Standing up to be counted: data quality challenges in Aboriginal and Torres Strait Islander higher education statistics

Neil Drew
Nulungu Research Institute, University of Notre Dame and
Australian Indigenous HealthInfoNet

Judith Wilks
Nulungu Research Institute, University of Notre Dame and
School of Education, Southern Cross University

Katie Wilson
Nulungu Research Institute, University of Notre Dame and
Victoria University of Wellington

Gillian Kennedy
Nulungu Research Institute, University of Notre Dame

Abstract: Data quality and availability in relation to Aboriginal and Torres Strait Islander students’ higher education participation and pathways remains a persistent challenge. In this paper we identify that, to date, there has been no systematic attempt to conceptualise and summarise many important aspects of data quality. The research reported in this paper, enabled through funding from an Office for Learning and Teaching seed grant, redresses this and proposes a conceptual framework for identifying and understanding the impacts of matters of data quality. We argue that the pursuit of a shared statistical literacy is best viewed through the dual lens of whiteness and Aboriginal and Torres Strait Islander terms of reference. Borrowing from the health sector, we conceptualise data quality issues as upstream, midstream and downstream. This framework identifies the locus of responsibility and intervention as a catalyst for purposeful action to address data quality challenges at the national, sectoral and institutional levels. The benefits of applying the proposed framework include a conceptual lens through which cultural issues may be unmasked; enhanced sector-wide critical statistical literacy; and a systematic accountability framework for assessing efforts to improve data quality. Finally, it is proposed that key elements from this framework might be usefully applied to the development of sector-wide guidelines for the collection, interpretation, use and storage of quality data and statistics to enhance the transition, participation and retention experiences of Aboriginal and Torres Strait Islander higher education students.
Introduction
The access, participation, retention, completion and transition rates for Indigenous higher education students are significantly lower than those of non-Indigenous students (Behrendt et al. 2012; Bradley et al. 2008; Devlin 2009; Kinnane et al. 2014; Wilks and Wilson 2015), yet understanding accurate participation statistics is very difficult as the two lead agencies for data collection, collation and analysis differ in the manner in which parity is conceptualised and calculated. For example, the Behrendt et al. (2012) review of Indigenous higher education suggested a population parity rate of 2.2 per cent as the aspiration, reflecting the proportion of the population aged between 15 to 64 years of age that is Aboriginal and/or Torres Strait Islander (based on 2006 Australian Bureau of Statistics (ABS) population statistics now ten years old). The Department of Education, Employment and Workplace Relations, on the other hand, argued for a parity rate of 3.1 per cent as an estimate of the proportion of Australian students expected to be Aboriginal or Torres Strait Islander peoples within non-Indigenous frameworks in Australia. Walter (2010:50), for example, cites recent policies and projects such as the Council of Australian Governments’ (COAG) ‘Closing the Gap’, in which, she argues, statistics are deployed for the purposes of ‘fixing the Indigenous problem’.

Further compounding efforts for statistical accuracy is that participation figures supplied by universities are reliant upon self-identification as Aboriginal and/or Torres Strait Islander by enrolling students. The resulting figures are almost certainly an underestimation given that some students may choose not to identify for a range of reasons (Wilks and Wilson 2015). Further, the Department of Education and Training statistics depend on the internal reporting processes of universities, which are known to vary by institution.

Statistics can be very persuasive yet are arguably one of the least understood of all contributions to debates of national importance. Indeed, statistics are often poorly understood by the very interests that invoke them, regardless of whether they are invoked in good faith or for the purpose of obfuscation or some other deceptive or manipulative intent. British Prime Minister Benjamin Disraeli’s oft-quoted statement that there are ‘lies, damned lies and statistics’ is not without merit. Since the advent of the abacus, statistics have been used to persuade; and they are easily manipulated. Nonetheless, statistics are crucial to our understanding of complex issues. This presents a dilemma — namely, we cannot do without data and statistics yet we are poorly equipped to manage both their use and their consumption.

In the Australian Aboriginal and Torres Strait Islander context, the use (and abuse) of data and statistics presents further layers of complexity. Researchers (for example Rowse 2009; Taylor, J 2011; Walter 2010; Walter and Andersen 2013) have interrogated motivations, ‘social interests’ (Watts 2003 cited in Rowse 2009:195) and practices associated with collecting data about Aboriginal and Torres Strait Islander peoples within non-Indigenous frameworks. Walter (2010:50), for example, cites recent policies and projects such as the Council of Australian Governments’ (COAG) ‘Closing the Gap’, in which, she argues, statistics are deployed for the purposes of ‘fixing the Indigenous problem’.

To date, the ‘objective’ statistical data and analysis in relation to Aboriginal and Torres Strait Islander peoples continue to generally reflect non-Indigenous epistemological, ontological and axiological viewpoints. These persist despite the efforts of key Indigenous advisory and representative groups such as the National Advisory Group on Aboriginal and Torres Strait Islander Health Information and Data and the National Indigenous Reform Agreement Performance Information Management Group. Gathering statistics, and the uses for which they are deployed, is a discursive act that serves social, political, cultural and historical functions.

As mentioned, data and statistics relating to Aboriginal and Torres Strait Islander participation in higher education are obtained from two main sources: the Department of Education and Training participation in higher education are obtained from two main sources: the Department of Education and Training (previously the Department of Education and the Department of Education, Employment and Workplace Relations) and the ABS. However other sources are used to supplement them; for example, the Department of Industry (previously the Department of Industry, Innovation, Climate Change, Research, Technology and Science), the Tertiary Education Quality and Standards Agency (TEQSA), individual universities through data collections contributing to reports such as the annual Indigenous Education Statements, and state Vocational Education and Training and private providers.
Australia is not alone when it comes to data quality challenges, and there are a number of international examples of the issues in higher education data quality and statistics (cf. Bonaccorsi et al. 2007; Deacon et al. 2009).

In the sphere of education, the manipulation of data and statistics for political agendas has found a particular resonance with the advent of ‘league tables’ and comparative indicators (Broadfoot 2004:4). Yet such data are often produced and presented in ways that many authoritative sources find dubious (Adelman 2010:1). In the Australian higher education sector many similar issues emerge but have not been systematically addressed until recently (Kinnane et al. 2014; PhillipsKPA 2012; Wilks and Wilson 2015).

Walter (2010:48, 53) identifies the embeddedness of ‘racial capital’ and the ‘presumptive objectivity’ in the collection of data relating to Aboriginal and Torres Strait Islander peoples in Australia. One example concerns the problems that can arise from data aggregation because of the often small samples and the relatively small numbers of Aboriginal and Torres Strait Islander peoples living in remote locations, which can render analyses vulnerable to claims of being ‘unreliable and not generalisable’ (Walter 2010:46). Moreton-Robinson et al. (2011:40) recommended a reporting template that measures percentages of Aboriginal and Torres Strait Islander students against the size of the Indigenous population in the university’s home state, rather than nationally, monitored by an Indigenous higher education body in order to provide clearer population parity.

Data collection in higher education ‘has evolved over the years in a somewhat haphazard way with little formal planning, subject only to occasional, ad hoc reviews’ (PhillipsKPA 2012:71). In the following discussion we question the extent to which statistics are accurate, reliable and valid in the mathematical or statistical sense but also, and importantly, their validity, reliability and representativeness in Aboriginal and Torres Strait Islander cultural contexts. We identify, summarise and analyse the main statistical and data quality challenges in higher education with a particular focus on Aboriginal and Torres Strait Islander people; and, finally, we offer a conceptual framework for understanding and analysing Aboriginal and Torres Strait Islander higher education data and statistics. It is proposed that this framework will aid clarity, analysis and interpretation of data and, importantly, provide a way forward for institutions to resolve the challenges identified in this paper. Ultimately, our aim is to begin a conversation whereby a set of data quality principles and guidelines to inform policy and practice balances the data needs of Aboriginal and Torres Strait Islander peoples with those of the existing non-Indigenous instrumentalities.

Project approach

In this work we recognise Aboriginal and Torres Strait Islander sovereignty, which values and privileges knowledges, voices and perspectives in research, and the importance of demonstrable community benefit flowing from research (AIATSIS 2012; Moreton-Robinson and Walter 2009; Nakata 2007; NHMRC 2007; Smith 2012).

The Office for Learning and Teaching project was implemented in four phases. Phase 1 involved a desk audit of available literature on data quality issues. In Phase 2 a draft discussion paper was developed as a trigger document for an expert panel consultation. Indigenous and non-Indigenous experts in the fields of statistics, demography, economics and higher education administration identified in the desk audit were invited to critically evaluate the findings of the draft paper. Further to this, three Indigenous and four non-Indigenous senior educators participated in a consultative conversational interview seeking their views regarding the key issues of data quality and the key challenges facing the higher education sector in this field. A satiation search strategy guided the recruitment of participants until no further substantive issues emerged. In Phase 3 a revised discussion paper was presented to Aboriginal and Torres Strait Islander higher education sector representatives and other stakeholders for comment, discussion and revision. And in Phase 4, following publication of the report by the Office for Learning and Teaching, the discussion paper, including a proposed draft data quality conceptual model, was made available for wider dissemination and feedback throughout the sector. This paper outlines the findings of our research using a conceptual framework for understanding and addressing the data quality challenges that were identified.

The case for elaborating data quality

Attempting to ‘count’ Aboriginal and Torres Strait Islander peoples is a surprisingly recent enterprise. According to the 1901 Australian Constitution,
Aboriginal and Torres Strait Islander peoples were counted but were excluded from the census (Martin, D et al. 2004). It was only after the 1967 referendum that an Aboriginal Enumeration Strategy was developed by ABS (see also AIHW and ABS 2006). This led to the development of an Indigenous self-identification question, and the shift from biological to sociological determinants for enumerating the ‘Aboriginal population’ (Rowse and Smith 2010:90). The exclusion of Aboriginal people from the census was not an apolitical act; it was an ‘active colonial practice’, and the ‘uncounted counting’ was emblematic of the assertion of colonial power and authority (Walter 2010:46).

Assumptions of Western knowledge, methodologies and research underpinning data collection and statistics have tended to exclude Aboriginal and Torres Strait Islander values and perspectives (Bodkin-Andrews and Carlson 2016; Martin, K 2003; Rigney 1999; Smith 2012; Walter and Andersen 2013). Located within such practices, Walter (2010:48, 53) identifies the embedment of ‘racial capital’ and the ‘presumptive objectivity’ in the collection (and subsequent statistical treatment) of data relating to Aboriginal and Torres Strait Islander peoples in Australia.

The so-called impartiality of Aboriginal and Torres Strait Islander statistical data collection in Australia is a problematic assumption, not least because it is implicated by the political and racial assumptions and values of those gathering data and framing the questions (Walter 2010). John Taylor (2011), in his analysis of postcolonial demography, is highly critical of the collection of Indigenous statistics for comparative purposes and benchmarking with non-Indigenous populations, suggesting that such statistics can be misleading since they do not reflect Indigenous worldviews or geographies.

The problem is compounded as ‘the more the story of the data [is] told from a non-Indigenous standpoint, the more evidence there is [to] embed that world view as the truth’ (Kukutai and Walter 2015:322). Bodkin-Andrews and Carlson (2016:793) identified what they termed ‘epistemological racism’ relating to minority groups and cultures, whereby ‘researchers are waylaid within a stereotypical notion’ perpetrated through the social history of dominant cultures.

Reflecting on the issues discussed above, we argue that it is important for non-Indigenous and Indigenous participants in the sector to (re)position themselves with a clear reflective dual lens (Drew et al. 2010) of whiteness and Indigenous terms of reference (Oxenham 2000). The dual lens promotes simultaneous reflection on the implications of white privilege (and the associated colonising practices) and Indigenous worldviews for understanding this contested and complex domain (Nakata 2007; Walter 2010). This mitigates against the dangers of ‘separate practice’ and promotes a deeper understanding of the intercultural space (Nakata 2013:6). The dual lens is not about ‘resistance to, and rejection of, Western theory and knowledge, theory and practice’ (Nakata 2013:8), but rather about seeking a shared space and adjusting the lens (Bodkin-Andrews and Carlson 2016:971).

A commitment to privileging Aboriginal and Torres Strait Islander knowledge is a recognition that data and statistics are not value free; they are socio-cultural, historical and political constructions that serve particular agendas (Walter 2010). When situated within the intercultural space in the way argued by Nakata (2013), they become a site for authentic critical reflection on practices that serve to reproduce rather than transform.

Strongly related to this is the importance of developing a strengths-based narrative (Gorringe 2011; Kukutai and Walter 2015; Sarra 2011). Discourses within the Aboriginal and Torres Strait Islander domain are replete with deficit-based thinking that has infused public policy and debate for many years. Walter (2010:45) described many data collections as the ‘statistical portrayal of Indigenous dysfunction’. Perversely, deficit thinking has served to promulgate policies, practices, services and programs that further marginalise Aboriginal and Torres Strait Islanders in Australian society.

The deficit approach underlies much comparative data and binary or dichotomous thinking (Kukutai and Walter 2015; Nakata 2013; Walter 2010). For instance, Walter (2010:53) provides an excellent example of strength-based versus deficit-based thinking by contrasting the question ‘How do Indigenous people grow strong?’ with ‘How poorly do Indigenous children perform compared to non-Indigenous children?’

The use and abuse of statistics

At the simplest level data are indicators or building blocks, and nothing more (Broadfoot 2004). However, once converted into statistics they assume the power to persuade, and as Walter (2010:53) points out,
'Statistics do not lie, but neither do they always tell the truth.' Statistics, despite their illusion of ‘value-neutrality’ (Jordan et al. 2010:352), exist within the frameworks in which they are interpreted and used to inform cultural, social, economic and/or political understanding and action.

The underlying factors and knowledge bases involved in the understanding (and misunderstanding) of statistics indicate the complexities of this domain (Gal 2002). Broadly speaking, as enunciated by Broadfoot (2004:4), the issues relating to the misleading or deceptive use of statistics may be categorised as one or more of the following:

- naive fallacy — a failure to properly understand
- deceptive fallacy — the manipulation of statistics, whether deliberate or otherwise, towards some desired end
- scientific fallacy — the belief that statistical or numerical data are inherently trustworthy or more worthy than other forms of knowing.

Succumbing to these fallacies will invariably 'pervert the route' towards understanding the complex and nationally vital issue of Aboriginal and Torres Strait Islander participation in higher education (Broadfoot 2004:6). Walter (2010:49) presented a similar categorisation of the 'political realities in which Indigenous statistics reside'.

**Interpretive/epistemic communities**

In the higher education sector there are multiple intersecting communities of interest. For example:

- professional (within which there are multiple communities — demographers, statisticians, social scientists and so on)
- bureaucrats and administrators (within government, the sector and community)
- higher education leadership
- lay community members.

To understand the complexity and diversity of stakeholders in the field we draw on the work of Throgmorton (1991, 2000) and others on the rhetoric of policy analysis. Various stakeholder groups can be conceived as different epistemic or interpretive communities — networks of experts who share ‘causal beliefs and policy goals’ (Cross 2013:142), each with their own unique language, syntax, rules, beliefs, motivations and cultures that lead them to comprehend issues, data and phenomena differently. These differences culminate in perspectives that are experience distant from one another despite an often superficially shared language (Throgmorton 1991). One goal of seeking a shared language is to move from experience distant to experience near. Experience near is most often found at the intersections of epistemic/interpretive communities. We seek to facilitate a shared statistical literacy that enhances experience near. However, this requires a deal of goodwill (or what Throgmorton (2000:376) called a ‘sustainable economy of spirit’). It is common for members of one interpretive community to devalue or even denigrate the beliefs and views of members of another interpretive community, usually based on stereotypical inferences (Throgmorton 2000). In this sense interpretive communities are also often socially constructed with respect to particular issues (Leiserowitz 2007). Indigenous and non-Indigenous participants in the higher education sector also constitute socially, historically, culturally and politically constructed interpretive or epistemic communities, and in this instance the experience near will be found in the intercultural space, as described above (Nakata 2013).

**Evidence-based decision making**

We are currently in an era where evidence-based policy making, or data-informed decision making (Allen 2002), is crucially important (Prout 2010). Data and statistics concerning the same populations are obtained from a variety of sources. Most of these sources have shortcomings, which are enumerated below. These data issues have proven to be frustratingly persistent despite the best efforts of many government, industry and research groups (AIHW and ABS 2006; Gilbert 2010; Trewin 2002). Further, frustration can arise from the transition of data into statistics, and discrepancies in the reporting of data for particular purposes (Jordan 2012). Biddle (2014:5) has noted that despite an ‘array of data about Indigenous Australians’, many gaps persist in ‘our understanding of Indigenous demography, health, socioeconomic status and wellbeing’.

Nevertheless, the ABS (2010) argues that ‘evidence-based decision making requires a systematic and rational approach to researching and analysing available evidence to inform the policy making process’. The ABS (2010) further states that evidence-based decision making has the following advantages, as it:

- helps ensure that policies are responding to the real needs of the community, which in turn, can
lead to better outcomes for the population in the long term
• can highlight the urgency of an issue or problem which requires immediate attention. This is important in securing funding and resources for the policy to be developed, implemented and maintained
• enables information sharing among other members of the public sector, in regard to what policies have or haven’t worked. This can enhance the decision making process
• can reduce government expenditure which may otherwise be directed into ineffective policies or programs which could be costly and time consuming
• can produce an acceptable return on the financial investment that is allocated toward public programs by improving service delivery and outcomes for the Australian community
• ensures that decisions are made in a way that is consistent with our democratic and political processes which are characterised by transparency and accountability.

It is incumbent on those utilising an evidence-based approach to ensure that the evidence is of sufficiently high quality to render it unimpeachable and particularly to interrogate it for signs of cultural insensitivity or blindness.

The need for a shared critical statistical literacy
Clearly the conceptual and methodological issues discussed above indicate the need for a shared critical statistical literacy. Walter’s (2010:45) observation that ‘statistics provide an interpretive mechanism for societal understandings across social, cultural, economic and political dimensions’ is a rallying call for enhanced critical statistical literacy. The ABS has a strong commitment to enhancing statistical literacy and describes four competencies that underpin the concept: data awareness; understanding statistical concepts; analysing and evaluating statistical information; and communicating statistical information. We do not think this goes far enough. Nor does it fully articulate the complexities of statistical literacy in cultural contexts. Writing from the health sector, Zarcadoolas et al. (2003) describe four key domains of health literacy: fundamental, scientific, community (civic) and cultural. While the first two are essential and foundational, the last two domains are crucial in the pursuit of culturally appropriate critical health literacy, as they are in pursuit of critical statistical literacy. Community (or civic) literacy refers to an ability to identify the motivations and agendas behind health policy and practice in order to name and interrogate them. This is known as the literacy environment. Cultural health literacy involves recognising cultural knowledge, world-views, customs and practices and understanding how they impact health and health-related behaviour. So, drawing on these principles, critical statistical literacy may be defined as:

The evolving skills and competencies needed to find, comprehend, evaluate and use statistical information and concepts to make educated choices. A statistically literate person is able to apply concepts and information to novel situations. A statistically literate person is able to participate in ongoing public and private debates about statistics, scientific knowledge, and cultural beliefs. This debate, in turn, advances statistical literacy, individually and collectively.

(Adapted from Zarcadoolas et al. 2003:119–20)

The important consequences of enhanced statistical literacy are:
• empowerment, providing opportunity for voice, agency and self-determination (De Walt et al. 2010)
• improved data-informed decision making
• enhanced capacity for cross-jurisdictional understanding
• enhanced capacity to recognise, avoid and refute naive, scientific and deceptive fallacies.

While many authors (cf. Gal 2002; Gundlach et al. 2015; North et al. 2014) have researched the idea of statistical literacy, the link between critical statistical literacy and critical health literacy pays useful homage to the insights offered by the experiences within the health sector regarding data quality issues.

Assertoric knowledge
As noted above, gathering statistics and the uses for which they are deployed is a discursive act that serves social, political, cultural and historical functions. We interrogate the discursive practices within the field of Indigenous higher education statistics and, in doing so, we do not seek ultimate truth per se but, rather, to create items of assertoric knowledge (Polkinghorne 1983).
Assertoric knowledge is asserted to the communities of interest in ways that make it available for debate rather than foreclosing on a preferred immutable truth. Assertoric knowledge guards against the development of uncritical polarities or binaries to create an authentic discourse in the middle ground. Assertoric knowledge claims, by their nature, are more open to ‘complex analysis, less certain conclusions and open to all ideas and politics’ (Nakata 2013:9–10).

The findings: statistical and data quality issues
The following section summarises our main findings in relation to statistical and data quality challenges as they relate to Aboriginal and Torres Strait Islander student participation in higher education. We separate these into three categories: upstream (those that derive from social structures and social policies, norms and practices), midstream (those that impact or emanate from the community) and downstream (those factors that have a direct impact). These are summarised in Table 1.

**Upstream**

**Data needs/scope**
What is the fundamental purpose for gathering data and deriving statistics? The answer to this is inevitably a values choice. In the higher education sector there are, of course, multiple operational purposes, including attraction, retention, transition, completion, course experience and funding. Recently, however, as one respondent noted, there has been a shift to higher order motivations, which include (or ought to include) institutional status, reward or profile linked to the achievement of these higher order aspirations. The participation of Aboriginal and Torres Strait Islander students in education is key to social inclusion and is a core goal of the Closing the Gap initiatives, and data must be tied to these overarching goals (Pechenkina and Anderson 2011). However, this association should not be uncritical. Jordan et al. (2010:40) note that while the Closing the Gap targets recognise the importance of culture, almost all the measures are a predetermined set of socio-economic indicators to the exclusion of ‘Indigenous life projects’.

If this is indeed the core purpose, then it suggests, and in fact demands, a strengths-based approach. This values-centred position has become an important feature of attempts to close the gap between Indigenous and other Australians, and quality data should serve this agenda and guide the selections and development of indicators.

**Data misuse and abuse**
Data quality should be judged against the fallacies outlined by Broadfoot (2004), above. The naive, the scientific and the deceptive fallacies should be identified, named and addressed. This, of course, is a central goal of culturally appropriate critical statistical literacy. For many authoritative figures, Australian Indigenous statistical treatments are ‘fraught with contradictory exclusionary practices’ (Walter 2010:46). Any data collections or treatments that serve to further marginalise, exoticise, pathologise, subjugate or ghettoise Aboriginal and Torres Strait Islander people through ignorance or Machiavellian intent are the antithesis of data quality and must be resisted. The ‘dual lens’ provides an opportunity for reflexivity in this respect. This challenge must be met through active engagement in the intercultural space (Nakata 2013).

**Data consistency**
Inconsistent datasets and processes for collection can lead to difficulties with understanding and planning for the successful transition of Aboriginal and Torres Strait Islander peoples into, and through, higher education (COAG Reform Council 2012:69; Kinnane et al. 2014). For example, analysis of the Australian Government’s higher education provider of undergraduate access and attainment statistics reveals different outcomes for a population-parity model at the state level compared to a national population-parity model.

One other source of frustration for users is the changes in classification systems from census to census (ABS 2013:Section 7). This means that there is an inevitable break in any time-series analyses. For example, classification systems have changed from 2006 to 2011, including occupations, industrial, cultural and ethnic groups, language, religious groups and countries. A final limitation of the census data is the lag between census dates and data releases. In a sense, all these limitations are inevitable and the ABS has an array of products and services designed to streamline access and suitability of data. So the task is perhaps to minimise, rather than eliminate, data limitations and better understand their impact on quality decision and policy making. The general
Drew et al.  

Standing up to be counted: data quality challenges

Table 1: Summary of data quality issues

| Upstream | Data needs/scope | What do we need to know?  
|          |                  | Why do we need to know it?  
|          |                  | This relates to the issue of strengths-based versus deficit-based thinking.  
|          | Data misuse and abuse | The political, social, cultural, racial motivations, including the impact of: cultural issues colonisation ghettoising/exoticising.  
|          | Data consistency | Jurisdictional differences among states. Incomparability between private and public sector representatives (such as schools and universities).  

| Midstream | Data completeness | Lack of Indigenous data from some sources. Missing or inconsistent data from some providers. Under-representation linked to self-identification practices.  
|          | Data collection methods | Culturally inappropriate data collection methods.  
|          | Data definitions | Misunderstanding or inappropriate definitions and nomenclature.  
|          | Data appropriateness | The appropriateness and sensitivity of data to cultural/community concerns.  
|          | Data levels of analysis | What is an appropriate level of measurement? Individual vs group vs sector.  

| Downstream | Data availability | Online access. Timeliness of reporting. Data for under-represented groups are scarce and not easily accessible.  
|           | Data type | Quantitative vs qualitative.  
|           | Data standards | Quality of indicators: reliability and validity sampling sample size.  
|           | Data storage and security | This includes the important issue of ownership and stewardship of data.  
|           | Data determinism | Data determinism refers to the tendency to use a 'scattergun' rather than a strategic and informed approach to data collection.  
|           | Data changes over time | Lack of longitudinal data.  

issues associated with the reliability and validity of census data are exacerbated for Aboriginal and Torres Strait Islander peoples (Martin, D et al. 2004). Evidence of inconsistency and lack of coordination between former departments integrated into the Department of Industry, Innovation, Science, Research and Tertiary Education in 2012 and between other state and Commonwealth government agencies has been noted. PhillipsKPA (2012:71–2) considered this led to ‘poor definition and design and multiple reporting of similar information’. The advent of the National Advisory
Group for Higher Education Data and Information was an important step by the Department to improve co-ordination and consistency across the sector, although this group has now been absorbed into the TEQSA Advisory Council and replaced by a Higher Education Data Committee (Department of Education 2013), which is overseeing the current departmental project creating a national repository for higher education data.

**Midstream**

**Data completeness**

Pechenkina and Anderson (2011) have noted that data relating to Indigenous student characteristics and socio-economic indicators were missing or inconsistent for some providers (COAG Reform Council 2012:69).

Nakata (2013) provides an excellent example of the issue of data completeness in higher education as part of his critiques of the politics of knowledge and knowledge production in the intercultural space. He contends that uncritical decolonising practices in pursuit of self-determination in Indigenous Education Units leads to the failure of some centres to utilise Western practices of statistical accounting of their student experiences, even at the demographic level. This failure stymies attempts to fully articulate Indigenous student experiences of higher education. In this failure of data quality, Indigenous Education Units and their leadership may become complicit in the reproduction of relative disadvantage for Indigenous students in the guise of self-determination and decolonising practice, as the deserved funding is not being attracted (Nakata 2013).

Self-identification is crucial to data completeness and has been a vexing issue for many jurisdictions, including higher education. Reliable and valid data and statistics in higher education depend on Aboriginal or Torres Strait Islander students’ self-identification at the time of enrolment, and therefore are almost certainly an underestimation given that some students choose not to identify for a number of reasons. Further, Biddle (2014:5) points out that the contexts within which self-identification questions are asked vary, the implication of this being that ‘an individual may be recorded as having a different Indigenous status across collections or through time, making comparisons quite complex’.

Data completeness is complicated by the politics of Indigenous identification, with contested and shifting perspectives on who can claim to have Indigenous heritage (Wilson and Barnes 2007). This is evident in the higher education sector, particularly in relation to scholarships and benefits (Kinnane et al. 2014), and also raises the issue of the contextual or circumstantial drivers of whether to identify or not. People may feel comfortable identifying in some circumstances but not in others; for example, for fear of discrimination or a lack of trust around how the data may be used (Kinnane et al. 2014).

**Data collection methods**

The literature is replete with examples of inappropriate data collection methods in Indigenous communities. This has led to widespread suspicion and mistrust regarding the motives of non-Indigenous researchers (for example Bodkin-Andrews and Carlson 2016; Martin, K 2003; Walter 2010; Walter and Andersen 2013).

The same is true of the motives of those who gather data for administrative collections. Genuine partnerships with Aboriginal and Torres Strait Islander communities are essential to gathering quality data (Kukutai and Walter 2015). In Indigenous contexts data collection methods and approaches require the building of relationships of trust (Taylor, J et al. 2012), the incorporation of cultural considerations, and the scoping of variations in definitions and terminology across different geographic areas and among peoples (Doyle and Prout 2012; Martin, D et al. 2004). Developing trust means engaging in authentic relationship building for the long term; ‘fly-in, fly-out’ methods will not yield optimum results. Authentic capacity building for local research-based practices is better suited. This was amply demonstrated in a recent demographic survey of Yawuru people in the West Kimberley (Taylor, J et al. 2012), in which the researchers were local community members trained and supported to gather data according to high standards of research practice. The net result was greater participation and significantly enhanced trustworthiness and credibility in the data, and thus significantly enhanced data quality.

**Data definitions**

The way that data are defined can have particular implications for Aboriginal and Torres Strait Islander collections. For example, the self-identification question for students at the point of
enrolment may be too coarse, and not sufficiently enabling for the definition of people's unique and complex heritages to be counted and to capture kinship or language groups, culturally appropriate relationships, historical associations and community representations. Conversely, there are other problems to be taken into consideration when coarse-grain definitions and statistics are applied, particularly for people who are unable to state more defined language group or clan group associations.

Many data collection practices do not distinguish 'Indigenous' from 'Aboriginal and Torres Strait Islander' (or use them interchangeably), nor do they clarify the distinctions among 'Aboriginal', 'Torres Strait Islander', 'Aboriginal and Torres Strait Islander'. Nakata (2013) has further questioned this by noting that both 'Indigenous' and 'Aboriginal and Torres Strait Islander' are essentially Western constructs.8

In the higher education sector specifically, slight differences have been noted between collections, adding to the 'workloads and levels of frustration within Universities' (PhillipsKPA 2012:73). We found that ambiguous or inconsistent definitions can also significantly affect the capacity to produce comparative data across the sector and further impact quality by lessening opportunities for benchmarking.

Data appropriateness

The appropriateness of data is an important quality consideration. Prout (2010) noted, for example, data on attendance may be insensitive to issues such as mobility, which is a particular issue for many Aboriginal groups in rural or remote communities.

Data levels of analysis

The levels at which data were collected (local versus sectoral versus national) are also implicated (and conflated with other issues discussed) in data quality. Data collected at too high a level of abstraction run the risk of homogenising Aboriginal and Torres Strait Islander groups. Conversely, data collected at the local level encounters accusations associated with difficulties of generalisability.

Data availability

The availability of data in the higher education sector also presents ongoing difficulties. For example, no data or statistics in higher education relating to Indigenous participation were available online prior to 2004 (Pechenkina and Anderson 2011). Data for under-represented groups are scarce and not easily accessible (Kinnane et al. 2014). Adelman (2010) has suggested that comparative statistics should be disaggregated based on a wider range of inclusion measures. This could, perhaps should, include under-represented groups such as prisoners and people with a disability.

At the sector level, PhillipsKPA (2012) identified timeliness as an issue affecting the availability of data, with reporting requirements changing as policies and programs change. Universities report that they are required to provide large amounts of data to the department but were then not provided equivalent access to data held by it in a timely manner. PhillipsKPA (2012:73) noted that 'universities would be far more willing to provide information if they had better access to it and if it could be used in ways which are useful to the institutions themselves'.

Data type

While the methodological bias against the contribution of qualitative data in favour of quantitative data has been largely ameliorated, decision makers in the higher education sector still display a preference for quantitative data to support their data-informed decisions. Appropriately balanced mixed methods offer the summary and comparative advantages of quantitative data and the richness and depth of qualitative approaches. Together they enhance the overall data quality, which has the potential to inform decision making. Myriad examples of high quality data are available from qualitative studies of Indigenous higher education (see Kinnane et al. 2014; Kippen et al. 2006; Rigby et al. 2011; Willems 2012).

Data standards

The impact of data standards is highlighted by Adelman (2010) in his illuminating critique of international comparisons. He argues that indicators misrepresent the situation when they fail to take a truly international perspective on the analysis. His analysis of the poor quality of indicators of participation, completion and study pathways demonstrates that a relative lack of sophistication in the analysis can have a profound impact on the conclusions drawn, which are likely to be almost entirely an artefact of the analysis rather than reflective of authentic trends. He shows that not including population trends over
time (for example) can seriously skew the data. In some countries, the declining proportion of youth means that the denominator falls and, over time, with a lag for numerator changes, will culminate in participation rates rising dramatically, while in those countries where there is a rise in the youth population, the opposite will result (Adelman 2010). Both are simply an artefact of the analysis.

It is also important to acknowledge the impact of the consumers on data quality standards. Statistically illiterate consumers of data and statistics are essentially poorly calibrated instruments with the capacity to significantly undermine data quality by promulgating the naive, the scientific and the deceptive fallacies. This degradation of the trustworthiness and credibility of the data can spread unquestioned throughout interpretive/epistemic communities, and in time assume a veneer of truth that is unsustainable under a more critically literate gaze. For example, to quote figures for self-identification in higher education without a full appreciation of the cultural, definitional and statistical issues underpinning the concept may lead to misapprehensions that perpetuate deficit approaches to Aboriginal and Torres Strait Islander participation.

At a more technical level, few consumers of data and statistics take the time to fully comprehend the data assumptions and treatments that underpin many (if not all) reported statistics. All statistical analyses embody choice. The choice not to report small sample sizes for statistical or ethical reasons has implications for the cultural sensitivity of the measure regarding community preference for analysis at the level of kinship or language groups, for example. In statistical literacy terms, it is not the choice per se that affects data quality in the public realm; it is the failure to appreciate the implications of often unavoidable choices.

**Data storage and security**

Data storage and security is essential to data quality and has significant cultural dimensions that are a consequence of poor research and data collection practices. Aboriginal and Torres Strait Islander peoples and communities have been ‘strip-mined’ for research purposes for decades, leading to very low levels of trust and confidence in data and statistics that are used to define their communities (Walter and Andersen 2013). This issue clearly goes beyond the administrative and legal requirements for confidentiality to include the culturally sensitive issue of ownership and stewardship of data that impacts Aboriginal and Torres Strait Islander peoples. An authentic sense of ownership and stewardship is needed to enhance trust and confidence.

**Data determinism**

‘Data determinism’ refers to the tendency to use a scattergun rather than a strategic and well-informed approach to data collection. We should strategically and purposefully decide what data are required and then set about seeking that data to the highest possible standard. Do not let the data lead. As one wry observer noted, ‘if you torture the data long enough it will confess’ (Coase 1981:27). A scattergun approach encourages data torture and eventually, inevitably, misuse and abuse of data. In higher education there is evidence across the sector of a tendency to a scattergun approach (PhillipsKPA 2012). By way of contrast, a directed, hypothesis-driven approach can also be susceptible to methodological biases (see Walter 2010; Walter and Andersen 2013). A balanced approach ensures the data are explored from multiple standpoints and methods to ensure consistency and to uncover misleading generalisations.

**Data changes over time**

A lack of longitudinal and cross-sectional data limits our understanding about transition from school to higher education (Andersen et al. 2008; Biddle and Cameron 2012; Biddle and Yap 2010; Wijesekere 2008; Wilson and Barnes 2007), and improved data are needed to understand school student mobility and how this affects educational participation and outcomes (Taylor, A and Dunn 2010; Wijesekere 2008). For example, the Department of Education and Training’s reporting methods have changed over time, making comparisons difficult if not impossible (Pechenkina and Anderson 2011). In their analysis of Indigenous student performance utilising enrolment, retention and completion rates, Pechenkina and Anderson (2011) note that in the 1990s completions were reported by field of study rather than by the provider.

The Department of Education and Training has undertaken reviews of reporting and data collection (Department of Education 2013; PhillipsKPA 2012). These reviews focused largely on the broader higher education reporting landscape and identified several areas of concern including lack of co-ordination; accumulation of reporting requirements over time; definitions and documentation; scale and
proportionality; instability in reporting requirements and inadequate planning for change; and accessibility and timeliness (PhillipsKPA 2012:71).

In 2013 the Minister for Education responded to the PhillipsKPA review, accepting all 27 recommendations relating to the issues identified above. The Department of Education is continuing with a four-year project seeking to streamline the collection and storage of higher education data and to aggregate data into a single repository. As noted by one respondent, this may be a ‘slow build’, with universities requiring considerable lead-in time and support to ensure the data that is supplied to the repository is robust. The TEQSA Advisory Council is the advisory body with this responsibility; the Department of Education and Training has established a Higher Education Data Committee to advise the department on higher education data collections and to act as a higher education data clearing house for the Commonwealth and states and territories (Department of Education and Training 2016). It is understood that this project also involves the development of a comprehensive higher education data dictionary and the development of a mechanism to cross-reference data collected by various agencies and organisations. There is no evidence to date of Indigenous membership, representation or consultation.

Discussion

Our proposed framework for disaggregating issues into upstream, midstream and downstream provides a clear signpost for understanding and responding at the right level of analysis or site of intervention. Clearly, different actors are implicated at the different levels. The upstream level will require the demonstration of leadership at the international, national and executive institutional level. Questions of leadership and commitment at this level are crucial and include the vitally important cultural commitments regarding the use and abuse of Aboriginal and Torres Strait Islander data; of what needs to be known and why; and of shared agreement across jurisdictions regarding the nature and scope of a shared critical statistical literacy. Without these commitments any effect on the other levels is likely to be piecemeal and ad hoc. Without strong leadership those in the midstream and downstream may lack motivation, be peremptory in their efforts and perhaps even cynical, as pointed out by one of our respondents. Compliance to liberate funding is clearly insufficient motivation and sends a very clear message that mitigates against positive enactment of data quality practices at all levels.

At the midstream level it is important to engage the Aboriginal and Torres Strait Islander community. A lack of purposeful commitment and culturally respectful motivation from those with the power and agency upstream has been identified as a problem by many Aboriginal and Torres Strait Islander observers. At the midstream level the concepts of intercultural space are clearly important. In this space our research findings suggest that reflecting on current practices through the dual lens of whiteness and Aboriginal and Torres Strait Islander terms of reference has the capacity to position people in the intercultural space with the potential for developing authentic and trusting relationships.

Downstream is the engine room for data quality. That engine room can only function effectively with the right types of guidance and support from the system that developed and engaged it. The pragmatics of data quality, including access, timeliness, reliability, validity, sampling data security, and the balance of cross-sectional and longitudinal methods, can only be assured by the right signals being sent from the midstream and upstream agents.

The above elements have been brought together with the intention of beginning a discussion among stakeholders about how to mitigate the challenges and capitalise on the opportunities that exist in the field of Aboriginal and Torres Strait Islander data and statistics in higher education settings. We contend that the conceptual framework described in this paper provides an opportunity to stimulate debate and action in this important domain. Rather than simply naming the issues, we have attempted to present them in a way that offers stakeholders a lens through which they can examine their own institutional practices systematically. A continuation of this important conversation throughout the sector, with subsequent measures being put in place, could make a significant contribution to the wider and important goal of instilling culturally appropriate and responsive policies, practices and procedures across the sector.

The overarching goal in the development of a framework for understanding data quality issues in Aboriginal and Torres Strait Islander higher education data and statistics is to develop a greater and more sophisticated shared statistical literacy among the diverse stakeholder groups in order to
support culturally appropriate and data-informed decision making nationally in the sector.

This will require four essential components:
1. a deeper understanding of the issues that impact data quality in the sector
2. a national commitment to culturally competent and safe practices in the development, collection, interpretation and use of Indigenous higher education statistics
3. a shared and agreed understanding of key indicators and their collection, interpretation and application
4. a national sectoral commitment to unmask and name issues that serve to obfuscate and politicise understanding of Indigenous engagement in higher education with the (mis)use of statistics and data

The outcomes of this research offer a range of purposeful and strategic pathways to advance both the debate and the practice of pursuing sector-wide collaborative efforts to address the issues of data quality in Aboriginal and Torres Strait Islander higher education statistics. We have offered a fresh and expansive conceptual framework for identifying, thinking about and responding to the issues. It is important that efforts be directed towards developing shared critical statistical literacy across the sector. It is evident that the wide range of epistemic or interpretive communities implicated in the sector has developed ways of representing and discussing the issues of data quality. The lack of a shared critical statistical literacy has culminated in a general incapacity to find a ‘common language’ at all levels — upstream, midstream and downstream. This ‘experience distant’ relationship (Throgmorton 2000) exacerbates communication difficulties that could be significantly ameliorated by the active pursuit of ‘experience near’. As noted earlier this would require a ‘sustainable economy of spirit’ (Throgmorton 2000:376) and a willingness to contribute items of assertoric knowledge in ways that make them available for debate rather than positional strategies in pursuit of disciplinary ascendancy.

**Conclusion and next steps**

In this paper we have considered a number of challenges, assumptions, complexities and possible approaches to improving Aboriginal and Torres Strait Islander data and statistics in higher education. We have outlined the context for improving data quality as it relates to Aboriginal and Torres Strait Islander higher education, and explored some specific statistical and data quality issues with reference to upstream, midstream and downstream elements.

We have to a degree, and for the purposes of clarity, artificially disaggregated the data quality issues into upstream, midstream and downstream but clearly they are integrated into a complex interplay of social, cultural, political and pragmatic factors that have the capacity to impact on intentions and actions. Our purpose was to offer a conceptual framework that provides an approach to systematically address data quality issues. Our higher order purpose is to stimulate debate in pursuit of national data quality standards and guidelines to inform policy and practice in this important domain.

In a forthcoming paper we will lay the foundation for a data quality framework. In anticipation of that paper some clear indicators towards the development of guiding principles have emerged from the investigation outlined in the current paper. We conclude that a data quality framework should (at least):
1. Name, interrogate and challenge the cultural insensitivity and inherent ‘whiteness’ of many current practices.
2. Avoid the ‘one-size-fits-all’ solution (Broadfoot 2004:4) and remain aware of initiatives in which progress is ‘far too subtle and context specific to be amenable to global targets’.
3. Focus on success indicators, as well as failure indicators, as part of a commitment to developing a strengths-based narrative in the higher education experience of Aboriginal and Torres Strait Islander peoples.
4. Name, interrogate and challenge the political context of data collection, interpretation and use (Walter 2010).
5. Agree on national standards for reporting. The ABS has a lead role in this respect.
6. Ensure quality management in data services — that is, what Allen (2002:1) describes as data that are ‘fit for purpose’.
7. Remain aware of the distinctiveness of context while not promulgating homogenisation of Indigenous experience/culture in pursuit of standardisation.
8. Agree on definitions. There is a need for an agreed standard question or approach to ascertaining Indigenous identification to enhance consistency and comparability across jurisdictions and sectors. It is also important to have articulated
processes to follow if a student does not want to answer such questions (AIHW and ABS 2006).
9. Establish continuing professional development opportunities to enhance shared statistical literacy.
10. In line with international human rights standards, ensure a commitment to Aboriginal and Torres Strait Islander self-determination in data gathering, including priority setting and decision making about the data to be collected and the methodologies and methods used. This also includes obtaining the free, prior and informed consent of those whose data is to be collected (Kukutai and Walter 2015:323).

In this current paper we have offered a point of provocation to challenge the dominant discursive agendas around the collection and use of data and statistics relating to Aboriginal and Torres Strait Islander peoples. At the very least it is hoped that in doing so we have stimulated a conversation among stakeholders about Aboriginal and Torres Strait Islander higher education data and statistics as they currently stand.

NOTES
1. In this paper we use the terms ‘Indigenous’ and ‘Aboriginal and Torres Strait Islander’ interchangeably. We recognise and, in fact, name this as a data quality issue in the paper. However, in the absence of evidence to support the change of nomenclature in sources we cite, we are ethically bound to use the terminology from the source reports. This is a rather ironic example of the need for data quality issues to be addressed systematically.
2. Attributed to Disraeli by Mark Twain but not verified in any of Disraeli’s writing thus often attributed to Twain.
3. The project received ethics approval from the University of Notre Dame Australia Human Research Ethics Committee on 1 August 2013 (reference number 013105F).
4. A key distinction exists between data and statistics. For example, Schield (2004) defines this difference as statistics summarising data, with the statistical summaries influenced by the contextual collection, selection and manipulation of data to output for specific reasons.
5. In this review we acknowledge that we are often conflating data with statistics; however, there is an indivisibility of data and statistics when it comes to articulating the issues.
6. Many, if not all the issues discussed, have been identified in other sectors such as health. Some have been explicitly identified within the higher education sector directly, while others impact collaterally as sources of data that inform higher education statistics. In order to present as comprehensive a picture as possible, all are discussed in this section. Clearly though, the locus of responsibility for their resolution will be shared across jurisdictions.
7. The issues are artificially disaggregated for the purposes of clarity and analysis. They are clearly interrelated in a complex web of (mis)understanding.
8. See earlier note regarding use of nomenclature in this paper.

REFERENCES
AIHW (Australian Institute of Health and Welfare) and ABS 2006 Recent developments in the collection of Aboriginal and Torres Strait Islander health and welfare statistics 2005, AIHW cat. no. IHW 15; ABS cat. no. 4704.0.55.001, AIHW and ABS, Canberra.
Behrendt, Larissa, Steven Larkin, Robert Griew and Patricia Kelly 2012 Review of higher education access and outcomes for Aboriginal and Torres Strait Islander people: final report, Department of Industry, Innovation, Science, Research and Tertiary Education, Canberra.
Standing up to be counted: data quality challenges


Deacon, Roger, Ruksana Osman and Michelle Buchler 2009 *Audit and interpretative analysis of educational research in South Africa: What have we learnt?*, Centre for Education Policy Development, Johannesburg.


Gorringe, Scott 2011 ‘Honouring our strengths — moving forward’ (keynote paper from the 2010 Annual SPERA Conference, University of the Sunshine Coast, Queensland), *Education in Rural Australia* 21(1):21–37.

Gundlach, Ellen, Clarence Maybee and Kevin O’Shea 2015 ‘Statistical literacy social media project for the masses’, *Journal of Faculty Development* 29(2):71–79.


Kinnane, Stephen, Judith Wilks, Terri Hughes and Sue Thomas 2014 *Can’t be what you can’t see’: the transition of Aboriginal and Torres Strait Islander students into higher education*, <www.olt.gov.au/project-transition-aboriginal-and-torres-strait-islander-students-higher-education-2011-0> accessed 14 June 2014.


Standing up to be counted: data quality challenges

Drew et al.


—— and Chris Andersen 2013 *Indigenous statistics: a quantitative research methodology*, Left Coast Press, Walnut Creek, CA.


Wilson, Tom and Tony Barnes 2007 'Continuing challenges in attempting to measure the size, and changing size, of Australia’s Indigenous population', *People and Place* 15(3):12–21.


**Professor Neil Drew** is director of the Australia Indigenous HealthInfoNet (www.healthinfonet.ecu.edu.au) and an adjunct professor at the Nulungu Research Institute. Prior to joining HealthInfoNet he was a professor within the School of Arts and Sciences at the University of Notre Dame, Australia (UNDA). He has a background in social and community psychology with over 25 years of experience working with a diverse range of communities and groups. Prior to joining UNDA, he was director of the University of Western Australia’s Institute for Regional Development. He was the program head of the Aboriginal Youth and Community Wellbeing program in the East Kimberley, established in 2006. He is co-author of the text *Social psychology and everyday life*. In his current role he has a strong interest in knowledge transfer/exchange.

<n.drew@ecu.edu.au>

**Dr Judith Wilks** is an adjunct associate professor at the Nulungu Research Institute at the University of Notre Dame, Australia and in the School of Education at Southern Cross University. Her research interests stretch across a number of fields including access and participation in higher education for Aboriginal and Torres Strait Islander students; participatory methodologies for children and young people; civics and citizenship education; and child-friendly cities. With colleagues from the Nulungu Research Institute Judith has completed a number of major projects for the Australian Government’s Office for Teaching and Learning, documenting successful and culturally appropriate practices in the enrolment, retention and course completion of Aboriginal and Torres Strait Islander students across the Australian university sector.

<judith.wilks@scu.edu.au>

**Dr Katie Wilson** is a researcher and librarian whose research interests focus on the experiences of Aboriginal and Torres Strait Islander children, young people in schooling and adults in higher education; qualitative, post-qualitative, participatory, collaborative and Indigenous research methodologies; and policy and program areas of Indigenous education and curriculum. Her recent doctoral research centred on the perspectives of Aboriginal and Torres Strait Islander students in relation to learning Indigenous knowledges and perspectives in NSW schools. Katie was a researcher in the School of Education, Southern Cross University, and has held a range of positions in academic libraries in Australia, the United States and New Zealand. Katie was born in Wellington, Aotearoa New Zealand, and is a descendent of the Te Atiawa people of Taranaki. She is currently providing strategic research and publication advice and support at Victoria University of Wellington Library.

<katie.wilson@vuw.ac.nz>

**Gillian Kennedy** is a researcher at the Nulungu Research Institute and co-ordinator of Arts and Sciences programs at the University of Notre Dame, Australia, Broome campus, where she lectures in Aboriginal studies. Gillian has a background in primary school education and community and international development. She has also worked in remote Aboriginal communities on ranger/Caring for Country programs and her research interests include Aboriginal and Torres Strait Islander education, working on country and community sustainability.

<gillian.kennedy@nd.edu.au>