middle of the pre-school years (Rice et al., 1994). As might happen within a naturalistic context, if QUIL failed, this would be followed by providing children with semantic cues towards identifying the targeted word. If this was still not sufficient, the QUIL presentation would be simplified by presenting the word in a single simple phrase, facilitating bootstrapping and fast mapping further. The revised protocol adopted composite pictures which, by definition, targeted highly imageable words, making them suitable for an assessment of children's ability to attaching meaning to words

using both linguistic and extralinguistic (pictorial) cues (Gillette, Gleitman, Gleitman & Lederer, 1999).

In adopting the new protocol, our main aims were:

(1) To explore the reliability of the new DA protocol - the dynamic assessment for word learning (DAWL)
(2) To establish how the measure performs in relation to standardised static tests and in relation to teacher reports of language use (correlational validity).

Method

Participants

The study was carried out in two nurseries in inner-city London. Participants consisted of fifteen children who had recently joined nursery. Five of them were typically developing children aged between 3;06 and 4;01 selected randomly from children within this age group and for whom no concerns about language had been raised. These typically developing children were assessed first in order to pilot the new DAWL procedure and ensure that children were able to engage with the materials for the duration of the assessment (approximately thirty five to forty minutes). Most of the analyses refers to the other ten children (see below), although the typically developing children were included in an exploration of the correlation between DAWL and BPVS scores (table 2).

Ten children, aged between 3;07 and 4;03, who had been referred to speech and language therapy were identified across the two nurseries. Three of the children had been referred prior to starting the nursery by their parents. The remaining children were identified as having possible language difficulties by teachers and/or support staff within the school. All of the children only received speech and language therapy services within the school, according to the local Speech and Language Therapy provision arrangements.

All of the referred participants were children for whom the speech and/or language difficulties were the primary cause for concern. Children with an identified primary cause for a language difficulty (e.g. hearing impairment) were excluded. No other exclusionary criteria were adopted.

The DAWL procedure was designed to be administered in English and a basic receptive vocabulary in English was required. In both nurseries direct observation of the children made it possible to ascertain that all of the children, whether within the referred group or the normally developing group, used English within the nursery, although some of them also used other languages at home. This allowed for the inclusion of seven children for whom English was an additional language (EAL), five of whom were in the referred group and 2 in the typically developing group. Children with EAL were not considered as a separate group, but rather were included as they constitute a population for whom standardised assessments are particularly problematic and for whom information on learning ability was particularly relevant. Beyond the fact that speech and/or language was the primary cause for concern, there was no a priori reason for the participants to constitute a group of similar children. To some extent, the contrary is true. The hope was that different children would respond differently to the DAWL, and that this would reflect different learning potential and patterns of growth over time. This was the main focus of the research and most of the analyses refer to this group of (ten) children.

Measures

The assessment battery adopted in this research consisted of a combination of static and dynamic assessments which could be used within a single interaction lasting approximately 35 to 40 minutes. The BPVS (Dunn et al., 1997) was adopted as a static pretest of receptive vocabulary as well as a means of identifying vocabulary

items which a child had difficulty with. Some of these items were targeted during the interactive phase of the assessment, which is described further below. The Block Building task (BBT) and the Picture Similarities Task (PST) from the British Ability Scales (Elliott, 1996) were adopted as (static) measures of non-verbal ability. The former of the two also acted as an ice-breaker, whereas the latter was carried out between the interactive phase and the posttest phase. The assessment battery was carried out in the following order: BBT, BPVS, interactive phase, PST, posttest.

The interactive phase of the DAWL involved conversational interactions in which the child and assessor viewed a number of composite pictures together. Each composite picture presented a scene which included a number of objects and actions, one of which was a target vocabulary item which the child had difficulty identifying in the BPVS (see Fig. 1). The target item was represented at least twice within the composite picture. The aim of the interactive phase was to assess the child's ability to establish a match between the targeted word and referent (element within the picture) and to demonstrate the extent to which they could retain that word for expressive and/or receptive purposes. The child was assigned quantifiable scores for each of these component abilities (further details below).

The child was initially given an opportunity to explore the picture and describe it in their own words. The first level of prompting involved the assessor using open questions (e.g. "What can you see in this picture?"), followed by probes (e.g. "Where is the woman?"). If the child was able to use the word expressively in this naturalistic setting, it would clearly indicate that the child had a good understanding of the word but had been unable to demonstrate this in the standardised assessment. If the child did not use the target word expressively, he/she was subsequently given increasingly assistive settings within which to establish a link between the target word and the

referent. The next level (level 2) consisted of giving a three-part description of different aspects of the composite picture. For example, with the composite picture, targeting the word "balcony", the assessor might say:

"The girl is standing outside the house.
The woman is standing on the balcony
…and she is calling out to the girl."

Figure 1 inserted about here

At this stage the assessor would be careful not to provide clear and specific non-verbal cues in relation to the target words, in order to establish whether the child was able to identify the target from the linguistic and pictorial context, without requiring the adult to establish joint reference with the child. If the child were able to establish a word-referent match within this context, he/she would be engaged in an interaction where the semantic features of the word were discussed and the child's ability to generalise their learning would be assessed by establishing whether he/she could identify another occurrence of the same referent within the composite picture. For example, with the target word "balcony", the assessor might say:

"A balcony is part of a house, but it's outside.
You can go out on the balcony
…and you can keep plants and flowers on the balcony"

The amount of assistance required for the child to find a second occurrence of the referent would also be recorded. A pre-school child with adequate word-learning ability would be expected to pick up the word from the stream of speech using QUIL, engage in a discussion on the semantic features of the word and generalise easily to other occurrences of the word/referent.

A child might be unable to pick out a word from a stream of words and establish an initial link between word and meaning (from the linguistic context), but might possess conceptual/semantic knowledge which could be drawn upon to identify the referent correctly. So, if the child was unable to pick up the new word from the stream of words, the child's semantic knowledge of the word would be adopted/assessed by giving the child semantic information about the word (as above) and observing whether this assisted him/her towards identifying the targeted word. In other words, the child's semantic knowledge might act as the starting point in establishing a word-referent match. This is comparable to the verbal mediation provided by Alony and Kozulin's (2007) in their research on DA of receptive vocabulary with children with Down syndrome.

A child's ability to identify a first referent at this third level of prompting, and to generalise to a second referent at this stage would indicate that they possessed semantic knowledge associated with the word, but had not yet established a complete lexical representation for the word, for either receptive or expressive purposes.

If the child were still unable to identify the targeted word it could be presented in a syntactically simple context and accompanied by social and non-verbal cues, such as eye-gaze (e.g. "Look: The woman is standing on the balcony"). Camilleri and Law's (2007) research on DA, as well as research on fast mapping and word learning indicates that children with language impairment would be expected to be able to respond at this fourth level of input. However, if the child were still unable to identify the referent correctly, this could be made explicit by physically pointing it out to the child in temporal contiguity with the word. This would constitute the fifth and final level of prompting. If a child were to consistently depend on the fifth level of assistance, it would indicate that they were at a very basic word learning level –

mainly dependent on the processes of perceptual salience and temporal contiguity, as is the case with normally developing infants during their first months of word-learning.

During the interactive phase of the DA, each child was shown up to eight composite pictures, each targeting a specific vocabulary item. Each conversational interaction around a particular picture was driven by the child's contributions. The assessor was guided by scripts for each picture, but these were adopted flexibly according to the individual child. The main reason for scripting was to provide a consistent hierarchy of assistance through the interaction. The graded process of the interactive phase is depicted in figure 2 below.

The interactive phase could be evaluated quantitatively, in terms of the amount of prompting needed for each word. The child obtained a weighted dynamic score for each targeted word, on the basis of the amount of prompting needed for the child to match the word with the referent. Five points were assigned to the first, least assistive level of prompting and one point was assigned to level five. This scoring system is similar to the one adopted by Bain and Olswang. Using this weighted scoring system, the child was given credit for levels of prompting below the ones adopted to help differentiate children functioning at different levels. The current study adopted a mean weighted dynamic score (MWDS) for comparability across children who had been exposed to different numbers of composite pictures and target words.

Children's ability to generalise word learning across referents within the composite picture was also assigned a weighted score, referred to as the mean weighted generalising score (MWGS). Children were given a score of three if they spontaneously identified a second occurrence of the referent or if they identified it without assistance when asked to do so. They were assigned two points if they

required assistance in the form of linguistic input and a score of one if they were physically directed to the second occurrence. By adopting weighted scoring, children could achieve a MWGS between 1 (if they were physically directed to the second occurrence of each targeted word) and 6 (if they successfully generalised all words without assistance).

Figure 2 inserted about here

Besides considering the amount of assistance children needed to establish an initial link between a targeted word and referent and to generalise to a second occurrence of the referent, the DAWL incorporated measures of children's ability to establish and retain representations of the word in memory for both receptive and expressive purposes. These measures were incorporated within the interactive phase and also included in the posttest phase as the "posttest of content". Once the child had identified the two occurrences of the target item, he/she was given opportunities within the conversational interactions to use the word expressively. If they were unable to do this, they were subsequently given the opportunity to point it out. The extent to which children were able to use the targeted words expressively and receptively was scored. If the child was able to use a word appropriately expressively, demonstrating knowledge of both the form and the semantics of the word, the child was assumed to be able to also use it receptively, and credited with a score on the receptive scale.

As a rule children were shown eight composite pictures. However if the assessor felt it was appropriate, he/she could target fewer vocabulary items. For this reason, children's score on the posttest of content was expressed as a proportion. For example a child who was able to use four out of eight targeted items expressively

would achieve a score of 0.5. This method of scoring allowed comparability, irrespective of the number of items targeted. Two of the referred children were in fact targeted with six items instead of eight because they lost focus on the activity and would perhaps not have responded to further items in a way which was representative of their abilities. The other eight children engaged fully with the materials and eight target items were used.

The posttest phase also included a "posttest of process". A process similar to the interactive phase was carried out with four words which were selected on the basis that they consisted of advanced vocabulary items ("syringe", "easel", "glider", "fern") which preschool children would not be expected to have knowledge of. This "post-test of process" was therefore a dynamic posttest, which also included graded assistance on the part of the assessor, similar to Campione and Brown's (1985) use of dynamic maintenance/transfer tasks, which allowed the investigation of children's ability to adopt similar learning strategies across items of similar or increased complexity. The scores obtained at posttest are referred to as "mean weighted dynamic posttest" (MWDP) and "mean weighted generalising posttest" (MWGP) respectively.

Sequence of research activities

Ethical approval from the local NHS Ethics Committee was sought and obtained. Additionally, children's parents/carers gave written consent for their children's participation prior to the start of the research study. The first stage of research (time 1) took place approximately two months after the start of the school year. Typically developing children were only assessed at time 1. Referred children were followed up approximately six months later (time 2), before the end of the school year.

Referred children were assessed using the full battery of static and dynamic assessments at both times.

All of the referred children received speech and language therapy input, provided by the local health authority, within the nursery setting between time 1 and time 2. This led to the opportunity of making use of the information derived from the DA to inform service delivery to individual children. Written reports provided therapists with quantitative as well as qualitative information regarding the referred children's responses within the DA. Additionally, meetings were held with the therapists and key worker, during which the findings from the DA and ways in which this could inform their interactions with the individual children were discussed.

At time 2, children's key workers (the learning support assistant who was assigned main responsibility for the child) were given a questionnaire in which they were required to rate children's progress in vocabulary and language, their ability to learn new words in different settings, their ability to generalise language learning and the extent to which they would continue to require support when they moved from nursery to reception class.

Results

Reliability

Inter-tester reliability was investigated after the first few children were assessed with the DA procedure. The semi scripted nature of the different phases meant that qualified practitioners following the guidelines and scripts should be able to carry out the assessment in a similar fashion. Naturally one would expect a range of variations as would occur when different participants are involved in conversational interactions. However the degree of structure and scripting was intended to allow different

assessors to achieve similar responses when assessing a particular child. Although the primary researcher was the only assessor present at each DA procedure, inter-rater reliability on scoring of the DAWL could be evaluated thanks to video-recordings of the assessment sessions. A speech and language therapist was instructed on the procedure and scoring system and subsequently watched a random sample of four of the assessments carried out. Agreement was high across all of the weighted components (MWDS: 90.6%; MWGS: 96.88%; MWDP: 93.75%; MWGP: 93.75). The agreement on the more objective receptive and expressive components was 100%. This indicated that the DAWL possessed high inter-rater reliability, as far as scoring the assessment was concerned.

Internal consistency was established through examining correlations between scores on the different components of the DAWL (see table 1). Spearman's correlation coefficients were adopted given that the measures were on different scales. Positive correlations were found across all of the scores derived from the DAWL. A number of correlations were statistically significant. For example the MWDS and the MWGS were found to be highly significantly correlated (.829; p 0.002).

Insert table 1 here

Although the post-test of process cannot be strictly considered as providing a measure of test-retest reliability it does give an indication of consistency of process for each child and controls against the task tapping into content-specific fluctuations in performance. Highly statistically significant correlations were found between scores on analogous tasks across the interactive phase and the posttest of process phase of the assessment, with different vocabulary items. For example the MWDS (obtained during the interactive phase using vocabulary items selected individually for

each child) was highly correlated with the analogous score obtained during the posttest of process (the MWDP), when a common set of "harder" words were adopted with all children (0.878; $p < 0.001$). Similarly, the MWGS (during interactive phase) and the MWGP (during the posttest phase) were also significantly correlated (.685; p 0.014).

Insert table 2 and table 3 here

Associations with other measures of language functioning (Correlational Validity)

Appropriate evidence to assess the validity of a new test varies (Kane, 2006). One important contribution to this process is the examination of association between the novel measure and existing measures of functioning. In DA, assessing validity is particularly difficult because we hypothesise that the DA will be more sensitive than other measures. Thus, it is important to note that there is a tension between having high correlations for static and dynamic measures (indicating validity) versus lower correlations (indicating perhaps that the assessments are managing to give additional information when used together).

Correlational validity was explored in the first instance by examining the relationship between scores derived from the DAWL and the BPVS scores concurrently. Table 2 shows the correlations (Spearman's rho) between BPVS scores and the different scores derived from the DAWL across both referred and non-referred groups. Positive correlations were found between the BPVS score and all other measures. All correlations were statistically significant, with the exception of the correlation with the generalising score at post-test, although this correlation also

was close to statistical significance. This is indicative of the concurrent validity of the DAWL measures.

Similarly, positive correlations were found between the BPVS score and all other measures when the referred group was considered separately (see table 3). Four of the six measures derived from the DAWL were found to correlate significantly with BPVS score. The moderate correlations seen here may suggest that the DAWL is measuring skills not tested by the BPVS.

As far as the DAWL's predictive validity was concerned, the MWDS was found to be positively correlated with the full range of measures taken at time 2 of children's progress and status (see table 4). The first of these measures was the change in BPVS score between time 1 and time 2. MWDS was found to be significantly correlated with the change in percentile score on the BPVS and with keyworkers' overall ratings of children. Of the different components on the rating scale, keyworkers' ratings of children's ability to generalise learning and of children's need for support for language and learning were statistically significantly correlated with children's MWDS. The analogous measure taken during the posttest of process was also positively correlated with measures of referred children's status at time 2 (see table 5).

Insert table 4 and table 5 about here

By comparison, the MWGS was found to be significantly correlated with LSA's overall ratings of referred children, but not with the change in BPVS scores (see table 6). The positive correlations between the dynamic measure and LSAs ratings for "generalising new language" and "support for language and learning" were also very close to statistical significance. Generally positive correlations were also

found between the MWGP score and all other measures, but none of them were statistically significant.

Positive correlations were also found between the proportion of words named by children during the interactive phase and measures of referred children's status at time 2. However none of them were statistically significant, maybe due to small numbers of participants.

A rather different set of correlations was found when the second expressive task was considered (see table 7). This was the expressive task carried out during the posttest of content and reflected children's ability to retain sufficiently specified lexical representations of targeted words in memory for the duration of the assessment. No correlation was found between this measure and changes in BPVS score. On the other hand, generally positive correlations were found between this measure and keyworkers' ratings of children's status as word learners. In particular, a significant correlation was found between children's score on the expressive task and the rating of their ability to generalise word learning.

Insert table 6 and table 7 here

Discussion

The challenge in developing this DA of word learning was to develop a procedure which was psychometrically robust, and yet would possess ecological validity and provide a range of measurable (i.e. quantitative) responses/scores. Furthermore, the measure aims to provide qualitative information which would inform teachers and key workers in addressing children's needs.

The DAWL showed good levels of internal consistency, inter-rater- and test-retest reliability as well as correlational concurrent and predictive validity with a variety of measures. To our knowledge this makes it the first dynamic assessment of receptive vocabulary to be developed in this way. The semi-scripted conversational approach which was adopted achieved a balance between flexibility and responsiveness to the learner and the need for reliable outcomes and interpretations, as advocated by Haywood and Lidz (2007). From the child's perspective, the interaction approximated the naturalistic circumstances of word use in school settings and other contexts. However, in the graduated prompting tradition rather than the mediational one, the assistive levels of cueing were predetermined.

Evidence for the concurrent validity of the DAWL can be found in the positive correlations between each of the measures derived from the DAWL and the static BPVS scores. Positive correlations, most of which were statistically significant, were found whether the whole group of fifteen children was considered or whether the referred group alone was considered. This is not to say that the measures obtained from the DA were redundant. These measures reflected different aspects of children's abilities to learn new words whereas the BPVS was a performance measure of children's word knowledge. It is to be expected that measures of children's word

learning abilities (e.g. the DAWL) and measures of word knowledge (e.g. BPVS) are correlated when group data is considered. With larger, longitudinal samples, future studies will be able to provide sensitivity and specificity data to add to the evidence surrounding validity of DA measures. Similarly, it is to be expected that different aspects of word learning are also related. This was in fact found to be the case, with measures derived from the DAWL being highly correlated to each other. This is further evidence that the DAWL measured different aspects of the same construct, and is therefore evidence of the measure's internal consistency.

Children do not learn language in isolation, or via the type of static test paradigm used in the BPVS, but rather use cues from the situation and from their conversational partners to narrow down a word meaning (Bloom, 2000; John-Steiner, Panofsky, & Smith, 1994), and then later to extend this to other examples. The naturalistic context as well as the range of tasks adopted within the DAWL meant that ecologically valid aspects of word learning were explored. With the exception of the receptive task within the "post-test of process", children produced a range of responses on the different components of the DA. This meant that it was possible to look into the link between children's scores on the DA and their progress over the six-month period of the study.

Notably the weighted dynamic score (interactive phase) at time 1 was found to be significantly correlated with both of the key measures of children's status at time 2 - the change in percentile score on the BPVS between time 1 and time 2 and key workers' overall ratings of children's status at time 2. The analogous scores derived during the posttest of process were also consistently positively correlated with measures of children's progress/status at time 2. In their pioneering research adopting a graduated prompting approach to assess cognitive skills, Campione and Brown

(1987) found that learning and transfer scores possessed predictive validity, providing additional information which could "play a role in the *identification* component of diagnosis" (Campione & Brown, 1987, p. 99). The indications are that the same could be said on the findings regarding the weighted dynamic scores, the first of which constituted a measure of learning and the second of which could be considered a measure of transfer of that same process of learning. Positive correlations were also found between both weighted generalizing scores and keyworkers' overall ratings.

These strong correlations between weighted DAWL measures and measures of children's progress were found in spite of the small number of children involved. While it would be inappropriate to overstate the clinical significance of these findings, the data does provide some evidence of the predictive validity of the weighted scores (particularly the weighted dynamic scores) which provide a measure of children's ability to establish new word-referent matches in interactive contexts.

Expressive scores were also generally positively correlated with measures of children's progress. Interestingly, the second expressive score (during the posttest) which can be considered to give a measure of children's ability to retain lexical representations beyond the immediate learning situation, was found to be significantly correlated with keyworkers' rating of children's ability to generalize their language learning to contexts beyond the immediate learning one. This suggests the validity of the expressive measure. It is worth noting that children's retention of the specific words targeted in the DAWL was not assessed beyond the posttest. While children may well have retained some of the words and continued to add further information to the initially fast-mapped representation, the scope of the assessment was to measure children's potential to learn words, rather than their longer term learning of those specific words. Further studies are now needed to extend and confirm the

psychometric properties of the DAWL as a valid and reliable measure of children's word-learning potential.

Clearly, the measures derived from the DAWL, key workers' ratings and the static BPVS are qualitatively different. Nevertheless the positive correlations among them point towards a related underlying construct. In particular, the correlations between DAWL scores and measures of children's medium term progress (over the six-month period) validates the procedure as a measure of children's word learning potential.

One of the key criticisms of static measures is that while they might possess predictive validity in a group or correlational sense, standardised tests are less useful when applied to prediction in individual cases (Haywood et al., 1992). Additionally, Brown and Ferrara (1985) also argued that static measures are particularly poor as predictors of later outcome when used with preschool-age children. This is certainly true of the assessment of young children's language, given that individuals vary considerably in the rate and route of development (Law, Boyle, Harris, Harkness, & Nye, 1998; Leonard, 1998; Tomblin, 2008).

In this respect it is important to look at individual cases when considering whether DA does a better job than static assessments when applied to these cases. A detailed analysis of individual children's profile of static and dynamic scores, as well as their progress over time, is beyond the scope of the current paper. However, it is important to mention that for most (but not all) of the children, the quantitative measures derived from the DAWL were found to be consistent with their progress over the duration of the study.

The main exception was Pablo (pseudonym), a child with EAL, whose responses within the DAWL were indicative of good word learning potential. In his

case, while his responses within small group activities mirrored his performance on the DAWL, this was not followed by more generalised improvements in language abilities over the six-month period. In this case, the quantitative measures derived from the DA were not followed by the expected progress. Nevertheless the DAWL constituted a useful source of qualitative information for school staff who worked with him in one-to-one and small group contexts. A commonality across children was therefore that the DA informed the process of planning interactions aimed at improving their vocabulary and their language skills more broadly. This exploratory research was carried out with a relatively small sample of children precisely to allow the possibility of using the qualitative information obtained to plan individualised therapy and school activities for all of the referred children involved.

Overall the quantitative measures derived from this DA did provide additional information regarding the likelihood that a child would make progress in the short to medium term, as demonstrated by group correlational data. Perhaps inevitably, when individual cases were considered, the extent to which children in fact made progress which was consistent with their DA performance varied. In this respect, the DAWL should be considered as a source of information which complements standardised assessments and other methods of information gathering as part of a comprehensive assessment repertoire, rather than some foolproof method for classifying children.

The use of DA has implications for subsequent therapy content and style. For example, one of the implicit aims of any DA is to provide qualitative information not available from a static test score. In this particular example, therapists and teachers are able to identify not only which words the child does not understand, but also the patterns of error and success. In other words, the professional can note the types of cues which facilitate the child's identification of a vocabulary item, whether

generalisation to other examples is occurring and whether the child has any insight into their own vocabulary learning strategies. Depending on these patterns a number of changes can be implemented into clinical and educational practice. Firstly, teachers and therapists could utilise graduated prompting techniques in teaching/learning situations, similar to the ones which were found to be successful in the DA context. Second, where appropriate, they could focus more directly on teaching strategy use alongside a metalinguistic approach, whereby children are encouraged to make their strategies explicit and conscious. For example, children can be asked 'what other bits of the picture helped you there?'. However, it is important to note that the 'assumed' benefits to intervention have yet to be proven in a convincing way. Ongoing work by Hasson suggests that therapeutic implications of DA are likely to be multi-factorial and dependent on how information is communicated from assessment to therapist/teacher, on the training provided to the therapist/teacher and on the individual children themselves (Hasson & Botting, 2010).

Research on dynamic assessments of language which were designed within both the mediational/metacognitive model (Peña et al., 1992; Peña et al., 2001; Peña et al., 2006, Ukrainetz et al. 2000.) and the graduated prompts approach (Bain & Olswang, 1995; Olswang & Bain, 1996; Glaspey & Stoel-Gammon, 2007) have tended to emphasize the unique quantitative information, which was provided by the dynamic assessments and which was not available through the use of static measures. Given the limitations of static tests when assessing young children, particularly as far as the assessment of language is concerned (Rutter, 2008), it is unsurprising that researchers have sought to obtain additional quantitative information which can inform classification, or the "identification component" (Campione & Brown, 1987, p. 100)

of diagnosis. Within the context of the provision of health services, where the question as to *whether* to provide intervention is fundamental, quantitative measures derived from dynamic assessments might play an important role. On the other hand, while dynamic assessments might *improve* classification, it is important to be aware that no quantitative measure, whether derived from static or dynamic assessments, can be accurate all of the time and with all cases. There are several reasons for this, which include the fact that there is no gold standard against which to evaluate this classification, that children's responses within a test situation (static or dynamic) change, and, perhaps most importantly because the individual's underlying cognitive or language abilities are also constantly changing as a result of maturational forces in interaction with social and environmental factors. What may be more relevant, particularly in educational contexts is how dynamic assessments can inform the "prescriptive component" (Campione & Brown, 1987, p. 99) of diagnosis, in informing intervention or educational provision.

The DAWL, trialled in this study, produced additional information about children's vocabulary skills on both quantitative and qualitative levels. The findings support the view that standardised and dynamic assessments should be viewed as complementary tools which, in combination, can serve different purposes. When a child achieves low scores on a static assessment, the DAWL may be used to improve both the accuracy of the differential diagnosis and the understanding of the processes underlying the child's performance. While some questions are best addressed using static assessments and others using dynamic assessments (Haywood & Lidz, 2007), it is ultimately the professional, not the assessment tool that makes informed decisions about whether intervention is warranted and what the nature of the intervention should be. The DAWL procedure provided additional information that can help the

practitioner make informed decisions on both of these levels. The goals of improving diagnostic classification and of collecting qualitative information for the purpose of designing individualised intervention are not mutually exclusive and are probably best achieved through a variety of sources, which can include both standardised and dynamic assessments.

Reference List

Alony, S. & Kozulin, A. (2007). Dynamic assessment of receptive language in children with Down syndrome. ***Advances in Speech-Language Pathology, 9,*** **323-331.**

Bain, B. A. & 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,*** **81-91.**

Bishop, D. V. M. (1997). ***Uncommon Understanding - Development and Disorders of Language Comprehension in Children*****. Bath: Psychology Press.**

Bloom, L. (2000). The intentionality model of word learning. In R.M.Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), ***Becoming a word learner - A debate on lexical acquisition*** **(pp. 19-50). Oxford: Oxford University Press.**

Brown, R. (1957). Linguistic determinism and the part of speech. ***Journal of Abnormal and Social Psychology 55,*** **1-5**

Brown, A. L. & Ferrara, R. A. (1985). Diagnosing zones of proximal development. In J.W.Wertsch (Ed.), ***Culture, Communication and Cognition: Vygotskian Perspectives*** **(pp. 273-305). Cambridge: Cambridge University Press.**

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, 9,* 312-322.

Campione, J. C. & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C.S.Lidz (Ed.), *Dynamic assessment* (pp. 82-115). New York: The Guilford Press.

Dollaghan, C. (1987). Fast mapping in normal and language impaired children. *Journal of Speech and Hearing Disorders, 52,* 218-222.

Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale*. Windsor: NFER - Nelson.

Dunn, L. M. & Dunn, L. M. (1982). *Peabody Picture Vocabulay Test - revised*. Circle Pines MN: American Guidance Service.

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

Elliott, J. (2003). Dynamic Assessment in Educational Settings: Realising potential. *Educational Review, 55,* 15-32.

Feuerstein, R., Falik, L. H., & Feuerstein R. (2003). The Learning Potential Assessment Device. In *Feuerstein's theory and applied systems: A reader* (pp. 51-97). Jerusalem: The International Center for the Enhancement of Learning Potential.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73,* 135-176

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. ***Language Speech and Hearing Services in Schools, 30,*** **196-206.**

Guthke, J. & Wingenfeld, S. A. (1992). The Learning Test concept: Origins, state of the art, and trends. In C.H.Haywood & D. Tzuriel (Eds.), ***Interactive testing*** **(pp. 64-93). New York: Springer-Verlag.**

Harris, M. (1992). ***Language experience and early language development - From input to uptake*****. Hove, East Sussex: Psychology Press.**

Hasson, N. & Botting, N. (2010). Dynamic assessment of children with language impairments: A pilot study. ***Child Language Teaching and Therapy, 26,*** **249-272.**

Hasson, N., Camilleri, B., Jones, C., Smith, J. & Dodd, B. (2013). Discriminating disorder from difference using dynamic assessment with bilinigual children. ***Child Language Teaching and Therapy, 26***, 57-75

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

Haywood, C. H. & Lidz, C. S. (2007). ***Dynamic assessment in practice: Clinical and educational applications*****. New York, NY: Cambridge University Press.**

John-Steiner, V., Panofsky, C. P., & Smith, L. W. (1994). ***Sociocultural approaches to language and literacy*****. Cambridge: Cambridge University Press.**

Kane, M. T. (2006). Validation. In R.L.Brennan (Ed.), ***Educational Measurement*** **(pp. 17-64). Westport, CT: American Council on Education.**

Lantolf, J. P. & Poehner, M. E. (2004). Dynamic assessment of L2 development: bringing the past into the future. ***Journal of Applied Linguistics, 1,*** **49-72.**

Law, J., Boyle, J., Harris, J., Harkness, A., & Nye, C. (1998). Screening for speech and language delay: A systematic review of the literature. ***Health Technology Assessment, 2*****.**

Law, J. & Camilleri, B. (2007). Dynamic assessment and its applications to children with speech and language learning difficulties. ***Advances in Speech-Language Pathology, 9,*** **271-272.**

Law, J., McBean, K., & Rush, R. (2011). Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage. ***International Journal of Language & Communication Disorders, doi: 10.1111/j.1460-6984.2011.00036.x*****.**

Leonard, L. (1998). ***Children with specific language impairment*****. Cambridge, MA: The MIT Press.**

Lidz, C. S. (1991). ***Practitioner's guide to dynamic assessment*****. New York: The Guilford Press.**

Olswang, L. B. & Bain, B. A. (1996). Assessment information for predicting upcoming change in language production. ***Journal of Speech and Hearing Research, 39,*** **414-423.**

Peña, E. & Gillam, R. B. (2000). Dynamic assessment of children referred for speech and language evaluations. In C.S.Lidz (Ed.), ***Dynamic assessment: Prevailing models and applications*** **(pp. 543-575). New York, Elsevier Science.**

Peña, E., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C. et al. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. ***Journal of Speech, Language, and Hearing Research, 49,*** **1037-1057.**

Peña, E., 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. & Quinn, R. (1997). Task familiarity: Effects on the test performance of Puerto Rican and African American children. ***Language Speech and Hearing Services in Schools, 28,*** **323-332.**

Peña, E., Quinn, R., & Iglesias, A. (1992). The application of dynamic methods to language assessment: A nonbiased procedure. ***The Journal of Special Education, 26,*** **269-280.**

Peña, E., Resendiz, M., & Gillam, R. B. (2007). The role of clinical judgement of modifiability in the diagnosis of language impairment. ***Advances in Speech-Language Pathology, 9,*** **332-345.**

Poehner, M. E. & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. ***Language Teaching Research, 9,*** **233-265.**

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.**

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. ***Journal of Speech and Hearing Research, 37,*** **106-122.**

Sternberg, R. J. & Grigorenko, E. L. (2002). ***Dynamic testing: The nature and measurement of learning potential*****. Cambridge: Cambridge University Press.**

Tomasello, M. & Farrar, M. (1986). Joint attention and early language. ***Child Development, 57,*** **1454-1463.**

Tomblin, J. B. (2008). Validating diagnostic standards for specific language impairment using adolescent outcomes. In C.Frazier Norbury, J. B. Tomblin, & D. V. M. Bishop (Eds.), ***Understanding developmental language disorders*** **(pp. 93-114). Psychology Press: New York.**

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

Table 1 Correlations across DA measures at time 1 for referred children

		MWDS t1	*MWGS t1*	*MWDP t1*	*MWGP t1*	*proportion of correctly named items1 t1*	*proportion of correctly named items2 t1*
MWDS t1	*Correlation Coefficient*[*]	1.000					
	Sig. (1-tailed)	.					
	N	10					
MWGS t1	*Correlation Coefficient*[*]	.829[**]	1.000				
	Sig. (1-tailed)	.002	.				
	N	10	10				
MWDP t1	*Correlation Coefficient*[*]	.878[**]	.711*	1.000			
	Sig. (1-tailed)	.000	.011	.			
	N	10	10	10			
MWGP t1	*Correlation Coefficient*[*]	.730[**]	.685*	.731[**]	1.000		
	Sig. (1-tailed)	.008	.014	.008	.		
	N	10	10	10	10		
proportion of correctly named items1 t1	*Correlation Coefficient*[*]	.496	.558*	.521	.631*	1.000	
	Sig. (1-tailed)	.072	.047	.061	.025	.	
	N	10	10	10	10	10	
proportion of correctly named items2 t1	*Correlation Coefficient*[*]	.617*	.604*	.369	.321	.640	1.000
	Sig. (1-tailed)	.029	.032	.147	.183	.023	.
	N	10	10	10	10	10	10

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

Table 2 Correlation between BPVS and DA scores at time 1

			Percentile score on BPVS t1
Spearman's rho	*Mean weighted dynamic score t1*	*Correlation Coefficient*	.782**
		Sig. (1-tailed)	.000
		N	15
	Mean weighted generalising t1	*Correlation Coefficient*	.850**
		Sig. (1-tailed)	.000
		N	15
	MWDP t1	*Correlation Coefficient*	.710**
		Sig. (1-tailed)	.002
		N	15
	Mean weighted generalising posttest t1	*Correlation Coefficient*	.436
		Sig. (1-tailed)	.052
		N	15
	proportion of correctly named items1 t1	*Correlation Coefficient*	.499*
		Sig. (1-tailed)	.029
		N	15
	proportion of correctly named items2 t1	*Correlation Coefficient*	.598**
		Sig. (1-tailed)	.009
		N	15

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

Table 3 Correlation between BPVS and DA scores at time 1 (referred group)

			Percentile score on BPVS t1
Spearman's rho	*MWDS t1*	*Correlation Coefficient*	.648*
		Sig. (1-tailed)	.021
		N	10
	MWGS t1	*Correlation Coefficient*	.804**
		Sig. (1-tailed)	.003
		N	10
	MWDP t1	*Correlation Coefficient*	.652*
		Sig. (1-tailed)	.020
		N	10
	MWGP t1	*Correlation Coefficient*	.275
		Sig. (1-tailed)	.221
		N	10
	proportion of correctly named items1 t1	*Correlation Coefficient*	.453
		Sig. (1-tailed)	.094
		N	10
	proportion of correctly named items2 t1	*Correlation Coefficient*	.624*
		Sig. (1-tailed)	.027
		N	10

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

Table 4 Correlation between mean weighted dynamic score at time 1 and measures at time 2

			MWDS t1
Spearman's rho	*bpvs difference*	*Correlation Coefficient*	.549*
		Sig. (1-tailed)	.050
		N	10
	progress in language	*Correlation Coefficient*	.207
		Sig. (1-tailed)	.283
		N	10
	learning language	*Correlation Coefficient*	.522
		Sig. (1-tailed)	.061
		N	10
	generalising new language	*Correlation Coefficient*	.599*
		Sig. (1-tailed)	.034
		N	10
	support for language and learning	*Correlation Coefficient*	.694*
		Sig. (1-tailed)	.013
		N	10
	overall rating	*Correlation Coefficient*	.720**
		Sig. (1-tailed)	.009
		N	10

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

Table 5 Correlation between mean weighted dynamic posttest score at t1 and measures at t2

			MWDP t1
Spearman's rho	*bpvs difference*	*Correlation Coefficient*	.596*
		Sig. (1-tailed)	.035
		N	10
	progress in language	*Correlation Coefficient*	.188
		Sig. (1-tailed)	.302
		N	10
	learning language	*Correlation Coefficient*	.307
		Sig. (1-tailed)	.194
		N	10
	generalising new language	*Correlation Coefficient*	.375
		Sig. (1-tailed)	.143
		N	10
	support for language and learning	*Correlation Coefficient*	.686*
		Sig. (1-tailed)	.014
		N	10
	overall rating	*Correlation Coefficient*	.541
		Sig. (1-tailed)	.053
		N	10

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

Table 6 Correlation between weighted generalising score at time 1 and measures at time 2

			MWGS t1
Spearman's rho	*bpvs difference*	*Correlation Coefficient*	.285
		Sig. (1-tailed)	.213
		N	10
	progress in language	*Correlation Coefficient*	-.012
		Sig. (1-tailed)	.486
		N	10
	learning language	*Correlation Coefficient*	.432
		Sig. (1-tailed)	.106
		N	10
	generalising new language	*Correlation Coefficient*	.546
		Sig. (1-tailed)	.051
		N	10
	support for language and learning	*Correlation Coefficient*	.546
		Sig. (1-tailed)	.051
		N	10
	overall rating	*Correlation Coefficient*	.612*
		Sig. (1-tailed)	.030
		N	10

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

Table 7 Correlation between proportion of words expressed correctly during the posttest phase at time 1 and measures at time 2

			proportion of correctly named items2 t1
Spearman's rho	*bpvs difference*	*Correlation Coefficient*	.050
		Sig. (1-tailed)	.445
		N	10
	progress in language	*Correlation Coefficient*	-.220
		Sig. (1-tailed)	.271
		N	10
	learning language	*Correlation Coefficient*	.497
		Sig. (1-tailed)	.072
		N	10
	generalising new language	*Correlation Coefficient*	.577*
		Sig. (1-tailed)	.040
		N	10
	support for language and learning	*Correlation Coefficient*	.215
		Sig. (1-tailed)	.275
		N	10
	overall rating	*Correlation Coefficient*	.415
		Sig. (1-tailed)	.116
		N	10

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).