Productivity Commission
Indigenous Evaluation Strategy Project

Oral Submission

Date of Oral Submission: 25 August 2020

Submission provided by: Dr Helen Moore, AM, on behalf of the Australian Council of TESOL Associations (ACTA)

This is an oral submission to the Productivity Commission's Indigenous Evaluation Strategy project. The information contained therein was provided by videoconference by the organisation for Commission staff to assist in documenting in written form. The submitter has agreed that this is a correct reflection of their views and has approved it for publication on the Productivity Commission website.

Opening comments

Thank you for allowing the Australian Council of TESOL Associations (ACTA) to make a second submission. We appreciate the opportunity to make the submission orally. The submission will first make some specific points about the Strategy and Guidance documents, then raise more general issues that apply across all documents, and finally comment on a couple of important chapters.

Overall, we congratulate the Commission on this impressive work, which we sincerely hope will have considerable impact. The following comments focus on suggestions for improvement but any criticisms should not be taken to imply a lack of warm endorsement for the thrust and achievements of this important project.

PART 1: Comments about particular pages/wording in the Strategy and guidance documents

These documents are mostly very good at their level of generality, but opportunities exist for some improvements in wording. Some changes we suggest relate to more fundamental problems in the background paper, to which I'll return later.

Page 22 of the Strategy document and **page 10** of the guidance document include a 'cross system priority' of:

> *Addressing racism, discrimination and social inclusion, healing and trauma, and the promotion of culture and language for Aboriginal and Torres Strait Islander people.*

We would recommend that you change the last bit of this to read the *'strengthening of cultures and languages'* (using the plural). (We make a wider point re diversity in our general comments below.)

Further, we cannot find anything in the three documents that talks about *promoting* languages or directing evaluations to consider whether or not languages and language learning (of English and/or Aboriginal & Torres Strait Islander languages) are promoted. I'll come back to this point in my more general comments (section 2.1).

On **page 24** of the Strategy, in the list of actions, there's nothing (as far as we can see) on where Commonwealth-State responsibilities intersect and how recommendations might carry forward into State/Territory jurisdictions. Especially with mainstream policies, this is a huge and complex issue,

not least in education, health and justice. So we wonder if it's possible for the Commission to consider actions that will require Commonwealth-State co-operation and how this might best be achieved.

On **page 26** of the Strategy, regarding building capability, as far as we can see, there's nothing there (or in the whole set of documents) about traditional training institutions (notably, universities and the VET sector). The bulk of capability is recruited from and developed in these institutions. Bodies like ANZSOG also have a role to play in training. The discussion of building capability should include how to impact on traditional training institutions.

There is also a need to consider professional associations (such as ACTA) and advocacy groups and the networks that these associations have and are capable of developing. Most of the work of ACTA's member associations is professional development. An example of the strategic capability building that professional/advocacy groups can do is the Multicultural Youth Advocacy Network's Framework to assist organisations in working with young people from refugee backgrounds. (Go to: https://myan.org.au/resources/national-youth-settlement-framework/). These kinds of resources could be developed by professional associations and advocacy groups – in fact, this work is probably best done by them because they are close to the coalface.

Figure 1 on **page 2** of the guidance document outlines the Strategy's principles. But the role of expert knowledge is unclear. The diagram outlines four principles that should underpin evaluations: they should be credible, useful, ethical and transparent. But which of those headings includes the role or contribution of expert knowledge in guiding evaluations? It might come under 'credible'. But the elaboration of what's meant by credible doesn't really cover expert knowledge. The downplaying and sidelining of expert knowledge is a criticism that extends across all the documents, which I'll come back to in my more general comments.

On **page 7** of the guidance document, there is a reference to the *translation* of surveys. What is meant by translation seems to be inconsistently explained throughout the document. Sometimes it seems to mean translation from one language to another, but at other times not. Further, translation from English to an Aboriginal or Torres Strait Islander language may actually be a waste of time and money. The question is: what are the most effective ways of *communicating*, not just translating. Are the most effective ways oral or written? Should they rely on literacy? Should they be pictorial? Should they be in English, or should they be in a language or dialect other than English? *Who* should do the 'translating' (in both the literal and more extended sense), that is, which speakers of a given language should be asked to make translations?

Box 4 on page 15 of the guidance document could include mention of *the evaluator's* credibility. We've heard from people who work in the field that evaluators are often simply not credible, including and maybe especially 'fly in-fly out' evaluators.

In that same box, you could consider including something that allows for 'in process' feedback, which is talked about elsewhere in the documents. Feedback as an evaluation proceeds is something that was really emphasised to us by people working in the field. An evaluation can start out with clear objectives and goals, but when you hit the ground, you can find that something unanticipated is going on. So you need to allow for feedback and change during evaluations.

Page 17 of the guidance document describes 'questions to consider when deciding who will conduct the evaluation'. You might consider building more into the question 'Are evaluators culturally and technically competent?' It should certainly be revised to read: 'Are evaluators culturally, <u>linguistically</u> and technically competent <u>and</u> knowledgeable?' Another question is: 'Are they credible in this particular context?'

Page 25 of the guidance document includes a box on collecting baseline data. If baseline data is irrelevant or lacks key indicators, the evaluation is useless. For example, if baseline data in education or health or justice does not take account of a person's linguistic starting point —their dominant language, or the language in which they are most comfortable — or if it assumes English is the same as literacy (which it is not), then the evaluation is useless. More needs to be said about deciding on what constitutes relevant data.

The issue of baseline data applies to two indicators that are ubiquitous in education, including evaluations, viz. whether students are *Aboriginal or Torres Strait Islanders* and whether they are from a *Language Background Other than English*.

Neither of these indicators tells us anything about whether or not a student is *an English language learner*.

Students identified as Aboriginal and Torres Strait Islander may or may not be fluent in English. They may or may not speak Aboriginal or Torres Strait Islander languages and/or non-standard varieties of English.

Similarly, data on 'Language Background Other than English' (unfortunately known as LBOTE) tells us nothing about students' English language proficiency. Students with a language background other than English may come from a home where, say, the mother is a native English speaker (even monolingual) and the father is a Chinese speaker. Children from these households may be monolingual English speakers, *or* partial or full bilinguals in English and another language, *or* partial or full bilinguals in two languages neither of which is English (as is true for some speakers of Aboriginal languages), *or* monolinguals in a language other than English, *or* fluent in non-standard varieties of English.

Currently, date based on both these identifiers are combined with data that measures disadvantage, which further obscures the achievements and extent of need for those who are learning English. Currently, we have no nationally consistent picture of English language learners in Australian schools.

ACTA has argued over many years that baseline data should directly address whether or not students are English language learners. That would require a valid and reliable measurement tool that maps known pathways in learning English. The Australian Curriculum and Assessment Authority (ACARA) has developed such a tool but its acceptance has been stalled in COAG for several years.

Page 32 of the guidance document includes a box on presenting evaluation findings for different audiences. Factsheets are given as an example. But factsheets won't work for some communities, some parents and some community elders, so that needs elaboration. Other types of presentation may be more appropriate — for example: visual modes of presentation; radio programs; and drama.

Page 33 of the guidance document contains a series of questions on what to consider when reporting evaluation findings. Here elaboration is needed on what is meant by different formats and by 'knowledge translation' (which is so abstract, we're not sure we know what it means). Likewise, cultural sensitivity does not necessarily include linguistic sensitivity.

PART 2: More general comments

These comments apply across all three documents, but are based on our reading of the background paper.

2.1 The phrase *'Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges'* and its intended meaning need elaboration

What is meant by *'Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges'* needs explanation. Nowhere in the discussion document (as far as we can see) is the meaning of this phrase systematically spelt out. Consequently, it increasingly seems like a kind of mantra.

Taking the first key noun, *'people'*: somewhere in the document the danger of generalisations about 'Aboriginal people' is acknowledged but generally a unified and homogeneous group is implied.

For example, the discussion of representation on the Indigenous Evaluation Council says that the majority of people should be from an Aboriginal or Torres Strait Islander background. But that is quite vague. How should/might representation from very different kinds of Aboriginal and Torres Strait Islander people be decided? What is meant by *'representation'* when there is such enormous diversity among Aboriginal and Torres Strait Islander people, for example, between urban, rural and remote people? Or, from a linguistic perspective, how will representation allow for the different speakers of Aboriginal and Torres Strait Islander languages in different parts of Australia? How will monolingual speakers of Indigenous languages be represented? How will people who speak different non-standard varieties of English be represented?

Regarding the last group: the complex differences between *Standard Australian English* and *Aboriginal and Torres Strait Islander non-standard varieties of English* have crucial implications in domains such as education (for example, in literacy teaching and testing, and teacher education)[1] and the administration of justice (for example, in misinterpretations of verbal evidence and wrongful self-incrimination)[2]. Different meanings can attach to the same word/use of language, which can cause fundamental misunderstandings, incorrect assessments of the facts, injustices, discrimination and debasing self-assessments (as when Torres Strait Islanders say they speak 'rubbish English').

The need to unpack what's meant by 'Aboriginal and Torres Strait Islander people', and their differences and commonalities, specifically applies in the background paper section on data and the discussion of how people identify themselves. Section 9.2 on data collection needs more elaboration on what Aboriginal and Torres Strait Islander identification both tells us and ***doesn't*** tell us. In regard to language use, the following distinctions may be relevant:

---

[1] See the submission from Denise Angelo et al.

[2] See pioneering work on language and the law by Diana Eades and others (e.g. http://classic.austlii.edu.au/au/journals/PrecedentAULA/2015/12.html)

- monolingual and bi/multilingual **speakers of traditional and emerging Aboriginal and Torres Strait Islander languages**

- monolingual and bi/multilingual **speakers of English and traditional and emerging Aboriginal and Torres Strait Islander languages**

- **learners of Standard Australian English**, and literacy in English, as an additional language

- Aboriginal and Torres Strait Islander **learners of traditional Aboriginal and Torres Strait Islander languages**.

Sometimes references to including Aboriginal and Torres Strait Islander people seem almost patronising. It shows up in the phrase *'cultural safety'*, which often seems problematic (although not always - for example, there is a section about children where it is appropriate). The more *'cultural safety'* appears in the text, the more feels infantilising — as though it's referring to timid little children. We suggest a word search of all three documents to ask if *'appropriate'* isn't preferable, at least in some places.

This issue relates closely to what we see as overtones of '*them'* (= alien others/infantilised/weak Aboriginal and Torres Strait Islander people) and *'us'* (= white people/evaluators /enlightened experts/the norm, etc.).

An example is in Box 5.11 on page 185 of the background paper. The box says that a barrier to accessing government services is a 'lack of cultural safety'. The explanation in terms of disrespect and insensitivity is acceptable but the word 'safety' gives the impression that we're talking about frightened children who need to be protected. 'A lack of cultural appropriateness' would imply an equal footing on all sides.

If we turn some of the points in Box 5.11 on their heads to make Aboriginal and Torres Strait Islander people equal players, it becomes clear that a lack of 'cultural safety' doesn't take enough into account. For example, the list includes a *'lack of awareness – people do not know what services are available in the first place'.* This framing tends to blame the victim. But the explanation could also be framed as *'services are not made available to people where they are and how they need them'*. That phrasing places clear responsibility for getting information to people on the information **providers**, not its ignorant recipients.

The next point in the list is another example: *'language and literacy – access to information or service delivery can be inadequate for Aboriginal and Torres Strait Islander people who speak English as a second language'.* If this were rephrased to say that information should be given in the languages and ways people speak, Aboriginal and Torres Strait Islander people would be placed on an equal footing. It would displace the implication/assumption that 'we' white, benevolent evaluators who mean well (who, *of course,* speak English) are going out 'there' to 'help' those 'others' who (unfortunately) don't speak English.

The homogenisation of 'Aboriginal and Torres Strait Islanders' also shows up in the lack of recognition of the potential for differences and even conflicts in people's preferences and priorities. An example is the sentence on page 172: *'measures of value should reflect Aboriginal and Torres Strait Islander people's preferences'*, which presents these preferences as incontestable and

uncontested. Determining preferences among *any* group of people is incredibly tricky – there is (nearly) always some conflict. The them/us assumption is unmasked if we compare that sentence with a (potential) recommendation that 'measures of value should reflect white people's preferences', which is clearly absurd. So why can this recommendation apply to Aboriginal and Torres Strait Islander people?

Regarding *'knowledges',* as far as we can see there is no unpacking of what is meant by knowledges. None. Does it include language? In some places, language is clearly included and the term 'knowledges' works well. But elsewhere 'knowledges' is elaborated but language(s) is/are not mentioned. 'Knowledges' is a great word, but what does it mean in these documents?

To sum up: what is meant by *each* noun in the phrase 'Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges' should be explained at the outset. This explanation should systematically apply throughout the discussion. Explanations shouldn't just appear where someone has realised that an explanation or elaboration is needed.

2.2 Consideration of language, language issues, multilingualism, English language proficiency, and English language varieties is inconsistent and mostly missing

All three documents seem to reflect an assumption that mention of culture (or knowledges) will allow language issues to be addressed.

The discussion of priorities early in the discussion paper refers to the United Nations Declaration on Indigenous Rights. But no reference is made to the Declaration's paragraphs on languages. We made extensive reference to these paragraphs in the ACTA submission. Their implications are substantive.

We are disappointed that our extensive submission arguing that language(s) and language issues should be taken account of in evaluations does not seem to have been attended to in any part of the Commission's documents. There is just one reference to the ACTA submission (re data that's included and omitted – we don't think this is the most important thing we said). We also can't see any evidence that attention was paid to other submissions that focussed on language issues, notably the submission by Denise Angelo and others, the submission by the ARC Centre for Excellence for the Dynamics of Language and the submission by Audiology Australia (deafness is a huge issue in Aboriginal communities, with a huge impact on educational achievement for not just Aboriginal children and adults but everybody).

Somewhere in the background document is a quote that has words to the effect of 'You can't see what you're not looking for' (p. 161). We think that observation clearly applies to the Commission's own discussion. If you're not looking for language issues, you don't see them. This criticism applies not just to the Commission's analysis – the submissions and frameworks referenced also say very little or nothing about language.

But language is really important. At the beginning of ACTA's submission (p. 9 ff.), we reported on an unpublished paper on policies and evaluations that do and don't consider language in its own right, and what gets overlooked when you don't consider it (see also Appendix B). We commend those sections of our submission to you. (The paper may have been published by now.)

There's a very good paragraph somewhere in the background document about how minorities are overlooked and the tendency to preference the majority. Attention is paid to how to balance minority perspectives (in this case, Aboriginal and Torres Strait Islander perspectives) against those of the majority. This principle should extend to the submissions the Commission receives. The majority of submissions say nothing about language, but maybe it's worth asking: 'Is there an important unexamined bias in how evaluations and professional evaluators work?' Just because evaluators never think about language does not mean it's not important in the real world.

Your recommendations include creating a Centre for Evaluation and building an evaluation culture – we find this a little scary in some ways, because it could end up being another silo, where evaluators only go to evaluator conferences and only talk to other evaluators. This means that the assumptions that exist amongst evaluators will be perpetuated. Something needs to be built in that breaks that.

If the principles in the Commission's analysis are to be applied to its own way of framing things – viz. that minority perspectives have a place, and that evaluators need to try and look for things that are *there*, rather than just what they are initially primed to look for – then language(s) might gain more attention in the final version of these documents.

Because language is *so* fundamental, it's frequently taken for granted – like breathing or the heart beating. It's only when something goes wrong that people pay attention. Then these things are encountered as problems: emphysema or heart failure. Likewise, language issues mostly gain attention with reference to problems, barriers or impediments, for example when people are described as not speaking English or illiterate. But if we think about language(s) as assets or strengths, a different way of understanding language issues emerges. Policies need to take account of the way that languages actually work in society and for individuals, just as doctors need to understand how the heart works. Policy evaluations need to focus on whether, to what extent and how policies have understood and responded to how language(s) work, and what can and should be done to build on language strengths and language assets. That goes far beyond saying 'People don't understand English, therefore the policy is not working.'

Page 114 includes a discussion about what works and does not work, but it seems to be taken for granted that we all know what works. But the question should be: 'what works *for whom* and *according to what criteria*?' For example, with Cashless Debit Card you could say that this policy is 'working' for those who think it is a good idea. Likewise, the abolition of bilingual programs in the Northern Territory has 'worked' for policy makers who think that bilingualism holds kids back. But many language educators consider the termination of (good) bilingual programs to be a key factor in why the Closing the Gap targets have not been met. So what works for whom and why is a question that needs to be asked.

Page 114 makes no mention of what we've learned about including or excluding language issues from consideration. But a huge amount of evidence (and the United Nations Declaration) supports incorporating the mother tongue at least as a starting point for learning in the formal school system. With respect to the Closing the Gap target for kindergarten, children are unlikely to maintain school attendance if they're confronted with a language that they don't understand. If data is sought from schools where (among other things) the starting point is the child's mother tongue (and maybe even

where it is promoted with pride alongside learning English), it might be found that those schools are doing better in meeting Closing the Gap targets.

NAPLAN is a prime example of the impact of mainstream polices that take no account of the language proficiency of Aboriginal and Torres Strait Islander children. NAPLAN sets many of these children up to fail. A chilling illustration of how this process works is documented in the film *In My Blood It Runs*. The teacher tells her class of predominantly Aboriginal students (in Alice Springs) not to worry that they can't do the test. They should just tick any box, because at least if they tick a box, they give themselves a 25 per cent chance of getting a mark. The impact of this advice on these students' self-esteem and commitment to schooling requires no imagination – it is evident in their faces.

A comment we've received from an educational consultant who works out of Darwin and spends a lot of time in remote communities was that evaluation of school education for Indigenous-language speakers and Aboriginal English speakers (and they are not the same) in the Northern Territory focusses on looking for *the same outcomes at the same stage of schooling* as for students who speak Standard Australian English. As already discussed regarding base-line data, the assessment tools used to measure the achievement of fluent English users do not accurately reflect the progress and achievement of students who are *learning* English at school, learning other content *through* English and who use other languages beyond the school. Their learning pathways – and therefore benchmarks for progress – are not the same. Learners' actual knowledge and skills should be benchmarked against what is known about *the steps in learning English*, not a mother tongue (English) *literacy* scale.

(To illustrate: I'm a reasonably fluent and literate speaker and writer of English. I learned German at school and in first year university and can understand a bit. In a test of reading German or geography or numeracy in German, I would do appallingly. But that wouldn't mean I have literacy or numeracy problems or don't know any geography. A valid and reliable test would have to check out where I am in my German learning, which would need to be taken into account in any assessments of my knowledge and skills in geography, reading and numeracy.)

To sum up, a review of the three documents should systematically apply the following question to all sections: **would adding the word *'language(s)'*, or the phrase *'language issues'* or *'language use'* or *'language proficiency'* make any difference to the recommendations about how evaluations should be conceptualised and operationalised?'** In some places adding these words will make no difference. But in others including these words will make a substantive difference.

2.3 The background paper has a narrow view of stakeholders, other participants in policy and program evaluation, and established knowledge

2.3.1 Stakeholders

The Commission documents rightly place huge emphasis on including Aboriginal and Torres Strait Islander stakeholders and participants in the evaluation process. The 'perspectives, priorities and knowledges' of Aboriginal and Torres Strait Islander people are given value throughout the document (although what this means is not specified), but this value is not accorded to those of other stakeholders in these evaluations. The emphasis on Aboriginal and Torres Strait Islander

participation, which we absolutely endorse, perpetuates the 'them'/'us' dichotomy when the stakes and roles of others in evaluations remain an undefined 'us' and are not explicitly examined.

Other potential stakeholders and contributors include:

- **academics with formal expert knowledge** about Aboriginal and Torres Strait Islanders and their perspectives, priorities and knowledges, for example linguists, criminologists, health researchers and specialists in language learning.

- **experts 'on the ground'**, that is, professionals and practitioners who work in Aboriginal and Torres Strait Islander contexts – including teachers, doctors, nurses, legal aid lawyers, etc.

- **politicians and policy makers** are given some attention in chapter 7 but not nearly enough. They have significant stakes in the policies and programs that they initiate and for which they are responsible. If they have not moved on, they may have equal stakes in how these policies and programs are evaluated. I'll come back to this group in discussing Chapter 7.

- **voters/taxpayers and the general public**, whose interests are often invoked, often to serve particular purposes.

'Experts' can be academics who spend their lives acquiring disciplined formal knowledge and skills in thinking about Aboriginal and Torres Strait Islander 'perspectives, priorities and knowledges'. Experts are also the professionals and practitioners with deep experiential knowledge, who have built up trust in Aboriginal and Torres Strait Islander communities. These different types of experts may work closely together, for example where linguists and teachers work in schools where Aboriginal and Torres Strait Islander languages are spoken. Both kinds of experts include Aboriginal and Torres Strait Islanders.

Academic experts and practitioners may or may not see their interests as closely aligned. The criticism we received of the example of evaluation in the discussion paper in box 4.1 on page 123 (which we endorse and will discuss in Part 3) reflects the profound hostility that can exist between practising teachers and academic researchers. In contrast, the submissions on languages issues that the Commission received (mentioned earlier) reflect long-standing and close working relations between academic researchers and teachers.

There needs to be systematic consideration of the knowledge, skills and networks (both among each other and in the field and communities) of these two kinds of experts. They have very real stakes in evaluation outcomes. Their interests in the outcomes of an evaluation may be quite different from those of a politician or public servant.

For example, do these experts have any role in the Office of Indigenous Policy Evaluation? Or in particular steering committees? Should the Strategy include a place for specialised sub-committees or short-term committees for particular evaluations?

In short, we recommend that the revised documents should systematically address the question: *who are the stakeholders – across the board – and how should they be positioned and represented* in the evaluation process?

2.3.2 Knowledge

Over and above the stakes that different groups have in evaluations, and the different 'perspectives, priorities and knowledges' they can bring to this process, there must also be a place for **_knowledge per se_** – that is, what has been established as fact through consistent research findings and practical experience, and supported by credible theory building.

The discussion document places considerable emphasis on process, capabilities and competency but makes hardly any reference to actual knowledge and the need for those contributing to evaluations to be knowledgeable. How can assessments be made as to whether a policy or program is worthwhile (or not) if those designing and undertaking these assessments are not knowledgeable in one way or another about a policy/program (and its predecessors) and its context? (See also our earlier comments re credible evaluators.) This is not an abstract question. Evaluations of programs are frequently designed by government officials with little or no substantive knowledge about the actual programs and policies they seek to evaluate. I'll return to this point in discussing Chapter 7.

As asked earlier: where does knowledge fit into the principles that are described as informing evaluations? Does it come under 'credible'? We recommend that the background paper – and the whole conceptual framework of the Strategy – needs revision to systematically include reference to bodies of expert knowledge and research, and the specialist skills underpinning this knowledge, for example, in linguistics, education, medicine and the law.

PART 3: Comments on three particular chapters

Chapter 4 – Types of evaluation

We suggest a scan through the discussion document to determine if the description of evaluation types, approaches and methods is consistent, and especially regarding whether quantitative methodologies are assumed as the norm or preferenced. In places it seems as if different people have written different sections, or someone had a good idea in one part and added it there, when it is quite important across the board.

On page 122, there is a quote at the top of the page from the New Zealand Evaluation Association:

> *At the end of an evaluation process, an evaluation needs to be able to say whether something is any good, or not, and why.*

We would also add 'What are the criteria determining the *why*?' This issue carries through into section 4.1, where the report raises three questions:

1) *how well has the policy or program been delivered?*

2) *what difference did the policy or program make?*

3) *do the benefits of the policy or program justify the costs?*

These questions will not yield answers that inform good judgements about a policy or program. For example, you could ask these questions in relation to bilingual programs (or the forced removal of Aboriginal children from their parents):

1) *How well has the stamping out of bilingual programs and replacing them in the Northern Territory been delivered?*

Answer: it's been delivered well. But it's bad policy.

2) *What difference did the policy or program make?*
   Answer: It has made a huge difference, but the differences are undesirable.

3) *Do the benefits outweigh the costs*?
   But there's another question: The benefits to whom? and the costs to whom?

The key question is: 'Is the policy or program a good policy or program?' or 'Does the policy or program pursue desirable goals?' These questions need to follow from and be grounded in other parts of the Strategy that set out the values espoused by the Commission. Answers to these questions are not content-neutral or independent of value judgements and contexts.

In box 4.1 on page 123, we don't think firm distinctions can be made between 'evaluation type', 'evaluation approach, 'evaluation method', and 'data collection' in the way that seems to be implied. These distinctions don't work particularly well in education contexts. For example, if you wanted to ask 'how well is a bilingual education program working?', these distinctions would play into each other more than is implied. The elements in box 4.1 would be better shown as a circle with arrows going each way.

In box 4.1 it also says that the type of evaluation is dictated by the question, but very often the type of evaluation is dictated by the theoretical and epistemological stance of the evaluator. A commitment to ethnographic perspectives and approaches will determine the evaluation question(s) and the types of data collected.

On page 125, it's said that:

> *One of the key challenges for impact evaluation is coming up with an estimate of the actual effects of a policy or program. To do this, evaluators need to estimate the effects of a program on policy and program participants (the 'treatment group'), and compare this with non-participants (the 'control group') who represent the counterfactual (box 4.3).*

This approach doesn't work very well in educational contexts because too many variables are in play. The example given in Box 4.4 on page 127 is a case in point. It's a very good example of very bad evaluation in education. The evaluation set out to assess the effectiveness of a web-based learning tool compared to regular instruction. But the multiplicity of variables in both 'treatments' makes such an assessment impossible. In our view, attempting it is fruitless.

To start with, the 'control group' is described as follows:

> *A control group of 148 students received literacy education through regular classroom instruction*

Among the crucial variables that are not considered but could, each on their own, have crucial effects are: the quality of the 'regular' instruction (both the teaching and the pedagogic resources they teachers had at their disposal); the focus of the teaching (which can vary wildly); and where the 'regular' classroom instruction took place (the children's normal school? a laboratory or other special situation?); whether the 'regular' instruction was appropriate to the children's English

language proficiency; the students' ages and gender; the number of children in the 'regular' classroom compared to the treatment group; and the student-teacher ratio in each group. We are told that the students in the treatment group had trained teachers. But what level and focus of training did the teachers in the regular classroom have? Did the students in the regular classroom have teachers they liked (— student-teacher relationships have a huge bearing on student learning)?

We are told that:

> The study found that all treatment group students made significant gains compared with their control group peers in two key reading competences – phonological awareness and phoneme-grapheme knowledge.

If you train any group of people in two specific skills, they will probably get better at them, no matter whether or not it is web-based. Were these same competencies also the focus of the 'regular' classroom instruction? Or did the classroom instruction focus on other skills that, if tested, might have shown improvement? The fact that one group were withdrawn to have fun with 'rich animations' and 'game-like learning activities' no doubt helped their learning. But was the regular teaching as interesting or was it routine and boring? Was it the rich animations and game-like learning activities that made the crucial difference or the fact that these activities were web-based? The results prove nothing about the teaching tool but they might indicate something about making teaching and learning engaging, especially for learners who are struggling.

Crucially, the results don't tell us whether the treatment students were actually better *readers* than the regular group. If properly taught, the regular group may well have developed key reading skills at a higher level than those in the treatment group.

The most interesting result was that:

> Aboriginal and Torres Strait Islander students and students who spoke English as a second language were also more likely to drop out of the trial.

So: was speaking English as a second language a random variable not taken into account? And what is meant by 'students who spoke English as a second language'? We are told that the students were pre-tested for literacy but not whether they pre-tested for English language levels? Were the students who dropped out fluent English users or were they English language learners? Did they drop out because of unfamiliar cultural content in the trial tool, or because of English they couldn't understand, or both?

In our view, this kind of educational intervention and its supposed evaluation attempt to emulate laboratory experiments in the hard sciences. Educational settings are not conducive to such approaches.

Unfortunately, these kinds of pseudo-scientific educational evaluations are common. While some are simply misguided, others are frequently undertaken and reported by those with vested interests in promoting (and even directly selling) particular fixes for literacy and numeracy problems. Education authorities in Aboriginal and Torres Strait Islander contexts, along with others, are particularly open to these one-shot solutions promoted by educational shysters and carpet-baggers.

Box 4.8 on page 134 provides an example where the research focus has determined the outcome. The evaluation:

> *identified four different family types served by the RSAS, each requiring different kinds of support.*

But the focus on families of this evaluation has set families up to take the blame for children's poor school attendance. Clearly, it's important to identify and address families' aspirations for their children. But (at least as described) the research seemed to ignore a key question regarding what the schools were offering these families. 'Unsure families' were described as:

> *Not convinced that the education would lead to opportunities for their children*

If these families were in remote communities where English is rarely or never used in everyday life and sent their children to schools that took no account of their language, delivered the curriculum in English and taught literacy assuming English proficiency, these families are highly likely to be unsure about what schooling can offer. But if the children's schools started from where the children were at, and interacted more with families, maybe the families would cease being unsure. This is a potential example of where the research focus leads to 'blaming the victim' rather than interrogating the context.

Chapter 5 – Evaluating for quality results

Parts of this chapter seem biased in favour of quantitative approaches, and apply quantitative criteria to qualitative methods.

Observational methods are inconsistently referenced. In educational contexts, observational methods can be effective. The evaluator simply sits and watches what's going on in a classroom – with either a structured, or semi-structured, or an open or theory driven approach to what he/she is observing.

At the bottom of page 161, there is a quote from Patton:

> *Our very process of taking in information distorts reality.*

As I have already elaborated, if evaluators are not interested in language issues, they won't notice them. The next paragraph talks about cultural bias, where evaluation findings are influenced by the evaluator's own cultural values and perspectives. Blindness to language issues can also bias findings, for example, if the effects of using English in a given context are ignored or if it is assumed that written texts are the an effective mode of communication.

Chapter 7 – Embedding a culture of evaluation and learning

This was a really interesting chapter but could probe much further.

Here we come back to politicians as stakeholders. No doubt there are constraints on what the Commission can and can't say about politicians, government ministers and government officials/public servants. But these people are huge stakeholders in policy and program evaluation. We suggest that more attention could be paid to how and why political imperatives frequently override evidence or prompt its highly selective use – a cynical view would be that it's the norm. In

regard to literacy, it's almost predictable that every five or six years, an Education Minster will discover a literacy crisis. The evidence to support claims of this crisis and the consequent implementation of simplistic solutions is highly selective. It is in this context that pseudo-scientific evaluations like the one critiqued earlier gain traction.

The termination of bilingual programs in the Northern Territory is a prime example. That followed from a report showing that Territory Aboriginal children were failing dismally in NAPLAN tests. The NT Education Minister – who was, in fact, an Aboriginal woman – took fright, and decided that bilingual programs were the cause of the problem. (See what I said earlier about homogenising Aboriginal and Torres Strait Islander people and their knowledges, perspectives and priorities). Evidence and evaluations were irrelevant. She replaced bilingual programs with tightly focussed approaches to literacy and phonics, which, as the Closing the Gap data reveal, have also failed. The Northern Territory National Emergency Response is another example of a knee-jerk political response to a report revealing child abuse in Aboriginal communities.

The discovery of these crises in education and other domains is spurred on by the media, where pressure to produce an eye-catching story places politicians under pressure and acts against considered approaches to evaluation. Even where more in-depth and serious investigative media reporting is done, the fact that journalists lack knowledge (see above) means their investigations remain relatively superficial and often get crucial factors wrong (or half-right). We suggest this section might say more about the effect of the media, among other forces, in undermining a culture of evaluation.

A related issue is the impetus for incoming governments to change policies and programs. Elections are driven by the promises of oppositions to make changes. Promising to continue successful programs and policies is not good strategy. When new governments take office, they are compelled to make changes, and evaluations can be irrelevant. (An example from provision for adult migrants is the $4 million worth of citizenship materials developed when Philip Ruddock was Immigration Minister. Evaluations showed that teachers and students loved them. The incoming Labor Government scrapped them. These materials now moulder in retired teachers' cupboards.)

For a program to be properly evaluated, it needs to run for a while. Obviously, if a program is terrible, it should go, but programs are often abandoned just when they are bedded down and people have got used to working with them.

The institution of New Public Management approaches in the public service since the late 1980s have undermined commitments to a culture of evaluation. Career and promotion structures were instituted to incentivise individuals to move between policy areas. These incentives undermine deep institutional knowledge and learning from experience. I've encountered public servants with no knowledge of previous evaluations of policies and programs for which they have become responsible. The result is constant reinvention of the wheel and repeated evaluations that go nowhere. It is to be hoped that the Review of the Public Service referred to in the discussion document has tackled this problem. Embedding a culture of evaluation will require confronting the incentives for policy makers that work against this culture.

A cornerstone of New Public Management is outsourcing and delivery of programs through competitive contracting that has created a dynamic that is hostile to any genuine evaluation culture.

Outsourced programs allocated through competitive contracting are governed by contracts that seek to enforce compliance against detailed specifications and KPIs, which, by definition, can't be questioned by service providers. Commercial-in-confidence requirements and confidentiality agreements shield both these specifications and provider performance from scrutiny. On the other hand, contracting out absolves government from responsibility for what actually is delivered. (Thus, for example, the current contracts for English programs for adult migrants rely on paper-based audits of student assessments, which teachers openly acknowledge they sometimes fabricate in order to keep students who need English in their classes. However, the responsible public servants cannot tell Senate Estimates whether the methods of assessing students are valid and reliable, how many teachers are employed in these programs or their levels of qualification – the Quality Assurance provider is responsible for these matters.)

The ideologies that underpin program delivery need to be taken into account. An evaluation culture requires a commitment to professionalism and honesty that trumps self-interest. It requires trust and open communication between parties. It requires a willingness to report and admit mistakes and failures. If outsourcing and competitive contracting cannot be questioned – if competition is seen as the prime motivation for human behaviour – as has been the case at least until recently, an evaluation culture is impossible.

The question for this chapter might be: *how can an evaluation culture be established that counters these undermining forces?*

PART 4: The information requests

Re information request 6.1, which is about priorities, the top priority for ACTA is an in-depth quantitative and qualitative evaluation of the effects of NAPLAN on Indigenous children, which includes investigating how NAPLAN data reflects and does not reflect their educational achievements and needs, but also factors such as the impact students' self-esteem and commitment to schooling. NAPLAN testing plays a major role in mainstream schooling – schools advertise themselves around NAPLAN results; parents select their children's schools based on NAPLAN results; many schools teach to improving achievements in NAPLAN tests; private schools exclude students with poor NAPLAN results; and, as already indicated, policies that affect Aboriginal and Torres Strait Islander students are instituted on the basis of NAPLAN data.

A second priority would be a quantitative and qualitative longitudinal evaluation of bilingual teaching/learning approaches for Aboriginal and Torres Strait Islander students, especially in regional and remote locations, conducted by knowledgeable and culturally aware evaluators.

Re information request 7.3, which is about who should be members of the central evaluation body, ACTA recommends that academic and practical experts should be included at all stages of policy development and evaluation. We would be pleased to suggest individual experts.

****