IMPROVEMENTS AND ALTERNATIVES TO THE DISADVANTAGED LBOTE MEASURE

Prepared by NSW Department of Education and Communities

for SPWG meeting April 11 2013

2 April 2013

Funded by the Department of Education, Employment and Workplace Relations

Table of Contents

List of Tables

List of Figures

List of acronyms

ACARA	Australian Curriculum Assessment and Reporting Authority
AEEYSOC	Australian Education, Early Childhood Development and Youth Affairs Senior Officials Committee
EAL/D	English as an Additional Language or Dialect
ELP	English Language Proficiency
ESL	English as a Second Language
LBOTE	Language Background Other Than English
NAP	New Arrivals Program
NAPLAN	National Assessment Program for Literacy and Numeracy
NSW	New South Wales
NSW DEC	New South Wales Department of Education and Communities
SES	Socio-Economic Status
SPWG	Strategic Policy Working Group
TESOL	Teachers of English to Speakers of Other Languages

1. Introduction

The report of the Gonski review of funding for schooling recommended a new schooling resource standard that includes "a series of loadings for student- and school-based sources of disadvantage" (p 153). Limited English Language Proficiency (ELP) is one such source of disadvantage. The Gonski report noted that while state systems are able to measure ELP through their own data systems, there is no nationally consistent measure of ELP currently available.

In its modelling of a new schooling resource standard, the Gonski review instead used the only measure which was available nationally, the 'disadvantaged LBOTE (Language Background Other Than English)' measure, as a proxy measure of educational disadvantage within the broader LBOTE population. The disadvantaged LBOTE measure was developed by the Australian Curriculum, Assessment and Reporting Authority (ACARA). It combines LBOTE with parental education equivalent to Year 9 or below.

However, it is generally acknowledged[1] that the disadvantaged LBOTE measure is a poor approximation for limited English language proficiency.

In September 2012, the Australian Government agreed to fund two projects (endorsed by the Strategic Policy Working Group (SPWG) in July 2012) to:

1) Investigate any improvements that could be made to the 'disadvantaged LBOTE' measure; and,
2) Conduct a cost-benefit analysis of a trial of a more accurate, nationally consistent ELP measure.

This report presents the findings of the first project, which was completed by the NSW Department of Education and Communities (NSW DEC).

2. Existing Accuracy and Coverage of Disadvantaged LBOTE

The NSW DEC analysis firstly provides evidence that the disadvantaged LBOTE measure is inadequate as a proxy measure for limited English language proficiency for three reasons:

a) It underestimates the absolute size of the cohort needing support (low size equivalence)
b) It fails to identify many of the students needing support (limited coverage)
c) It includes too many students who do not require support (lack of accuracy)[2]

[1] Acknowledged in the Gonski Report (p. 118); demonstrated in NSW DEC analysis provided to the SPWG meeting May 24 2012, and acknowledged by SPWG at the July 26 2012 meeting.
[2] The latter two criteria for evaluating the disadvantaged LBOTE proxy measure (i.e., coverage and accuracy) are similar in concept to the two statistical measures used in Epidemiology for assessing medical tests – Sensitivity (relating to the rate of true positives) and Specificity (relating to the rate of true negatives).

These conclusions were reached by comparing disadvantaged LBOTE students to those with limited ELP as gauged by the NSW DEC measure of English as a Second Language (ESL). This section describes the data and analysis that supports this conclusion.

2.1. NSW DEC assessment of ESL need

In the NSW government school system, students requiring ESL support are categorised into one of three broad phases based on the descriptors shown in Table 1 below.

Table 1: ESL phases used in NSW DEC

Phase	Proficiency	Progression	Number of students*
1	Students whose understanding and use of English is obviously limited in all social and educational situations	With ESL support, students generally progress out of this phase after nine months of instruction	17,176
2	Limited to familiar social and educational circumstances	Students generally progress out of this phase after three years of instruction with ESL support	54,636
3	Students who occasionally need assistance in specific educational situations	Students generally progress out of this phase after seven years of instruction with ESL support	62,405

** Source: NSW DEC Annual ESL Survey 2011; for a full description of the different ESL phases, refer to: https://www.det.nsw.edu.au/policies/student_serv/equity/comm_rela/d04_23_ESL_Guidelines.pdf*

Each school's level of ESL need and ESL teacher allocation is determined by the numbers of ESL students in each phase as reported in the NSW DEC Annual ESL Survey. In determining ESL allocation, the greatest weighting is given to Phase 1 ESL learners who are beginning in English. Relative to Phase 1 students who receive a weighting of 100%, Phase 2 students receive a weighting of about 60% and Phase 3 students receive a weighting of 10%.

2.2. Determining the size of the target cohort for analysis

In order to examine the accuracy and coverage of the disadvantaged LBOTE measure as a proxy for limited ELP, it was necessary to first determine the target cohort of students with limited ELP, to which the proxy measure can then be compared. The following paragraphs describe the rationale used to establish the target cohort for subsequent analysis.

In 2011[3] there were a total of 134,217 students assessed as ESL Phase 1, 2, or 3. While all ESL students require support, they do not all require the same level of support. For the purposes of the current analysis, a conservative estimate was adopted of the number of students requiring substantial support for limited English language proficiency, as well as a method for selecting the cohort of students for the analysis. Further, because the proposed funding model and the current measure of disadvantaged LBOTE do not have a graduated process for including students requiring different levels of support (i.e. students are either 'in' or 'out'), it was necessary to consider how best to account for students who require different levels of support: an 'equivalent full-time support' count was necessary.

The funding arrangements in NSW DEC reflect that students in Phase 1 require the highest level of support, and that Phase 2 students also require significant support, receiving a funding weight that is slightly over half the weight for Phase 1 students. Phase 3 students, attracting only one-tenth the weighting of Phase 1 students, for the most part no longer need substantial support for limited ELP.

Using the NSW DEC funding arrangements as a guide, a conservative estimate of the limited ELP cohort (in 'equivalent' full-time support terms) would include 100% of ESL Phase 1 students and 50% of ESL Phase 2 students.

Support for this estimate is also provided by a NSW DEC analysis of the level of educational disadvantage associated with ELP, which was contained in the NSW DEC submission to the Gonski review. The analysis indicated that, given equivalent student family SES background, students assessed as ESL Phase 1 and 2 have significant educational disadvantage (corresponding to approximately 1.5 and 0.7 NAPLAN bands respectively) relative to ESL Phase 3 students and LBOTE students not requiring ESL support. On average, students who are assessed as having very limited English proficiency (ESL Phase 1) have twice the level of disadvantage as students assessed as ESL Phase 2.

With ESL Phase 2 students having approximately *half* the educational disadvantage as Phase 1 students and receiving slightly over *half* the funding weight received by Phase 1 students, therefore half of the Phase 2 students were included in the ELP cohort.

In conclusion, this analysis has conservatively estimated the size of the NSW government school cohort requiring substantial support for limited English language proficiency as 100% of ESL Phase 1 students and 50% of ESL Phase 2 students. This represents approximately 45,000 NSW government school students.

[3] As the main analysis for this report was performed in September 2012, it relied on data from 2011 NSW ESL Annual Survey because that was the only ESL data available at the time. However, it is not expected that the main patterns presented or conclusions made in this report would change if the more recent 2012 ESL data were used in the analysis.

2.3. Methodology to assess alignment of disadvantaged LBOTE and limited English proficiency

The analysis used the 2011 student background data for NSW government schools that was provided to ACARA for the 2012 MySchool website launch. The disadvantaged LBOTE measure was constructed similarly to ACARA's methodology (i.e. LBOTE students where parental school education level is Year 9 or below).

Student ESL phase assessments for 2011 were obtained from the 2011 DEC annual ESL survey. As the target cohort defined as "limited English proficiency" includes 50% of ESL Phase 2 students, it was necessary to randomly select half of the Phase 2 students for the analysis. To ensure robust estimates of the alignment between the two measures, 10 separate random samples were drawn of 50% of ESL Phase 2 student records. Each of these samples was combined with ESL Phase 1 student records to create 10 sets of students with "limited English proficiency".

Each set was then separately merged with the student background data and analysed to provide an estimate of the number of students who were both disadvantaged LBOTE and of limited English proficiency. Results reported represent the average across the 10 sets of analyses.

Results are presented in Venn diagrams to show the size of the respective groups and the extent of overlap between them.

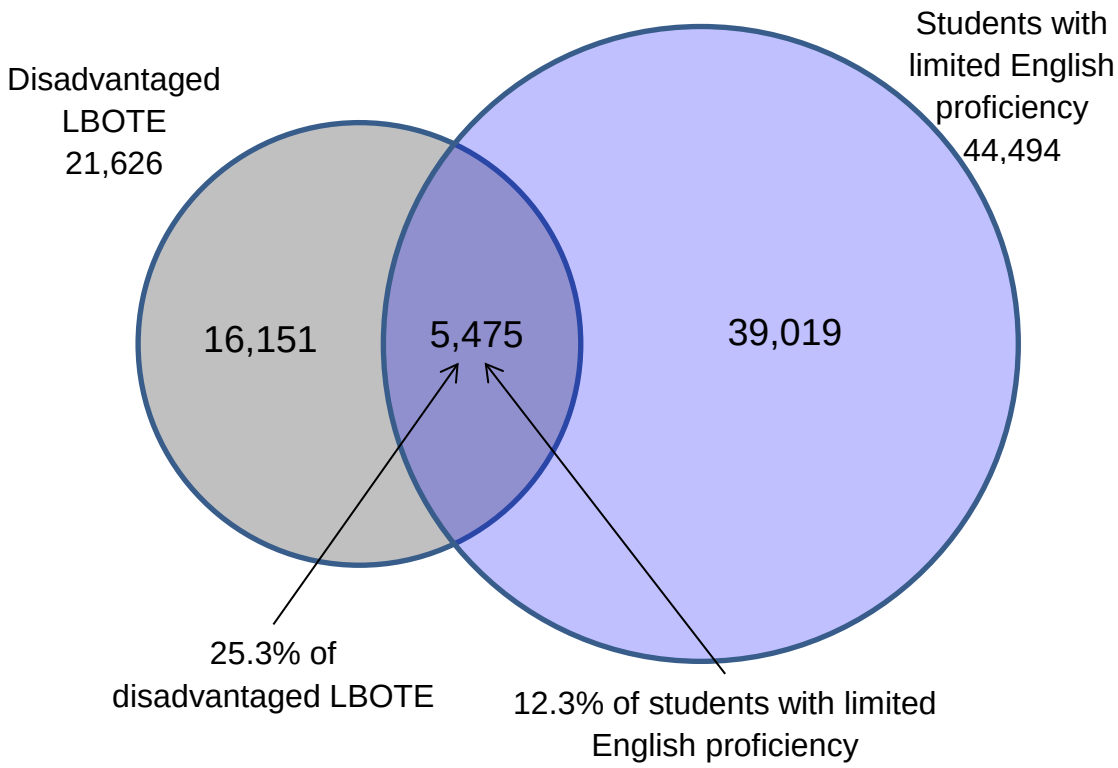
2.4. Extent of alignment between disadvantaged LBOTE and limited English proficiency

In 2011 in NSW government schools, 21,626 students were classified as disadvantaged LBOTE. This represents only half the number of students identified as having limited English proficiency (44, 494), providing immediate evidence that the disadvantaged LBOTE measure inadequately estimates the size of the cohort requiring support.

The disadvantaged LBOTE measure also lacks accuracy and coverage in identifying the right students requiring support. As shown in Figure 1, of the 21,626 students classified as disadvantaged LBOTE, only 5,475 (25%) actually have limited English language proficiency. In fact, the disadvantaged LBOTE measure only correctly identifies 12% of the total number of students identified as limited ELP.

Therefore, the disadvantaged LBOTE measure not only significantly underestimates the size of the cohort needing support but it also does not capture the right students.

Figure 1: Misalignment between disadvantaged LBOTE and limited English proficiency



Source: NSW DEC annual ESL survey 2011; 2011 student background data

The mismatch between students with limited English proficiency and those identified as disadvantaged LBOTE can also be seen in a cross-tabulation of student ESL status and parental school education level. Table 2 demonstrates that low parental education (the basis of the disadvantaged LBOTE measure) bears little relationship to a student's level of English proficiency.

Table 2: Misalignment between disadvantaged LBOTE and ESL status

		Parental School Education Level					
		Unknown year level	**Year 9**	**Year 10**	**Year 11**	**Year 12**	**TOTAL**
ESL status of all LBOTE students	Not required	14,250	6,523	12,547	4,522	50,440	88,282
	Phase 1	2,368	2,354	1,687	668	10,099	17,176
	Phase 2	7,113	6,287	6,473	2,643	32,120	54,636
	Phase 3	10,071	6,052	7,609	3,023	35,650	62,405
	Unknown/To be assessed	537	410	413	182	2,090	3,632
	TOTAL	34,339	21,626	28,729	11,038	130,399	226,131

Source: NSW DEC annual ESL survey 2011; 2011 student background data based on Parent 1 in the DEC student enrolment system

Table 2 shows that:

- The parents of Phase 1 and Phase 2 students do not necessarily have low levels of school education. In fact, more than 50% of parents of Phase 1 and Phase 2 students have high levels of school education, having completed Year 12 or equivalent.
- A corollary is that many disadvantaged LBOTE students either require no ESL support or are in ESL Phase 3 (12,575 in total, or 60% of students classified as disadvantaged LBOTE).

Therefore, parental education levels bear little relationship to students' English language proficiency.

2.5. Impact at the school level

The mismatch between disadvantaged LBOTE and students with limited ELP is even greater at the school level. Simulation analysis of the change in funding using disadvantaged LBOTE shows that 10% of all NSW government schools would receive funding even though they have no students with limited ELP. Further, almost 20% of schools (with students identified as limited ELP) would lose all funding, and a further 40% of schools would experience changes in excess of $10,000.

2.6. Summary

Disadvantaged LBOTE is in fact a measure of low socio-economic status (SES), not an indicator of limited English proficiency. Therefore the measure should not be used to assess students' eligibility for the ELP loading in the new schooling resource standard because it does not represent the cohort that actually requires support.

3. Improvements and Alternatives to Disadvantaged LBOTE

The following analysis explores possible improvements and alternatives to the disadvantaged LBOTE measure as a proxy for identifying students with limited English proficiency. While no measure is expected to achieve 100% identification of students with limited ELP, a measure which is 'good enough' would avoid the need to develop and implement a new national collection of English language proficiency which would be costly for all jurisdictions and impact on the workloads and resources of schools and teaching staff.

To be effective as a basis for distributing ELP funding equitably, however, a proxy measure would need to:

- capture at least 80% of the target cohort (coverage criterion),
- not include more than 20% of students who don't actually require support (accuracy criterion), and
- not produce a cohort that is significantly larger or smaller in size than the target cohort requiring support (size equivalence criterion).

3.1. Using data elements available nationally

At the national level, there are only 2 data elements relating to language background that are consistently available across all jurisdictions:

- Main language other than English spoken at home; and
- Country of birth.

(see *Data Standards Manual: Student Background Characteristics*, ACARA, October 2012)

The disadvantaged LBOTE measure is derived from the first measure, in conjunction with parental school education level. As discussed in the previous section, this results in a measure of SES rather than English language proficiency.

3.1.1 Combining LBOTE and country of birth

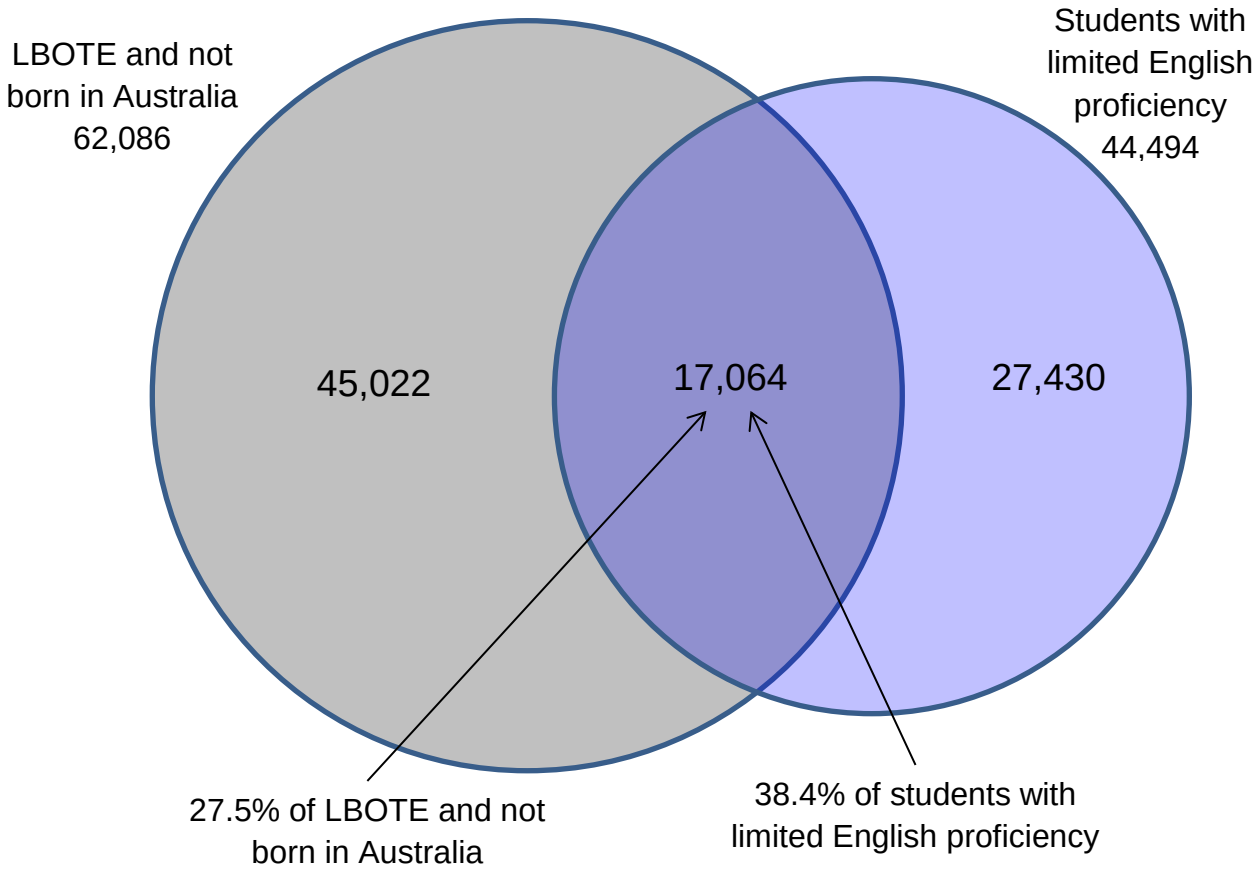
One alternative to the disadvantaged LBOTE measure is to define a measure using both nationally available measures, i.e. "LBOTE students not born in Australia". In NSW the majority of LBOTE students (73%) were born in Australia. It may be argued that students whose language background is not English and who were not born in Australia are more likely to require support for ELP than those born in Australia.

Using the same methodology as in the previous section, the size and overlap of the group of students defined as LBOTE and not born in Australia was analysed relative to those identified as having limited English proficiency.

In 2011 in NSW government schools 62,086 LBOTE students were born in a country other than Australia. This number is approximately 50% higher than the number of students with limited ELP hence this measure overstates the size of the target cohort.

This measure also lacks accuracy and coverage in identifying students needing support. As shown in Figure 2, only 27% of LBOTE students not born in Australia have limited English proficiency, and they cover only 38% of the target cohort.

Figure 2: Alignment between LBOTE students not born in Australia and students with limited English proficiency



Source: NSW DEC annual ESL survey 2011, 2011 student background data

Although this measure is an improvement on disadvantaged LBOTE which captured only 12% of the target cohort, it does not provide a sufficiently large enough improvement to warrant being recommended as the basis of the ELP loading in the new schooling resource standard.

3.1.2 Other alternatives

The measures of disadvantaged LBOTE and LBOTE students not born in Australia are based on *all* languages other than English and *all* countries of birth other than Australia. Alternative measures could be based on a reduced set of language backgrounds and/or countries of birth that are more likely to represent students requiring ELP support.

However, targeting specific languages or countries of birth is not recommended because:

a) Many languages/countries have too few students to provide statistically reliable results;
b) Defining limited English proficiency by language or country of birth could entrench cultural stereotypes;
c) Obtaining stakeholder agreement about which languages/countries of birth should be included in such a measure is likely to be fraught; and
d) While empirical research could suggest, at the system level, which languages/countries of birth are more related to limited English proficiency, there is likely to be huge variation at the school level, leading to inappropriate loadings for many schools.

In summary, the data elements available nationally are insufficient to derive proxy measures of English language proficiency with any accuracy.

3.2. Using other data elements available for NSW government schools in combination with disadvantaged LBOTE

The analysis further considered whether other data elements available for NSW government schools, in conjunction with disadvantaged LBOTE, would better identify students with limited ELP. If so, these data elements may be able to be developed and collected nationally more easily and efficiently than a direct measure of English proficiency. The data elements considered included:

- Length of time in an Australian school;
- Refugee student; and
- Students in the New Arrivals Program (NAP[4]).

[4] To be eligible for support under the NSW ESL New Arrivals Program a student must:
• speak a language other than English as their first language
• be in need of intensive ESL tuition
• be newly arrived in Australia (enrolling in school within six months of arrival or, for Kindergarten students, within 18 months of arrival)
• be enrolling in an Australian school for the first time or transferring within six months of arrival
• be an Australian citizen, a permanent resident, an approved provisional visa holder or an approved temporary visa holder with an Authority to Enrol form issued by the Temporary Residents Unit.

The following analyses demonstrate, however, that none of these data elements sufficiently improved the disadvantaged LBOTE measure.

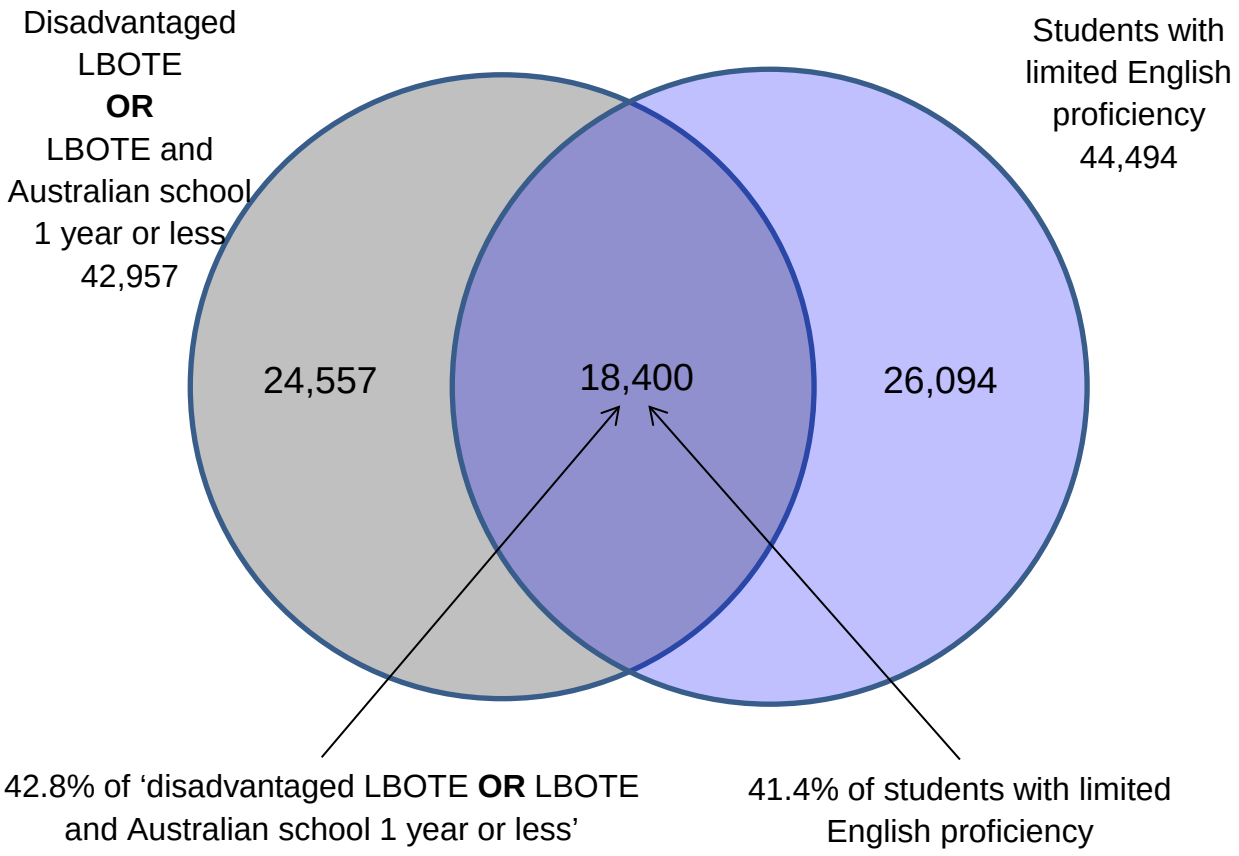
3.2.1 Combining disadvantaged LBOTE with length of time in an Australian school

The first analysis considered students who were disadvantaged LBOTE, or were LBOTE and had spent one year or less in an Australian school.

This measure increases the size of the cohort from 21,626 (disadvantaged LBOTE only) to 42,957 (either disadvantaged LBOTE, or LBOTE and time spent in an Australian school is one year or less). This cohort size is a close approximation of the size of the limited ELP cohort (see Figure 3), indicating that the new measure has improved size equivalence.

However, accuracy and coverage of the measure are still an issue. Although the revised measure includes a larger proportion of the target cohort (41.4%), it still captures less than half of all students needing support for limited ELP.

Figure 3: Disadvantaged LBOTE, or LBOTE and time spent in an Australian school is one year or less

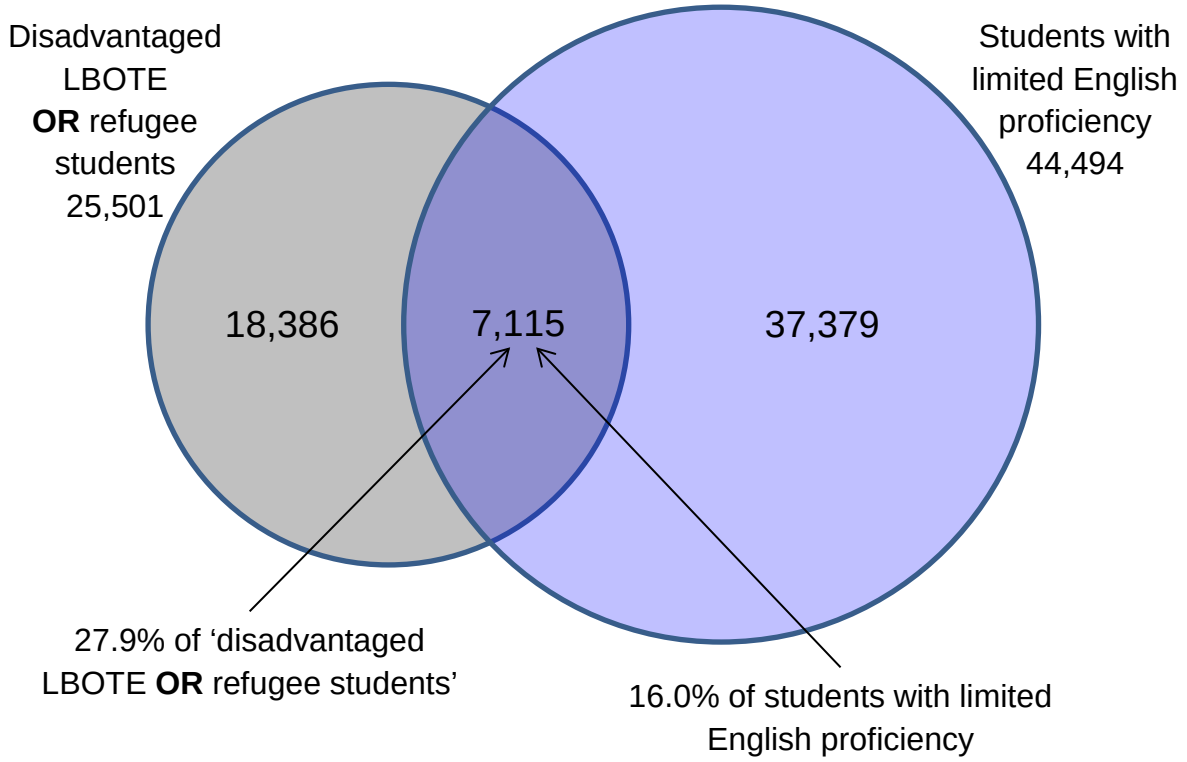


Source: NSW DEC annual ESL survey 2011; 2011 student background data

3.2.2 Combining disadvantaged LBOTE students with refugee students, or New Arrivals Program students

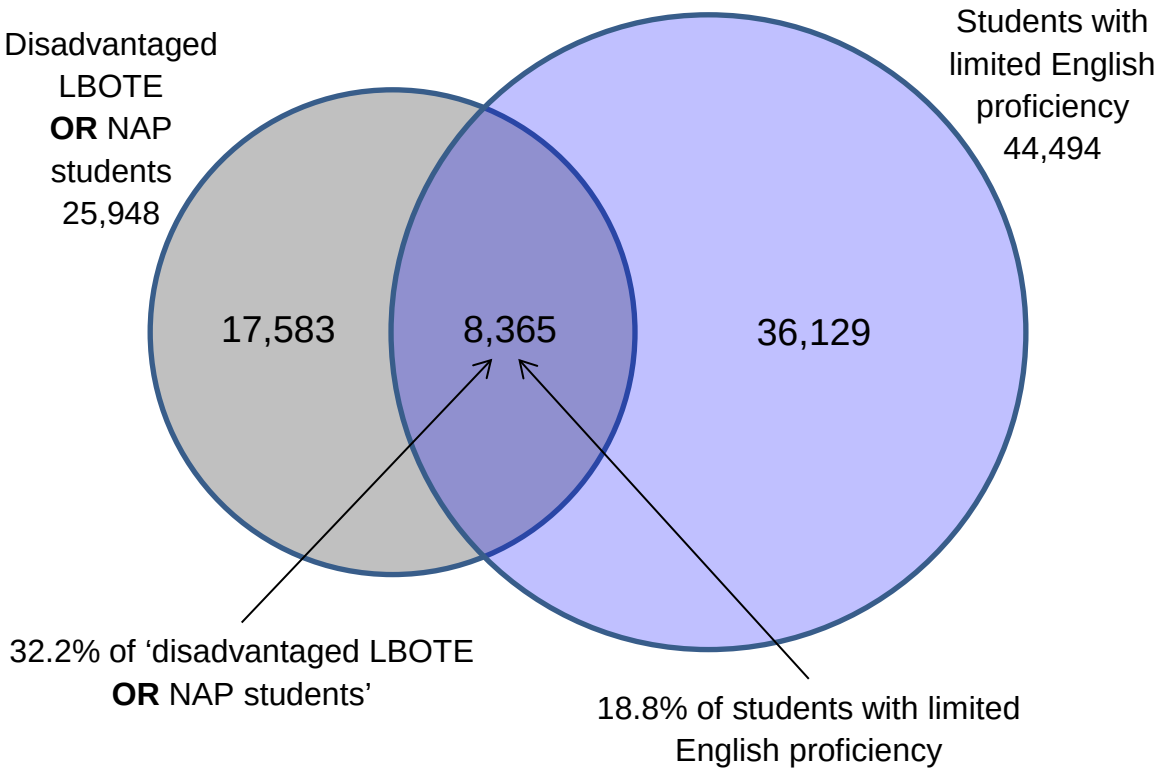
Including refugee students or those on the New Arrivals Program (NAP), also fails to improve the accuracy of disadvantaged LBOTE (see Figure 4 and Figure 5). They also fail to identify the majority of students who need support, capturing only 16–19% of the target cohort.

Figure 4: Disadvantaged LBOTE or refugee students



Source: NSW DEC annual ESL survey 2011; 2011 student background data

Figure 5: Disadvantaged LBOTE or NAP students



Source: NSW DEC annual ESL survey 2011; 2011 student background data

Furthermore neither of these alternatives adds much to the size equivalence aspect of the disadvantaged LBOTE measure as the size of the cohort needing support for limited ELP is still significantly underestimated.

In summary, there does not appear to be any other measure that can be combined with disadvantaged LBOTE which sufficiently improves the coverage, accuracy and/or size equivalence of the measure as a proxy measure for limited English proficiency.

3.3. Using other data elements available for NSW government schools

NSW DEC also investigated whether other existing measures, either singly or in combination, could identify students with limited ELP significantly better than the disadvantaged LBOTE measure.

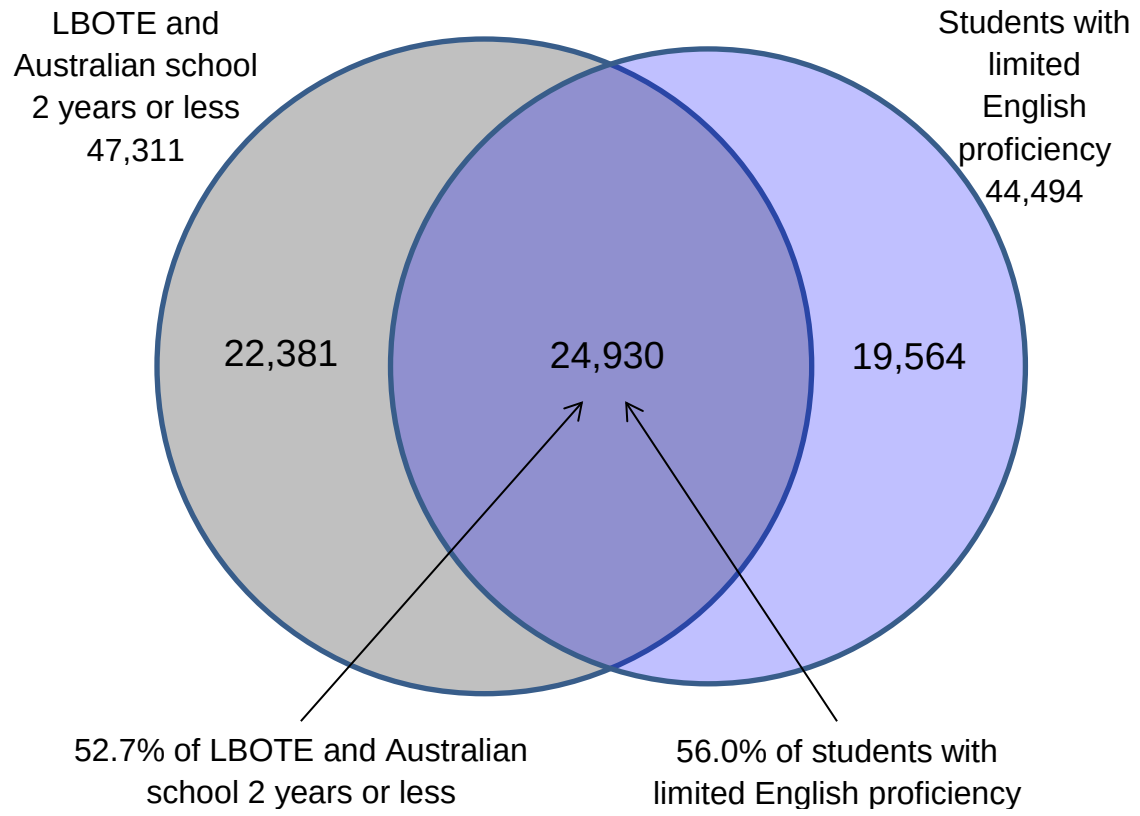
The following series of analyses show that while some measures offer somewhat improved accuracy, coverage or size equivalence, they do not reach the desired level of efficacy required to justify inclusion in the new schooling resource standard, or to warrant the time and cost necessary to develop them as nationally consistent measures.

3.3.1 Length of time in an Australian school

Counting LBOTE students who have spent 2 years or less in an Australian school gives a better measure than disadvantaged LBOTE as it captures more of the target cohort (56%), and is roughly similar in size to the target cohort (see Figure 6). However, capturing 56% of the target cohort is still considerably below the 80% required for adequacy as a proxy measure.

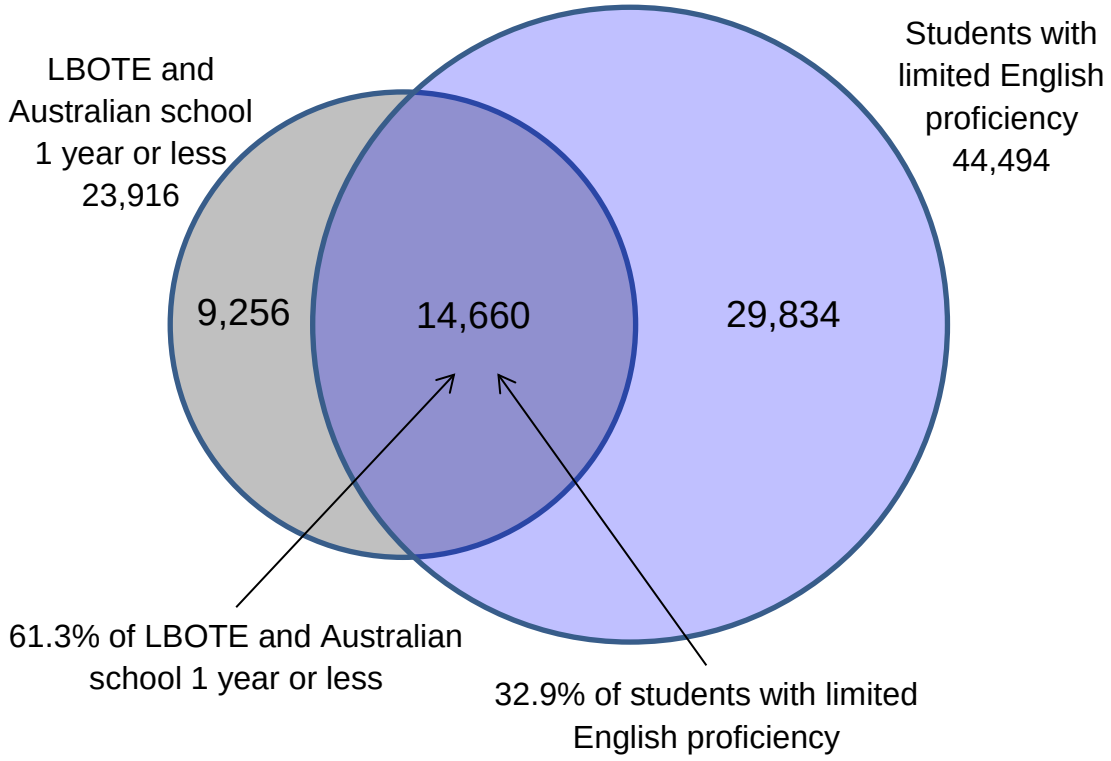
Reducing length of time in an Australian school to one year or less does not adequately improve the fit, as it underestimates the size of the cohort requiring support (see Figure 7).

Figure 6: Time spent in an Australian school is two years or less



Source: NSW DEC annual ESL survey 2011; 2011 student background data

Figure 7: Time spent in an Australian School is one year or less

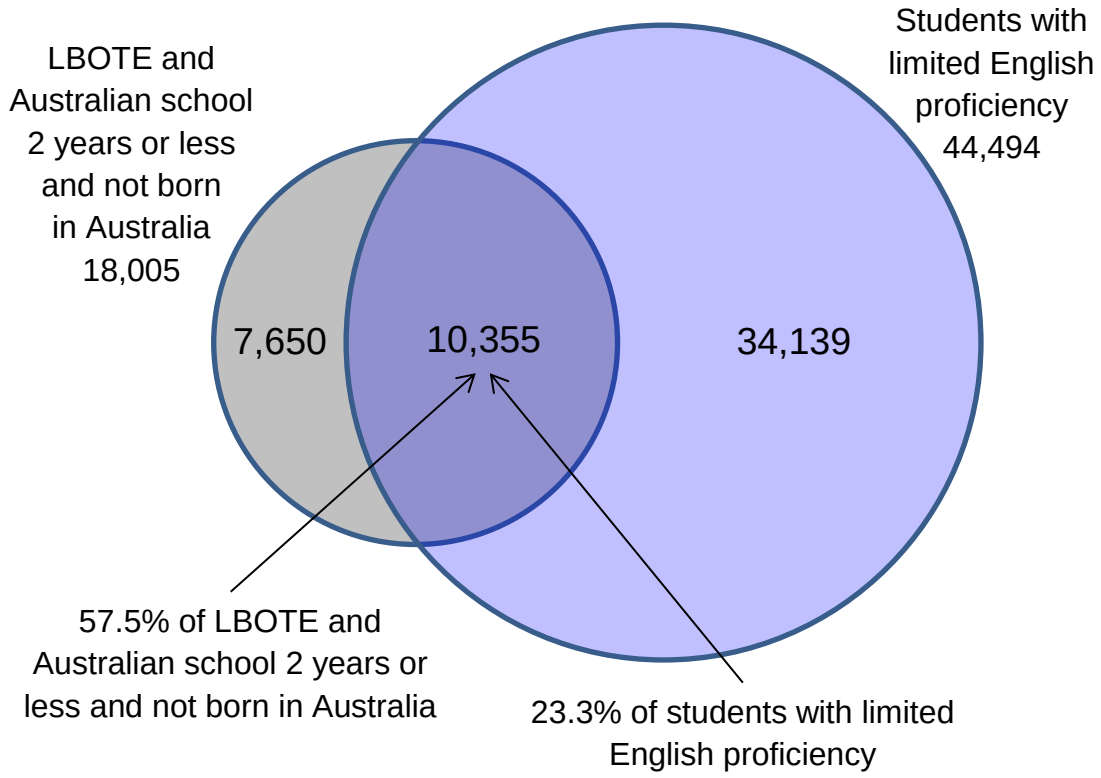


Source: NSW DEC annual ESL survey 2011; 2011 student background data

3.3.2 Length of time in an Australian school combined with not born in Australia

The next analysis considered LBOTE students who had spent 2 years or less in Australian schools AND were not born in Australia. While 57% of these students were also students identified as having limited ELP, this measure fails on both coverage and size equivalence aspects of proxy measure efficacy, identifying only 23% of the target cohort as well as significantly underestimating the size of cohort requiring support (see Figure 8).

Figure 8: Time spent in an Australian School is two years or less and not born in Australia

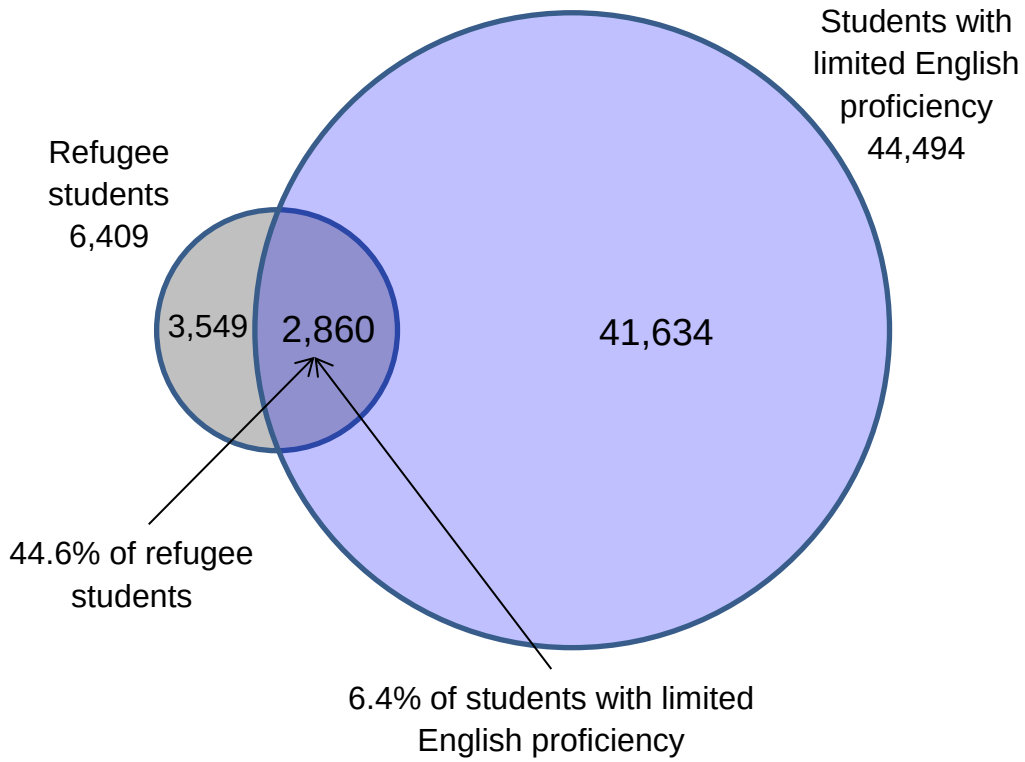


Source: NSW DEC annual ESL survey 2011; 2011 student background data

3.3.3 Refugee status, and/or NAP participation

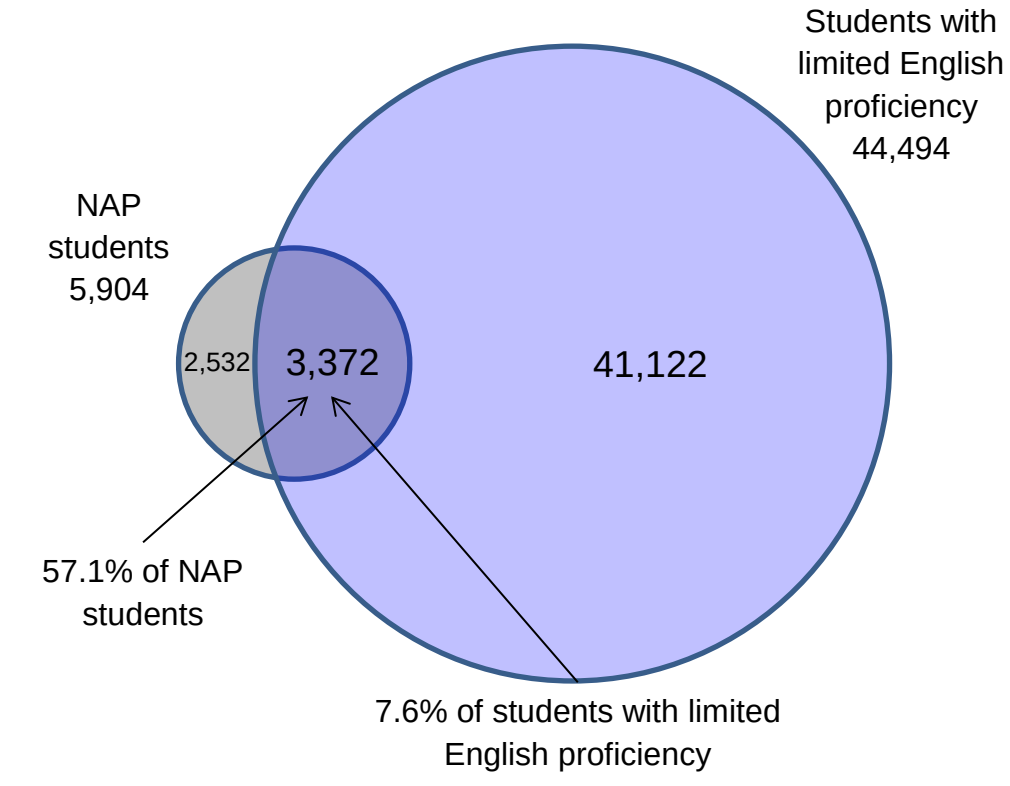
The following 3 analyses considered refugee status, New Arrivals Program (NAP) participation and a combination of the two (Figures 9-11). As all analyses indicate, these measures considerably underestimate the target cohort, and fail to identify a sufficient number of students identified as limited ELP.

Figure 9: Refugee students



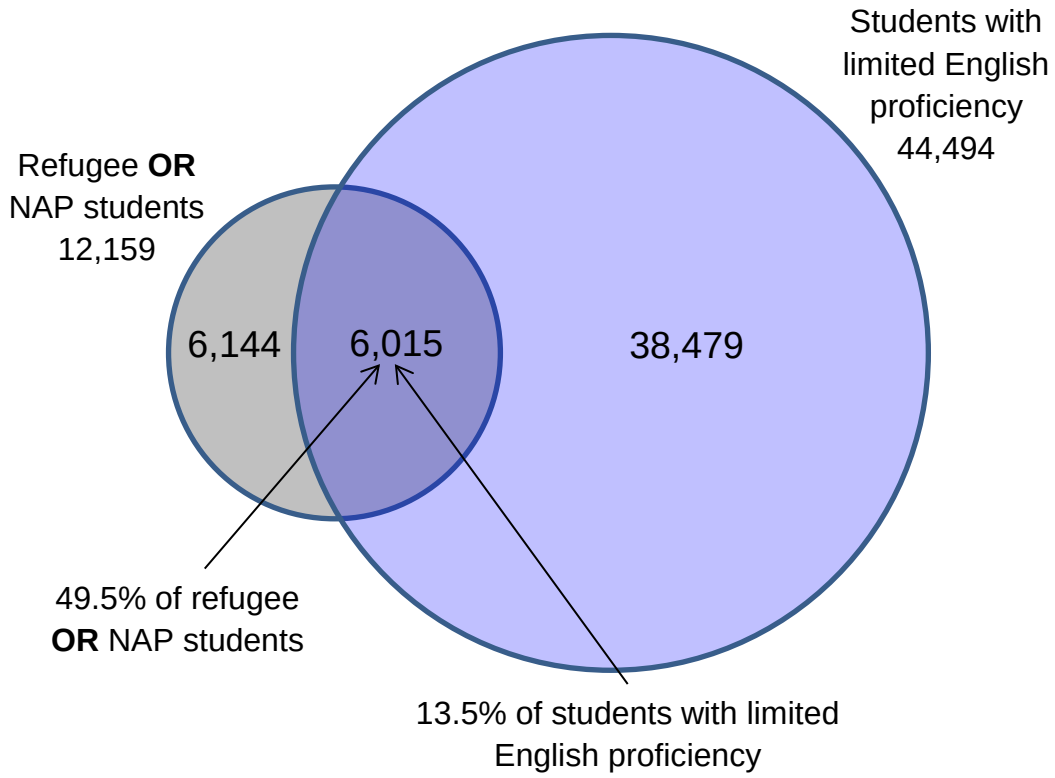
Source: NSW DEC annual ESL survey 2011; 2011 student background data

Figure 10: NAP students



Source: NSW DEC annual ESL survey 2011; 2011 student background data

Figure 11: Refugee or NAP students



Source: NSW DEC annual ESL survey 2011; 2011 student background data

3.4. Summary of improvements and alternative measures

A range of possible improvements and alternatives to disadvantaged LBOTE were analysed as potential proxy measures of students with limited English proficiency. Most displayed insufficient size equivalence, coverage and accuracy to be considered as appropriate proxy measures for inclusion in a new schooling resource standard as a loading for limited English proficiency. Table 3 provides a summary of the efficacy of the measures considered.

The measure that showed the greatest improvement was 'LBOTE and <=2 years in an Australian school'. This measure had good size equivalence (106%) along with reasonable, albeit insufficient, improvement in coverage (56%) and accuracy (52.7%). One other measure showed similar size equivalence, but less of an improvement in the other aspects: 'Disadvantaged LBOTE, **OR** LBOTE and <=1 year in Australian school'. However, both these measures included data elements not currently available at the national level (i.e., length of time in an Australian school).

Table 3: Summary of all measures considered

Group	Size of group	Size equivalence (group size as % of target cohort size)	Intersect (no. in group also in target cohort)	Coverage (% target cohort represented)	Accuracy (% of group accurately identified)
Target cohort: Students with limited English proficiency (44,494)					
Disadvantaged LBOTE *(Figure 1)*	21,626	49%	5,475	12.3%	25.3%
LBOTE and not born in Aust *(Figure 2)*	62,086	140%	17,064	38.4%	27.5%
Disadvantaged LBOTE, **OR** LBOTE and <=1 year in Aust school *(Figure 3)*	42,957	97%	18,400	41.4%	42.8%
Disadvantaged LBOTE, **OR** refugee students *(Figure 4)*	25,501	57%	7,115	16.0%	27.9%
Disadvantaged LBOTE **OR** NAP students *(Figure 5)*	25,948	58%	8,365	18.8%	32.2$
LBOTE and <=2 years in Aust school *(Figure 6)*	47,311	106%	24,930	56.0%	52.7%
LBOTE and <=1 year in Aust school *(Figure 7)*	23,916	54%	14,660	32.9%	61.3%
LBOTE and <=2 years in Aust school and not born in Aust *(Figure 8)*	18,005	41%	10,355	23.3%	57.5%
Refugee students *(Figure 9)*	6,409	14%	2,860	6.4%	44.6%
NAP students *(Figure 10)*	5,904	13%	3,372	7.6%	57.1%
Refugee **OR** NAP students *(Figure 11)*	12,159	27%	6,015	13.5%	49.5%

3.5. Conclusion

The NSW analysis has shown that the disadvantaged LBOTE measure fails as a proxy measure for limited English proficiency in three ways: insufficient size equivalence, poor coverage and lack of accuracy.

None of the other measures considered displayed sufficient improvement in all three aspects to be considered as appropriate proxy measures for inclusion in a new schooling resource measure as a loading for limited English proficiency.

It is noted that two of the measures considered in this analysis did show close size equivalence and some improvement to coverage and accuracy:

- 'LBOTE and <=2 years in an Australian school', or
- 'Disadvantaged LBOTE, **OR** LBOTE and <=1 year in Australian school',

However, both include a data element that is not currently available nationally (i.e., length of time in an Australian school). If either of these measures were to be considered as replacements for disadvantaged LBOTE, then the generalisability of the NSW results to other jurisdictions would first need to be examined for those jurisdictions that already hold this data element. If the results prove generalisable, then a nationally consistent data collection would need to be developed. These two steps are likely to require considerable time and cost, with the end result being a measure that still falls short in identifying the cohort of students requiring support for limited English proficiency.

Ultimately, a new and nationally consistent measure is needed that is both more accurate and has greater validity in targeting funding to limited ELP students.

4. NSW Trial of the EAL/D

4.1. Background

The *English as an Additional Language or Dialect (EAL/D) Learning Progression* was developed by ACARA in 2011, with input from content experts and academics, as part of a teacher resource package to assist with the development of new programs for the national curriculum.

The *Learning Progression* describes the development of English language typical of students learning English as an additional language or dialect (EAL/D). It includes broad descriptions of the characteristics of learner groups at each of four phases of English language learning (Beginning, Emerging, Developing and Consolidating) across each of the four language modes (listening, speaking, reading and writing).

As envisaged by ACARA, the purpose was to indicate the level of support needed for students to access learning area content. Teachers could use the progression to:

- understand the broad phases of English language learning that EAL/D students are likely to experience
- identify where their EAL/D students are located on the progression and the nature of their listening, speaking, reading and writing skills
- monitor the linguistic progression of their EAL/D students.

In 2012, NSW DEC investigated the reliability and validity of the EAL/D measure as a way of identifying students with limited English language proficiency. The remainder of this paper presents the findings of that analysis.

4.2. Methodology

The trial included 97 specialist ESL and class teachers with diverse prior ESL teaching and assessment background from 56 schools, including primary, secondary, central schools as well as Intensive English centres. Most of these schools were from metropolitan regions, with a few from provincial areas.

Table 4 provides information on the range of the teaching, training and qualifications background of the teachers, based on the responses provided by 74 of the 97 teachers involved.

Table 4: Background of teachers in the trial

	Background category	No.	%
All Teachers		74	
Current position in the school	ESL teacher	46	62%
	Class teacher	16	22%
	Executive teaching	9	12%
	Other specialist teacher	2	3%
	Executive non-teaching	1	1%
Training and qualifications	TESOL* or ESL pre-service training	49	66%
	TESOL* or ESL postgraduate qualification	38	51%

Note: Based on the responses provided by 74 of the 97 teachers to the evaluation survey
TESOL: Teachers of English to Speakers of Other Languages

A stratified sample of 944 students, across key target grades (Kindergarten, Years 3, 5, 7 and 9), gender groups, sub-demographic groups (i.e., Aboriginal, international student, refugee), and representing the range of English proficiency levels based on the current NSW ESL phase assessment tool, was selected by participating teachers for inclusion in the trial (see Table 5).

Table 5: Background of students in the trial

		No of students	% of all students
All Students		944	
Gender	Girls	452	47.9%
	Boys	492	52.1%
ESL Phases	Phase 1	345	36.5%
	Phase 2	369	39.1%
	Phase 3	201	21.3%
Grade	Kindergarten	150	15.9%
	Year 3	161	17.1%
	Year 5	161	17.1%
	Year 7	229	24.3%
	Year 9	233	24.7%
	Other Grades (Year 8, 10, 11)	10	1.1%
Aboriginal Students		29	3.1%
Intensive English Centre		168	17.8%
Refugee		197	20.9%
International Student		28	3.0%

Due to the small sample size of students in some demographic groups (e.g., Aboriginal or international students), the generalisability of results for these groups of students will need to be confirmed by a larger trial, preferably by a national trial involving students of more diverse cultural and linguistic background than those of NSW EAL/D students.

A detailed program of professional learning prepared teachers for the trial. A double-marking process was utilised: 639 of the total 944 students were assessed by two teachers on every language mode (listening, speaking, reading and writing) using the EAL/D instrument. Various types of inter-rater statistics, such as exact and adjacent agreement rates, Kappa rates and Dependability Index (score reliability coefficient) were examined to assess the reliability of teachers' judgements using the EAL/D instrument. Various aspects of construct validity of the instrument were also analysed.

Following the trial an online evaluation survey was conducted to collect and analyse teacher feedback on the utility of the instrument and the value and adequacy of support provided for teachers throughout the trial.

4.3. Results

4.3.1 Reliability of the EAL/D instrument

Reliability refers to consistency in teachers' judgements from one assessment scenario to another and the following types of reliability statistics were investigated:

- Exact agreement rate – the proportion of times two teachers agreed exactly
- Adjacent agreement rate – the proportion of times two teachers agreed within one phase (Beginning, Emerging, Developing and Consolidating)
- Kappa rates – Agreement rates adjusted by chance agreement
- Dependability Index – score reliability coefficient

Exact agreement rates

On average, 80% of the time two teachers' judgements on a mode of language for the same student matched exactly, using EAL/D (see Table 6).

Table 6: Exact agreement rates across modes and groups of students

Student Groups	Exact Agreement			
	Listening	Speaking	Reading	Writing
All Students	80.5%	78.7%	81.8%	82.4%
Girls	81.5%	80.5%	81.5%	83.6%
Boys	79.5%	77.0%	82.2%	81.3%
Aboriginal	100.0%	83.3%	88.9%	100.0%
ESL Phase 1	88.1%	84.4%	90.4%	87.2%
ESL Phase 2	78.7%	75.1%	81.5%	80.3%
ESL Phase 3	69.4%	75.5%	68.5%	76.4%
Kindergarten	84.4%	77.1%	86.3%	84.4%
Year 3	87.4%	72.7%	80.2%	88.2%
Year 5	85.2%	80.0%	80.9%	83.5%
Year 7	78.7%	81.3%	82.0%	82.7%
Year 9	71.5%	80.3%	80.9%	75.9%
Refugee Students	82.1%	78.8%	85.4%	88.1%
International Students	57.1%	71.4%	92.9%	71.4%
Intensive English Centre Students	89.7%	85.0%	96.3%	88.8%

Note: Total number of students included in the double-marking process: 639.

Teacher consistency does differ across groups of students and modes. The data indicates that teachers can assess reading and writing modes more reliably than they do listening and speaking, with speaking being identified consistently as the mode most difficult to assess reliably.

The data also demonstrates that the comparatively lower reliability associated with assessing speaking is more of a problem for assessing boys than for girls. There is no clear explanation for this difference between boys' and girls' results. Further analysis of data through a national trial with a larger cohort might shed light on the source of difference and/or confirm the generalisablity of these results.

In addition to the above, there is evidence that teachers could assess students of lower English proficiency levels more consistently than they could students of higher proficiency levels. Greater consistency in teachers' judgements when assessing students of limited English proficiency might have also contributed to the higher than average agreement rates observed for assessments concerning students at the Intensive English Centres (IEC) and refugee students, since these students are more likely to be of lower English proficiency than the rest of the student population in the sample.

A further factor contributing to the observed higher consistency in teachers' assessments of IEC students is the capacity of teachers in these centres to develop a close knowledge of their students' English language proficiency. IEC teachers have a detailed knowledge of their students because they work with smaller groups

of students for extended periods of the school day and they are required to assess and report student progress each term to determine class placement or high school readiness.

Adjacent agreement rates

Table 7 shows the average adjacent agreement rates (i.e. the proportion of time two teachers agreed within one phase) across all students and separately for different groups of students.

Table 7: Adjacent agreement rates across modes and student groups

Student Groups	Adjacent Agreement			
	Listening	Speaking	Reading	Writing
All Students	97.8%	97.9%	98.4%	99.0%
Girls	98.7%	98.0%	99.0%	100.0%
Boys	97.0%	97.9%	97.9%	98.2%
Aboriginal	100.0%	100.0%	100.0%	100.0%
ESL Phase 1	99.1%	98.6%	99.5%	100.0%
ESL Phase 2	97.2%	98.0%	98.0%	98.4%
ESL Phase 3	96.5%	96.5%	97.2%	98.6%
Kindergarten	97.9%	95.8%	98.9%	97.9%
Year 3	99.1%	98.2%	97.3%	99.1%
Year 5	94.8%	97.4%	96.5%	100.0%
Year 7	98.0%	98.0%	99.3%	98.7%
Year 9	98.7%	99.4%	99.4%	99.4%
Refugee Students	99.3%	98.7%	100.0%	99.3%
International Students	100.0%	100.0%	100.0%	100.0%
Intensive English Centre Students	100.0%	100.0%	100.0%	100.0%

Note: Total number of students included in the double-marking process: 639.

On average, the proportion of times when two teachers' EAL/D phase judgements for the same student on the same mode differs by more than one phase level varied from 1.0% when assessing writing to 2.2% when assessing listening. Further analysis is required to explore whether comparatively higher than average occurrences of significant variations in teachers' judgements produced for the kindergarten students' speaking and Year 5 students' listening phases were attributable to random factors or any systematic issues in the teachers' assessment processes.

Kappa rates

Acknowledging that two teachers can agree by chance alone, Kappa rates, which adjust for chance agreement, are provided in Table 8.

Table 8: Kappa rates across modes for selected student groups

Student Groups	Listening	Speaking	Reading	Writing
All Students	0.71	0.69	0.74	0.74
Kindergarten	0.75	0.64	0.76	0.71
Year 3	0.81	0.58	0.70	0.80
Year 5	0.78	0.68	0.71	0.75
Year 7	0.68	0.72	0.72	0.71
Year 9	0.54	0.70	0.70	0.61

Kappa is a measure of the difference between the observed agreement and the expected agreement by chance alone, standardised to be on a -1 to +1 scale. Using well-established criteria, the overall agreement rates suggest that teachers achieve a substantial level of agreement between each other, when using the EAL/D to make an assessment of a student's English language proficiency level.

Dependability Index

If the EAL/D is rolled out to schools for either resource or program planning, it is most likely that only one teacher (either a classroom teacher or an ESL teacher) will be employed to make judgements for one student. Table 9 shows that the reliability of teachers' judgements across all modes in this single-marking scenario reach the conventionally desired level of score reliability (i.e., 0.8) for high-stakes tests.

Table 9: Dependability Index for teachers' judgements using EAL/D

	Listening	Speaking	Reading	Writing
Dependability Index	0.84	0.84	0.86	0.86

Note: The Dependability Index is estimated based on a single-marking scenario

The Dependability Index for reading and writing aspects of language proficiency is around 0.86, slightly better than that for the conversational aspects of language (i.e., listening and speaking).

Differences in owner teacher judgement and non-owner teacher judgement

One relevant question is whether owner teachers (i.e., those who have greater opportunities for ongoing interaction with the students assessed) would be systematically more lenient or harsh in their judgements as compared to non-owner teachers (those who have limited ongoing interaction with the students and who would rely primarily on collected work samples to make assessments). If there is no evidence of significant difference, the tool may be used by teachers who might not be the owner teachers.

To investigate this, non-parametric tests were employed to examine the significance of the median difference between owner and non-owner teachers' judgements. The tests show that, for each language mode, the median difference is not statistically significant. While there are small differences in owner teacher and non-owner teacher judgements, there are no apparent patterns in these differences – that is, there is no evidence of owner teachers tending to assign either a higher or lower EAL/D phase for the same student than the non-owner teachers.

Summary

Results in this section have demonstrated that teachers can achieve a desirable level of consistency when using the EAL/D to make judgements on a student's language proficiency levels. Any occurrences of significant variations in teacher assessments are shown to be rare. Furthermore, there is no evidence of systemic differences in judgements made by teachers who have a direct teaching relationship with the student being assessed and those who don't. This finding suggests that the use of the EAL/D instrument in a school may not need to be limited to a particular group of teachers, so long as adequate training is provided.

4.3.2 Validity of the EAL/D tool

Validity is a multi-dimensional concept and the following tests were conducted to assess the four aspects of it:

• Discriminant validity	Can the four language modes of the EAL/D instrument be empirically discriminated by teachers? For example, is there any evidence of teachers' judgements on one mode being confounded by students' assessed performances on other modes?
• Concurrent validity	Is the relationship between assessments using the EAL/D instrument and those from other external constructs (e.g., NAPLAN) as expected?
• Measurement validity	Is there evidence demonstrating the measurement quality of the assessments? For example, is there evidence of the four modes measuring a single underlying ability (i.e. the English language proficiency of EAL/D students)? Are the four rating scales used to assess the four modes functioning as intended?

- Structural validity — Is the internal structure of the assessments produced by teachers using the EAL/D consistent with the expected interrelations among the different modes of language, as derived from the relevant second language acquisition theory or consistent with those from prior empirical studies?

Discriminant validity

Discriminant validity requires evidence that one teacher's judgement on one mode correlates more highly with another teacher's judgement of the same mode than it does with the alternate teacher's judgements on any other modes, for the same student. This is because different modes are intended to measure different aspects of the language proficiency, and there needs to be empirical evidence that teachers can use the EAL/D instrument to effectively discriminate the various conceptually distinct traits of the latent proficiency in an appropriate manner.

Table 10 provides a matrix of inter-correlations between two teachers' judgements, on the same mode and on different modes, for the same student. It shows there is a satisfactory level of discriminant validity evidence for the EAL/D instrument. This can be verified from the table, as each diagonal value is higher than any other values lying in its row or column. For example, Teacher 2's judgements on the listening mode correlated with Teacher 1's judgements on the same mode for the same students at 0.85. This correlation is higher than the correlations they have with Teacher 1's judgements on any other modes, for the same students (0.75, 0.70 and 0.68 for reading, speaking and writing, respectively).

Table 10: Inter-correlations between two teachers' judgements on the same or different modes

		Teacher 2 judgement			
	MODE	**LISTENING**	**READING**	**SPEAKING**	**WRITING**
Teacher 1 Judgement	**LISTENING**	.85**	.69**	.69**	.68**
	READING	.75**	.87**	.70**	.75**
	SPEAKING	.70**	.70**	.84**	.67**
	WRITING	.68**	.75**	.67**	.87**

Note: **Correlation is significant at the 0.01 level (2-tailed).

On the other hand, Table 10 provides additional evidence to support the conclusion made in the previous section – i.e., teachers are able to make similar judgements on the same aspect of language proficiency for the same students. The table shows strong correlations between two teachers' EAL/D phase judgements on the same mode, ranging from 0.84 for speaking, to 0.87 for reading and writing.

Concurrent validity

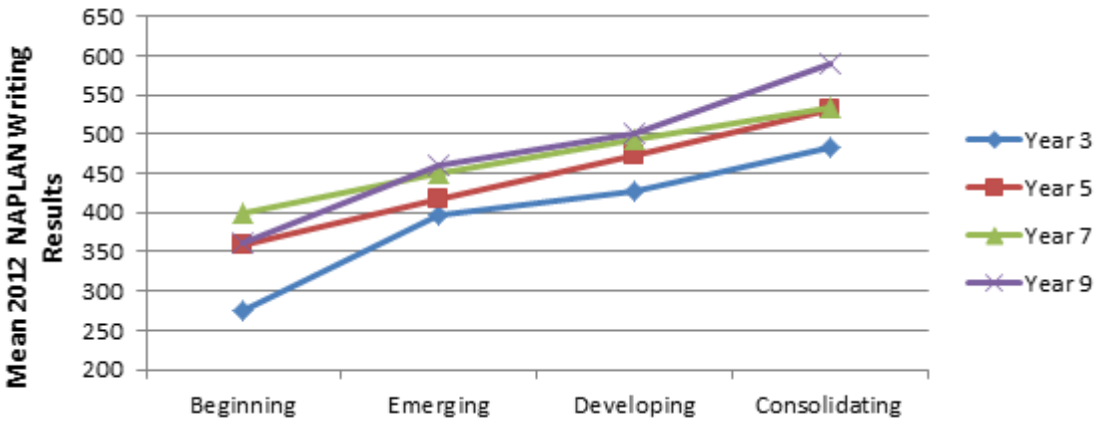
There is a strong relationship between the EAL/D and the NAPLAN reading and writing scores.

While NAPLAN and the EAL/D were designed for different purposes, the expectation is that assessments from these two instruments on reading or writing aspects of language should exhibit a reasonable level of concordance, particularly when the two sets of assessments were undertaken at a similar time (the trial data was collected in May/June, while the NAPLAN tests were administered in May) and when both sets of assessments were attempting to capture a similar aspect of language proficiency (either reading or writing) for the same students.

For this analysis, 90.4% of Years 3, 5, 7 and 9 students from this trial were able to be matched to the 2012 DEC NAPLAN dataset (i.e., 708 matched students for writing and 706 for reading). Of the students matched, 88 were exempted, 16 were absent and 3 were withdrawn from the NAPLAN reading tests. Similarly, 88 were exempted, 11 were absent and 3 were withdrawn from the NAPLAN writing tests.

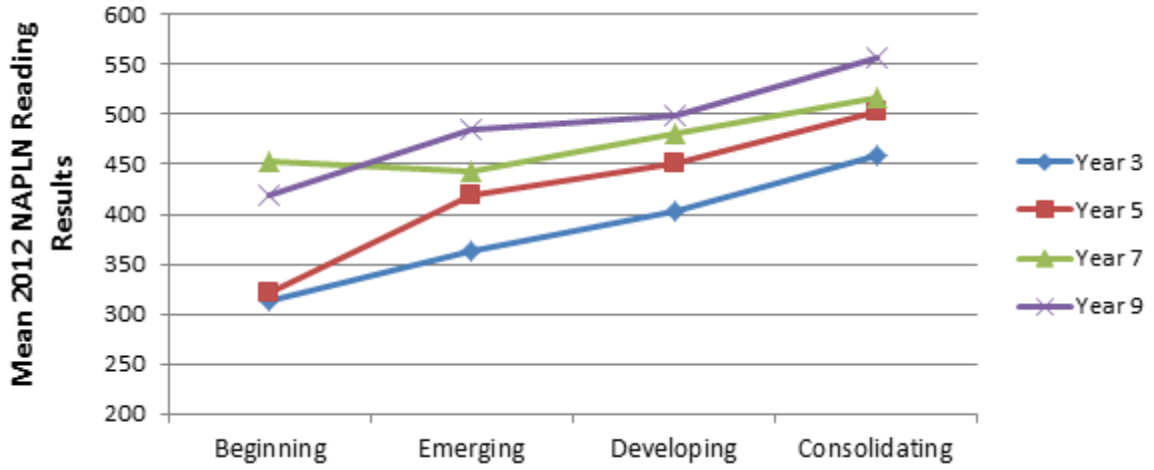
Figure 12 and Figure 13 depict the mean NAPLAN scores of students at each assessed EAL/D phases, for writing and reading separately.

Figure 12: Relationship between NAPLAN writing results and EAL/D writing phases



Notes: A total of 606 matched students were included in the analysis. For those students who were double-marked, the EAL/D phases used were those determined by the owner-teachers.

Figure 13: Relationship between NAPLAN reading results and EAL/D reading phases



Notes: A total of 599 matched students were included in the analysis. For those students who were double-marked, the EAL/D phases used were those determined by the owner-teachers.

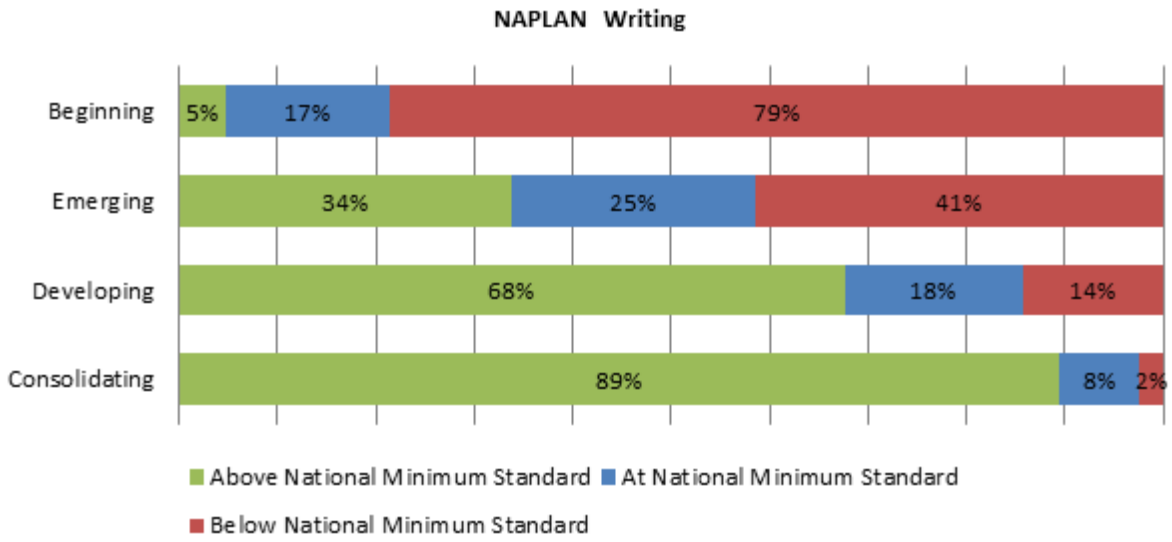
Since most of the exempt students are assessed at the lowest proficiency level, the mean NAPLAN scores reported here represent inflated estimates for those students at the Beginning EAL/D phase. Bearing that limitation in mind, Figure 12 and Figure 13 nonetheless show that the expected relationship between the EAL/D writing and reading phases and NAPLAN writing and reading scores are empirically confirmed. For each grade cohort, higher EAL/D phases are associated with higher mean NAPLAN scores, and lower EAL/D phases with lower mean NAPLAN scores, as expected.

However, for Year 7 reading results (Figure 13), the average NAPLAN reading score of students at the Emerging phase is slightly lower than that of students at the Beginning phase. This anomaly is most likely due to the exclusion of exempt students in the analysis.

To correct for this bias, the relationship between EAL/D and NAPLAN was re-examined using NAPLAN results as referenced to the National Minimum Standards. Consistent with national reporting rules, exempt students were coded as having achieved below National Minimum Standards.

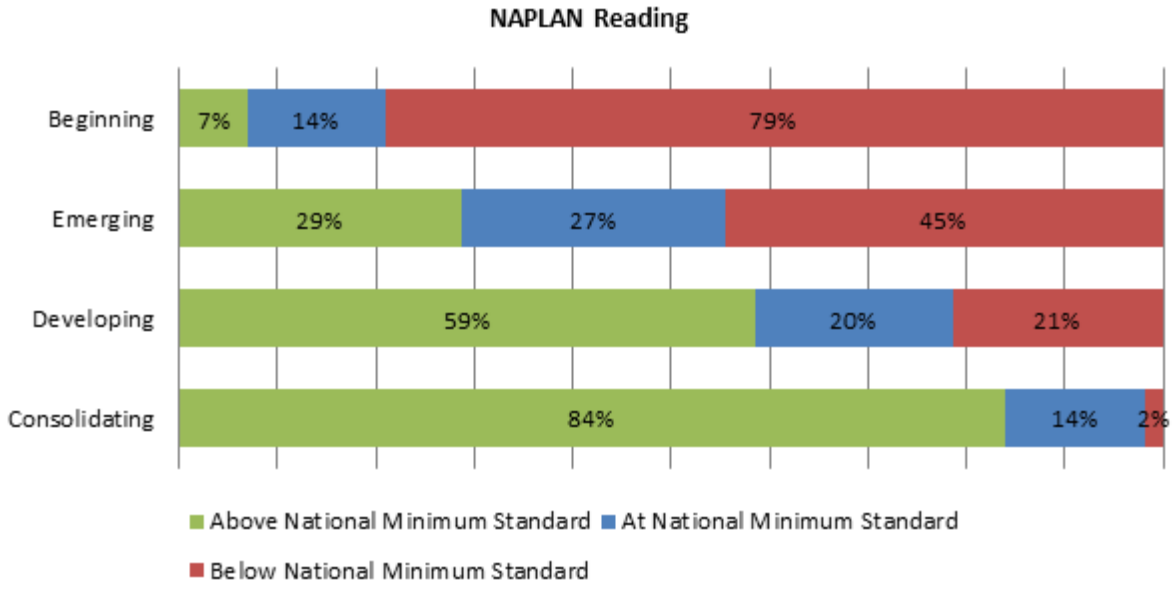
Figure 14 and Figure 15 report the proportions of matched Years 3, 5, 7 and 9 students at each EAL/D phase who are either below, at or above National Minimum Standards, for writing and reading separately.

Figure 14: Proportions of students at each EAL/D writing phase who are either below, at or above the National Minimum Standard



Notes: Exempt students included, and absent and withdrawn students are excluded from the analysis.

Figure 15: Proportions of students at each EAL/D reading phase who are either below, at or above the National Minimum Standard



Notes: Exempt students included, and absent and withdrawn students are excluded from the analysis.

Figure 14 and Figure 15 demonstrate that, as expected, students who were assessed at the Beginning EAL/D phase are most likely to achieve below the National Minimum Standard for their grades, and students who were at the Consolidating levels were mostly likely to achieve above the National Minimum Standard. For example, while nearly 80% of the students assessed at the Beginning phase on the reading mode were below the National Minimum Standard in the NAPLAN reading test, only 21% of the students at Developing and 2% of the students at the Consolidating phase were below the National Minimum Standard.

The extent of concordance between the NAPLAN scores and the EAL/D phases demonstrated from Figure 12 to Figure 15 provides evidence to support the intended interpretations of the four ordered EAL/D phases.

Measurement validity

The NSW trial investigated the following two questions to examine the measurement qualities of the assessments made by teachers using the EAL/D instrument:

1) *Are the four language modes measuring a single underlying ability?* and
2) *Are the four rating scales used to assess the four modes functioning as intended?*

If there is sufficient evidence of the four modes measuring a single ability, and of the rating categories on the scales being used meaningfully and as intended, the EAL/D assessments across the four modes can be summarised to a single score as an indicator of the student's overall English language proficiency level. This has important implications for the utility of the EAL/D instrument, as it is desirable to have one single assessment for each EAL/D student for the purpose of resource allocation.

The NSW trial used the Rasch Partial Credit Model to investigate both measurement questions. Evidence of the four modes measuring a single underlying ability was first collected from analysing the distributions of EAL/D mode assessments that were unexpected by a uni-dimensional model. Various statistical indicators such as INFIT and OUTFIT mean squares and their associated t-statistics were then examined for each language mode[5]. Results from these analyses confirm that teachers' mode judgements obtained from using the EAL/D instrument contribute to the development of a single ability continuum. There is no evidence of modes under-fitting the uni-dimensional measurement model, which would have suggested other significant sources for explaining variations in assessed performances across the modes. Nor was there any evidence of modes over-fitting the measurement model, which would have indicated that they were operating too similarly to each other, even though they were meant to measure different aspects of the latent ability.

[5] For detailed interpretations of the Rasch analysis, please see the full EAL/D report which is published at http://www.dec.nsw.gov.au/documents/15060385/15385042/Report26Februaryfinal.pdf

For the second measurement question, Table 11 (see page 33) reports the average ability measure for each EAL/D phase as well as the Rasch-Andrich thresholds, across the four modes, for each grade dataset separately.

For each grade, across all modes, the average person ability measures advance with the sequential categories on all rating scales, which is consistent with the intention that a higher category on a rating scale should indicate more of the underlying ability.

Further visual inspections of Category Probability Curves reveal that the Rasch-Andrich thresholds are spread across the latent continuum, and that they are neither too close nor too far apart. This means that collectively, all the phase categories help in defining distinct points on the latent construct being measured. Higher ability persons are more likely to score in a higher category than lower ability persons, across the continuum, as expected.

To summarise, there is sufficient evidence to indicate that the four rating scales are functioning properly, with a higher phase on a rating scale corresponding to a higher overall ability level, and vice versa. Taken together with the strong evidence of the four modes measuring a single underlying ability, it means that, for each student, the EAL/D assessments made by teachers across the four modes can be summarised into a single score, for the practical purpose of comparing a single ability.

Structural validity

This aspect of validity requires that the internal structure of EAL/D assessments is consistent with what is known about the structural relations inherent in the underlying construct.

Combining the previous Rasch analysis with Principal Component Analysis, the NSW DEC study found a contrast between the academic aspects of the language (reading and writing) and the conversational aspects of language (listening and speaking), with writing being consistently identified as the most difficult mode to develop or achieve, as compared to other modes.

Table 11: Average ability measures and Rasch-Andrich thresholds

Mode	EAL/D Level	Kindergarten		Year 3		Year 5		Year 7		Year 9	
		Average Ability Measure	Thresholds	Average Ability Measure	Thresholds	Average Ability Measure	Thresholds	Average Ability Measure	Thresholds	Average Ability Measure	Thresholds
LISTENING	1	-9.2	na	-10.28	na	-12.86	na	-8.82	na	-11.88	na
	2	-3.98	-8.19	-4.29	-8.21	-4.07	-9.73	-2.46	-6.76	-2.7	-10.35
	3	2.28	0.43	2.73	-0.85	3.6	0.83	1.86	0.87	3.56	2.42
	4	7.94	7.76	10.63	9.06	9.4	8.91	6.44	5.89	8.67	7.93
SPEAKING	1	-8.96	na	-11.38	na	-12.86	na	-9.01	na	-11.01	na
	2	-3.91	-7.06	-4.39	-8.61	-4.49	-9.81	-2.69	-7.06	-2.84	-9.09
	3	1.71	0.37	2.75	-1.06	3.68	0.11	1.49	1.15	3.14	1.76
	4	6.9	6.69	10.62	9.67	9.78	9.7	6.32	5.91	8.33	7.33
READING	1	-7.91	na	-10.9	na	-13.61	na	-8.12	na	-10.38	na
	2	-1.78	-4.99	-3.95	-10.04	-3.86	na	-1.94	-6.56	-0.99	-8.38
	3	4.28	4.99	4.03	-0.11	3.8	-3.97	2.66	-0.13	4.35	1.23
	4	8.9	na	11.25	10.16	9.56	3.97	8.52	6.69	9.39	7.15
WRITING	1	-7.24	na	-10.49	na	-12.63	na	-8.08	na	-10.97	na
	2	-0.56	-4.67	-1.99	-10.57	-2.37	-11.06	-1.23	-7.01	-1.03	-9.47
	3	6.6	4.67	6.34	0.23	5.22	1.58	2.9	0.45	4.51	1.84
	4	8.82	na	11.67	10.35	9.97	9.49	8.04	6.56	9.02	7.62

Notes: Levels 1, 2, 3, 4 correspond to four ordered EAL/D phases: Beginning, Emerging, Developing and Consolidating.

This finding is not completely surprising, given findings from other studies relating to EAL/D students' progression along different dimensions of the English language.

For instance, other studies suggested that academic English proficiency is more difficult to develop than conversational aspects of language for EAL/D students. These studies demonstrated that, while it could take students up to 3 years to develop conversational competency, it could take between 7 to 10 years for them to develop academic language competency.

Research into second language development also shows that ESL students tend to acquire the receptive language skills before productive language skills (listening before speaking and reading before writing). Even with the content of the EAL/D already taking into account the known differential patterns of development of EAL/D students in English language proficiency, the data from the NSW trial still showed that, consistent with these findings, writing proficiency is more difficult to develop than any other language modes.

The fact that the structural patterns in the EAL/D assessments are consistent with findings from other studies on EAL/D students' progression along the different aspects of English language ability can be seen as a supporting piece of evidence that the *EAL/D Learning Progression* is operating as intended.

Further evidence of concurrent validity

The study also examined the relationship between the current tool used in NSW government schools – ESL phases 1, 2, and 3 – and the EAL/D.

In order to provide further evidence of the alignment between the two measures, the average ability estimates of students in the three ESL phases were mapped on the EAL/D Progression continuum. Figure 16 illustrates the results for the primary student data (Years 3 and 5 combined) and Figure 17 for the secondary student data (Years 7 and 9 combined).

On both figures, the Rasch average ability measures, estimated from the EAL/D assessments, of the Phase 1, 2, and 3 students are identified on the continuum. Also identified on the same continuum are the ability estimates of students with various typical EAL/D proficiency profiles (e.g., from students who were assessed Beginning across all four modes, to those who were assessed Consolidating across all the four modes). This makes it easier to understand the language developmental profiles of an average ESL Phase 1, 2 or 3 student, relative to major developmental milestones on the *EAL/D Learning Progression* continuum.

Figure 16 and Figure 17 demonstrate that higher ESL phases, on average, are associated with higher overall language proficiency levels on the EAL/D, which offers another indication that the EAL/D is operating as intended.

Figure 16: Average locations of the Phase 1, 2 and 3 students on EAL/D – Years 3 and 5

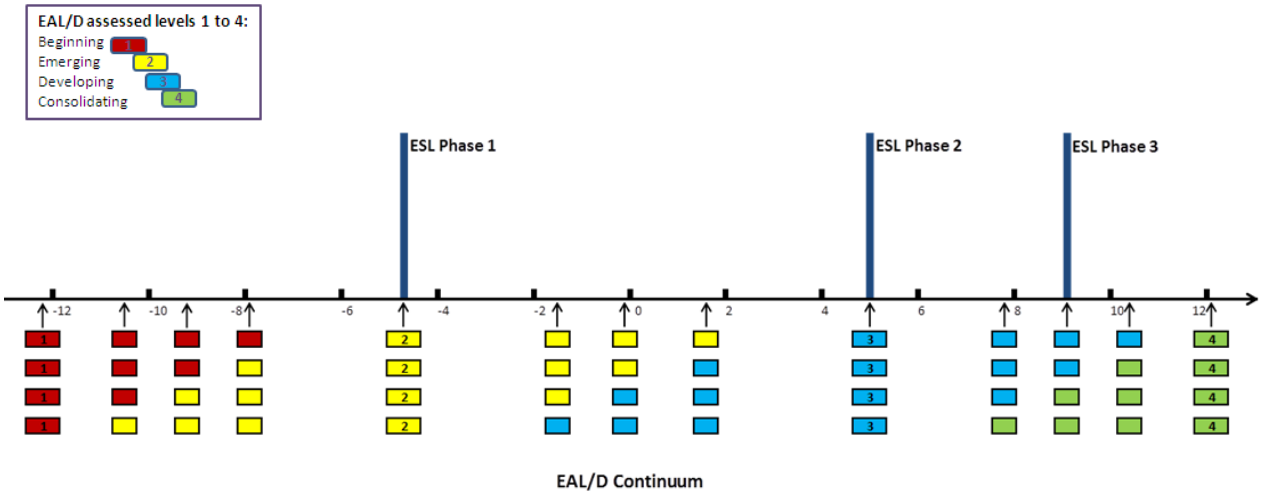
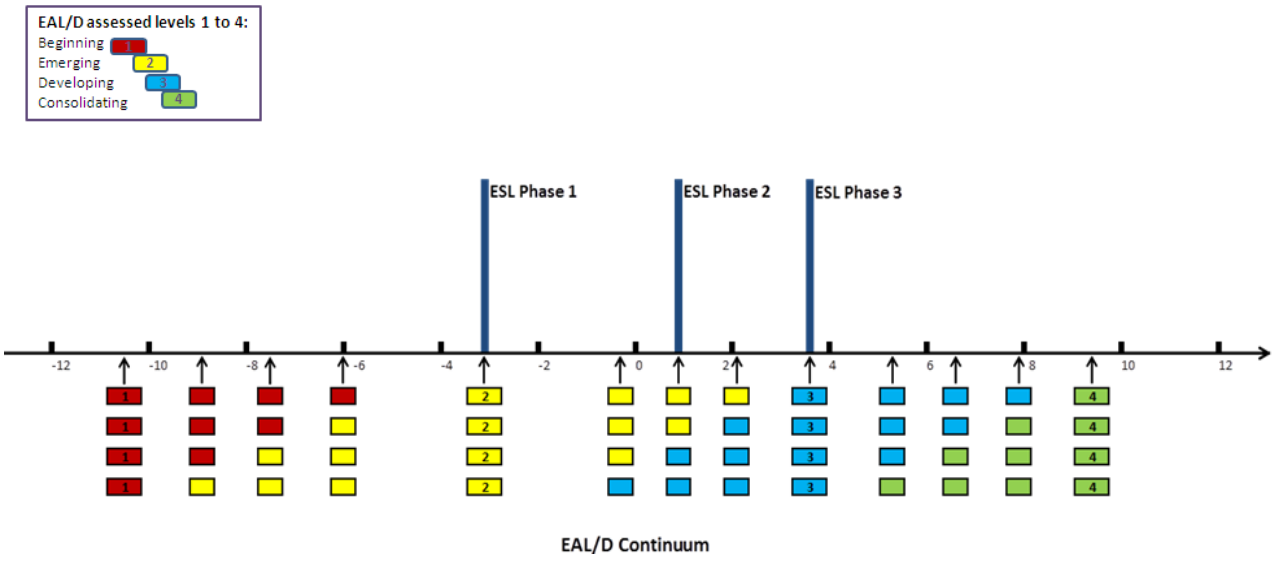


Figure 17: Average locations of the Phase 1, 2 and 3 students on EAL/D – Years 7 and 9



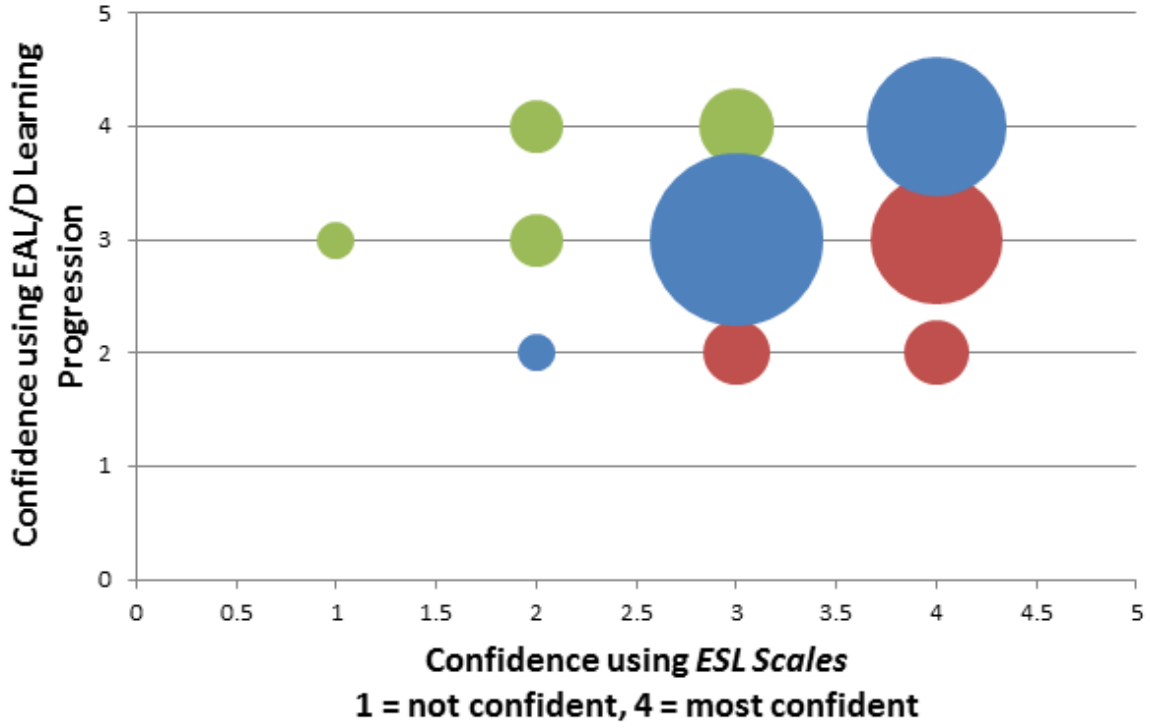
4.3.3 Teacher response to the EAL/D tool

A teacher feedback survey and focus group meetings provided quantitative and qualitative feedback on the utility of EAL/D, particularly compared with the existing ESL Phases tool. A total of 74 teachers (77% of the trial participants) responded to the survey.

The trial compared teacher confidence using the two tools and the results (shown in a bubble chart in Figure 18) indicate that most teachers had similar levels of confidence in using both the existing measure and the EAL/D. The confidence in using the existing measure is to be expected as it has been used in NSW since

1996. The equivalent level of confidence in using the EAL/D indicates how accessible this tool is after only two days' professional learning.

Figure 18: Confidence in using the *EAL/D Learning Progression* in relation to confidence using the ESL Scales

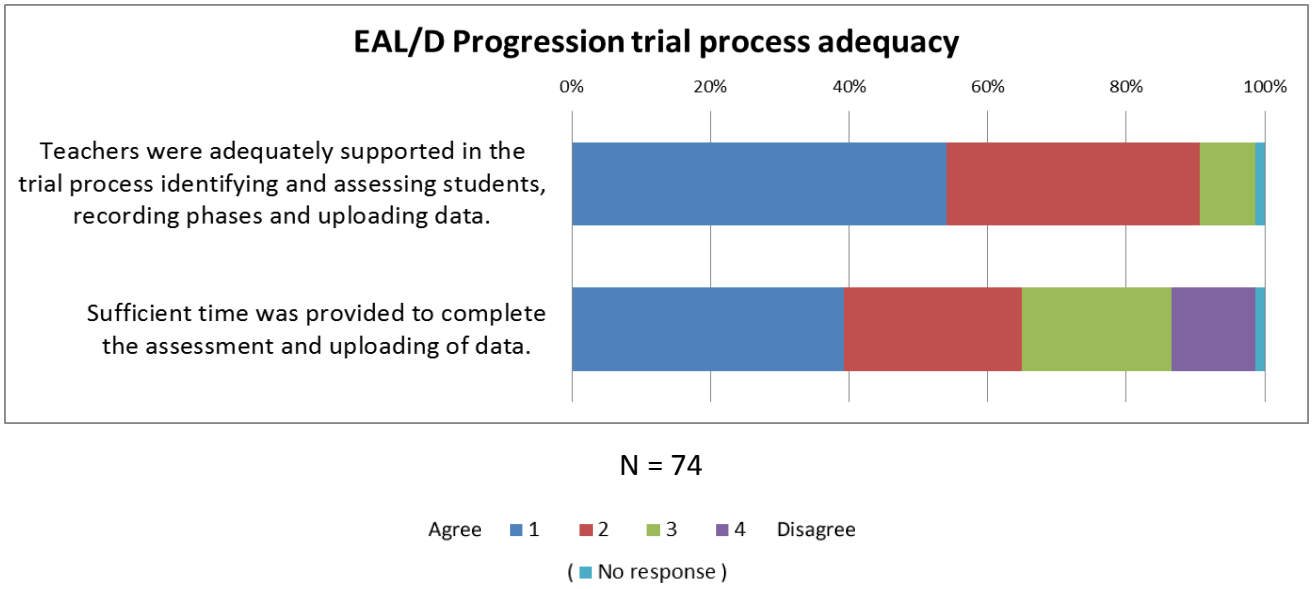


Notes: The size of the bubbles is proportional to the number of teachers in each category.
Only teachers who have used ESL scales are included in this analysis.
Red bubbles = teachers indicating greater confidence with ESL scales than EAL/D
Green bubbles = teachers indicating greater confidence with EAL/D than ESL scales
Blue bubbles = teachers indicating equal confidence with ESL scales and EAL/D

Teachers specifically supported the use of the EAL/D to replace the current ESL Phases for identifying a broad phase of English language proficiency, with 66% of respondents indicating their support.

Generally, teachers found the training program comprehensive and well-organised. The survey data showed that around 90% of teachers claimed they were adequately prepared for identifying and assessing students, recording phases and uploading data (see Figure 19). However only 65% of teachers agreed that they had sufficient time to complete assessments and upload data.

Figure 19: Teacher feedback on EAL/D Progression trial process



The feedback and survey data indicated the need for a program of mixed delivery mode for teacher professional learning that provides opportunity for professional dialogue and collaboration.

4.3.4 Summary of NSW trial

In conclusion, the NSW trial has provided sufficient reliability and validity evidence for EAL/D to be used in NSW government schools as a broad measure of ELP for resource allocation. In addition, the NSW trial has national implications in terms of informing the design and the associated cost-benefit analysis of a prospective national trial. Such a national trial would provide recommendations about the potential use of the EAL/D to report ELP across jurisdictions.

5. Conclusion

The aim of the current project was to determine 'what short-term improvements could be made to the disadvantaged LBOTE measure to improve its accuracy.'

The NSW DEC analysis found that disadvantaged LBOTE should not be used to assess students eligible for the ELP loading because it:

a) did not identify the right students, and,

b) bore little relationship to the size of the cohort needing support.

NSW DEC investigated other options to better measure low ELP, either by augmenting the disadvantaged LBOTE measure or replacing it. However, none of the options investigated reached the level of accuracy, coverage or size equivalence required to justify inclusion in the new schooling resource standard, or to warrant the time and cost necessary to develop them as nationally consistent measures.

Trying to improve the disadvantaged LBOTE measure, or find alternatives to it, leads to the conclusion that a new national measure is needed to accurately target funding to low ELP students.

The analysis also presented the results of a NSW DEC trial into the *EAL/D Learning Progression* which found that EAL/D had high levels of reliability and validity, as well as acceptance by teachers, and was superior to the current, internal NSW DEC measure of ELP, known as ESL Phase 1, 2 and 3, which was itself superior to the disadvantaged LBOTE measure. The full report of the NSW trial of the *EAL/D Learning Progression* is available at: http://www.dec.nsw.gov.au/documents/15060385/15385042/Report26Februaryfinal.pdf

In conclusion, a new, nationally consistent measure is needed that is both more accurate and has greater validity than disadvantaged LBOTE. This study has shown that the EAL/D measure is a clear candidate for that role. A second project conducted by NSW DEC will present a cost-benefit analysis of developing a nationally consistent ELP measure, focused on EAL/D.

6. References

Australian Curriculum, Assessment and Reporting Authority (2012). English as an additional language or dialect, Teacher Resource - Overview and EAL/D Learning Progression. ACARA, Sydney.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105.

Cummins, J. (1984). *Bilingualism and Special Education: Issues in Assessment and Pedagogy*. Clevedon: Multilingual Matters.

Garcia, G (2000). *Lessons from research: What is the length of time it takes limited English proficient students to acquire English and succeed in an all-English classroom?* NCBE Issue Brief No 5. national clearinghouse for bilingual education, September 2000. http://www.ncela.gwu.edu/pubs/issuebriefs/ib5.htm

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.), (pp.17–64). Westport, CT: American Council on Education.

Messick, S. (1996). Validity of performance assessments. In G. Phillips (Ed.), *Technical issues in large-scale performance assessment* (pp.1–18). Washington, DC: National Centre for Educational Statistics.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago.

Viera, A., & Garrett, J. M. (2005). Understanding interobserver agreement: The Kappa statistic. *Family Medicine, 37*(5), 360–363.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.