Development of the *Manipulating Sentences Test: Forms A & B* for Students in the Middle Years (Ages 9;0 – 12;11)

ABSTRACT

With increased awareness of the impact of oral language skills on literacy development, there is an increasing need for classroom language assessment tools that can be administered by teachers and school special needs coordinators within an individual, small group or whole class setting. The availability of assessment tools for students in the middle years is particularly limited. Following the consideration of available speech pathology oral language assessment tools, the aim of this study was to develop a sentence level assessment of syntactic awareness for use with students in the middle years.

The Manipulating Sentences Test was created with multiple forms for assessment in written or oral format at an individual level, or in written format at a group level. Pilot testing was then conducted with eleven students aged between 9;4 and 12;3. Each student received three matched tests, an existing standardised assessment and two forms of the researcher-developed assessment. Data was analysed with regard to correlation between test forms, test – retest reliability, assessment format (oral versus written) and context (individual versus small group). Item by item comparisons were also made between students and within students across different tests, as well as analysis of individual student errors. Pilot testing results reported in this study compare well with existing normative data for both the individual oral testing format and the group written testing format. Further testing and development using a broader population sample is indicated.

INTRODUCTION

Links between oral language and literacy development and educational attainment have been long established. Assessments of students at school entry have indicated that oral language ability gives the strongest prediction of educational attainment at 7 years (Wells, 1981). Many studies have considered links between the oral language skill of phonological awareness and reading, but oral language influences reading development much more broadly than just isolated skills such as phonological awareness or even vocabulary knowledge. (NICHD Early Childcare Research Network, 2005).

It has been found that "some aspects of grammatical and lexical structure in a child's oral language are important to his/her learning to read." (Torrence & Olson, 1984, p.176). Torrence and Olson also proposed that a student's ability to use complex syntax in his/her oral language may be a predictor of reading ability. The importance of syntactic awareness in reading has not been overlooked, with the use of syntactic clues during reading being used to teach new vocabulary (Sinatra and Dowd, 1991). One important aspect of syntactic awareness in reading is its use, along with other strategies, to predict words that the student is having difficulty reading. For example, when a student reads *a*, *the,* or *an*, he/she can be sure that the next word will be a noun or adjective and not a verb.

It is clear that given the links between oral language skills and literacy development these should be taken into account at the classroom level. This implies that teachers should be aware of the importance of evaluation of oral language and the teaching of oral language skills, as whilst students enter school with a certain level of oral language knowledge, this needs to continue to be built and grow through the educational process. Whilst formal testing of literacy skills to assist in intervention and placement in reading levels/groups has become the norm in the early years of schooling in recent years, it is not apparent whether these measures are being systematically used in the middle years (Years 5 - 8). Furthermore assessment and intervention in the area of oral language is often seen to be the exclusive domain of speech pathologists rather than teachers. There is a need for the use of oral language assessment at a school/classroom level, particularly for those students identified as at risk for reading difficulties and even more so for students in the middle years who have identified literacy difficulties.

However whilst teachers utilise reading and literacy assessments, they often lack the necessary tools to clearly identify oral language weakness. Such tools do exist and are utilised by speech pathologists, but are not in a form that makes them appropriate or easy to use in the classroom, therefore it was decided to adapt an existing assessment tool for these purposes. The assessment tool chosen evaluates syntactic awareness and sentence production; particularly for students in the middle years.

There are a number of assessments available to speech pathologists to formally assess students in the middle years in the area of sentence production. These include subtests within the Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4; Semel et al., 2003) and the Test of Language Development – Intermediate Third Edition (TOLD-I: 3; Hamill & Newcomer, 1997), the two most commonly used speech pathology assessments for middle years students. The above assessments contain the following subtests which relate to sentence production:

CELF – 4

- Formulated Sentences
- Recalling Sentences
- Sentence Assembly

TOLD-I:3

- Sentence Combining
- Word Ordering

In order to provide classroom teachers with an assessment tool assessing sentence production with a focus on syntactic awareness, it was determined to adapt and correlate a new test with one of the existing formats, but in determining this process it was important to consider what each of the above tests actually measures. Sabers (1996) suggests that "sentence production as a measure of oral expression can tap five components of language: phonology, syntax, morphology, semantics and pragmatics" (p.106). Further, the construction of the test and the administration guidelines can then determine the degree to which these components are measured. Another important aspect to consider alongside these components is demands placed on the student's short-term auditory memory (STAM).

Other considerations are also important, such as investigation into the assessment tools that teachers are likely to choose to use. Such investigations indicate that

teachers tend to utilise tests that are already in use or readily available. Further to this, choice of assessment is made on the basis of test scope (eg. results indicate strengths and weaknesses and can be easily adapted to program planning) and ease of use. (Johnson & Beauchamp, 1987)

Therefore the following criteria were established in choosing the characteristics of the assessment tool to adapt and modify: The test should be 1) primarily an assessment of syntax versus semantics and 2) memory load should not be a significant feature. It should be 3) easy to use and score; 4) with the possibility of adaptation for group administration. Finally, 5) it should be possible to analyse it with regard to strengths and weaknesses and to make links to program planning.

The above tests were analysed according to these criteria and given an overall score which can be seen in table 1 below. Further details regarding this decision making process can be found in Appendix 1. The Sentence Assembly subtest from the CELF-4 scored highest and was therefore chosen as the sentence production assessment tool to adapt and modify. Semel et al. (2003) have defined the objective of the Sentence Assembly subtest as being "to evaluate the student's ability to formulate grammatically acceptable and semantically meaningful sentences by manipulating and transforming given words and word groups." Further they indicate that the abilities evaluated by the test are required in classroom activities such as "sentence combining and sentence analysis" and relate to skills at the sentence and discourse levels of oral language such as "formulating descriptions, questions, responses, or conversation".

Table 1: Test Selection Criteria

Test	Syntax (vs semantics)	Low STAM load	Ease of use & scoring	Adaptability for group administration	Suited to Program Planning	Score
Formulated Sentences	0	+1	-1	-1	+1	0
Recalling Sentences	0	-1	0	-1	-1	-3
Sentence Assembly	+1	0	+1	+1	+1	+4
Sentence Combining	+1	-1	+1	-1	+1	+1
Word Ordering	+1	-1	+1	-1	+1	+1

Strong relationships exist between the Sentence Assembly subtest and transformational grammar literature by Chomsky (1957). On the Sentence Assembly task, students are required to transform groups of stimulus words into sentences in a similar way to the way that we theoretically transform deep structures into surface structures. The stimulus words/word groups form the constituents or the base of the sentence which can then be placed into known sentence frames. A student needs to have knowledge of a range of sentence frames/types in order to successfully complete this task After completing the first sentence, the student does not need to construct the second sentence from scratch, but rather utilise knowledge of the sentence frame to quickly transform the sentence by 1) switching the subject and verb to create a question from a statement (or vice-versa as the case may be), 2) switching the subject and object, or 3) by moving whole clausal structures. Examples of these types of transformations are given below and an example of deep and surface structures can be seen in diagrams 1 and 2 further below.

Examples of Transformations:

The boy was followed by the dog.

→ Was the boy followed by the dog? ***[1. Statement to question transformation]***

→ The dog was followed by the boy. ***[2. Switching subject and object transformation]***

She bought the car after she got the job.

→ After she got the job, she bought the car. ***[3. Clause level transformation]***

Diagram 1: **Deep Structure for:** ***The boy who won the contest was clever***

[CELF-4 Sentence Structure – Item 19]

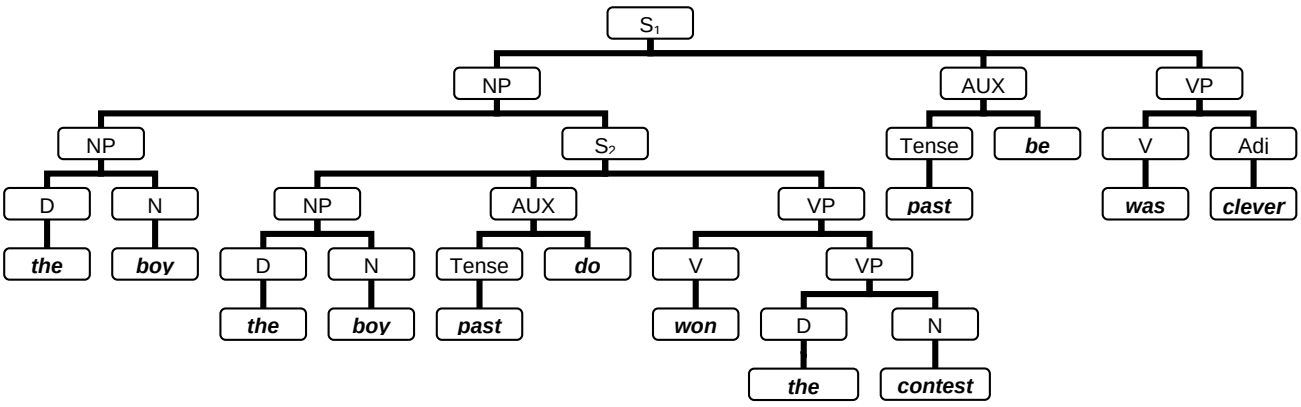
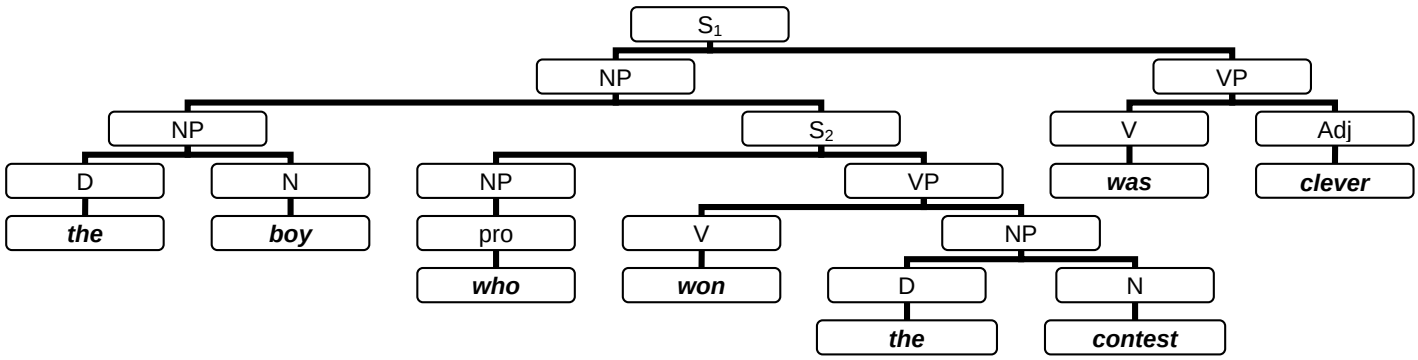


Diagram 2: Surface Structure for: *The boy who won the contest was clever*
[CELF-4 Sentence Structure – Item 19]



It can therefore be seen that the Sentence Assembly test has a strong emphasis on syntactic knowledge and whilst semantic skills are not entirely eliminated they are minimised through the choice of simple vocabulary and concepts. Memory load is minimised in the test by providing written stimulus materials and allowing the examiner to reread them if required (although it should be noted that only one repetition is allowed). Working Memory skills are however still an important factor in being able to chunk and group the words into the sentence frames. A strong advantage of Sentence Assembly over some other tests of sentence production is the ease of scoring since all possible correct responses are known and have been noted on the scoring form.

The remainder of this paper deals with the adaptation of the Sentence Assembly test into new formats for individual (oral or written) and group (written only) administration and the pilot testing of these formats.

METHOD

Preliminary Work:

Initial work was completed designing test items for Manipulating Sentences Test Forms A, B & C. The test items from the CELF-4 Sentence Assembly subtest were used to facilitate this work. Care was taken to ensure that for each new test item the following criteria were met:

1) The sentence structure of the original test item was maintained ie. content and structural words were changed to those of the same type.

For example:

The girl didn't put the keys in her pocket

→ The boy didn't leave the toy under his bed

At a constituent level:

girl	→ boy	animate nouns
put	→ leave	transitive verbs
keys	→ toy	inanimate nouns
in	→ under	prepositions
her	→ his	possessive pronouns
pocket	→ bed	inanimate nouns

The only constituents that remain exactly the same in both sentences are the determiners (the) and the negative auxiliary (didn't). Both sentences maintain the following surface structure: Noun Phrase (Determiner + Noun) Auxiliary + Verb Phrase (Verb + Noun Phrase [Determiner + Noun]) + Prepositional Phrase (Preposition + Noun Phrase [Possessive Pronoun +Noun]

2) The number of possible transformations remained the same. This involved also taking into account the logic or semantic content of the sentences. For example:

He caught the bus after he left the house → *2 possible sentences*

He drew the picture after he wrote the story → *4 possible sentences*

Despite both sentences having the same structure at a meaning level, the two events in the first example occur in a certain order, restricting the number of logical transformations, whereas in the second example the events can occur in any order. Therefore these two sentences are not appropriate item substitutes for each other.

3) The vocabulary was kept at a similar level as the original test items in order to increase readability and processing speed.

4) The constituents of each item were organised into the same order as in the original test, with stimulus words being approximately the same print size and laid out in the same way.

Forms A, B & C were all originally created as individually administered tests, however testing reported in this study has been completed only using Forms A and B. Multiple forms were created to avoid over exposure to items in test-retest and to allow for adaptation to whole class versus individual versions. It was also felt that creating extra sample sentences would assist if replacing test items if difficulties with consistency were found on specific items during initial testing.

The initial design of the group administration task involved the use of stickers however this was abandoned after testing with one student using Form C. Following this the current group administration task was devised utilising Form A. Further details of this are explained in the procedure section below.

Participants:

There were eleven participants, who ranged in age from 9;4 – 12;3 with at least one student represented within each age year between 9 and 12. Ten of the students attended primary school ranging from grades four through six and one student attended secondary school and was in year seven. The participants came from four different Catholic schools located in the north-eastern suburbs of Melbourne and one Catholic school in the outer-eastern suburbs of Melbourne. Eight of the students had no previously diagnosed language difficulties, one student had previously received funding for a severe language disorder, one had a mild oral language difficulties and delays in STAM, and one was last assessed in mid 2004 and found to have moderate oral language difficulties. At the time of the assessments five of the students were receiving speech pathology support for mild difficulties in the area of articulation, but did not have any other identified oral language difficulties. There were five females and six males represented in the study, all of whom spoke English as their main or only language. Characteristics of the participants are summarised in table 2 below.

Table 2 – Participant gender, age, year level and background.

Initials	Gender	Age	Year Level	Identified Oral Language Difficulties	Receiving Articulation Therapy	English Language Background
LL	F	9;4	4			Main
MR	F	10;5	4	✓	✓	Only
MK	M	11;2	5		✓	Main
AM	M	11;4	5		✓	Only
CB	M	11;4	5		✓	Only
AH	F	11;6	6			Only
DM	M	11;6	6			Only
SC	F	11;9	6			Only
ZM	F	11;9	6			Only
KW	M	12;3	6	✓		Only
NK	M	12;3	7	✓	✓	Only

The participants were chosen to provide a spread of ages between 9;0 and 12;11. More students were chosen between 11;0 and 12;11 as it was expected that these students would have a greater understanding of the task and be able to complete a greater number of items on the assessments (ie. to allow for comparisons to be made between all test items across the three different assessment forms). Participants without a history of identified oral language difficulties were also chosen for these reasons; however it also important to include some students with language difficulties to ensure that their response patterns between the different test forms were similar to those students without identified language difficulties. Only participants from backgrounds where English is the only or main language were chosen as it was felt that this would ensure greater stability with regard to knowledge of sentence structure and vocabulary. Whilst a number of the students selected had mild articulation difficulties, this was not expected to impact on their performance.

Materials:

The following materials were used to assess students. Each student involved in the study received three assessment tasks:

1) Manipulating Sentences: Form A (MSA) - Test written by the researcher to match the Sentence Assembly subtest from the CELF-4. *[refer to Appendix 2a for stimulus materials and Appendix 2b for scoring form]*

2) Manipulating Sentences: Form B (MSB) - Test written by the researcher to match the Sentence Assembly subtest from the CELF-4. *[refer to Appendix 3a for stimulus materials and Appendix 3b for scoring form]*

3) Sentence Assembly (SA) **–** One of 15 subtests from the CELF-4 for students aged 9;0 **–** 12;11 years. *[refer to Appendix 4a for stimulus materials and Appendix 4b for scoring form]*

Procedure:

Students were given the same instructions for each test, which consisted of one demonstration item, two practise items and nineteen test items, with the test items discontinued after consecutive zero scores on five items. Students were required to formulate two logical/meaningful sentences from groups of words. The words for each item were always read out to the student and individual items were discontinued if the student paused for more than 10 seconds in providing a response. After trial items were given, all students commenced at item one each test and continued until the discontinue rule was met. The tests were administered in accordance with directions for the *CELF-4 Sentence Assembly* subtest, full details of which can be found in Appendix 5a.

Nine students received the written format of the MSA test, with six students receiving this task as a group assessment. These students received the same instructions as per the other tests, but were provided with a pen and additional instructions related to the recording of their responses. Students undergoing group testing were allowed forty seconds to complete each item and were asked to commence their second sentence after twenty seconds. Refer to Appendix 5b for details of additional instructions given for the group assessment task.

To allow for any influence on results related to task familiarity half of the students in each group received the Sentence Assembly subtest first and then the Manipulating Sentences subtests and the other half the students received one of the researcher devised Manipulating Sentences tests first. The Sentence Assembly subtest was always either the first or second test administered. Refer to table 3 below.

Table 3 – Order of test administration
[and context (ie. Group versus 1:1) and format (ie. oral versus written)]

Student	SA	MSA			MSB
LL	2	1	O		3
MR	1	2	W		3
MK	1	3	W	G1	2
AM	2	3	W	G1	1
CB	2	1	O		3
AH	1	2	W	G2	3
DM	1	2	W	G2	3
SC	2	1	W	G2	3
ZM	2	1	W	G2	3
KW	2	3	W		1
NK	1	2	W		3
1st Test Administered	45.4%	36.4%			18.2%
		54.6%			

O=oral responses; W=written responses; G1=group 1; G2=group 2

Ten of the participants received all three assessments on the same day. Of these, six received all three assessments in one sitting, with the remaining four students receiving all three assessments within a ninety minute period, but not in one sitting. One student received two of the tests in the same sitting on one day and the third test seven days later. Each individual test took an average of eight minutes to administer, with the group task taking approximately twenty minutes to complete.

RESULTS

Correlation between test forms

The results for each student on each of the three assessments administered are summarised in table 4 below, along with analysis of the correlation between test forms. Scaled score data from the CELF-4 was used to calculate scaled scores and confidence intervals across all three tests. A confidence level of 68% was used. Scores between tests were considered to correlate where the confidence ranges between results on different tests overlapped. This process for considering *Standard Errors of Measurement* (SEM) relating to chance or measurement error has been referred to by Hutchinson (1986) and others. For example CB's confidence range for the SA test was scaled scores of 7-9 and for the MSA test was scaled scores of 5-7. Therefore these scores have been considered to correlate. As can be seen from an examination of the data in table 4, correlation occurred on 28 out of 33 tests (85%) and for eight out of the eleven participants across all three tests. Correlation Coefficients were calculated using *Pearson's r* and indicate high correlation (r > .70) between all three assessments.

Table 4 – Results and Correlation between test forms

	Results (Standard Score ±Confidence Interval [68%])			**Correlation between Test Forms** (Based on overlap in confidence range when comparing test results)		
Student	**SA**	**MSA**	**MSB**	**SA-MSA**	**SA – MSB**	**MSA-MSB**
LL	9 (±1)	9 (±1)	9 (±1)	✓	✓	✓
MR	5 (±1)	6 (±1)	5 (±1)	✓	✓	✓
MK	9 (±1)	10 (±1)	9 (±1)	✓	✓	✓
AM	5 (±1)	6 (±1)	6 (±1)	✓	✓	✓
CB	8 (±1)	6 (±1)	10 (±1)	✓	✓	N
AH	10 (±1)	10 (±1)	10 (±1)	✓	✓	✓
DM	6 (±1)	5 (±1)	6 (±1)	✓	✓	✓
SC	4 (±1)	7 (±1)	8 (±1)	N	N	✓
ZM	12 (±1)	10 (±1)	12 (±1)	✓	✓	✓
KW	5 (±1)	4 (±1)	4 (±1)	✓	✓	✓
NK	3 (±1)	6 (±1)	6 (±1)	N	N	✓
				r = .76	r = .75	r = .80

Test – Retest comparisons

Test-Retest comparisons have also been made, particularly as students received repeated tests of the same type on the one day. This information is recorded in table

5 and graphically in chart 1. As outlined earlier, the order of administration of the tests was varied across students, as it was expected that as a student gained greater familiarity with the task, his/her performance would improve. Therefore this data has been organised in order starting with the first test administered. As expected, the results indicate overall improvement in scores with each test administered despite the variations in the order of administration. The average change in raw scores across the three tests (1.32) compares favourably with test-retest results from the standardisation of the CELF-4 Sentence Assembly test which showed similar gains (1.37) in the mean across the age groups represented in this study. Similarly the Standard Difference (.42) and Correlation Coefficients (r=.78, .82, .79) also compare well with the standardisation results of the CELF-4 Sentence Assembly test [Standard Difference (.47); Correlation Coefficient (r=.82)].

Table 5 – Test - Retest Raw Score Comparisons

Student	Test 1	Test 2	Test 3	Test Average
LL	6	7	7	6.67
MR	3	4	3	3.33
MK	10	11	12	11.00
AM	4	3	4	3.67
CB	4	8	11	7.67
AH	12	12	13	12.33
DM	4	3	4	3.67
SC	5	2	7	4.67
ZM	13	15	15	14.33
KW	3	5	3	3.67
NK	2	6	6	4.67
Mean (±SD)	6.00 (±4.03)	6.91 (±4.26)	7.73 (±4.08)	6.88

Average increase in test mean	1.32 (CELF-4 SA: 1.37 [avg])
Standard Difference (Test 1-Test 3)	0.42 (CELF-4 SA: 0.47 [avg])
Correlation Coefficients	r = .78, .82, .79 (CELF-4 SA r=.82 [avg])

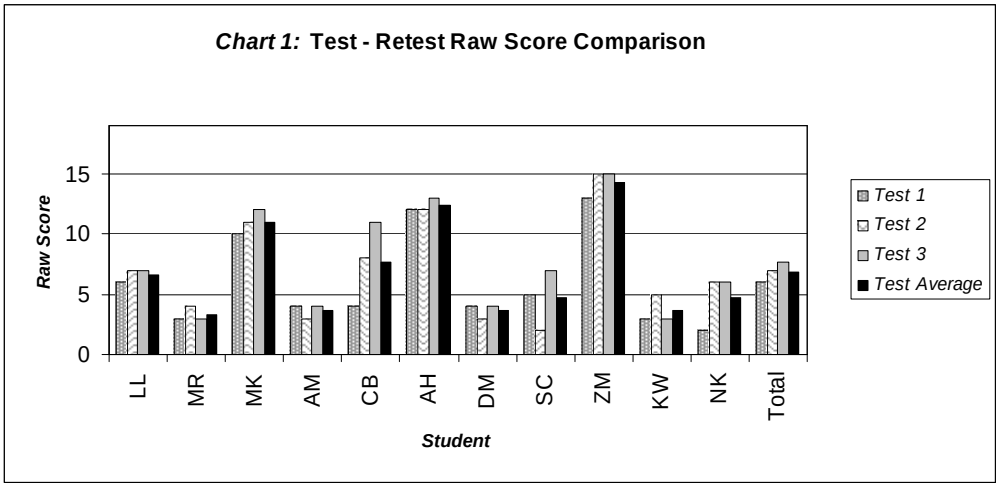


Chart 1: Test - Retest Raw Score Comparison
Raw Score
15
10
5
0
LL
MR
MK
AM
CB
AH
DM
SC
ZM
KW
NK
Total
Student
Test 1
Test 2
Test 3
Test Average

Item Analysis

Comparisons were made between individual test items across tests. Comparisons were made for consistency of response both between students (see table 6 and chart 2 below) and within students (see table 7 and chart 3 below). Full details of the item by item analysis appear in Appendix 7. Information from tables 6and 7 below can be used to re-evaluate the construction/difficulty of individual test items. Correlation between items were considered to be relatively high considering the size of the sample, however further investigation of the construction/difficulty of MSA Items 3, 4,15, 16, 17,18,19 and MSB Items 4,10,16,17,18,19 was indicated.

Table 6: Item by Item Comparisons (between students)

Item	SA-MSA	SA-MSB	MSA-MSB
1	100%	100%	100%
2	100%	100%	100%
3	36%	64%	73%
4	45%	64%	55%
5	55%	82%	82%
6	73%	82%	73%
7	73%	73%	82%
8	78%	78%	73%
9	71%	86%	89%
10	80%	83%	43%
11	50%	67%	83%
12	75%	60%	100%
13	100%	100%	100%
14	100%	80%	80%
15	100%	100%	100%
16	75%	80%	50%
17	0%	75%	33%
18	33%	50%	33%
19	33%	75%	67%
Average	67%	79%	74%

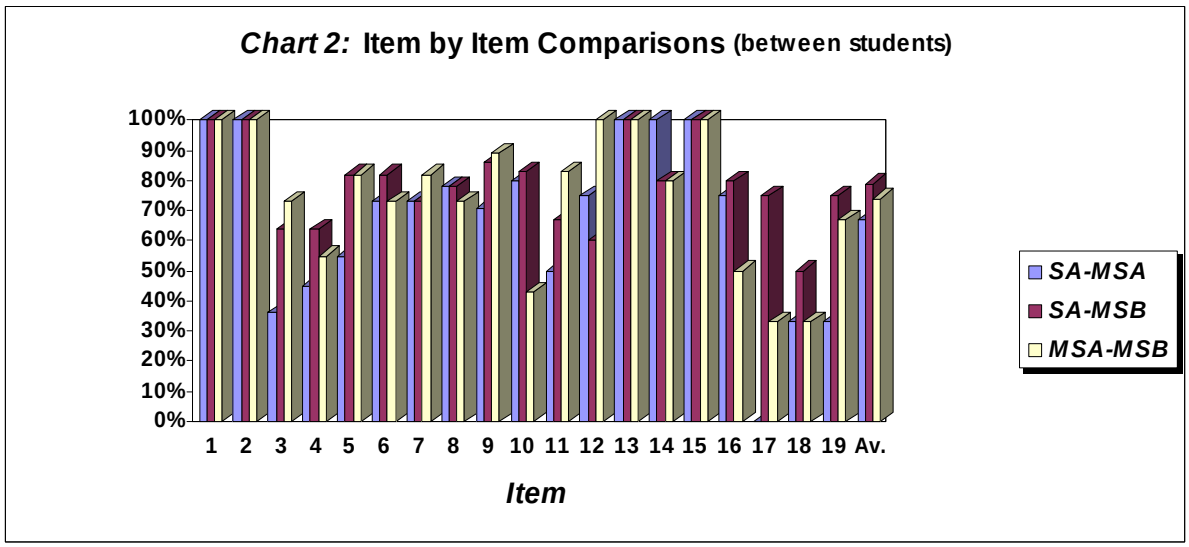
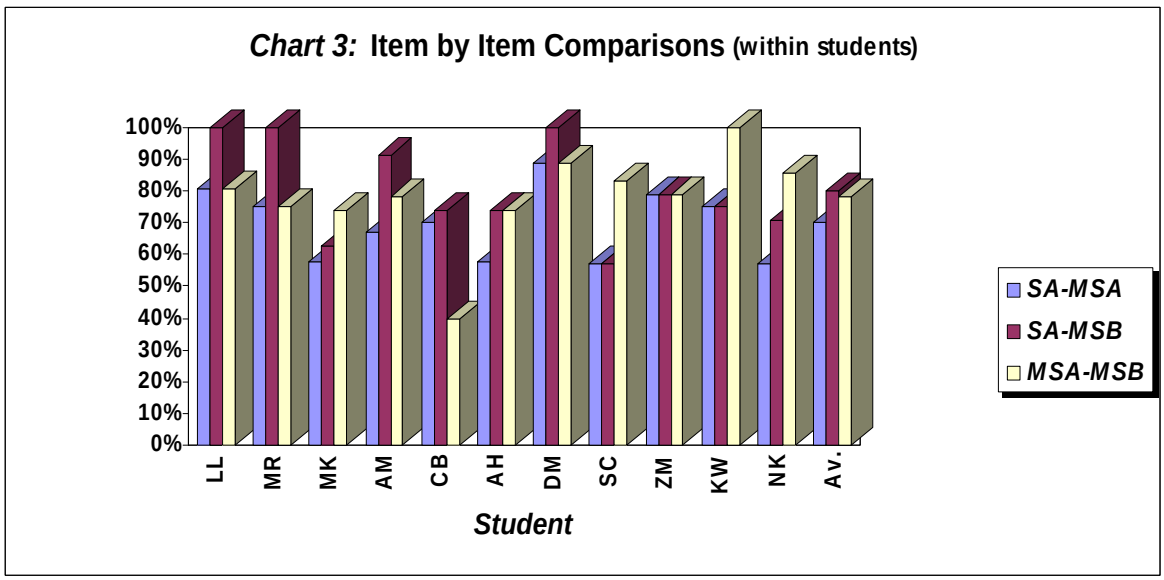


Table 7: Item by Item Comparisons (within students)

Item	SA-MSA	SA-MSB	MSA-MSB
LL	81%	100%	81%
MR	75%	100%	75%
MK	58%	63%	74%
AM	67%	91%	78%
CB	70%	74%	40%
AH	58%	74%	74%
DM	89%	100%	89%
SC	57%	57%	83%
ZM	79%	79%	79%
KW	75%	75%	100%
NK	57%	71%	86%
Average	70%	80%	78%



MSA result analysis:

Results on the MSA test have been compared with regard to presentation in a small group versus individual format and with regard to whether students were required to respond in a written versus oral format. These are between student comparisons and appear in table 9 below. Comparisons indicate negligible differences between group versus individual administration formats and small differences between oral versus written administration, with differences being well within one SEM at a 68% confidence level. However it should be noted that the sample size is small.

Table 9: Comparisons of 1:1 vs. group administration and oral vs. written responses on the MSA test

Context	1:1		Group
Format	Oral	Written	
LL	0		
MR		1	
MK			1
AM			1
CB	2		
AH			0
DM			1
SC			3
ZM			2
KW		1	
NK		3	
Average of differences between Scaled Scores on MSA and SA tests across subjects	1.0	1.7	1.3
1:1 versus *Group*	1.4		1.3
Oral versus *Written*	1.0	1.4	

NB. The numbers used in this table are the difference between each participant's scaled scores on the MSA and SA tests.

Student Comments and Test trends for individual students:

The nine students who were administered both written and oral formats of testing were asked for comments about differences between the test formats. A number of students (66%) commented that they felt the written task was harder because of issues relating to "getting the numbers in the right places". One student (NK) commented that it was harder completing the written task because "you don't get to say it out loud". Whilst these results would indicate that students felt more challenged by the written task, this was not reflected in the scores obtained, with six out of nine students actually scoring better on the written form of MSA in comparison

with the SA subtest. Also, the three students who scored lower on the written task were noted to have confidence ranges that overlapped on the SA and MSA tests.

Students were noted to make other perceptive comments during testing. MR commented that she was often able to construct one sentence, but not another. ZM was noted to identify items containing indirect objects and expressed frustration, saying “I can’t do this sort”. CB was highly interested in the nature of the tasks and made frequent comments, including that items containing before/after (subordinate clauses) were easier. In light of CB’s attempts to evaluate the task, it should also be noted that he made the greatest ongoing gains across each test administration.

Each student’s responses were evaluated for error patterns and error types using the *Error Analysis Form* which can be found in Appendix 8. This form records responses according to the following parameters:

1) Delayed response – this related to students who paused for more than 10 seconds and the item was discontinued. This information was coded with regard to whether this occurred following zero or one correct response.

2) Error – the student created a structurally (or semantically) incorrect sentence and did not acknowledge this or attempt self-correction. This was scored in terms of whether or not this occurred following the formulation of one correct sentence.

3) Error Type – a score was given regarding the number of correct responses for a particular sentence type, based on their occurrence within items in the test. The student was then given a grade of Pass, Emerging or Difficult for each sentence type. Sentence types are coded at two levels, the overall sentence category: *Declarative (D), Interrogative (I)* and *Imperative (C)* and substructures within the sentence: *Passive (P), Negative (N), Prepositional Phrase (PP), Infinitive (F), Indirect Object (IO), Subordinate Clause (SC)* and *Relative Clause (RC)*

Individual student characteristics following analysis on the Error Analysis Form are summarised in table 10 below:

Table 10: Comments regarding analysis of individual student performance

	Delayed Responses	***Errors***	***Error Types***
LL	50% of items administered were discontinued due to delayed responses. Movement from *No Response* to *One Correct* response across tests, possibly indicating increasing processing speed with task familiarity.	No errors across testing and demonstrating awareness of incorrect attempts.	Difficulty with Interrogatives particularly for sentences containing Negative and Infinitive structures.
MR	Significant increase in delayed responses on written task (MSA).	Similar across different test formats. Approx. 37% of test items administered.	Difficulty with all sentence categories (Declaratives, Interrogatives and Imperatives). Difficulty with Negatives, Prepositional Phrases, Infinitives and Indirect objects across all three tests. Increased difficulty with Subordinate clauses noted on MSA task.
MK	Similar delayed responses on SA and MSB tests, but decreasing when completing written responses.	Increased errors on written response testing associated with drop in delayed responses.	Difficulty with Indirect Objects and Relative Clauses across testing. Infinitives mastered after first test administered.
AM	Most responses delayed on SA and MSB tests. This number reduced on the MSA test.	Errors on approximately 20% of items administered.	Difficulty with Interrogatives and Negatives, Prepositional Phrases, Infinitives and Indirect Objects.
CB	60% of items administered were delayed with one item correct, on the first test administered, decreasing to 15% delayed across remaining tests.	No errors on first test administered (delayed), with reducing number of errors across remaining tests.	Improvements across most sentence types/structures across testing. Ongoing inconsistency with Prepositional Phrases and difficulty with Relative Clauses.
AH	Improvement with regard to discontinuation due to delayed responses across subsequent test administration: 16% → 5% → 0%.	Similar numbers of errors across testing.	Inconsistent skills with Indirect objects and Relative Clauses across assessments possibly indicating that the student is in the process of learning this language structure. Improvements in processing of Negatives and Prepositional phrases across assessments.
DM	Approximately 50% of items administered on SA and MSB tests were discontinued due to delayed responses (but few errors).	Approximately 50% of items in error on MSA test, but less delayed responses, possibly due to administration as a written assessment.	Difficulties were noted with Interrogatives, Negatives, Prepositional Phrases, Infinitives and Indirect Objects.
SC	Approximately 40% of items were discontinued due to delayed responses.	Similar numbers of errors were noted across assessments, being approximately 30% of items administered.	Difficulties were noted with Negatives, Infinitives, Indirect Objects, with inconsistent skills noted with Prepositional Phrases. Difficulty with Prepositional Phrases on SA test led to early discontinuation of the assessment and possibly to lack of correlation of results for this student.

ZM	Approximately 16% of items were discontinued due to delayed responses on the SA and MSB subtests (with few errors noted).	Approximately 30% of responses were in error on the MSA written task, but not on SA and MSB tests.	Consistent difficulty with Indirect Objects and emergent skills with Relative Clauses.
KW	Approximately 25% of items were discontinued due to delayed responses SA and MSB tests, but 50% of responses delayed on MSA written task.	Errors decreasing across testing: 37% → 20% → 14%.	Difficulties noted on Negatives, Prepositional Phrases, Infinitives, and indirect Objects. Inconsistent skills with Declaratives and Interrogatives.
NK	Over 50% of items administered were discontinued due to delayed responses.	Less than 6% of items in error across testing.	Difficulties with Interrogatives and emerging skills with Declaratives. Difficulties with Negatives, Prepositional Phrases, and Indirect Objects. Inconsistent skills with Infinitives.

Overall analysis of the types of student errors indicated consistency or even improvements across repeated testing with regard to the range of structural error types. Further, a general trend was noted that students receiving written format testing were more likely to make errors rather than have delayed responses, but the overall number of errors remained the same or decreased in line with test-retest considerations.

As expected, there was negligible difference between students receiving articulation therapy and the other participants. Also as expected, the students with oral language difficulties achieved the lowest scores on the assessments and with the exception of the student with mild language difficulties, made some of the least gains across retesting. It was also noted in the results, that MR and KW the students with moderate and severe oral language difficulties were the only students to increase in the number of delayed responses and the written format of MSA, the opposite of the general trend for other students.

DISCUSSION

The results obtained in this research indicated high correlation or criterion-related validity between the researcher-developed MSA and MSB tests and the existing SA test in both oral and written formats. As Hutchinson (1996) has stated, as "the results agree with the results of tests that users already accept as valid measures of the construct, we can have increased confidence in the test." (p.111). Once further evaluation has been done, the MSA and MSB tests should be useful for group or whole classroom assessment of students' metalinguistic syntactic awareness. This

will lead to better identification of student's who need support and teaching in this area and provide information regarding strengths and weaknesses.

Development of these tests has involved keeping in mind the importance of the scope of the tests and their ease of use (Johnson and Beauchamp, 1987). Ease of test use has been achieved through a combination of a simple scoring system and creation of a group written task that will allow for a decrease in overall assessment time required across students. Important in considering the scope of the test is the reported belief that teachers see assessment and teaching as linked and therefore assessment tools should provide links to teaching resources (Johnson & Beauchamp, 1987). Teaching resources are available for this area in the form of the Manipulating Sentences Program (CEOM Speech Pathology, 2000) and further links to teaching are given below. The available resources provide both classroom and home follow-up activities.

It was noted that some performances were below average for students (AM and DM) who have not been previously identified as having language difficulties. This highlights the importance of the development of tools such as this. If such tests are available, classroom teachers and Special Needs Coordinators will be more able to identify students who need support. These results may also indicate a need for greater focus on the explicit teaching of sentence structure and grammar in the classroom. If teachers gain access to tools such as the one described in this study, they will also need to become more aware of how to analyse testing results and consequently how to target identified weaknesses. Dowd and Sinatra (1991) have provided suggestions regarding syntactic cues and a framework for teaching grammar, which could in turn lead to specific teaching of grammatical constructions and further development of a student's comprehension, vocabulary and metalinguistic awareness. Supported use of Sentence Combining in an oral context could also be used to expand student knowledge of sentence structures.

The results for three students (CB, SC and NK) did not correlate across all three of the tests. In the case of CB, this is likely to be due to ongoing learning effects that occurred across assessments. As referred to earlier, CB demonstrated a great deal of interest in the assessment process; in that he actively commented on test items and appeared to be using problem solving processes. These skills may have in turn

led to greater learning effects for him than the other students. It should be noted that it is his first and last tests administered that do not correlate, which would add some support to this theory. Whilst a similar effect may have occurred for NK, it does not explain the marked drop in performance on the second test for SC. SC's assessment on the SA subtest occurred immediately after the group written assessment at her school. SC was observed to be generally slower in forming her responses than other students on the written task and she may have felt singled out in staying back when the other students were sent back to class. She did appear to grow in confidence on the final task and it may be that factors regarding confidence and self-esteem impacted on her performance. Other factors for SC may relate to error types as described in Table 10.

Whilst it was expected that the student's with identified language difficulties (KW, MR and NK) would achieve lower scores on the assessment, an unanticipated finding was the increased difficulty that these students had with providing any response within the allotted time on the written versus oral assessment tasks. This may be due to changes in the measurement method having a greater impact on these students. Although literacy assessment was not part of this study, due to their oral language difficulties these students are also more likely to have delays in their literacy development. It is possible that despite having the words read out to them, they were less efficient at processing the text to record their responses. NK's comment that he found the written task more difficult because he was unable to say the words, may also be relevant, as students with language difficulties may be more likely to require auditory feedback or have to "think out loud" in order to process sentences syntactically.

In relation to the above, it was anticipated that students having to provide responses in a written format might result in reduced scores, however this was not found. There has been some debate amongst speech pathologists about whether assessments, which include written stimulus materials, like the Sentence Assembly subtest are purely oral language. The written response format of the MSA test has taken this a step further, in that whilst syntactic processing of the test items may commence when they are first heard orally, there is no longer opportunity for students to provide oral responses. The impact of this may be most significant for those students who tend to "think out loud". The above issues relate strongly to discussion by Sabers (1996)

regarding the blurred distinctions that can occur between trait and measurement aspects. Whilst MSA written responses are not considered to be a trait part of the construct being measured, they are part of the measurement method that was ultimately chosen and therefore the impact of these changes, particularly for certain cohorts of students should be evaluated further in any future studies of this test format.

One limitation of this study is the small sample size, which would indicate that any findings should be interpreted with caution. This is particularly apparent when considering item-by-item comparisons. Different participants meeting the discontinue rule at different points in the assessment, with approximately half of the eleven students having discontinued on one or more tests by item 9. This meant that for later items small changes in consistency, which may have even been related to measurement error, led to large effects on the overall results. Therefore poor comparisons between items 16 through 19 should be interpreted with caution as results from only 3 or 4 students could be used in this analysis and inconsistencies that arose may have even related to these student having only emerging skills with the levels of the structural forms being assessed.

Some analysis of items with regard to future changes has already occurred. For example, in MSA Item 3 the words *did the ironing* might be changed to *made the beds*, as there may have been insufficient semantic separation between *ironing* and *washing* leading to several students to provide *She did the dishes after she washed the ironing* as their response. The challenge of MSB Item 10 may relate to the change of the auxiliary from *shouldn't* (MSA and SA) → *can't* (MSB) and therefore should be replaced with the former. It was not noted until well into the testing phase that MSB Item 5 and MSA Item 11 are almost identical (*The teacher did read the <u>boy</u> the book* versus *The teacher did read the <u>students</u> the book*) and one of them will need to be changed. Despite this, MSB Item 5 scored well when responses across the three tests were compared, whereas MSA Item 11 scored poorly. Further to this, of the seven students who were administered both items five students were consistent in their responses and the other two were incorrect on the first exposure to the item, but successfully completed the item on their second attempt. This may be due to any of the following:

1) a learning effect for those students;

2) the fact that measurement errors do occurs; or
3) the difference between *boy* and *student* was critical for some participants.

It is expected that the first two reasons given above are the most likely.

Future Directions/Research:

From this study, a number of future directions can be highlighted. Firstly, further development of the test should consider:

1) re-evaluation and changes to some individual test items where students results were not consistent across matched test items.
2) adjustments to the timing and presentation of items for the group written task.

Following this, research involving larger numbers of participants could be conducted. Such research should also involve classroom teachers trialling the assessment and providing feedback.

Possible development of the written format of the test into a self-administered computer program could also be explored. The test could then be administered without taking up valuable teacher time and the computer program would be able to analyse the results and then provide interactive teaching/practise based on skill deficits.

Secondly, there is a need for further development of therapy/teaching resources. The *Manipulating Sentences Program* is currently available, but should be revised to:

1) facilitate practise with a wider range of transformation types and to target specific sentence structures as the current program teaches the skills more generically.
2) provide more resources for classroom based activities versus individual parent directed support.

Further research should then be conducted to evaluate the development of syntactic awareness through the use of these therapy resources and to measure for positive impacts on literacy skills.

In conclusion, this study has provided steps towards a new assessment tool for teachers that can be administered in a group written format to assess syntactic awareness in the middle years. Pilot data and analysis has indicated high correlation or criterion related validity with an existing standardised assessment in this area. These finding also suggest future directions for both assessment and intervention.

REFERENCES

Chomsky, N. (1957) *Syntactic Structures.* The Hague: Mouton.

Hamill, D. D., Newcomer, P.L. (1997). Test of Language Development – Intermediate (3rd Edition). Austin. TX: Pro-Ed

Hutchinson, T. A. (1996). What to look for in the technical manual: Twenty questions for users. *Language, Speech, and Hearing Services in Schools*, 27,2 ,109-121

Laurence, J. J. & Beuchamp, K. D. F. (1987). Preschool assessment measures: What are teachers using? *Journal of the Division of Early Childhood*, 12, 1, 70-76.

NICHD Early Child Care Research Network (2005). Pathways to Reading: The role of oral language in the transition to reading. *Developmental Psychology*, 41, 2, 428-442

Sabers, D. L. (1996). By their tests we will know them. *Language, Speech, and Hearing Services in Schools*, 27, 2, 102-108

Semel, E., Wigg, E. H. & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals (4th Edition)*. San Antonio TX: The Psychological Corporation

Sinatra, R. & Dowd, C. A. (1991). Using syntactic and semantic cues to learn vocabulary. *Journal of Reading*, 35, 3, 224-229

CEOM Speech Pathology (2000). Manipulating Sentences Program. (Program available through the Catholic Education Office Melbourne - Speech Pathology Service)

Torrence, N. & Olson, D. R. (1984). Oral language competence and the acquisition of literacy. In A.D. Pellegrini & T. Yawkey (Eds).). *The development of oral and written language in social contexts* (167-181). Norwood, NJ.: Ablex

Wells, G. (1981). Some antecedents of early educational attainment. *British Journal of Sociology of Education*, 2, 2, 181-199

APPENDICES

Contents:

Appendix 1	Comparisons of sentence production tasks from CELF-4 & TOLD-I: 3
Appendix 2a	Manipulating Sentences – Form A (Group Task) – Stimulus Pages (Sample only)
Appendix 2b	Manipulating Sentences – Form A (Group Task) – Scoring Form
Appendix 3a	Manipulating Sentences – Form B – Stimulus Pages (Sample only)
Appendix 3b	Manipulating Sentences – Form B – Scoring Form
Appendix 4a	CELF 4 Sentence Assembly – Stimulus Pages (Sample only)
Appendix 4b	CELF 4 Sentence Assembly – Scoring Form
Appendix 5a	Administration Directions
Appendix 5b	Additional Administration Directions for Group Assessment Task
Appendix 6	CELF-4 Sentence Assembly – Scaled Scores (Ages 9;0 – 12;11)
Appendix 7	Item by Item Comparison Tables
Appendix 8	Error Analysis Form

Appendix 1 – Test Selection Criteria

Test	Syntax (vs semantics)	Low STAM load	Ease of use & scoring	Adaptable for group administration	Suited to Program Planning	Score
Formulated Sentences	Balanced between semantic and syntactic considerations	Little or no memory load on this task as the student only needs to remember the current word they are placing into a sentence	Easy to use, but there are challenges with regard to training and difficulties with inter-rater reliability for scoring	This test would involve adding other significant measurement methods such as writing and spelling and administration time if administered as a group assessment	Program planning implications arise directly from this assessment	
	0	+1	-1	-1	+1	0
Recalling Sentences	Both semantic and syntactic considerations	Significant and increasing STAM load as the student needs to remember the whole sentence	Easy to use, with some training required for scoring purposes.	Not possible to be adapted for group administration (as above)	Interpretation skills are required and results do not lead directly to teaching goals (ie. Skills involved in the assessment can not be directly taught through practice)	
	0	-1	0	-1	-1	-3
Sentence Assembly	Predominant focus on syntax	STAM load is minimized through use of written stimulus materials which are read to the student	Easy to administer and score	Possible to adapt for group administration with less significant measurement method changes than for other tests.	Program planning goals arise directly from the assessment	
	+1	0	+1	+1	+1	+4
Sentence Combining	Predominant focus is on syntax	Increasing significant and increasing STAM and working memory load as the student needs to remember and manipulate the words	Easy to administer and score	Possible to adapt for group administration, but would result in significant changes to methods.	Program planning goals arise directly from the assessment	
	+ 1	-1	+1	-1	+1	+1
Word Ordering	Predominant focus is on syntax	Increasing significant and increasing STAM and working memory load as the student needs to remember and manipulate the words	Easy to administer and score	Adaptations for group administration would result in the same assessment as that developed in this study, but reliability and validity of normative data would be likely to be reduced, due to too few items and rapid increase in level of difficulty. Poor correlation with existing normative data also expected due to significant changes in measured methods.	Program planning goals arise directly from the assessment	
	+1	-1	+1	-1	+1	+1

Appendix 2a: Manipulating Sentences – Form A - Group Task – Stimulus Pages

[NB. Cut pages in half and staple into a booklet]

3	1	2
sweet	**the apple**	**is**
3	**2**	**1**

Demo

helped	**the woman**	**the man**

Manipulating Sentences – Scoring Form A

Student: _____ DOB: __/__/__ DOAx: __/__/__

School: _____ Age: _____

1.	__ The boy was chased by the girl. __ The girl was chased by the boy. __ Was the girl chased by the boy? __ Was the boy chased by the girl?	1 0	11.	__ Did the teacher read the students the book? __ The teacher did read the students the book. __ The students did read the teacher the book. __ Did the students read the teacher the book?	1 0
2.	__ He read the book before he watched TV. __ He watched TV before he read the book. __ Before he watched TV he read the book. __ Before he read the book he watched TV.	1 0	12.	__ The man didn't drive the car through the tunnel. __ Didn't the man drive the car through the tunnel?	1 0
3.	__ She washed the dishes after she did the ironing. __ She did the ironing after she washed the dishes. __ After she did the ironing she washed the dishes. __ After she washed the dishes she did the ironing.	1 0	13.	__ He wore it even though it was too big. __ Even though it was too big, he wore it.	1 0
4.	__ You will put the milk in the fridge. __ Will you put the milk in the fridge? __ Put the milk in the fridge, will you?	1 0	14.	__ Didn't you see where you threw it? __ You didn't see where you threw it.	1 0
5.	__ Did the woman bake the man a cake? __ The woman did bake the man a cake. __ Did the man bake the woman a cake? __ The man did bake the woman a cake.	1 0	15.	__ He reached it even though he was short. __ Even though he was short, he reached it.	1 0
6.	__ Could you eat that with a spoon? __ You could eat that with a spoon. __ With a spoon, you could eat that.	1 0	16.	__ Can't you say if you liked it? __ You can't say if you liked it.	1 0
7.	__ The man isn't going to buy the book. __ Isn't the man going to buy the book?	1 0	17.	__ He didn't buy the CD because it was scratched. __ Because it was scratched, he didn't buy the CD. __ Because the CD was scratched, he didn't buy it. __ He didn't buy it, because the CD was scratched.	1 0
8.	__ The man didn't put the hat on his head. __ Didn't the man put the hat on his head?	1 0	18.	__ She turned off the tap after she washed the plate. __ After she washed the plate, she turned off the tap	1 0
9.	__ The woman doesn't want to buy the shoes. __ Doesn't the woman want to buy the shoes?	1 0	19.	__ The girl who borrowed the book wore glasses. __ The girl who wore glasses borrowed the book.	1 0
10.	__ They shouldn't leave the books there. __ Shouldn't they leave the books there?	1 0		**Subtotal**	
				Raw Score	
	Subtotal			**Standard Score**	

big	the house	is

Manipulating Sentences – Scoring Form B

Student: _____ DOB: __/__/__ DOAx: __/__/__

School: _____ Age: _____

1.	__ The deer was seen by the lion. __ The lion was seen by the deer. __ Was the lion seen by the deer? __ Was the deer seen by the lion?	1 0	11.	__ Did the man paint the lady a picture? __ Did the lady paint the man a picture? __ The lady did paint the man a picture. __ The man did paint the lady a picture.	1 0
2.	__ He ate the cake before he drank the milk. __ He drank the milk before he ate the cake. __ Before he drank the milk he ate the cake. __ Before he ate the cake he drank the milk.	1 0	12.	__ The boy didn't kick the ball at the park __ Didn't the boy kick the ball at the park?	1 0
3.	__ He drew the picture after he wrote the story. __ He wrote the story after he drew the picture. __ After he drew the picture he wrote the story. __ After he wrote the story he drew the picture.	1 0	13.	__ They bought it even though it was expensive. __ Even though it was expensive, they bought it.	1 0
4.	__ You will buy a book for your friend. __ Will you buy a book for your friend? __ Buy a book for your friend, will you?	1 0	14.	__ Didn't you think before you said it? __ You didn't think before you said it.	1 0
5.	__ Did the boy read the teacher the book? __ Did the teacher read the boy the book? __ The teacher did read the boy the book. __ The boy did read the teacher the book.	1 0	15.	__ They finished it even though it was hard. __ Even though it was hard they finished it.	1 0
6.	__ Can you build this with bricks? __ You can build this with bricks. __ With bricks, you can build this.	1 0	16.	__ Can't she wait until she finishes it? __ She can't wait until she finishes it.	1 0
7.	__ The girl isn't trying to finish her homework. __ Isn't the girl trying to finish her homework?	1 0	17.	__ She didn't read the book because it was too long. __ Because it was too long, she didn't read the book. __ Because the book was too long, she didn't read it. __ She didn't read it, because the book was too long.	1 0
8.	__ The boy didn't leave the toy under his bed. __ Didn't the boy leave the toy under his bed?	1 0	18.	__ He put on his uniform before he went to school. __ Before he went to school, he put on his uniform.	1 0
9.	__ The student isn't going to play the drums. __ Isn't the student going to play the drums?	1 0	19.	__ The man who sold the car was friendly. __ The man who was friendly sold the car.	1 0
10.	__ He can't kick the ball here. __ Can't he kick the ball here?	1 0		**Subtotal**	
	Subtotal			**Raw Score**	
				Standard Score	

tall	**the boy**	**is**

Sentence Assembly – Scoring Form

Student: _____ DOB: __/__/__ DOAx: __/__/__

School: _____ Age: _____

1.	__ The man was followed by the dog. __ The dog was followed by the man. __ Was the man followed by the dog? __ Was the dog followed by the man?	1 0	11.	__ Did the student send the manager an application? __ Did the manager send the student an application? __ The student did send the manager an application. __ The manager did send the student an application.	1 0
2.	__ He finished his homework before he played hockey. __ He played hockey before he finished his homework. __ Before he played hockey he finished his homework. __ Before he finished his homework he played hockey.	1 0	12.	__ The woman didn't put the lamp on the table. __ Didn't the woman put the lamp on the table?	1 0
3.	__ She bought the car after she got the job. __ She got the job after she bought the car. __ After she got the job, she bought the car. __ After she bought the car, she got the job.	1 0	13.	__ She ate it even though it was hot. __ Even though it was hot, she ate it.	1 0
4.	__ You will put the ball in the basket. __ Will you put the ball in the basket? __ Put the ball in the basket, will you?	1 0	14.	__ Don't you know where you put it? __ You don't know where you put it.	1 0
5.	__ Did the boy buy the girl an ice-cream cone? __ The boy did buy the girl an ice-cream - cone. __ Did the girl buy the boy an ice-cream cone? __ The girl did buy the boy an ice-cream cone.	1 0	15.	__ She kept it even though it was broken. __ Even though it was broken, she kept it.	1 0
6.	__ Could you fix that with glue? __ You could fix that with glue. __ With glue, you could fix that.	1 0	16.	__ Can't he remember if he mailed it? __ He can't remember if he mailed it.	1 0
7.	__ The restaurant isn't going to deliver the pizza. __ Isn't the restaurant going to deliver the pizza?	1 0	17.	__ He didn't finish his homework because it was difficult. __ Because it was difficult, he didn't finish his homework. __ Because the homework was difficult, he didn't finish it. __ It was difficult because he didn't finish his homework.	1 0
8.	__ The girl didn't put the keys in her pocket. __ Didn't the girl put the keys in her pocket?	1 0	18.	__ He caught the bus after he left the house. __ After he left the house he caught the bus.	1 0
9.	__ The runner isn't going to win the race. __ Isn't the runner going to win the race?	1 0	19.	__ The boy who won the contest was clever. __ The boy who was clever won the contest.	1 0
10.	__ We shouldn't cross the street here. __ Shouldn't we cross the street here?	1 0		**Subtotal**	
				Raw Score	
	Subtotal			**Standard Score**	

Appendix 5a: Administration Directions

General Directions:

Depending on the student's age[†] and the testing situation, you may point to the words in the boxes as you read them. You may do this for the demonstration trial and test items. To record responses to each item, check the blanks on the scoring form that correspond to the student's responses. Circle 1 for a correct response or 0 for an incorrect response. The student must give TWO correct responses for each item to be credited as correct. If the student provides an incorrect response, write it VERBATIM on the scoring form for later analysis.

Starting point:	All ages start at item 1
Discontinue items:	If the student pauses for more than ten seconds
Discontinue testing:	After 5 consecutive zero scores

Demo:

Show the student the stimulus page - *Demo* and say, **"Here are some words that can be made into two different sentences: *tall, the boy, is*[‡] [pause]. *The boy is tall*. This sentence tells something. That's one way of doing it. Here's another sentence with the same words** [pause]. ***Is the boy tall?*** **This sentence asks something."**

Practise 1:

Show the student the stimulus page – *Practise 1* and say, **"Now I want you to try it. Make two sentences using these words: *saw, the girl, the boy*. Use only those words.**

If the student produces one sentence, say, **"Now make another sentence with the words."**

If the student requests a repetition, responds incorrectly, or pauses for more than ten seconds, say, **"Remember the words are *saw, the girl, the boy*. Make a sentence** (or **a different sentence**) **with those words."**

If the student cannot produce a sentence, say, **"You could have said…"** [Present an option.]

Practise 2

Show the student the stimulus page - *Practise 2* and say, **"Here are some more words. Now make a sentence with these words: *is, on the hair, the kitten.*"**

[The same correction strategies were used as per *Practise 1*]

Test Items

Prior to turning to the first test item, introduce the test items saying, **"Now let's do some more. Each time you will make two sentences using the words I show you. Both sentences must be logical and make sense.**"

† ***This test should only be given to student's aged 9;0+***

‡ ***Throughout*** *the administration directions, substitute the appropriate words as per the test form being administered.*
(ie. If administering Manipulating Sentences – Form A – Demo, substitute the words: sweet, the apple, is*.)*

Appendix 5b: Additional Administration Directions for Group Assessment Task

General Directions:

[As per Appendix 5a]

Demo:

Give each student an assessment booklet and a pen/pencil. Show the students the stimulus page - *Demo* and say, **"Here are some words that can be made into two different sentences: *sweet, the apple, is* ‡** [pause]. ***The apple is sweet*. This sentence tells something. The words have been placed into order by placing numbers above the boxes. That's one way of doing it. Here's another sentence with the same words** [pause]. ***Is the apple sweet tall?* This sentence asks something. The numbers written under the boxes show the correct order for the words."**

Practise 1:

Show the students the stimulus page – *Practise 1* and say, **"Now I want you to try it. Make two sentences using these words: *helped, the woman, the man*. Use only those words. To show your answers, write numbers above the word boxes for one sentence and then numbers below the boxes for another different sentence."**

Allow the students up to 20 seconds to complete the first sentence. At the end of twenty seconds say, **"Please start your second sentence if you haven't already done so."**

Allow the students up to another 20 seconds to complete the second sentence. At the end of 20 seconds say (ie. 40 seconds for the whole item), **"Stop! Now let's look at your answers."** Select one of the students to read one of his/her answers. If his/her sentence is correct check with the group that they all have this sentence marked, otherwise provide a correct response for the group. Then ask a different student in the group to tell you the other sentence. If his/her sentence is the other correct response indicate this to the group, otherwise provide the correct response.

[NB. It is likely that the students will finish well within the time limit on the practise items. Therefore if this occurs, start discussion of the item as soon as all students have finished. Do not penalise students for starting the second sentence immediately after the first, but do not allow students to turn the page until you say they can.]

Practise 2

Show the student the stimulus page - *Practise 2* and say, **"Here are some more words. Now make a sentence with these words: *is, in the cage, the bird.*"**

[The same procedure should be used as per Practise 1]

Test Items

Prior to turning to the first test item, introduce the test items saying, **"Now let's do some more. Each time you will make two sentences using the words I show you. Both sentences must be logical and make sense. Use your pen to number above the boxes for the first sentence and below the boxes for your second sentence. I will let you know when to start your second sentence on each page and when to stop and turn the page.** *[pause]* **Turn the page.**"

Appendix 6: CELF-4 Sentence Assembly – Scaled Scores (Ages 9;0 – 12;11)

Scaled Score	9;0 - 9;11	10;0 - 10;11	11;0 - 11;11	12;0 - 12;11	Scaled Score
19	-	-	-	-	19
18	-	-	-	-	18
17	19	19	-	-	17
16	18	18	19	-	16
15	17	17	18	19	15
14	15-16	16	17	18	14
13	14	15	16	17	13
12	12-13	14	15	15-16	12
11	10-11	12-13	14	14	11
10	8-9	10-11	12-13	12-13	10
9	6-7	8-9	10-11	10-11	9
8	5	6-7	8-9	9	8
7	4	5	6-7	8	7
6	3	4	4-5	6-7	6
5	2	3	3	4-5	5
4	1	2	2	3	4
3	0	1	1	2	3
2	-	0	0	1	2
1	-	-	-	0	1
Confidence Level	**Scaled Score Points for Building Confidence Intervals**				**Confidence Level**
68%	1	1	1	1	68%
90%	2	2	2	2	90%
95%	2	2	2	2	95%
	Test – Retest Data				
Test (Mean ± SD)	9.9 ± 3.0	10.4 ± 3.1	-	10.1 ± 2.9	Test (Mean ± SD)
Retest (Mean ± SD)	11.3 ± 2.9	11.8 ± 3.2	-	11.4 ± 2.4	Retest (Mean ± SD)
Difference in means	1.4	1.4	-	1.3	Difference in means
Standard Difference*	.47	.44	-	.49	Standard Difference*
Correlation Coefficient	.81	.79	-	.85	Correlation Coefficient

* **NB. Standard Difference = Difference in means/average of SD's**

Reference: Semel, E., Wigg, E. H. & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals (4th Edition)*. San Antonio TX: The Psychological Corporation

Appendix 7: Item by Item Comparisons

Item by Item Comparison: SA - MSA

	LL	MR	MK	AM	CB	AH	DM	SC	ZM	KW	NK	Comparison between participants
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
3	✓	✓	✓	✗	✓	✗	✗	✗	✗	✓	✗	36%
4	✗	✗	✓	✗	✗	✓	✓	✗	✓	✗	✗	45%
5	✓	✗	✓	✓	✗	✗	✓	✓	✗	✗	✓	55%
6	✗	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	73%
7	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✗	73%
8	✓	✓	✗	✓	✓	✗	✓		✓	✓		78%
9	✓		✗	✓	✗	✓	✓		✓			71%
10	✗		✓		✓	✓			✓			80%
11	✓		✗			✗			✓			50%
12	✓		✗			✓			✓			75%
13	✓		✓			✓			✓			100%
14	✓		✓			✓			✓			100%
15	✓		✓			✓			✓			100%
16	✓		✗			✓			✓			75%
17			✗			✗			✗			0%
18			✗			✗			✓			33%
19			✓			✗			✗			33%
Comparison within participants	81%	75%	58%	67%	70%	58%	89%	57%	79%	75%	57%	67%

Item by Item Comparison: SA – MSB

	LL	MR	MK	AM	CB	AH	DM	SC	ZM	KW	NK	Comparison between participants
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
3	✓	✓	✓	✗	✓	✗	✓	✗	✓	✓	✗	64%
4	✓	✓	✗	✓	✓	✓	✓	✗	✓	✗	✗	64%
5	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	82%
6	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓	✓	82%
7	✓	✓	✗	✓	✗	✗	✓	✓	✓	✓	✓	73%
8	✓	✓	✗	✓	✗	✓	✓		✓	✓		78%
9	✓		✗	✓	✓	✓	✓		✓			86%
10	✓		✓	✓	✗	✓			✓			83%
11	✓		✓	✓	✗	✗			✓			67%
12	✓		✗		✗	✓			✓			60%
13	✓		✓		✓	✓			✓			100%
14	✓		✗		✓	✓			✓			80%
15	✓		✓		✓	✓			✓			100%
16	✓		✓		✓	✓			✗			80%
17			✓		✓	✗			✓			75%
18			✗		✓	✓			✗			50%
19			✓		✓	✓			✗			75%
Comparison within participants	100%	100%	63%	91%	74%	74%	100%	57%	79%	75%	71%	79%

Item by Item Comparison: MSA - MSB

	LL	MR	MK	AM	CB	AH	DM	SC	ZM	KW	NK	Comparison between participants
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
3	✓	×	✓	✓	✓	✓	×	✓	×	✓	✓	73%
4	×	×	×	×	×	✓	✓	✓	✓	✓	✓	55%
5	✓	✓	✓	✓	×	×	✓	✓	✓	✓	✓	82%
6	×	✓	✓	×	✓	×	✓	✓	✓	✓	✓	73%
7	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	×	82%
8	✓	✓	✓	✓	×	×	✓	×	✓	✓	✓	73%
9	✓		✓	✓	×	✓	✓	✓	✓		✓	89%
10	×		✓		×	✓		×	✓		×	43%
11	✓		×			✓		✓	✓		✓	83%
12	✓		✓			✓		✓	✓		✓	100%
13	✓		✓			✓			✓		✓	100%
14	✓		×			✓			✓		✓	80%
15	✓		✓			✓			✓			100%
16	✓		×			✓			×			50%
17			×			✓			×			33%
18			✓			×			×			33%
19			✓			×			✓			67%
Comparison within participants	81%	75%	74%	78%	40%	74%	89%	83%	79%	100%	86%	74%

Appendix 8: Error Analysis Form

Student: **DOB: __/__/__** **DOAx: __/__/__** **Age:**

Item	Test 1: _____ *Scaled Score: __*						Test 2: _____ *Scaled Score: __*						Test 3: _____ *Scaled Score: __*						Error Pattern across tests
	Delayed		Error		Error Type		Delayed		Error		Error Type		Delayed		Error		Error Type		
	NR	1✓	0✓	1✓			NR	1✓	0✓	1✓			NR	1✓	0✓	1✓			
1					D I	P					D I	P					D I	P	
2					D	SC					D	SC					D	SC	
3					D	SC					D	SC					D	SC	
4					D I C	PP					D I C	PP					D I C	PP	
5					D I	IO					D I	IO					D I	IO	
6					D	PP					D	PP					D	PP	
7					D I	N F					D I	N F					D I	N F	
8					D I	N PP					D I	N PP					D I	N PP	
9					D I	N F					D I	N F					D I	N F	
10					D I	N					D I	N					D I	N	
11					D I	IO					D I	IO					D I	IO	
12					D I	N PP					D I	N PP					D I	N PP	
13					D	SC					D	SC					D	SC	
14					D I	N					D I	N					D I	N	
15					D	SC					D	SC					D	SC	
16					D I	N					D I	N					D I	N	
17					D	N SC					D	N SC					D	N SC	
18					D	SC					D	SC					D	SC	
19					D	RC					D	RC					D	RC	
Totals					**D /19** **I /11** **C /1**	**P / 1** **N / 8** **PP / 4** **F / 2** **IO / 2** **SC / 6** **RC / 1**					**D /19** **I /11** **C /1**	**P / 1** **N / 8** **PP / 4** **F / 2** **IO / 2** **SC / 6** **RC / 1**					**D /19** **I /11** **C /1**	**P / 1** **N / 8** **PP / 4** **F / 2** **IO / 2** **SC / 6** **RC / 1**	