The New Zealand Articulation Test norms project

Jayne Moyle
Group Special Education
Lower Hutt

Abstract

The purpose of the New Zealand Articulation Test (NZAT) norms project was to provide normative data on the test for junior primary school children in New Zealand. The NZAT consists of 82 picture stimuli which elicit single word responses. Fifty-six Speech Language Therapists administered the NZAT to 1013 children aged 5 to 8 years from 53 schools throughout New Zealand. Results showed older children had more accurate articulation of words. Girls were found to score higher than boys in the younger age groups. Differences were also found by ethnicity at younger ages. Significant effects were found by socioeconomic status. There were no significant differences found by geographic region or population density (rural / urban).

Early sound acquisition studies

The first studies to gather normative data for speech sound acquisition for American children were conducted by Wellman, Case, Mengert and Bradbury (1931) and Poole (1934). These authors published ages of acquisition for speech sounds. These studies were followed by others in America, Australia and the United Kingdom (Anthony & McIsaac, 1971; Arlt & Goodban, 1976; Grunwell, 1981; Ingram, 1976; Prather, Hedrick & Kern, 1975; Sander, 1972; Templin, 1957). More recent studies, (e.g. Dodd, Holm, Hua & Crosbie, 2003; Goldman & Fristoe, 2000; Porter & Hodson, 2001 and Smit, Freilinger, Bernthal, Hand & Bird, 1990), have found a lower age of speech sound acquisition than the original studies by Poole (1934), Templin (1957) and Wellman et al. (1931).

Local articulation development studies

The first New Zealand speech sound production data was published by Justin, Gibson and Silva (1980) and Lawn and Silva (1983). Justin et al. (1980) assessed 681 children aged three to six years, using the Dunedin Articulation Test (DAT). The DAT uses a mixture of single words, phrases and automatic speech to assess 43 phonemes and 14 consonant blends. Results were presented as mean total scores for the whole test and age at which 50 % and 90 % of children articulated each speech sound correctly.

Speech production norms were also developed from the Dunedin Articulation Check (DAC), which was administered to 1640 children aged seven to nine years. The DAC was designed to screen large numbers of school aged children to supplement other channels of referral.

The test took 90 seconds for an experienced examiner to administer and assessed 20 later developing consonants and blends in a mixture of single word and sentence contexts. Results were presented in terms of mean total scores and percentage passing rates at ages seven and eight years.

While local norms are important to ensure valid comparisons during the assessment process, the vocabulary items and materials used in these tests are now outdated. The philosophy behind the DAC is also dated, as the government funded SLT service no longer screens large numbers of children, but encourages schools to meet the needs of children with mild speech difficulties. Special Education (SE) is the government funded special education service for New Zealand schools. Services are divided into initiatives targeted at specified groups of children with special needs. The Communication Initiative provides a speech language therapy service for the most severe one percent of children with communication disorders. Consequently, screening tools now used in Special Education must be able to determine if a child fits within that one percent bracket. Normative data plays an important role in this decision alongside a variety of factors such as language skills, cognition, social development, motor development, parent and teacher concern, number and type of errors, consistency of errors, overall intelligibility and stimulability of the child. The present research was undertaken to provide normative data for the New Zealand Articulation Test (NZAT), a current assessment designed for SE speech language therapists' use within the New Zealand context.

Difficulties with speech sound development data currently used

Clinical decisions in New Zealand on whether a sound is significantly delayed are usually based on the first classical studies by Templin (1957) and Wellman et al. (1931). Wellman et al. (1931) studied the speech development of 204 children aged two to six years. Templin (1957) studied 480 children aged between three and eight years. Both investigations used spontaneous and imitative methods of speech testing and did not include children with speech disorders. Participants in Wellman et al.'s (1931) study were recruited from the University child care facilities, assumedly having higher than average intelligence. Contrastively, Templin's (1957) sample was drawn from local schools and was balanced for socioeconomic status.

The most commonly used summary of the above studies was produced by Sander (1972). The author presented the results of these early studies (Wellman et al., 1931 and Templin, 1957) in graphic form showing the median age of customary articulation and the age at which 90 % of children were customarily producing the sound. This chart is useful as a quick visual guide of speech sound acquisition and is wisely used in New Zealand. A chart summarising recent local norms would make a useful replacement.

While such a reference chart would be useful, more detailed description of the sound development process is needed to empower clinical decisions. Sander (1972) demonstrated the arbitrary nature of the ages of acquisition set by each researcher. The author showed that the upper age limit for /t/ could change from age six to age three by altering the mastery criteria to correct production of the phoneme in two out of three positions rather than in all three positions. This is due to /t/ being "flapped" until later ages in medial position when it was correct earlier in other word positions. Data on the percentage of children producing sounds correctly in each position is therefore needed.

Another difficulty with reporting norms within Sander's chart is that error type is not taken into consideration. Different errors are acceptable for sounds at different ages, as children typically go through successively closer approximations of a consonant or consonant cluster's correct production over time. For example, a child may say /dim/ for "scream" at age two, /stim/ at age three, /skim/ at age four, and achieve correct production at age six with /skrim/. Therefore a six year old child saying /d/ for /skr/ is a concern, whereas the same child saying /sk/ for /skr/ is not a concern. This distinction is lost on many parents and teachers when using Sander's chart, as /skr/ reaches 90 % criterion at age seven. A summary of the error types for each sound in each position by age would be useful in assisting clinical decisions, as called for by Smit (1986). Consequently, this information has been included in the NZAT manual (Moyle, 2004) as raw data, but is beyond the scope of this research paper to analyse.

In summary, New Zealand therapists require local norms for speech sound production in all word positions, so that distinction between acquisition data of a sound's production between different word positions can be made. Charts summarising this information would be widely used.

Factors affecting speech sound acquisition

Factors such as age, gender, socioeconomic status, geographic region and ethnicity have been reported in the literature as potential effects on speech sound acquisition.

It is important to examine the effects of such factors so that normative data can be applied correctly to different populations. Age is a well established factor in articulatory proficiency. On the other hand, there have been conflicting reports on the effect of gender, socioeconomic status and ethnicity on speech sound acquisition. Finally, geographic region is not proven to have an effect on articulatory proficiency.

All studies reported agree that articulatory proficiency increases with age until articulatory mastery is achieved at about age eight to nine years. Dodd et al. (2003) summarised the literature regarding gender and socio economic status. Generally females were observed to develop articulatory proficiency, (particularly of interdental fricatives and consonant blends), faster than boys. However, studies vary on the ages that these differences were reported. For example, Smit et al. (1990) reported differences at ages four, four and a half and six years. Dodd et al. (2003) and Poole (1934) reported differences only after age five years. Other studies (e.g. Wellman et al. 1931) found no differences between five and six year old boys and girls. Boys have also been reported to be at higher risk for speech disorders than girls. (Weinrich, Jennen-Steinmnetz, Laucht, Esser & Schmidt (1998).

Socioeconomic status has been reported to contribute to some variation in articulatory development, with children of a lower socioeconomic status acquiring phonemes and phonological abilities more slowly (e.g. Templin, 1957). Smit et al. (1990) on the other hand, reported no effect on speech sound acquisition by socioeconomic status. Socioeconomic status has also been found to affect other areas of speech and language development such as vocabulary acquisition (Bates et al., 1994), phonological abilities, cognitive, language and pre-reading measures (Robertson 1998) and phonological awareness (Burt, Holm and Dodd, 1999). Law (1992) cautions against making a straightforward association between poor development and socioeconomic status, concluding that it is unlikely to be responsible for more than a slight variation in language development. Dodd et al. (2003) concluded in their literature review, that care needs to be taken in examining the role of socioeconomic status on speech development as different methods of measuring socioeconomic status, (e.g. occupation, education, income), can lead to different outcomes on its role in children's development.

Local norms are preferred so that comparisons can be made on a like to like basis. However, norms from other geographical regions are frequently used where local norms are not yet established. Goldman and Fristoe (2000) tested a sample of 281 English-speaking Canadian children on the Goldman-Fristoe Test of Articulation-2 (GFTA-2) to compare their results with the American normative sample and concluded no significant differences between the two populations. It was thus concluded that the American norms were appropriate for English-speaking Canadian children. Similarly, Smit et al. (1990) did not report differences by geographic region. If the GFTA-2 (Goldman & Fristoe, 2000) was not culturally biased in some of its word choice, a small study comparing a sample of New Zealand children with the American norms may have sufficed as normative data for New Zealand. However, several of the test words in the GFTA-2 are not commonly used in New Zealand English. Words included in articulatory testing should at least be in the passive vocabulary of children and ideally be actively used making the GFTA-2 inappropriate for use in New Zealand (Higgs, 1970).

There were no published studies found that have reported on the effect of ethnicity on articulation development. Ethnicity is closely associated with dialectal differences in spoken English in many countries. In New Zealand, however, there is little approaching dialectal variation in terms of phonemes, aside from Maori accented English (unaspirated initial /t/, final /z/ devoicing) (Robertson, 1996) and the rhotic /r/ of Southland. However, if differences were found in patterns or rates of articulation development, between ethnic groups, separate norms may be needed.

Few studies have examined the differences in children's articulation development comparing rural and urban populations. Smit et al. (1990) did not report any differences in articulatory proficiency by population density. This area will also be investigated by the current study.

Research questions

Sander's (1972) tables of normative ages for articulation development, devised from Wellman et al. (1931) and Templin's (1957) classical research, are the most commonly used norms in New Zealand clinical practise. This study seeks to replace these with local norms for the ages five to eight years for a commonly used test (NZAT). The results are published in the manual of the NZAT as percentages of children achieving each sound correctly in each position tested and charts summarising the age at which 90% and 95% of children achieve sounds (averaged across word positions tested). This study also adds to the body of data on whether gender, socioeconomic status, ethnicity, and geographic region or population density have an effect on total score on the NZAT. If significant

differences are found, this may indicate the need for separate norms for different groups, so that comparison can be made on a like to like basis.

Method

The NZAT normative sample was completed between February and June 2004, with a representative sample of 1014 children aged five to eight years tested at 53 schools nationwide.

Site and sample selection

Fifty-six examiners were recruited from every region in New Zealand. They were all employed by Special Education as Speech Language Therapists, familiar with articulation testing and proficient in transcribing responses using the International Phonetic Alphabet (IPA) (revised 1993). Administration guidelines were provided to each therapist to ensure the reliability of data collection. Contact was maintained between the research coordinator and examiners by newsletter, email and telephone.

Each examiner was given a sample assignment which described the demographics of the children to test. Examiners approached local schools for permission to assess children. Schools chose whether they required permission forms or informed consent for their students to participate. Therapists obtained the school roll for the ages concerned and selected the children matching the sample assignment using a randomised procedure. Every fifth child on the roll was selected starting from the tenth name on the list. In certain cases, it was not possible to use a randomised procedure to select the children, as there were not enough children in smaller schools with the target demographics. In these cases, all the potential subjects were used, provided they met the other subject criteria. The therapist tested the children at school during school hours.

The normative sampling procedure was simple. The consonant and blend sections of the test were administered to the children. If the child articulated the sound correctly, a dot was marked on the response form. If the child misarticulated, this was transcribed using IPA transcription. Diacritics were not recorded except for nasalized sounds. Response forms were returned to the research coordinator for analysis.

Characteristics of the sample

Gender and age

The standardisation testing plan is summarised in Table 1. The sample was divided primarily by gender and age, with a minimum of 100 children in each group, due to the established age effects and likely gender differences in articulation development. Due to the relatively rapid changes in articulation development at age five years, a six month interval was used for this younger sample. Intervals of one year were used for the six to eight year olds. Subjects could be any age within their respective age band. For example, six year olds were aged from 6 years, 0 months to 6 years 11 months. There was no attempt to balance the spread of ages within each band.

Table 1. Total number of subjects in the NZAT normative sample by gender and age

Age	Male (*n*)	Female (*n*)	Total
5;0 – 5;5	101	100	201
5;6 – 5;11	90	109	199
6;0 – 6;11	108	93	201
7;0 – 7;11	94	112	206
8;0 – 8;11	117	89	206
Total	510	503	1013

Geographic region

The population of geographical regions was proportionally represented in the sample. Table 2 shows the population of each region compared to the percentage in the normative sample. Proportions of rural versus urban children also reflected census statistics. Twenty-five percent of the sample came from communities with a population of 1000 or fewer with the remaining 75% from urban areas.

Table 2. Comparison of the percentages of the New Zealand population and NZAT sample population by region

Region	NZAT sample (*n*)	% NZAT Sample	% NZ Population*
Northland	38	3.75	3.7
Auckland	324	32.0	32
Bay of Plenty	64	6.31	6.2
Waikato	38	3.75	9.9
Manuwatu-Wanganui	82	8.09	6.2
Taranaki	16	1.58	1.2
Hawkes Bay	68	6.71	3.7
Wellington	124	12.23	11.2
Marlborough	14	1.38	1.2
Tasman	7	0.69	1.2
Canterbury	138	13.61	12.5
West Coast	7	0.69	1.2
Otago	51	5.03	4.9
Southland	25	2.47	2.5

** Note: NZ population data from Statistics New Zealand census information, 2003*

Socioeconomic status

An estimate of socioeconomic status was controlled for by splitting the schools into decile bands of low (1-3), mid (4-7) and high deciles (8-10). Decile ranking is based on the extent to which a school draws its students from low socioeconomic communities. Decile 1 schools are the 10% of schools with the highest proportion of students from low socioeconomic communities, whereas decile 10 schools are the 10% of schools with the lowest proportion of these students. There are six factors which determine the socio economic ranking: household income, occupation, household crowding, educational qualifications, income support, and ethnicity. The latest decile information available (Ministry of Education, 2003) was used. Table 3 shows the target percentage for each decile band compared to the normative sample. The percentage of the normative sample differed from target percentage by up to nine percent, with low decile schools being slightly over represented compared with the middle decile band.

***Table 3.** Percentage of NZAT normative sample drawn from low, mid and high decile bands*

Decile Band	n	Target %	% of sample
1 - 2	300	20	29
3 - 7	431	50	43
8 - 9	282	30	28

Ethnicity

Examinees were classified into the following groups: European / Pakeha, Maori, Pacific Island, Asian or other. Table 4 shows the proportion of the sample from each ethnicity compared with the population of New Zealand. Minority ethnicities were well represented in the normative sample.

***Table 4.** Comparison of the percentage of population from different ethnic backgrounds with the sample population*

Ethnicity	% of NZ Population	% of Sample Population *
European	77	63
Maori	14	24
Pacific Island	6	10
Asian	6	3
Other	1	7

** Note: NZ population data from Statistics New Zealand census information, 2003*

Special education status

Children who had Ongoing and Reviewable Resourcing Scheme (ORRS) Funding (verified as being the 1% of the population with the highest educational needs), permanent hearing loss, cleft palate and diagnosed syndromes which adversely affect learning were excluded from the normative data due to the link between articulation ability and these conditions. Children with hearing loss, cleft palate and certain syndromes display a pattern of errors

caused by physiological or sensory differences which are not applicable to the typically developing child.

Children receiving speech therapy services for language and/or speech difficulties, but did not fit in the above categories, were included in the sample. This was because the full range of articulatory ability, excluding children with identified physiological deficits, needed to be represented in the sample.

Children with inadequate English to express themselves fluently were excluded. Fluency was judged by classroom teachers before the children were included in the potential pool of subjects. These children were excluded because their scores may have been affected by a lack of vocabulary knowledge and/or an incomplete knowledge of English phonology as compared to their other language(s).

The assessment tool

The NZAT was designed for use within Special Education the Communication Initiative, which provides a service for the one percent of the school population with the highest communication needs. As the target population is the severely disordered, the test was designed to assess each sound in the easiest phonetic context so children eligible for intervention were not over-identified. Words were carefully selected so target sounds were not part of morphemes, or have phonemic contexts which often induce assimilation errors. For example, velars were not tested in the context of alveolar sounds. Initial and final sounds were tested in single syllable words and medial sounds were tested in bi-syllable words. This was because accuracy of word initial fricatives and phoneme production consistency have both been shown to decrease with increased syllable complexity (Bader, 2002; Ingram, Christensen, Veach & Webster, 1980). In clinical practise, assessment of multisyllabic words, vowels, stimulability and connected speech are also important parts of a diagnostic assessment protocol. These areas of assessment are included in the NZAT, but were not included in the norming process.

The NZAT assesses all word initial, medial and final consonants except medial /h/, /w/, /j/, /zh/ and final /zh/, /l/, /r/,voiced 'th', voiceless 'th' and /w/. These sounds were either inapplicable to New Zealand English phonology or deemed clinically insignificant, as they have a limited impact on intelligibility and therefore are unlikely to be therapy targets. This is either because of their rarity of occurrence, particularly in the vocabulary of young children, or the unlikelihood they would be in error. Nineteen double and four triple consonant blends were assessed in initial position. Final and medial blends were not assessed, as they are not commonly therapy targets. A total of 82 phonemes were assessed using 82 picture stimuli. The word structure of the stimulus items is displayed in Table 5.

Table 5. Word structure of the test items in the NZAT

Monosyllabic Words		Disyllabic Words	
Structure	N	Structure	N
CV	4	CVCCVC	1
CVC	27	CVCVC	15
CVCC	1	VCVC	1
CCV	2	CVCV	6
CCVC	17	CCVCVCC	1
CCCV	3	CCVCVC	1
CCCVC	1	CCVCV	2

Each sound was assessed on one test item only. Variability in sound production is a typical part of speech disordered children's articulation (Dinnsen, 1997; Leonard & Leonard 1985). Despite this, it was decided that a "snapshot approach" to sampling would be employed as this would reflect current practice of screening assessments in NZ and be time efficient. Clinically, this method is employed as a time saver, with multiple context testing of sound production only being employed if the child has particularly high levels of speech inconsistency. Each word was used to assess a single sound only. This method of testing was employed to increase transcription accuracy and due to the popularity of the format amongst therapists.

While the NZAT as a whole includes a vowel assessment, this was not included in the normative testing as vowel errors are rare after age three. School aged children should be able to use all vowels and diphthongs correctly, with the possible exception of unstressed vowels in multisyllabic words (James, 2001; Templin, 1957).

A standard elicitation sequence was used to avoid direct imitation by the child when spontaneous production of a test item could not be obtained. If the child failed to name the item correctly, the therapist would give a semantic cue or point to the specific part of the picture to name. If the child was still unable to name the picture after several semantic cues, the picture was named by the therapist, an intervening sentence about the picture inserted and then the child was asked to name the picture again.

Coloured pictures are arranged in grids according to the position of the target sound in the word, general developmental order and class of sound. There were between six to twelve pictures per page dependent on the

above classifications. Higgs (1970) argued for the order of items in tests to be randomised. It was decided, however, to order the items to allow for quick analysis, reference and administration of the certain sections of the test. Semantic cues shown to successfully elicit the target word during the trial study were printed at the bottom of pictures which had required an imitation cue for ten percent or more of children during the test development. (For further details on the development of the test, refer to Moyle, 2004).

Recording responses

Therapists transcribed incorrect responses using the IPA. If a child self corrected on their second production of a word, only the initial error was recorded. The only diacritic marked was nasalized sounds. Other minor variations, such as dentalisation of /s/, unreleased plosives and partial voicing, were scored as correct as they were deemed clinically insignificant.

Children's responses were judged to be acceptable if they were perceived by the therapist as being the adult accepted norm for New Zealand English. For example, flapping medial /t/ was common and was noted as an accepted variant in the results rather than an error. Lack of a recorded response was exceptionally rare and was treated as an error.

Data analysis

Data analysis was carried out by the New Zealand Council for Education Research. The analysis generated using SAS software (1999- 2001) and the statistical package R (2003). The total raw error score was calculated for each subject. The mean and standard deviation of the each group's score by age, gender, ethnicity, geographic region and population density were calculated. The groups' results were compared for significant differences using chi-square tests. Charts showing when 90% and 95% of subjects produced a sound correctly over an average of all positions tested were produced for each gender. Percentile ranks and standard scores for the total error score on the NZAT were calculated by age and gender.

Reliability studies

Two studies measuring inter-rater reliability and test retest reliability also ran concurrently. Inter-rater reliability was assessed using two pairs of therapists and 51 subjects. One member of the pair initially tested the subjects according to the study's administration guidelines. The subjects were not informed they would be retested. The second examiner retested the subjects two to three weeks later.

Test-retest reliability was assessed by testing a sample of 47 subjects again by the same examiner two to three weeks later, according to the same guidelines as the rest of the normative sample. The subjects were not informed they would be retested. Four different therapists participated in this study.

Percentages of agreement between examiners' results for production correctness for each sound were calculated. Median percentage agreements for inter-rater reliability were 98% for initial, medial and final sounds and 92% for blends. Median percentage agreements for test retest reliability were 100% for initial, medial, final and blended sounds.

Internal Reliability

Internal reliabilities were computed using co-efficient alpha. This score is based on the inter-correlation among all comparable parts of the same test. The alpha reliabilities represent a measure of homogeneity of items throughout the test and are calculated using the variance of the dichotomously scored items (correct or incorrect production).

Table 6 lists the alpha reliabilities for the five age groups by gender. The alpha reliabilities range from 0.84 to 0.95 for males and 0.86 to 0.95 for females. The median reliability was 0.93 for males and 0.92 for females. This represents an acceptable level of internal reliability.

Table 6. *Alpha reliabilities for each age group by gender*

Age (years; months)	Male		Female	
	N	Alpha	N	Alpha
5;0 – 5;5	101	0.95	100	0.95
5;6 – 5;11	90	0.93	109	0.92
6;0 – 6;11	108	0.91	93	0.94
7;0 – 7;11	94	0.93	112	0.86
8;0 – 8;11	118	0.84	89	0.87
MEDIAN		0.93		0.92

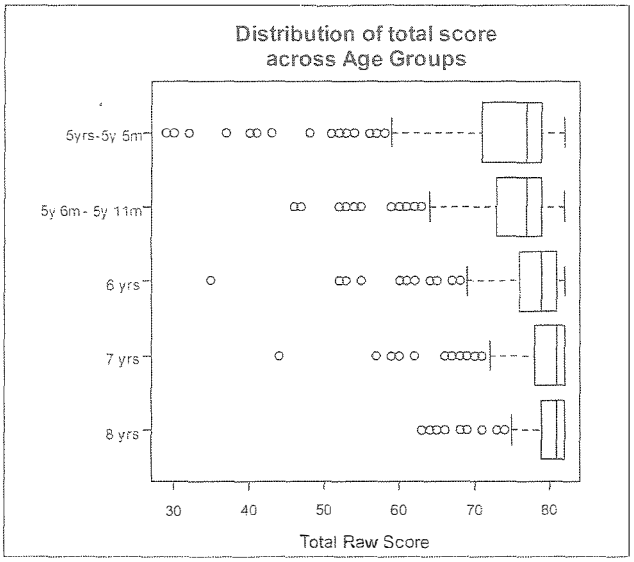
Results

The NZAT results of 1013 children aged between 5 and 8 years were analysed to obtain normative data. Older children had fewer articulation errors than younger children. The majority of children had developed most sounds at the youngest age tested. The test scores therefore show a skewed distribution. Most children got most items correct and very few children, particularly in the older age groups, scored any items incorrectly. As the distribution was skewed, neither ANOVA nor t-tests could be used successfully and chi-square tests were completed for the groupings below. Chi-square results are represented in Table 7 (below).

Testing for differences by age

Age group is significantly associated with the total raw score ($p = < 0.0001$). Older children were more able to articulate sounds in words more correctly than the younger children. Total scores for five to five and a half and five and a half to six year old age groups were not significantly different ($p = 0.941$). These two groups were therefore combined for the purposes of establishing associations by gender, ethnicity and region. The distribution of total scores across age groups is represented in Figure 1

Figure 1. Distribution of total scores across age groups



*Note: The boxplots show the distributions of raw scores across variables of interest. The box shows the inter-quartile range; the middle line indicates the median of the distribution, and the dotted lines to the whiskers show the range of the outer quartiles. Where data points are more than 1.5 * inter-quartile range away from the box they are considered to be extreme points and are marked with a single dot.*

Table 7. *Total score categories for chi-Square analysis contingency tables.*

Group	Total Raw Score (out of 82)	% of Observations in these groups	Number of Observations in the groups
1	72 and below	15%	160
2	73 – 77 inclusive	20%	222
3	78 – 80 inclusive	25%	285
4	81 and above	40%	346
Total		100%	1013

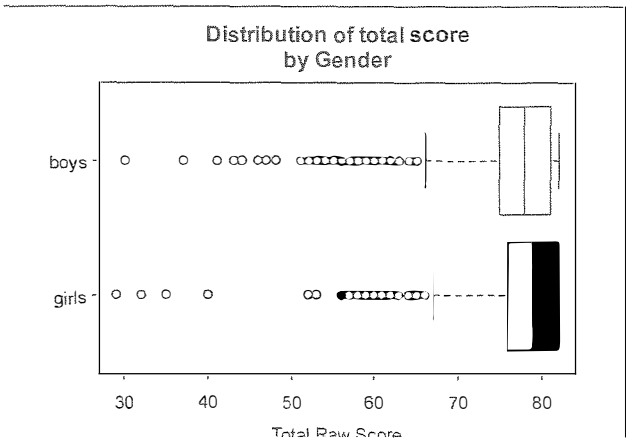
Results by gender

Significant gender differences in total score were found for 5 and 6 year olds, with the girls achieving higher scores. There was no gender difference, however, for the 7 and 8 year old age groups. These results indicate girls acquire articulatory proficiency more quickly than boys. The effect of gender on articulatory performance and gender distribution are represented in Table 8 and Figure 2.

Table 8. Results of chi-square tests for gender by age group

Age Group (years; months)	P Value
5;0 – 5;11	0.075*
6;0 – 6;11	0.006**
7;0 – 7;11	0.115
8;0 – 8;11	0.134
	* p < 0.05, **p<0.01

Figure 2. *Distribution of total score by gender*



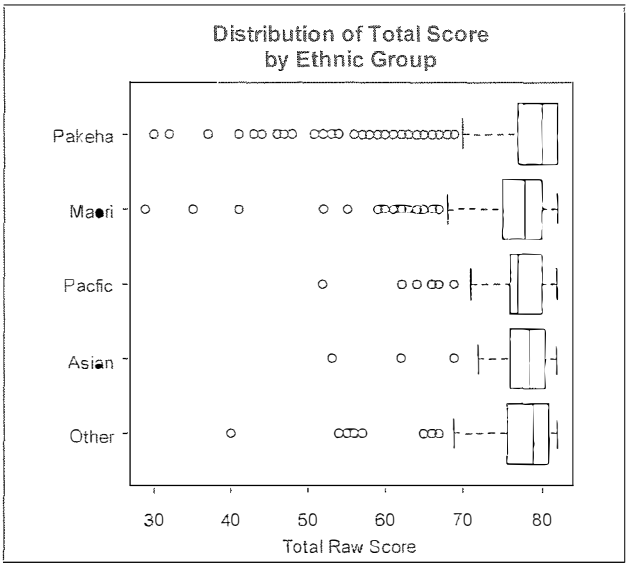
Results by ethnicity

As the sizes were very small for some groups, ethnic groupings were taken as Maori, Pacific, and neither Pacific nor Maori, to complete the test of no association between ethnicity and total score. The results show a convincing association at all but the 8 year old level, which was very close to significance at the 95% level (p = .051). At the earlier ages, Maori and Pacific children score significantly lower than their non-Maori non-Pacific peers. Differences are indicated to lessen by the 8 year old age group. These results are represented in Table 9 and Figure 3.

Table 9. Results of chi-square tests for ethnicity by age group

Age Group (years; months)	P Value
5;0 – 5;11	0.010**
6;0 – 6;11	0.004**
7;0 – 7;11	0.0003**
8;0 – 8;11	0.051
	* p < 0.05, **p<0.01

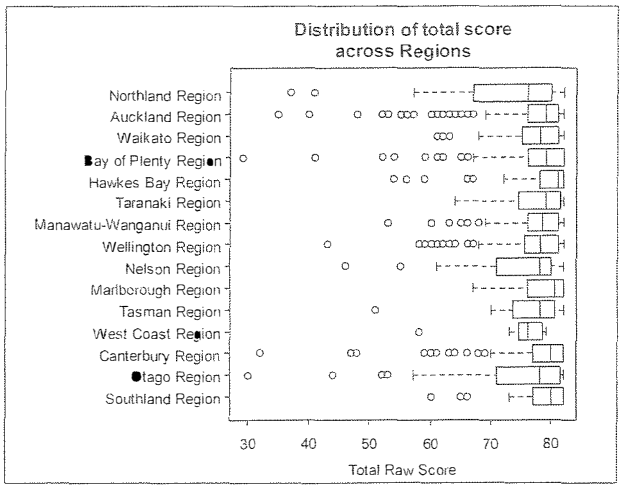
Figure 3. Distribution of total score by ethnic group



Results by region

The fifteen original regions were combined to form three main areas: upper North Island, lower North Island and South Island. There was no significant association between region and age group. Some regions look to have lower averages, although sample groups were too small to complete significance testing between individual regions. In addition, while each geographic region's sample was balanced for gender and age, they were not balanced for socioeconomic group and ethnicity, both of which were also significant factors in total score. The differences between regions are likely to reflect the decile level of the school the samples were drawn from, rather than represent the likely scores for the geographic region as a whole. The distribution of total score across regions is represented in Figure 4.

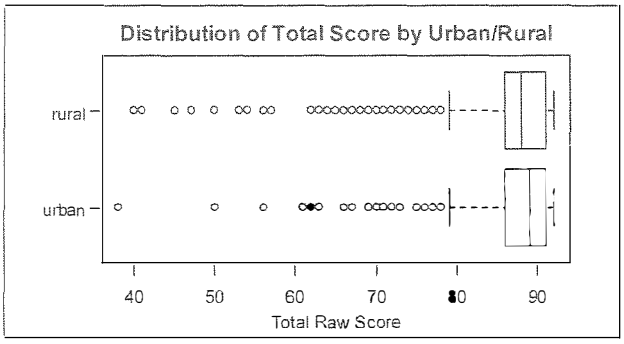
Figure 4. Distribution of total score across regions



Results by population density

Results were analysed to determine if population density had an effect on total scores. No significant association was detected in any age group between urban and rural scores. These results can be viewed in Figure 5.

Figure 5. Distribution of total score by urban/rural area



Socio-economic status

School decile rankings were used as an estimate of socio-economic status. Schools in deciles 1 and 2 were deemed "low" decile schools. Mid rankings were given to schools with decile ratings of 3 to 7 and high decile ranked schools had decile ratings of 8-10. Figure 5 presents the distribution of articulatory scores across socio-economic groupings.

Results were analysed to determine if there was a significant effect of decile band on total scores on the NZAT. Overall, (not taking age into consideration), there was a significant association between school decile (high, middle, low) and total score ($p = < 0.0001$). These results are represented in Table 11.

Figure 6. Distribution of total score by socio-economic group

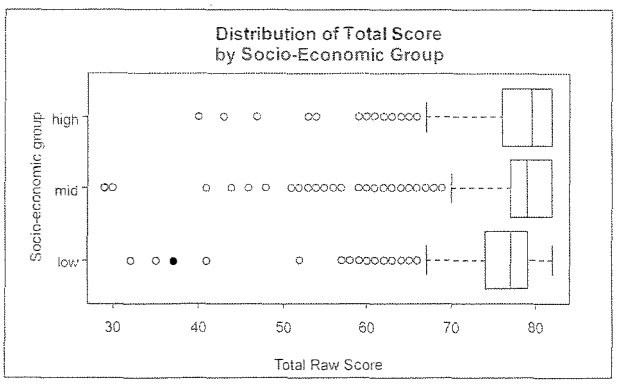


Table 11. Results of chi-square tests for school decile group by age group

Age Band	P Value
5; 0 – 5;5	0.181
5;6 – 5;11	0.040
6;0 – 6;11	0.002
7;0 – 7;11	0.001
8;0 – 8;11	0.054

Discussion

Overview of results

The findings showed that the vast majority of New Zealand children begin school at five years with adult speech sound patterns. A minority of children are still developing some affricates, glides, blends and fricative sounds. As expected, there was a clear pattern of older children achieving higher scores on the test than younger children. Also consistent with previous research, girls were seen to develop articulatory proficiency earlier than boys. Ethnicity and socioeconomic status were significant factors in younger children's performance in the NZAT. These factors have not previously been shown to have a significant effect on articulation development. Consistent with earlier findings, population density and region had no discernable effect on articulation proficiency.

Effect of significant factors on results

Gender

Overall, girls' scores on the test were significantly higher at ages five to six years but not at ages seven to eight years, indicating that girls develop sounds more quickly, but boys have largely "caught up" by seven to eight years.

Girls achieved 95% criterion for the following sounds before age five: /p/, /b/, /t/, /d/, /k/, /g/, /f/, /m/, /n/, /l/, /h/, /w/, /j/, /ng/, /sh/, /ch/ and /dZ/. Boys had achieved the same sounds except for /ng/, /sh/, /ch/ and /dZ/ by the same age. Both girls and boys had achieved /r/ at 90% mastery at age six years. Girls achieved 90% criterion for /r/ blends by age six, while boys achieved the same criterion between six to seven years. Initial triple consonant blends were achieved between ages six to seven years for girls and six to eight years for boys. While boys reached 90% criterion for /s/, /z/ and /s/ blends at age five and a half and showed a consistently increasing mastery of /s/, girls showed an apparent regression in /s/ mastery. Having entered school with 96% mastery of /s/, this dropped to 88% at age five and a half, then increased to 97% again for 6 year old girls. Kenney and Prather (1986), Poole (1934), Prather, Hedrick and Kern (1975) and Templin (1957) also reported a similar pattern of development, mainly for /s/ at preschool ages. Sax (1972) noted this trend amongst school aged children. Similarly, Goldman and Fristoe (2000) showed a six to ten percent dip in performance on /s/ for both boys and girls at school age, but only in some word positions. It is possible that some children, who acquired an acceptable /s/ early in development, then adopt an error pattern for a time before reverting to an acceptable production again (Smit et al., 1990). Kenney and Prather (1986) suggested that reversals affecting /s/ might be due to shifting standards of examiners, who may unconsciously tolerate greater deviations in younger than older children. If this postulation holds true, examiners in the current study must have only had higher standards for the girls at older ages, as the reversal trend was seen in the girls' sample alone despite both genders being tested in equal proportions. Another likely cause for this reversal is sample error or random fluctuation, as while /s/ is the most commonly affected phoneme, this trend has also been reported in other sounds (Smit et al, 1990). This issue requires further investigation in future research.

Girls showed greater proficiency with voiceless 'th' at age eight with 76% testing correct compared with 66% for boys. Girls reached 90% criterion for voiced 'th' at age eight with boys testing at 84% at the same age. Most studies have reported that girls achieve interdental fricatives faster than boys. For example, Smit et al. (1990) reported girls were producing voiceless 'th' correctly on 90% of attempts at age six and voiced 'th' at age four years six months. Boys reached 90% correct at eight and seven years respectively.

The current findings support past research indicating gender differences in articulation development. There are a number of possible explanations for gender differences: girls' fine motor skills develop earlier than boys, differences in brain maturation rates (Hyde & Linn, 1988), earlier maturation of speech organs (Darley & Winitz, 1961; Templin, 1957) and differences in socialisation (Moore, 1967).

Ethnicity

Children from Maori or Pacific Island ethnic backgrounds on average scored lower than children from other ethnic backgrounds. It is possible that these differences are in part caused by the greater susceptibility to ear infections by children of Polynesian descent (Cook et al., 1998; National Audiology Centre, 1998). Fluctuating hearing loss is a factor implicated in some speech sound development delays (Chalmers, Stewart, Silva & Mulvena, 1989; Shriberg, Friel-Patti, Flipsen & Brown, 2000).

Socio-economic status

School decile is a crude proxy measure for a child's socio-economic status. It is therefore important not to over interpret implications of the significance of any socio-economic findings. The results indicated a tendency for those in higher decile schools to score more highly in the older age groups. In the two younger age groups, there was no particular association between decile and total score. At age eight years it appears that all scores improve with age, but that improvement may be less consistent in the lower deciles. Given that ethnicity is a factor in deciding decile rankings of schools and that it was also found to be correlated to lower scores, the correlation between lower scores and decile band is not surprising.

Comparison of NZAT norms with Sander's summary chart

The results of the NZAT norming have been compared with Sander's chart, (based on studies by Templin, 1957 and Templin, 1957), which are the widely used norms in clinical practise in New Zealand. Overall, the developmental pattern is similar between the two sets of normative data. However, there are some differences in type of information presented and in which ages children reach 90% criterion.

Firstly, the NZAT norms are divided into male and female norms, whereas Sander's chart presents both genders together. The normative results of the NZAT indicated that boys and girls norms develop speech sounds at significantly different rates until at least age 8 years and therefore should have separate norms.

There was some difference in the sounds and ages tested. A selection of initial two and three sound consonant blends were assessed in NZAT but not in Sander (1972). The /z/ sound was not included in the NZAT but featured in Sander's chart. In addition, NZAT norms cover ages five to eight years, whereas Sander reported on ages two to eight years.

The ages at which children reach 90% criterion were almost all earlier in the NZAT norms compared with Sander's chart. Of the sounds which are still developing at ages five to eight, the NZAT norms place /t/, /ng/, /l/, /ch/, /sh/, /s/, /z/, /dZ/ and /v/ as reaching 90% criterion between six months to two years earlier than indicated on Sander's chart. Only /r/ met 90% criterion at the same time in both the NZAT norms and Sander's chart. This is consistent with the findings of other researchers who have collected local norms to replace earlier data and found that acquisition ages tend to have lowered over time (for example, Smit et al., 1990).

Voiceless 'th' was the only sound to develop later in the NZ norms. Seventy-one percent of NZ children tested at age eight had voiceless 'th' correct. Sander (1972) reported interdental fricatives reaching 90% criterion at eight years. A more recent American study (Goldman & Fristoe, 2000) reported similar results to Sander (1972). New Zealand children seem to be acquiring voiceless 'th'more slowly compared with American populations. Voiced 'th' results were similar between the three studies, with 87% to 93% of children having acquired this sound by age eight years. As this is the only sound to be developed more slowly in the NZAT results than in other articulation development studies, and the discrepancy is wide (19%), this may indicate a dialectal effect rather than purely developmental processes. The substitution of /f/ for voiceless 'th' is becoming more accepted by a wider group of NZ society, amongst people of non European ethnic backgrounds, non-English speaking backgrounds and lower socio economic classes.

Limitations of this research

Essentially, this research seeks to provide local norms for the NZAT, rather than to define customary ages of acquisition. It is limited in its breadth of study as the test looks at a single instance of production for each sound in single words only. The definition of customary ages of acquisition calls into question issues such as phonetic versus phonemic acquisition. The NZAT looks at phonemic acquisition, (i.e. whether a child uses the sound correctly in a word context), not if a child can simply produce the sound in any context. Reporting on the frequency of phonological processes is also beyond the scope of the current study.

The NZAT test is designed for ages five to eight years and is unsuitable for younger children due to the time needed to administer the test and the high number of small pictures per page. An articulation test needs to be developed and normed for preschoolers aged two to four years in New Zealand.

Implications for clinical practice

New Zealand clinicians are advised to use these local and current norms in conjunction with the NZAT test rather than continue to use Sander's chart. This because the Sander's norms are based on data from 1931 and 1957; on different populations in terms of ethnic mix and educational practises from NZ; tested on smaller sample groupings (circa 50 per age grouping), which did not include children with articulation disorders; and also used a different articulation test. It is also advised that separate norms be used for girls and boys. Clinicians should also bear in mind that children from Polynesian ethnic or lower socioeconomic backgrounds may score lower on the test especially at younger ages.

Acknowledgements

The author wishes to thank Special Education management for supporting and funding this research; the many therapists, schools and children around the country who participated in this project ; and Edith Hogden, Hillary Ferral and Jeremy Zwieglaar (New Zealand Council of Educational Research) who completed the statistical analysis of this study.

References

Aalt, P. B. & Goodban, M. J. (1976). A comparative study of articulationacquisition as based on a study of 240 normals, aged three to six. *Language, Speech and Hearing Services in Schools*, 7, 173-180.

Anthony, N. & McIsaac, M. (1971). Notes on patterns of development found by using the qualitative phonetic assessment sheet of the Edinburgh Articulation Test. *British Journal of Disorders of Communication*, 5, 148-164.

Badar, R., (2002). Factors affecting the consistency of word production: age, gender, word characteristics. Unpublished MSc dissertation, University of New Castle on Tyne, UK, as cited in Dodd et al., (2003).

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznik, S., Reilly, J. & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. In K. Pera, G. Collis and B. Riuchards (Eds), *Growing Points in Child Language*. Cambridge University Press, 85-124.

Burt, L., Holm, A. & Dodd, B. (1999). Phonological awareness skills of 4 year old British children: An assessment and developmental data. *International Journal of Language and Communication Disorders*, 34, 311-335.

Chalmers, D., Stewart, I., Silva, P. & Mulvena, A. (1989). *Otitis media with effusion in children* – the Dunedin study. London: Mac Keith Press.

Cook, L., Kirk, R., Bidwell, S., Hider, P., Weir, R. and Tolan, C. (1998). *Screening programmes for the detection of otitis media with effusion and conductive hearing loss in pre-school and new entrant school children*. New Zealand Health Technology Assessment: Christchurch School of Medicine.

Darley, F. & Winitz, H. (1961). Age of first word: review of research. *Journal of Speech and Hearing Disorders*, 26, 272- 290.

Dinnsen, D. D., (1997). Nonsegmental phonologies. In M. J. Ball and R. D. Kent (Eds). *The New Phonologies*. London: Singular Publishing Group Inc., 77-125.

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

Goldman, R. & Fristoe, M. (2000). *The Goldman-Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Service Inc.

Grunwell, P. (1981). The development of phonology: a descriptive profile. *First Language*, 3, 161-191.

Higgs, J. A. (1970). The articulation test as a linguistic technique. *Language and Speech*, 13, 262-270.

Hyde, J. & Linn, M. (1988). Gender Differences in verbal ability: a meta analysis. *Psychological Bulletin*, 104, 53-69.

Ingram D. (1976). *Phonological disability in children. Studies in Language Disability and Remediation* 2. Edward Arnold.

Ingram, D., Christensen, L., Veach, S. & Webster, B. (1980). The acquisition of word initial fricatives and affricates in English between 2 and 6 years. In G. Yeni-Komshian, J. Kavanagh and C. Ferguson, (Eds), *Child Phonology*. New York: Academic Press, 169-192.

James, D. (2001). An item analysis of Australian English Words for an articulation and phonological test for children aged 2 to 7 years. Clinical Linguistics and Phonetics, 15, 6, 457-485.

Justin, C, Gibson, M and Silva, P. (1980). The Dunedin Articulation Test: Clinical Instrument for Speech Therapists. Dunedin, New Zealand: The Otago Speech Therapy Association.

Justin, C, Lawn, L and Silva, P (1983). *The Dunedin Articulation Check: A Clinical Instrument for Speech Therapists*. Dunedin, New Zealand: The Otago Speech Therapy Association.

Kenney, K. & Prather, E. (1986). Articulation development in preschool children: consistency of production. *Journal of Speech and Hearing Research*, 29, 29-36.

Kilminster, G. and Laird, E. (1978). Articulation Development in Children aged three to nine years. *Australian Journal of Human Communication Disorders*, 6, 1.

Law, J. (1992). *The Early Identification of Language Impairment in Children*. London: Chapman and Hall. As cited in Dodd, Holm, Hua & Crosbie, (2003).

Leonard, L. B. & Leonard, J. S., (1985). The contributions of phonetic context to unusual phonological patterns: a case study. *Language, Speech and Hearing Services in Schools*, 16, 110-118.

McLeod, S., Hand, L., Rosenthal, J. and Hayes, B. (1994). The effect of sampling condition on children's productions of consonant clusters. *Journal of Speech and Hearing Research*, 37, 868-882.

Moyle, J. (2004). *New Zealand Articulation Test*. Tauranga Special Education, Elizabeth Street, New Zealand.

Moore, T. (1967). Language and Intelligence: A longitudinal study of the first 8 years. Part 1: patterns of development in boys and girls. *Human Development*, 10, 88-106.

National Audiology Centre, (1998). New Zealand hearing screening statistics: July 1996 - June 1997. Report to the Transitional Health Authority January 1998. Auckland: National Audiology Centre.

School decile information (2003). Retrieved 17 January 2004, from *www.minedu.govt.org/index.cfm?layout=document&documentid=769&data=1*

Shriberg, L., Friel-Patti, S., Flipsen, P. and Brown, R. (2000). Otitis media, fluctuant

hearing loss and speech-language outcomes: A preliminary structural equation model. Journal of Speech, Language and Hearing Research, 43, 100-120.

Poole, I. (1934). Genetic development of articulation of consonants sounds in speech. *Elementary English Review*, 11, 159-161.

Porter, J. and Hodson, B. (2001). Collaborating to obtain phonological acquisition data for local schools. *Language, Speech and Hearing Services in Schools*. 32, 165-171.

Prather, E. M., Hedrick, D. L. & Kern, C. A. (1975). Articulation Development in children aged 2 to four years. *Journal of Speech and Hearing Disorders*, 37, 55-63.

R Development Core Team, (2003). R: *A language and Environment For Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Robertson, K. L., (1998). Phonological awareness and reading acquisition of children from differing socio-economic backgrounds. Dissertation Abstracts International Section !: *Humanities and Social Sciences*, 58 (8-A), 3066.

Robertson, S. (1996). Maori English and the bus-driving listener: a study of ethnic identification and phonetic cues. *Wellington Working Papers in Linguistics*, 8, 54-69.

Sander, E. K. (1972) When are speech sounds learned? *Journal of Speech and Hearing Disorders*, 37, 55-63.

SAS software, version 8.1 of the SAS System for Windows. © 1999-2001. SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Sax, M. R., 1972. A longitudinal study of language change, *Language, Speech and Hearing Services in Schools*, 3, 41-48.

Smit (1986). Ages of Speech Sound Acquisition: Comparisons and critiques of several normative studies. *Language, Speech and Hearing Services in Schools*. 17, 175- 186.

Smit, A., Hand, L., Freilinger J., Bernthal, J. & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55, 779-798.

NZ population ethnicity information (2001). Retrieved 3 September, 2004 from www.stats.govt.nz/domino /external.pafull.pasfull.nsf/ web.

Templin M. (1957) Certain language skills in children: their development and interrelationships. *Institute of Child Welfare Monograph* 26. Minneapolis: The University of Minnesota Press.

Weinrich, D., Jennen-Steinmetz, C., Laucht, M., Esser, G. & Schmidt, M. (1998). At risk for speech disorders? Correlates and course of language disorders in preschool children born at risk. *Acta Pediatrics*, 87, 1288-1294.