

# Chapter 16

## Sources of Evidence on Student Achievement in Northern Territory Bilingual Education Programs

Brian Devlin

### Introduction

The purpose of this chapter is make some observations about how academic achievement has been measured in the Northern Territory (NT), both in relation to schools with bilingual programs, and more generally. Four particular measures of progress have been taken into account: Research projects, critical reviews of the relevant literature, external test results and accreditation reports. The questions I have posed are: What information of value have these sources of information contributed? Do they tend to support or refute claims made about the effectiveness of programs, and the academic performance of students? To maintain focus I have drawn attention to two claims about comparative student outcomes. I have avoided reference to the international research on bilingual education as summarised in Grimes (2009) and elsewhere, or the 1200 or more international case studies mentioned by Lo Bianco (2010). As valuable as that body of research on bilingual education might be, it is not the focus of this paper. Only selected research findings concerning Northern Territory (NT) programs are considered here. The reason for this is that such results have, from time to time, been ignored, denied or misinterpreted by decision makers in the NT.

#### **Brief background—The external testing program in the NT (1983–2007)**

The shift to standardised testing was slow to permeate practice in the NT. For example Collins and Lea (1999) report a lack of monitoring, reporting and record keeping in the education system. When I was a school principal in the mid 1980s, I remember flying quite regularly into Darwin from Elcho Island to attend meetings of the Primary Assessment Program (PAP) Committee, which was coordinating the development of tests that could be used across the Northern Territory. The aim of

---

B. Devlin (✉)  
Charles Darwin University, Casuarina, Australia  
e-mail: brian.devlin@cdu.edu.au

these teacher-developed assessment instruments was to provide confidential, moderated, systemic feedback to the Northern Territory Department of Education (NT DOE) and equally confidential feedback to particular schools to inform staff and parents. PAP test results were not to be revealed publicly. That was our clear understanding at the time. Cataldi and Partington (1998) observed that in Lajamanu the PAT reading tests were administered along with Marie Clay's Stones reading readiness and the Gap reading test.

The PAP tests were replaced by a Multilevel Assessment Program (MAP) test battery, which was a more sophisticated set of measures based on item banks and modern mathematical understandings. These tests were administered in August each year to students in Years 3, 5 and 7. The aim was to compare student achievement levels against nationally agreed benchmarks. Each year the NT Board of Studies would then prepare a report to the Education Minister on the results obtained. At different times politicians would also discuss these results in the Legislative Assembly; for example, on August 24, 2005 the Education Minister reported that

the number of indigenous students achieving the Year 3 national reading benchmark improved 31% since 2001, the Year 5 benchmark improving by 43%, and the Year 7 figures improving by 46%. The Year 3 numeracy benchmark has improved by 10%, Year 5 by 38% and Year 7 by 43%.

So, in the years leading up to 2008 there was a sense in the NT that progress was being made. Students in bilingual programs had been compared to those in English-only programs at different times.

From 2008 onwards much of the discussion about student achievement has been framed with reference to the *National Assessment Program—Literacy and Numeracy*, generally referred to as NAPLAN or NAPLaN (this volume). For example, in a report commissioned by the NT Government in 2013, the performance in students at five schools—Yuendumu, Lajamanu, Milingimbi, Maningrida and Yirrkala—was tracked from 2008 to 2013 using those national test results (Wilson, 2014, Appendix 7, p. 281).

### **Rationale**

The argument put forward in this chapter is this: Although there is some evidence available to help us determine whether bilingual programs in remote Aboriginal schools have ever been successful, effective, or of value to local people, these findings have often been ignored, or drowned out by ideological disagreements, especially for a few years after December 1, 1998, when the Northern Territory Government announced that it would be redirecting funding away from bilingual programs and then again after October 2008, when a later NT government ceased its support for the step-model of bilingual education in seven remote schools. The 'noise' associated with the resulting polarised debates has made it difficult to hear what the researchers have had to say.

It is always reasonable to ask whether a government-funded program has been worthwhile, especially if it has continued in one form or another for more than four decades, as bilingual programs have in locations such as Areyonga, Galiwin'ku and

Yirrkala. How we might establish a program's worth is a task requiring careful evaluation and, of course, there are a number of ways that might be done. For example, we could ascertain whether the educational program has been giving local people the skills and knowledge they need to help them realise their hopes for a better life (Kral, this volume). As part of that, we might want to know whether a particular bilingual education program has assisted with the introduction of modern scientific and technical knowledge in a remote Aboriginal community, and whether it has contributed to young people finding meaningful work in the modern sector as skilled entrepreneurs or employees. Questions such as these relate to program *impact*, and are necessarily complex, for their answers depend on some sensitive, longitudinal, socioeconomic research and an ability to make sound, on-balance judgements and connections. Alternatively, we could settle on an easier method, one that simply measures a few selected aspects of student performance on some (hopefully) valid and reliable tests. This is the *output*-focussed approach, which governments have generally adopted. When we seek to assess program impact, we are asking a question about effectiveness: "How has this program contributed to society more generally?" When we gauge program outputs, we are focussed on finding out what a program has achieved for a particular institution or agency. When our aim is to compare outputs across institutions the results invariably have political, funding-related ramifications.

My focus in this paper is not on the broader, economic and social impact of bilingual education, although that might be an important and interesting topic to explore (Disbray and Devlin 2017). Nor is it concerned with other outputs that have been compared and measured, such as student attendance. What I have set out to do is to single out one aim of bilingual education and to ask: Are there any available data that would tell us whether, at different times, in different locations, that particular aim has been achieved? My rationale is that it is more useful to consider the research findings concerning one of Program X's aims at Time Y than to join in the never-ending debate about whether Abstraction A, freely defined, is a better approach than Abstraction B, which only encourages disagreements about untestable generalisations or myths (Nicholls 1999).

Which aim to choose though? Since its inception in 1973, the NT Bilingual Program has had eight aims. As McKay (this volume) has shown, these eight aims were changed and reprioritised in 1980, when a formal evaluation program commenced. In 1975 the first two aims were to help children to believe in themselves and to feel proud of their heritage through "the regular use" of L1 in school and "learning about Aboriginal culture" and, secondly, to teach "each student how to read and write in his own language" (Australian Department of Education 1975, p. 1). These aims mirrored Whitlam's idealistic vision—where bilingual education would encourage "greater respect for aboriginal languages and culture" through programs that "when fully implemented ... will affect most aboriginal children in the Northern Territory and will be extended to tribal areas of northern Queensland, the Kimberleys in Western Australia and northern South Australia" (Whitlam 1972). Five years later the eight aims had been rewritten and reordered to ensure a sharper focus on proficiency on English literacy and numeracy. Now in first place

was the aim “to develop competency in reading and writing in English and in number to the level required on leaving school to function without disadvantage in the wider Australian community” (Northern Territory Department of Education 1980, p. 2; McKay, this volume).

The reason for that change in wording was that program evaluators wanted some more precise, testable objectives to measure when appraising the performance of students in remote Aboriginal schools with bilingual programs as part of the accreditation exercise that had been planned since 1979 (Devlin 1995). The focus of this paper is on identifying sources of evidence on achievement of the first (revised) aim (Northern Territory Department of Education 1980, p. 2). That task is taken up in a later section of this paper, “Two claims regarding bilingually educated students’ achievement in literacy and numeracy”. The following section deals with a few preliminary ideas that will help make sense of what follows.

## Which Perspective Counts When We Measure a Program’s Success?

What criteria should we use when making judgements about a program’s ‘effectiveness’ or its ‘value to parents and students’? Whose perspective counts as important? Some assume that it is the government’s viewpoint that should prevail. After all, the Australian government had made a considerable investment in setting up ambitious bilingual programs during the 1973–78 period, and the Northern Territory Government inherited the expense of keeping them going after July 1, 1978 when the *Northern Territory (Self Government) Act* came into operation. So, for elected parliamentary representatives, paid government officials, and for the voting public more generally, it was entirely appropriate at that time to take stock and ask: Are these programs that were initiated and paid for from Canberra worth maintaining? How do students in these schools with bilingual programs compare with others? Are they achieving better results in English and Mathematics? Such questions all relate to *program accountability*, the *interests of stakeholders*, and *measuring academic outcomes*. To ask them is entirely justified. At the same time it is useful to remind ourselves that such questions are *limited*; that is, from the perspective of a program evaluator, they focus on output (that is, what a school or program achieves for itself in the short term) rather than impact (what a school or program achieves for others in the long term). Secondly, it is too *narrow* to gather data just on student outcomes for English and Mathematics, for doing so not only ignores achievements in other subject areas, not to mention vernacular literacy (reading or writing in L1), but it completely excludes a second perspective, which concerns the *value* of such programs (Devlin 2009), as perceived by local people (such as the Aboriginal authors whose views are expressed in this volume: Banbapuy Ganambarr, Deminhimpuk Francella Bunduck, Dhungala Munungurr, Dorothy Gapanay, M. Munungurr, Tess Ross, Tobias Ngande, W.W. Wunujmurra,

Yalmay Yunupingu, and others.) The authors of *Learning Lessons* contrast these two perspectives as follows:

The bilingual program was begun in schools in the Northern Territory in 1973. Government and bureaucratic proponents of the program at the time cited improved school attendance and better outcomes in English literacy and numeracy among the primary aims and anticipated benefits. Indigenous support always centred on what was seen as the first real recognition by Government of the value of Indigenous language, culture and law. In other words, while there was common support for the program, it came from different perspectives. In many quarters these different perspectives have not changed in more than twenty-five years.

(Collins and Lea 1999, p. 121)

## **Two Claims Regarding Bilingually Educated Students' Achievement in Literacy and Numeracy**

In this section of the paper I introduce two claims that have been made about the achievement of students in relation to Aim 1 (Northern Territory Department of Education 1980, p. 2) and briefly assess the evidence that might support or refute them.

### **Claim 1**

Some NT bilingual education programs have been comparatively effective in improving student academic results (Silburn et al. 2011, p. 26).

#### ***Explanation***

This claim, which was made in another review commissioned by the NT Government, appears to be supported by the available evidence. To start with, some useful official findings are available. These are the result of a series of school-based evaluations conducted for accreditation purposes soon after the NT government took over responsibility for education from the Commonwealth. As Devlin (1995, p. 25) has explained:

Telegrams dispatched to regional offices in 1980 announced the NT Department's decision to introduce accreditation procedures as a way of evaluating bilingual education. The purpose of accreditation, which would set out to evaluate the performance of each bilingual school program using the official aims of bilingual education as a yardstick, was to make sure that programs were being effectively conducted. Participating schools were told of the benefits that accreditation would confer, namely official recognition and a permanent allocation of additional resources.

There was problem with evaluating Aim #1 though. Even though it had been reworded to make it as specific and measurable as possible, thereby assisting bilingual program evaluators, the task of coming up with conclusive findings proved to be quite elusive for the researchers concerned. That is, it was difficult for them to say with certainty that the students they tested had developed "competency in reading and writing in English and in number to the level required on leaving

school to function without disadvantage in the wider Australian community” (Northern Territory Department of Education 1980, p. 2). At the time there were no national standards or tests available to allow evaluators to determine what that level might be.

The second-best alternative was to use some proxy measures, and these were developed as part of a ‘sophisticated’ model of program evaluation (Harris and Jones 1991, p. 45; Northern Territory Department of Education 1991), one that included interviews, document analysis, comparisons between designated bilingual programs and about six ‘non-bilingual control schools’. In attempting to gather what useful data they could, however, evaluators came up with comparative findings that reflected quite well on bilingual programs (Harris and Jones 1991).

### *Evidence*

After examining all the accreditation reports two researchers concluded that the Bilingual Program had

produced on the whole statistically significant academic growth in English and Maths...but this growth has not been as great as predicted from the theoretical advantages of the bilingual approach....A statement true of all schools except Oenpelli would be that, in general, significant gains in academic terms had been demonstrated in comparison to the pooled results of a group of non-bilingual control schools. The independent measures recorded by Murtagh (1979) and Gale et al. (1981) corroborate this statement in relation to Barunga (Bamyili) and Milingimbi.

(Harris and Jones 1991, p. 45)

The independent measures referred to above are analysed later in this paper.

Devlin (1995) explained that accreditation had aimed at a comprehensive examination of the *variables* affecting the operation of bilingual programs in individual Aboriginal schools, especially those designated as operational, social/psychological and academic ones. Students in Years 5, 6 and 7 were assessed on criterion-referenced English and Maths tests which had been jointly devised by NT Department of Education curriculum advisers, staff from the Evaluation and Research section of the same department, and teachers from six Aboriginal schools. Before being administered, the tests were piloted then subjected to item analysis. A control group of at least six non-bilingual schools was chosen for comparative testing purposes. Devlin (1995) then analysed results obtained by three schools which were eventually accredited: Yirrkala, St Therese’s and Shepherdson College. He explained that what the accreditation teams found when they analysed the test results for these schools was that bilingually educated pupils had performed as well on the English and Maths tests as pupils in the reference group of non-bilingual schools and in some cases they had performed better. For example, tests administered at Shepherdson College in 1981–2 and 1984 indicated that students in the reference group did not perform significantly better than Shepherdson students on any test at any year level (Markwick-Smith 1985, p. 47). Those at Shepherdson College “performed significantly better in enough areas, particularly in Years 5 and 7, to suggest that overall they have greater proficiency in school work than pupils in the [English-only] Reference Group schools” (Markwick-Smith 1985, p. 49) (Table 16.1).

**Table 16.1** Accreditation reports: three examples to illustrate their scope

Name	Years conducted	Type	Details	Findings
Richards and Thornton (1981)	1981	Matched group comparison between one school with a bilingual program (Yirrkala) and a reference group of six schools without bilingual programs	Students in years 5-7 were assessed using criterion-referenced English and Maths tests developed by curriculum advisers, staff from the evaluation and research section, and teachers from six Aboriginal schools. The tests were trialled and then subjected to item analysis before being administered	No statistically significant difference in performance was noted except for the subtraction sub-test in which year 5 Yirrkala students attained significantly better results
Stuckey and Richards (1982)	1981	Matched group comparison between one school with a bilingual program (St Therese's) and a reference group of schools without bilingual programs	As above	Year 7 St Therese's students attained significantly better results on all English tests bar one and the multiplication test. Year 6 significantly outperformed the reference group in tests on addition, time, written production and oral production. Year 5 did significantly worse on addition and maths problems
Richards (1984)	1982 and 1983	As above	As above	In 1982 St Therese's students performed better than the reference group on all tests—significantly better for Years 5-7 subtraction, cloze, written production, Year 7 oral reading, Year 6 multiplication and Year 5 reading comprehension. In no year (1982 or 1983) did reference group students significantly outperform those at St Therese's

## Claim #2

Academic results for students in ‘step’ model bilingual programs were worse than those for English-only programs at similar, remote Aboriginal schools (NT DET 2008).

## Comment

This was a significant claim for it was made by the Education Minister in the NT Legislative Assembly on November 25, 2008. Also, it was on this basis that the NT Government decided to cease its support for the ‘step’ model of bilingual, biliteracy education.

## Alleged evidence

On the 20 national tests conducted in 2008, bilingual schools were said to have done comparatively worse than a group of similar non-bilingual schools, on all but one test: Year 9 numeracy (Devlin 2011; NT DET 2008). A document, *Data for bilingual schools in the Northern Territory*, was tabled in the Legislative Assembly the next day as evidence to support the NT Government’s decision to phase out ‘step’ model bilingual programs the previous month (NT DET 2008). It claimed that, compared to ‘non-bilingual’ schools, ‘bilingual’ schools achieved better academic outcomes on only three of the 20 items in the 2008 NAPLaN literacy and numeracy tests; namely, Year 3 Grammar, Year 3 Reading and Year 5 Grammar.

## Evidence

Claim #2 is not supported by the available evidence. Using MySchool data, Devlin (2010b) checked the accuracy of that claim and found that the authors had neglected to mention a few other cases where the bilingual group did as well or better (Table 16.2).

The analysis in Devlin (2010a, b), using official NAPLaN data on the

MySchool website, showed that Year 3 students in the Government’s ‘bilingual school’ sample had actually performed better than the comparison group on four out of five tests; namely, (1) Reading, (2) Spelling, (3) Grammar and Punctuation, and (4) Numeracy; only in Writing did they lag behind (cf. Wigglesworth et al. 2011).

Simpson et al. (2009) drew attention to the weakness of the NT Government case against the ‘Step’ model of bilingual education. It was important that these scholars took this stand, for people connected with bilingual programs had been transferred, staff at Literature Production Centres had been directed to work in classrooms, resources had been diverted and Indigenous teachers had been marginalised—all on the basis of dubious claims, the questionable interpretation of national test scores (NAPLaN 2008), incorrect basic arithmetic, and the selection of an invalid school sample for comparative purposes (Devlin 2011). The sample was invalid because it included a school that did not have primary grades (Xavier EC) in the control group of primary schools, and one was running a Language Revival (LR) program. This was a heritage language learning program, rather than one in which student were taught in their first language and English, which was true of the other bilingual programs (Table 16.3).



**Table 16.2** NAPLAN results for Year 3 spelling, Year 3 numeracy, Year 7 numeracy across 16 comparison schools selected by the NT Department of Education and Training

Comparison of Year 3 Spelling scores ( <i>Source MySchool 2008</i> )			
Non-bilingual	School means	Bilingual	School means
Alekarenge CEC	212	Lajamanu CEC	
Angurugu CEC	186	Maningrida CEC	202
Borrooloola	203	Milingimbi CEC	223
Gapuwiyak CEC	186	OLSH Thamarrur CEC	198
Kalkaringi CEC	197	Shepherdson Coll. CEC	195
Ngukurr CEC	229	Yirrkala CEC	
Raminging CEC	220	Yuendumu CEC	
*Xavier		*Numbulwar	297
Group School Mean	204.7	Group School Mean	223
Comparison of Year 3 Numeracy scores ( <i>Source MySchool 2008</i> )			
Non-bilingual	School means	Bilingual	School means
Alekarenge CEC		Lajamanu CEC	
Angurugu CEC	245	Maningrida CEC	201
Borrooloola	251	Milingimbi CEC	303
Gapuwiyak CEC	236	OLSH Thamarrur CEC	219
Kalkaringi CEC	271	Shepherdson Coll. CEC	224
Ngukurr CEC	251	Yirrkala CEC	282
Raminging CEC	206	Yuendumu CEC	
*Xavier		*Numbulwar	331
Group School Mean	243.3	Group School Mean	260
Comparison of Year 7 Numeracy scores ( <i>Source MySchool 2008</i> )			
Non-bilingual	School means	Bilingual	School means
Alekarenge CEC	239	Lajamanu CEC	363
Angurugu CEC		Maningrida CEC	399
Borrooloola	394	Milingimbi CEC	362
Gapuwiyak CEC	380	OLSH Thamarrur CEC	344
Kalkaringi CEC	386	Shepherdson Coll. CEC	417
Ngukurr CEC	366	Yirrkala CEC	323
Raminging CEC	386	Yuendumu CEC	311
*Xavier	368	*Numbulwar	361
Group School Mean	359.9	Group School Mean	360

\*The invalid entries have been asterisked

## Levels of Evidence

So far in this paper I have considered two claims in the light of some available evidence. Such an approach is fine as far as it goes, but it does not take account of the *quality* of the evidence that is put forward to support or refute a particular claim. One way to do this might be to adapt and apply National Health and Medical

**Table 16.3** Schools in NT DET's comparative sample by grade range and program type (*Source* NT DET 2008)

Non-bilingual	Grades (program)	Bilingual	Grades (program)
Alekarenge CEC	P-9	Lajamanu	P-9 (Bilingual)
Angurugu CEC	P-9	Maningrida CEC	P-12 (Bilingual)
Borroloola	P-9	Milingimbi CEC	P-9 (Bilingual)
Gapuwiyak CEC	P-9	Numbulwar	P-9 (Revitalisation)
Kalkaringi CEC	P-9	OLSH Thamarrur CEC	P-12 (Bilingual)
Ngukurr CEC	P-12	Shepherdson Coll. CEC	P-12 (Bilingual)
Raminging CEC	P-12	Yirkala CEC	P-12 1 (Bilingual)
*Xavier	7-11	Yuendumu CEC	P-9 (Bilingual)

Research Council guidelines, which have been prepared to help researchers rate the key components of any 'body of evidence'.

These components are:

1. The evidence base, in terms of the number of studies, level of evidence and quality of studies (risk of bias).
2. The consistency of the study results.
3. The potential clinical impact of the proposed recommendation.
4. The generalisability of the body of evidence to the target population for the guideline.
5. The applicability of the body of evidence to the Australian healthcare context.

(National Health and Medical Research Council, n.d., p. 4)

Using this levels-of-evidence approach one might imagine that the place to start looking for gold-standard evidence might be randomised, controlled trials. However, no such trials have been conducted in the Northern Territory to gauge the efficacy of any small, remote, bilingual programs, although there have been some non-randomised, experimental cohort studies. These have taken the form of a matched-group comparison between two schools (Murtagh 1979, 1982), a matched-group comparison within one school (Gale, McClay et al. 1981), and case-control studies involving one school and a control group of six or seven reference schools (e.g., Markwick-Smith 1985; Richards 1984; Richards and Thornton 1981; and Stuckey and Richards 1982). Some of this work has previously been reviewed (Devlin 1995; Silburn et al. 2011), but student attainment in English literacy and numeracy is worthy of reconsideration here, because of its relevance to political and policy-related decision-making.

### **The evidence is "limited, but consistent"**

A team of researchers from the Menzies School of Health was commissioned in 2011 to review the literature related to bilingual and English as a Second Language (ESL) approaches (Silburn et al. 2011). In undertaking this project the researchers used National Health and Medical Research Council guidelines (NHMRC, n.d., p.4) to help them distinguish high-strength evidence from research studies with a medium or low standard of evidence rating.

### The Menzies School of Health reviewers concluded that

In considering the effectiveness of different language acquisition approaches in the Australian Indigenous context of English as a foreign or additional language, no definitive conclusions are able to be drawn given the limited sample sizes of the available studies and/or their lack of internal or external reliability. Most of the available reports have evaluation design limitations which render comparisons of the outcomes of Northern Territory bilingual and non-bilingual programs inconclusive. These include poorly selected comparison groups and/or a lack of rigorous statistical analysis resulting in studies reporting weakly supported findings (Devlin 1995). Nevertheless studies by the NT Department of Employment, Education and Training (DEET 2004) and academic researchers (Batten et al. 1998; Devlin 1995; Lee 1993; McKay 1997; Gale et al. 1981; Murtagh 1982) offer limited but consistent evidence that some NT bilingual education programs have been comparatively effective in improving student academic results.

(Silburn et al. 2011, p. 26)

In reviewing 243 studies by evidence type and rating, the researchers found that

in the Australian Indigenous context, the bulk of the reported evidence comes from descriptive or quasi-experimental studies, case studies and reviews. There are 24 studies involving particular methodologies or specific approaches for SAuE language acquisition within the Australian Indigenous context. Of these only 15 included direct outcome measures of the efficacy for identified instructional approaches, including three studies reporting to DET system-level data in the Northern Territory context. However, in only three of these (Devlin 1995; Gale et al. 1981; Murtagh 1982) was the description of the study methodology considered sufficient for it to be rated for evidence of efficacy by the SPR Standards for Evidence (SPR 2009).

Of the three studies mentioned above, Devlin (1995) was a review of the research including official accreditation reports which until then had been buried in the grey literature. The others (Gale et al. 1981; Murtagh 1982) were empirical studies.

### High-strength evidence

Silburn et al. (2011) considered that several independent research studies were based on high-strength evidence, including those relevant to the NT (Gale et al. 1981; Murtagh 1979, 1982). The relevant studies are summarised in Table 16.2 and elaborated a little in the text that follows.

#### Murtagh (1979, 1982)

Edward J. Murtagh, a linguist from Stanford University, assessed the results obtained by 58 Year 1–3 students in two schools, one with a bilingual program, and one without. (The two schools were Bamyili and Beswick, located east of Katherine and about 450 km southeast of Darwin). This cross-sectional study was conducted over 10 weeks. Although Murtagh refers to the ‘experimental group’ and the ‘control group’, his was not an experimental study, in fact, since the students were not randomly selected (1982, p. 16). His study is best classified as an example of a post-test only, nonequivalent control group design. Murtagh reported that, “the results of this study indicate very definite trends towards the superiority of bilingual schooling over monolingual schooling for Creole-speaking [Kriol-speaking] students with regard to oral language proficiency in both the mother tongue, Creole

and the second language, English”. This was because the bilingually schooled students attained better results on measures of oral language proficiency in their first language (L1) and L2, and were found to be better able to separate the two languages.

### **Gale et al. (1981)**

Gale et al. (1981) undertook a longitudinal comparison of Year 5–7 students over four years using tests in oral English, English reading, English writing and mathematics. As such, their research could be classified as a multiple-group, time series design. Data-gathering instruments used by the researchers included the Peabody Picture Vocabulary Test and story-retelling for assessing oral English Proficiency; Dolch sight words, a cloze test and the Schonell Reading Test were used to gauge reading progress.

In regard to the bilingually schooled students the researchers reported results that were both consistently better and statistically significant at the Year 7 level. It was their considered opinion that

since the introduction of bilingual education the [Milingimbi] children are not only learning to read and write in their own language and furthering their knowledge and respect for their own culture, but they are also achieving better academic results in oral English, reading, English composition, and mathematics than they were under the former English monolingual education system

(Gale et al. 1981, p. 309)

When the research conducted by Gale et al. (1981) was reviewed 30 years later, as part of a systematic investigation into English language acquisition and instructional approaches for Aboriginal students with home languages other than English, it was praised for its “high strength of evidence” (Silburn et al. 2011, p. 88). These reviewers noted though that there had been several possible intervening variables

such as a possible Hawthorne effect, curriculum changes and progressive exposure to English and Western culture outside of the school. Possible confounders acting in the opposite direction are that the Indigenous teachers supporting the bilingual classes had little or no teaching experience at the start of the program and that the L1 curriculum and teaching resources were very limited in the early years of the program

(Silburn et al. 2011, p. 87)

The reviewers from Menzies reported that the study had been “well conducted” and although the cohort size of around 20 at each year level was small, it was “sufficient to achieve statistically significant results in several cases” (Silburn et al. 2011, p. 88). In addition they noted (pp. 87–88) that

Both groups performed equally on English vocabulary tests. Bilingual students had better results on story retelling which increased with year level, but this was not statistically significant at 5 % level. The three reading tests, particularly comprehension, showed a similar very significant pattern. At year 5 level the bilingually educated children were significantly behind the English educated children but by year 7 this was reversed. In most cases these results were statistically significant at 5 % level. The authors cite this as evidence of L1–L2 skills transfer.

The year 7 bilingually educated children also scored significantly higher than the English only group in written English composition and several arithmetic tests.

Overall on the ten tests at year 7 level, the English only students performed better on two tests (neither significant) while the bilingually educated students had better scores on eight, five of these at 5 % significance level and two at 10 %.

### Research with a medium standard of evidence rating

Two studies are listed in Table 16.4. These provided the evidence for refuting Claim #2, as discussed in the previous section.

It has been pointed out by Georgie Nutton, who coauthored the *Early years English language acquisition* literature review (Silburn et al. 2011) that

potentially one of the key and critical issues for readers may be understanding that comparison studies and quasi experimental studies can often be rated lower on standards such as the NHMRC ratings only because the qualities and specifications of the ‘bilingual’ and ‘non-bilingual’ programs are not well articulated or monitored—i.e. what children actually experience as each of these programs is not described.

(Georgie Nutton, personal communication, September 2015)

Since Devlin (2010a) is not relevant to the present chapter, as it concerns attendance, not achievement, it is not discussed further here. Devlin (2010b) pointed out that standard deviations and the possibility of standard errors in the measurement were ignored when NAPLaN test scores were presented in the *Data in Bilingual Schools* document (November, 2010). This was not in accordance with NAPLaN reporting protocols (Table 16.5).

### Research with a low standard of evidence rating

A departmental statistician undertook a comparative and longitudinal Logit analysis of Multilevel Assessment Program (MAP) test scores for the 2001–2004 period in the Northern Territory (Begg 2004). The findings, which were published in

**Table 16.4** Research studies with a high strength of evidence rating (Silburn et al. 2011)

Name	Years conducted	Type	Details	Findings
Murtagh (1979, 1982)	1979	Matched group comparison between two schools (Bamyili and Beswick)	Cross sectional comparison of 58 Year 1–3 students in two schools, one with a bilingual program, and one without	Bilingually schooled students attained better results on measures of oral language proficiency in L1 and L2, and were better able to separate the two languages
Gale et al. (1981)	1976–1979	Matched group comparison within one school (Milingimbi)	Longitudinal comparison of Year 5–7 students using tests in oral English, English reading, English writing and mathematics	Bilingually schooled students obtained consistently better results, statistically significant at the Year 7 level

**Table 16.5** Research studies with medium standard of evidence rating (Silburn et al. 2011)

Devlin (2010a)	2010	Matched comparison of attendance at 16 schools	The comparison schools were chosen by the NT Government	Average attendance was poor across all the schools, but it was better in those with bilingual programs
Devlin (2010b)	2010	Matched comparison of results at 16 schools	The comparison schools were chosen by the NT Government	Results were below standard across all the schools, but those in bilingual programs performed comparatively better on most NAPLAN tests

*Indigenous languages and culture in Northern Territory schools Report 2004–2005* (NT DEET 2005, p. 35), were thought to provide an “indication of positive results” (NT DEET 2006, p. 25).

Silburn et al. (2011, p.79) considered this comparative data analysis to be “very interesting”, but rated it low on their standard of evidence rating. They summarised it in this way:

The sample is students from 10 2-Way schools and 10 “like” schools. They combined MAP results over 4 years; a total of about 3000 tests in all. About half the data is for 2-Way students. Across all years the 2-Way students had better enrolment and participation. Their MAP reading scores were lower than the control group in year 3 but improved more rapidly scoring higher in years 5 and 7 when their English was better established. Results for both cohorts were markedly below the national benchmark across all domains and at every year level.

The report notes that the 2-Way schools are better resourced (by 20–30 %) and the matching is on similar student populations.

There are some indications that 2-Way learning improved outcomes and retention.

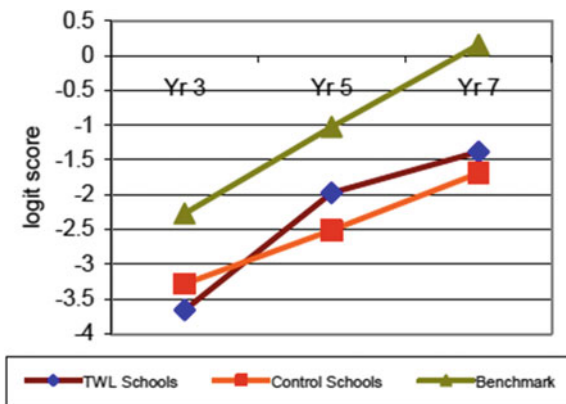
The report is very broad based making a number of important points about the need for an evidence base and need for improved consistency and quality in the delivery of 2-Way learning.

As Devlin (2011, pp. 270–271) has explained, Begg’s analysis had been done as part of a departmental review of bilingual education. A representative graph (NT DEET 2005, p. 35, chart 6) is included here to illustrate one of his findings (see Fig. 16.1). Students in schools with bilingual programs [‘2-way schools’] scored comparatively lower on MAP tests at the Year 3 level, but had moved ahead of the comparison group by Year 5 and maintained a slight lead by Year 7. The following conclusion was reached:

while the combined comparison of Two Way Learning and like school MAP reading scores supports the theory that students’ English literacy acquisition is accelerated through bilingual instruction, due to the smallness of the numbers of students with scores that are analysed relative to the whole school cohort, this data can only be taken as indicative rather than conclusive (DEET 2004, p. 35)

It should be noted though that both groups performed well below expected Benchmarks. Also “large numbers of students from both groups of schools...did not record any achievement in testing” which could lead one to ask “what educational benefit these students are gaining from school” (DEET 2004, p. 35).

**Fig 16.1** A comparison of combined mean reading scores on MAP Tests, 2001–4, attained by Two Way learning and ‘Like’ school students (Source NT DEET 2004, p. 35)



The label “TWL schools” in Fig. 16.1 refers to Two-Way Learning schools; i.e., schools with bilingual programs. Each point in this chart combines MAP test result data from 2001, 2002, 2003 and 2004.

Another study assigned a low standard of evidence rating by Silburn et al. (2011, p. 79) was Devlin (1995). Silburn et al. (2011) noted with interest “the move from quantitative appraisal to qualitative community based assessment” that had been analysed this study, adding that “This may explain why little data is available from the later Years” (Silburn et al. 2011, p. 81). They concede that “some qualitative data from 1984 on literacy/numeracy skills is included and discussed in detail”, but consider that “some of the conclusions are weakly supported by the data” (p. 81). They conclude though with the observation that “there is a pattern in the data of broadly equal achievement when comparing students in bilingual schools with an equivalent cohort in English immersion”.

## Conclusion

This chapter has considered a very specific question: What research evidence is available that would help tell us whether bilingual education in the Northern Territory has ever been effective in promoting better student attainment in English and Mathematics? One reason for choosing such a limited focus on Western-style academic performance in this chapter, rather than taking account of vernacular literacy, school-community relations, or some other wider indicator of achievement, is that it is important to draw a line in the sand. Some politicians and senior bureaucrats have specifically denied in recent years that *any* evidence in favour of NT bilingual programs exists (AAP 2008; Doyle 2009; Devlin 2011, p. 270; Freeman and Bell, this volume). For that reason it became the task of this chapter to show that some supporting evidence is available, though it is fairly sparse, and not all of it warrants a high strength rating (Silburn et al. 2011).

## References

- Australian Associated Press. (2008, November 17). Calma backs bilingual education in NT. Retrieved April 22, 2011, from <http://news.theage.com.au/national/calma-backs-bilingual-education-in-nt-20081117-6990.html>.
- Australian Department of Education. (1975). *Handbook for teachers in bilingual schools in the Northern Territory of Australia* (2nd ed.). Darwin: Australian Department of Education, Northern Territory Division, Bilingual Education Section.
- Batten, M., Frigo, T., Hughes, P., & McNamara, N. (1998). *Enhancing English literacy skills in Aboriginal and Torres Strait islander students: A review of the literature and case studies in primary schools*. Research Monograph 54. Melbourne: ACER Press. Retrieved July 26, 2014 from [http://research.acer.edu.au/indigenous\\_education/5](http://research.acer.edu.au/indigenous_education/5).
- Begg, B. (2004). *Reading logit scores*. [Powerpoint presentation on the findings from DEET.] Darwin: Department of Education and Training.
- Cataldi, L., & Partington, G. (1998). Beyond even reasonable doubt: Student assessment. In G. Partington (Ed.), *Perspectives on Aboriginal and Torres Strait islander education* (pp. 309–332). South Melbourne, Australia: Cengage.
- Collins, B., & Lea, T. (1999). *Learning lessons: An independent inquiry into Aboriginal education in the Northern Territory*. Darwin: Northern Territory Department of Education.
- Department of Education, Employment and Training. (2004). *Indigenous languages and culture in Northern Territory Schools Report 2004–2005*. Retrieved July 31, 2015 from [http://www.det.nt.gov.au/\\_\\_data/assets/pdf\\_file/0009/5130/ILCreport.pdf](http://www.det.nt.gov.au/__data/assets/pdf_file/0009/5130/ILCreport.pdf).
- Devlin, B. (1995). The evaluation of bilingual programs in the Northern Territory, 1973–1993. *International Journal of the Sociology of Language*, 113, 25–41. Retrieved December 12, 2015, from <http://goo.gl/4qSsnR>.
- Devlin, B. (2009). *Bilingual education in the Northern Territory and the continuing debate over its effectiveness and value*. Paper presented to an AIATSIS Research Symposium, Bilingual education in the Northern Territory: Principles, policy and practice, Canberra, Friday June 26, 2009. Retrieved November 19, 2009 from [http://www.abc.net.au/4corners/special\\_ed/20090914/language/](http://www.abc.net.au/4corners/special_ed/20090914/language/)
- Devlin, B. (2010a). *Attendance data—'bilingual' and 'non-bilingual' schools compared*. Retrieved from <http://groups.google.com.au/group/foblmail>.
- Devlin, B. (2010b). *Performance data—'bilingual' and 'non-bilingual' schools compared*. Retrieved from <http://groups.google.com.au/group/foblmail>.
- Devlin, B. (2011). A bilingual education policy issue: Biliteracy versus English-only literacy. In N. Purdie, G. Millgate, & R. Bell (Eds.), *Two way teaching and learning: Towards culturally reflective and relevant education* (pp. 49–69). Melbourne: ACER Press.
- Disbray, S., & Devlin, B. (2017). Imposed solutions versus local aspirations: Two recent policy decisions affecting remote education in the Northern Territory. In E. Baehr, Elisabeth, & B. Schmidt-Haberkamp (Eds.), “... and there'll be NO dancing”. *Perspectives on policies impacting Indigenous Australia since 2007*. Newcastle upon Tyre: Cambridge Scholars Publishing.
- Doyle, M. (2009, September 14). *Four corners*. [Television broadcast of Debbie Whitmont's Interview with Gary Barnes.] Sydney: Australian Broadcasting Commission. Retrieved December 11, 2015, from [http://www.abc.net.au/4corners/special\\_ed/20090914/language/](http://www.abc.net.au/4corners/special_ed/20090914/language/).
- Gale, K., McClay, D., Christie, M., & Harris, S. (1981). Academic achievement in the Milingimbi bilingual education program. *TESOL Quarterly*, 15(3), 297–314.
- Grimes, C. (2009). *Indigenous languages in education: What the research actually shows*. Darwin: Australian Society for Indigenous Languages.
- Harris, S., & Jones, P. (1991). The changing face of Aboriginal bilingual education in the Northern Territory: A 1990 update. *The Aboriginal Child at School*, 19(5), 29–53.



- Lee, P. (1993). *Bilingual education in remote Aboriginal schools: Developing first and second language proficiency*. Broome, WA: Catholic Education Office of WA, Kimberly Region, and the Commonwealth of Australia.
- Lo Bianco, J. (2010). *Indigenous languages in education: Do current policies match our needs?* Presentation to a Public Forum, "Indigenous Languages in Education: do current policies match our needs?", Charles Darwin University, 9 September 2010. Retrieved December 12, 2015, from <http://www.ausil.org.au/sites/ausil/files/Sept%2010%20LoBianco%20Darwin%20Indig%20Lgs%20Policy.pdf>.
- Markwick-Smith, V. (1985). *Bilingual accreditation ongoing report—Shepherdson College*. [Prepared for the Accreditation Panel, August 1985.] Darwin: NT Department of Education.
- McKay, P. (1997). The bilingual interface project report: The relationship between first language development and second language acquisition as students begin learning English in the context of schooling. *Queensland Journal of Educational research*, 15, 123–132.
- Murtagh, E. (1979). *Creole and English bilingual education: Some research findings*. Darwin: Summer Institute of Linguistics, Australian Aborigines Branch.
- Murtagh, E. (1982). Creole and English used as languages of instruction in bilingual education with Aboriginal Australians: Some research findings. *International Journal of the Sociology of Language*, 36, 15–33.
- National Health and Medical Research Council. (2000). *How to use the evidence: Assessment and application of scientific evidence*. [Handbook series on preparing clinical practice guidelines.] Canberra: Author. Retrieved August 3, 2015 from [https://www.nhmrc.gov.au/\\_files\\_nhmrc/publications/attachments/cp69.pdf](https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/cp69.pdf).
- National Health and Medical Research Council. (n.d.). *NHMRC additional levels of evidence and grades for recommendations for developers of guidelines*. Retrieved December 12, 2015, from [https://www.nhmrc.gov.au/\\_files\\_nhmrc/file/guidelines/developers/nhmrc\\_levels\\_grades\\_evidence\\_120423.pdf](https://www.nhmrc.gov.au/_files_nhmrc/file/guidelines/developers/nhmrc_levels_grades_evidence_120423.pdf).
- Nicholls, C. (1999). Ten common attitudes to, and myths about, the Northern Territory's bilingual education programs. *Ngoonjook*, 16, 77–88.
- Northern Territory (Self Government) Act 1978. (2013). *Northern Territory (Self Government) Regulations 1978*. [Statutory Rules No. 102, 1978 as amended.] Retrieved December 12, 2015 from <https://www.comlaw.gov.au/Details/F2013C00783>.
- Northern Territory Department of Education (1980). *Objectives for the development of bilingual education*. Darwin.
- Northern Territory Department of Education. (1991). *Appraisal/accreditation process for Northern Territory Bilingual Schools*. [Unpublished manuscript].
- Northern Territory Department of Education and Training. (2005). *Indigenous languages and culture in Northern Territory schools. Report 2004–2005*. Darwin: Author.
- Northern Territory Department of Education and Training. (2008, October 27). *Data for bilingual schools in the Northern Territory*. [Unpublished manuscript tabled in the Legislative Assembly by the Minister for Education on November 26, 2008].
- Northern Territory Department of Employment, Education and Training. (2006). *Indigenous education strategic plan 2006–2009*. Darwin: Northern Territory Government. Retrieved June 22, 2009, from [http://www.det.nt.gov.au/education/indigenous\\_education/strategic\\_directions/strategic\\_plan/index.shtml](http://www.det.nt.gov.au/education/indigenous_education/strategic_directions/strategic_plan/index.shtml).
- Richards, A. (1984). *Accreditation ongoing report—St Therese's School*. [Prepared for the Accreditation Panel, May 1981.] Darwin: NT Department of Education.
- Richards, A., & Thornton, C. (1981). *Report of the bilingual education program at Yirrkala Community School*. [Prepared for the Accreditation Panel, May 1981.] Darwin: NT Department of Education.
- Silburn, S.R., Nutton, G.D., McKenzie, J.W., & Landrigan M. (2011). *Early years English language acquisition and instructional approaches for Aboriginal students with home languages other than English: A systematic review of the Australian and international literature*. Darwin: Centre for Child Development and Education, Menzies School of Health

- Research. Retrieved August 3, 2015, from [www.education.nt.gov.au/\\_\\_data/assets/pdf\\_file/0020/20198/MenziesReport.pdf](http://www.education.nt.gov.au/__data/assets/pdf_file/0020/20198/MenziesReport.pdf).
- Simpson, J., Caffery, J., & McConvell, P. (2009). *Gaps in Australia's Indigenous language policy: Dismantling bilingual education in the Northern Territory*. AIATSIS Discussion Paper, 24. Retrieved June 3, 2016 from <http://aiatsis.gov.au/publications/products/gaps-australias-indigenous-language-policy-dismantling-bilingual-education-northern-territory>.
- Stuckey, P., & Richards, A. (1982). *Report of the bilingual education program at St Therese's School*. [Report prepared for the Accreditation Panel, May 1981.] Darwin: NT Department of Education.
- Whitlam, G. (1972, December). Aboriginal education assistance. [Prime minister's press statement, no. 16, December 14, 1972]. Retrieved December 11, 2015, from <https://pmtranscripts.dpvc.gov.au/release/transcript-2748>.
- Wigglesworth, G., Simpson, J., & Loakes, D. (2011). NAPLAN language assessments for Indigenous children in remote communities: Issues and problems. *Australian Review of Applied Linguistics*, 34(3), 320–343.
- Wilson, B. (2014). *A share in the future: Review of Indigenous Education in the Northern Territory*. [Final report]. Melbourne: The Education Business. Retrieved December 16, 2015, from [http://www.education.nt.gov.au/\\_\\_data/assets/pdf\\_file/0007/37294/A-Share-in-the-Future-The-Review-of-Indigenous-Education-in-the-Northern-Territory.pdf](http://www.education.nt.gov.au/__data/assets/pdf_file/0007/37294/A-Share-in-the-Future-The-Review-of-Indigenous-Education-in-the-Northern-Territory.pdf).