NATIONAL TEACHING WORKFORCE DATASET

PROJECT REPORT

JUNE 2014

Produced on behalf of the AEEYSOC National Teaching Workforce Dataset Working Group

Funded by the Australian Government through the Teacher Quality National Partnership
ISBN 978-1-74361-918-6 [Powerpoint]
ISBN 978-1-74361-917-9 [PDF]

© Commonwealth of Australia 2014

With the exception of the Commonwealth Coat of Arms, the department’s logo, any material protected by a trade mark and where otherwise noted all material presented in this document is provided under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Australia licence.

The details of the relevant licence conditions are available on the Creative Commons website (accessible using the links provided) as is the full legal code for the CC BY-NC-SA 3.0 AU licence.

This material must be attributed as the National Teaching Workforce Dataset Project Report 2014 authored by Mike Willett, Daniel Segal and Will Walford (Ernst and Young) under contract with the Commonwealth of Australia as represented by the Department of Education who is the copyright owner of the material.

Copyright requests and enquiries concerning further authorisation should be addressed to:

The Copyright Officer, Department of Education, Location code C10MT1 GPO Box 9880 Canberra ACT 2601 or emailed to copyright@education.gov.au.

The terms of use for the Commonwealth Coat of Arms are available from the It's an Honour website.

Where a copyright owner, other than the Commonwealth, is identified with respect to this material, please contact that third party copyright owner directly to seek permission.
This report is one of a series of three on the National Teaching Workforce Dataset. The other reports are:

- Data Analysis Report
- Data Dictionary
Executive Summary

Background

This report is one of three that represent the culmination of nearly three years of dedicated effort. This effort has been led and championed by numerous individuals working across Australia with a passion and commitment to improving teacher quality through the provision of nationally consistent data. The work undertaken as represented in this report is motivated to support those who teach, support and inspire Australia’s school children every day.

The outcome from the effort has been to develop an initial National Teaching Workforce Dataset (NTWD). The NTWD has been designed to provide understanding and insight into school teachers across Australia. Though the initial NTWD itself represents a three year effort, the antecedents for a teacher level data collection and understanding have been discussed for many years prior.

This Project Report outlines in depth the background to the NTWD, the methodology, benefits and future options. Analysis and insight from the data collected is held within the Data Analysis Report.

Process

To produce this analysis required engagement with more than one hundred stakeholder organisations. After a process of evaluating data held by employers and regulators around Australia, twenty one organisations were able to provide teacher level data for the initial NTWD and six provided aggregate data. Addressing and managing data protection and privacy concerns was critical to custodians having comfort in releasing their data for classification and integration to form the initial NTWD.

The initial NTWD differs from other collections in a number of key areas that are critical in understanding and interpreting the findings of this report. These include:

- Teacher level collection, rather than collection of pre-summarised information, enabling greater depth in the questions that can be asked of the data
- Leverages existing data stores, rather than creating a new collection mechanism for an already data fatigued sector
- Alignment of jurisdictional data definitions to defined national standards, which enables more extensive comparisons but does flatten the richness of jurisdictional differences and experience
- Data integration across regulator and employer data to provide a richer level of understanding of each teacher

The initial National Teaching Workforce Dataset

The initial NTWD holds data on 440,313 members of the teaching workforce across the country, consisting of 313,791 ‘known employed’ teachers and 126,522 ‘additional registrants’.

As could be expected, within such a large workforce, there is incredible diversity in the demographics, qualifications and employment amongst teachers.

The Data Analysis report presents analysis of all the 45 data items initially specified to be included in the initial NTWD. Each data item, where data is available, is profiled and presented. Further in depth analysis is then presented into six areas of focus. The intention of these is to provide a more thorough investigation of areas that are of higher value in understanding the teaching workforce. These six areas are:

- Teacher age
- Aboriginal and Torres Strait Islander status
- Principals
- Teachers in low SES schools
- Teacher qualifications
- Registered teachers that are not employed
The Data Analysis report provides insight into the nature of Australia’s largest profession. Understanding the complexity and diversity of Australia’s teaching workforce through a teacher level data collection can support improved and informed decision making and provide direction for future research questions. The data in the Data Analysis report suggests areas for further investigation and represents a further critical step in understanding and supporting teachers.

Key observations on the NTWD process

The collection of teacher level data is achievable with cooperation between sector employers and regulators. Managing privacy concerns of data custodians, while also enabling datasets to be integrated is critical to having a dataset that allows for richer analysis. In the absence of a national teacher identifier, the methodology adopted by the NTWD proved successful in integrating datasets while minimising the risk of identification of an individual.

This Project Report outlines some of the key observations for a teaching workforce data collection. Developing clarity on future purpose for a collection of this nature supports determining the data items necessary for collection and analysis. This then supports the establishment of national data standards and adoption of a common language around terminology. Jurisdictional or sector differences in how the teaching workforce will always exist but this enables a national level discussion and benchmarking.

With data collected from all regulators around Australia, there is confidence that the teaching workforce has been substantially captured in the initial NTWD. Limitations do exist across custodians and specific data items. These are documented in depth in the Data Analysis Report and options on how to improve the quality of teaching workforce data, across key dimensions of accuracy, completeness, depth, access and consistency, are outlined.

Custodians indicated that there is a considerable burden of data collection on them already. Consideration of appropriate rationalisation and alignment to existing collection events (such as the Schools Census or re-registration) may reduce this burden.

The NTWD affords an opportunity to take a longitudinal perspective of the teaching workforce. The creation of an NTWD identifier would, if future collections adopted the same approach, allow for analysis of movement and change over time of the teaching workforce.

Providing benefit to participating organisations was always an objective of the NTWD. This objective was fulfilled in a number of ways including:

- Provision of data from the NTWD to enable custodians to generate their own queries that may not have been presented in the Data Analysis Report. The nature and extent of data release was enshrined within confidentiality agreements signed between EY and custodians, and supported by a letter from the Australian Government Department of Education.
- Data Analysis providing jurisdictional comparisons on selected data items to support benchmarking activities.
- Some custodians adopted the data standards used by the NTWD, and adopted these within their own internal information systems.
- Data quality procedures performed to establish national consistency provided direction to custodians on areas to address within their own information systems.

The long term storage of the data is an area requiring further attention. Processes to minimise the risk of identification of an individual while maximising the utility of the data to multiple audiences, including employers, regulators, researchers and the public, can be introduced to further the value and insight from teaching workforce data.
# Table of Contents

## PART I  PROJECT BACKGROUND  9

## PART II  METHODOLOGY OF THE NATIONAL TEACHING WORKFORCE DATASET  13

- Project Governance  14
- Methodology  14
- Data Management framework  20
- Classification Details  23
- Data Linkage  24
- Data Enrichment  25
- Data Release  25

## PART III  IMPROVING THE NATIONAL TEACHING WORKFORCE DATASET  27

- Value of Teaching Data  28
- Summarising the Way Forward  29
- Options for Future Collection  30
- Improving the Completeness of Data  31
- Non-Government sector perspectives on Data Collection  32
- Aligning the NTWD to other collections  33
- Improving the Accuracy of Data  34
- Achieving National Consistency  35
- Collecting deeper data with a Longitudinal Perspective  37
- Access to Data  38
- Future Release of Data  39

## PART IV  EXTRACTING VALUE FROM THE NATIONAL TEACHING WORKFORCE DATASET  41

- Using Data to Improve Teacher Quality  42
- Teacher Supply and Demand  43
- A Framework for Modelling Teacher Supply and Demand  44
- Options for Implementation  45
- Utilising the NTWD with Other Datasets  46

## PART V  INDICATIVE SUPPLY AND DEMAND MEASURES  49
Part I: Project Background
Project Background

The appetite for a national teaching workforce dataset collection has existed for some time and been realised in different forms.

In 1999, the Australian College of Education, funded by the Australian Government, sponsored the Griffith University Centre for Leadership and Management in Education and the University of Queensland, to undertake a national teacher survey, the “Teachers in Australian Schools”. Its purpose was to “to provide a detailed picture of the Australian teacher workforce, and to gather information to assist in future planning of the workforce”. This study repeated and extended three previous studies (in 1963, 1979 and 1989) sponsored by the Australian College of Education.

In 1999, 2001, 2003 and 2004, the Ministerial Council for Education, Early Childhood Development and Youth Affairs (MCEECDYA), published reports prepared by the Australian Government on “Demand and Supply of Primary and Secondary School Teachers in Australia”. Using aggregated data sourced from a range of datasets, these reports profiled the teaching workforce at that time, considered labour market trends and investigated supply and demand factors.

Since 1989, the “National Report on Schooling in Australia” has been undertaken annually with a small range of teacher data. Since 2009, responsibility for production of these reports resided with the Australian Curriculum And Reporting Authority (ACARA). The majority of data for these reports is sourced through the National Schools Statistics Collection (NSSC). The Australian Bureau of Statistics (ABS) also draws data from the NSSC to publish its annual “Schools, Australia” report.

The “Teachers in Australian Schools” survey was succeeded by the current “Staff in Australian Schools” (SiAS) surveys. This research has been conducted by the Australian Council for Educational Research (ACER) and have been conducted in 2006-07, 2010 and 2013 and was commissioned by the Australian Government.

SiAS 2007, which also included separate research into approaches to workforce data collections and planning processes, made the initial recommendation to “develop a process to achieve common core data sets and definitions, and noting the potential of this to facilitate the pooling and sharing of workforce data by government and non-government systems in the long-term”. In 2007, the then MCEETYA agreed in principle to “develop a process to achieve common [workforce] core data sets and definitions” and subsequently agreed to the development of a national teaching workforce dataset. This work was later moved under the Teacher Quality National Partnership (TQNP) Facilitation reform to improve the quality and availability of workforce data, overseen by the Australian Education, Early Childhood Development and Youth Affairs Senior Officials Committee (AEEYSOC). A sub group was subsequently tasked with this responsibility.

The first activity of the sub group, in conjunction with ACER, included production of a report titled the “Development of a Framework for a Nationally Consistent Dataset concerning the teaching Workforce across Australia” (Framework Report). The Framework Report provided a conceptual methodology, and defined key principles and protocols for consideration in the development of a nationally consistent dataset. Critically, the Framework Report also defined 45 data items that were recommended for inclusion in the collection. It also provided international benchmarking and options for implementation.

In 2009, AEEYSOC established a Working Group to undertake the work and subsequently approved its work plan the following year.
The TQNP agreement aimed to deliver a range of reforms to raise the standard of teaching and learning in Australia through innovative approaches to pre-service teacher education, professional development, leadership and workforce planning and management. A key reform under the TQNP was to improve the quality and availability of teacher workforce data to inform workforce planning. Better data is needed to assess current teacher shortages, identify future career intentions and the impact of significant events (such as the global financial crises) on teacher labour markets, as well as potentially informing national teacher reforms being developed and implemented under the TQNP.

The National Dataset was intended to provide a robust, responsive system for the classification, collection, storage, analysis and reporting of national and jurisdictional data pertaining to the Australian teaching workforce. It was intended to generate reliable data to monitor and report on workforce trends and to inform decision-making nationally and within and across education jurisdictions and sectors on a range of workforce planning issues including current and projected supply and demand in the teaching workforce. The National Dataset was intended to consist of collection of related data about the teacher workforce rather than a single set of data, as a broader collection is more likely to meet the needs across a range of users and more able to respond to changing data demands over time. It required the effective and efficient collection of data, which has a clearly defined need, to reduce the overall burden, to report the key findings from this data and to make the data available for detailed analysis by stakeholders.

In 2011, EY was contracted by the Australian Government on behalf of the Working Group to commence work to collect and analyse the initial National Teaching Workforce Dataset. The remainder of this report documents the process and methodology undertaken by EY in this task.
Part II: Methodology for the Interim NTWD
Methodology for the Interim NTWD

Project Governance

The development of the interim NTWD was supported by varying degrees of governance. The governance structures of the NTWD were also used to support the Longitudinal Teacher Workforce Main Study.

Following a selection process, EY (formally Ernst & Young) was chosen to develop the initial NTWD.

A Working Group, reporting to AEEYSOC, was chaired by a representative from Victoria Department of Education and Early Childhood Development (DEECD). In addition to Victoria, other members of the Working Group included representatives from:

- Australian Government Department of Education
- Australia Capital Territory Education and Training Directorate
- New South Wales Department of Education & Communities
- Northern Territory Department of Education and Training
- Queensland Department of Education
- South Australia Department of Education and Child Development
- Tasmania Department of Education
- Western Australia Department of Education
- National Catholic Education Commission
- Independent Schools Council of Australia
- Australia Capital Territory Teacher Quality Institute (representing members of the Australasian Teacher Regulatory Authorities)

Supporting the Working Group, a Stakeholder Reference Group was established. This group met quarterly initially and then, as required, to address specific questions raised by the Working Group or to have input into the process. This group was also chaired by Victoria DEECD and included representatives from the following organisations:

- Australian Government Department of Education
- ACT Teacher Quality Institute (representing members of the Australasian Teacher Regulatory Authorities)
- Victoria Institute of Teaching (VIT)
- Australian Bureau of Statistics (ABS)

- Australian Curriculum, Assessment and Reporting Authority (ACARA)
- Australian Institute for Teaching and School Leadership (AITSL)
- Australian Council of Deans of Education (ACDE)
- Australian Secondary Principal’s Association (ASPA)

A smaller project group comprising the Australian Government Department of Education, DEECD and EY met each fortnight to progress the work.

Methodology

The following section outlines the key steps undertaken to build the National Teaching Workforce Dataset.

Working Group members advised on appropriate contacts or organisations that may have held data that was appropriate for the NTWD. In total, more than one hundred organisations were identified. These organisations are mapped opposite.
It became apparent that only 36 of these stakeholders may have access to necessary teacher level data. This included the Government employer, Catholic Education Office (or equivalent), Independent School Association, and regulator for each jurisdiction.

In addition, four national organisations were also identified, including the Australian Government Department of Education, Department of Immigration and Citizenship (DIAC), the ABS and the Australian Council for Educational Research, who may hold relevant data.

Each of the 36 potential custodians was issued with a data questionnaire that sought to understand the availability of data and its appropriateness for inclusion in the NTWD. The questionnaire sought to understand and evaluate the attributes of the data held by custodians. The attributes used to assess the data were:

- **Existence:** the data is captured
- **Compatibility:** data can be incorporated into the NTWD
- **Completeness:** all data is captured
- **Accuracy:** data is captured free of error
- **Currency:** data is up to date
- **Depth:** historical data is held
- **Access:** technical and compliance issues do not limit access to data
- **Consistency:** data definitions align to those specified in the Framework Report
The data questionnaire was developed in Microsoft Excel and a screen shot of it is shown below. Controls were built into the spreadsheet to minimise the risk of data entry error and support users in completion of the questionnaire. Controls included the use of drop down boxes, locking of cells and help tips.

The questionnaire was designed to be applicable both to organisations that might hold teacher/unit level data (e.g., employers and regulators), as well holders of aggregate data (e.g. the Australian Government Department of Education). In addition, for each of the 45 data items, the following information was sought:

- Is the data item collected
- Is it collected at a teacher or aggregate level
- How the item is collected (survey, registration etc)
- When is the item collected
- How is the item stored (paper based vs. electronic)
- What system is the data stored in
- Does the definition used by the data custodian align to the Framework Report definition
- Is the format of data held by the data custodian consistent with that in the Framework Report
- Is the data accurate and complete
- Is the data current
- Is any history held
- Are there any limitations on accessing the data

Below is a screen shot of part of the questionnaire.

The questionnaire also asked respondents for the contact details of the person who had completed the questionnaire as well as the sector, state/territory and school level coverage of the response.

The questionnaire asked for a lot of detailed information in relation to the availability of data, and some custodians had difficulty with completing it. This is understandable as each questionnaire required consideration of up to TBD values. Other custodians commented that there was often duplication in responses across different data items, but that this actually made it easier to complete as a “copy and paste” approach could be used.

In parallel to identification of potential data custodians and issuing of data questionnaires, information sessions were held in the capital city for each jurisdiction and all identified stakeholders were invited to attend. An information pack was distributed in advance and outlined the purposes of the initial NTWD, the background to the project, timelines and key deliverables, the proposed methodology. Following the information session, potential custodians were given the option to ask further questions on the data questionnaire that had been issued to them.

Upon receipt by EY, questionnaire responses were assessed for data quality. In a number of cases, further clarification was sought from custodians, where responses did appear consistent or did not provide enough depth to enable evaluation of data to be made.
The questionnaire also had a hidden sheet which structured the data provided by custodians in a more useable, tabular fashion. This enabled direct input of responses from custodians into analysis tools.

Analysis of all data was performed to determine its suitability for inclusion in the interim NTWD. This analysis was performed on a data item level, as opposed to a whole of custodian perspective. The purpose of this level of granularity was to enable the WG to understand the precise availability of data and enable a later targeted data extraction request to a custodian for only items of agreed quality.

The WG clearly articulated a desire to collect data into the initial NTWD, even if this had limitations associated with this, on any of the attributes against which that data was being assessed. The criteria agreed to by the WG for inclusion of a specific data item is shown below and was specific to each data attribute. The criteria was set that even if a data custodian had significant limitations with either the existence of a data item or its compatibility to the definitions of the dataset, this data item would still be requested. This was to ensure capture of custodians who did not have complete data (for example, qualifications data was limited to the qualification required to teach but may not hold all other qualifications held by a teacher). Issues with data compatibility were to be addressed to Ernst & Young in handling collected data. This process of classification is detailed later in this report.

The WG did require the completeness, accuracy, currency, depth, access to and consistency of available data only have minor limitations or else not be included in the NTWD.

Once the data had been assessed, summaries were produced to understand the availability of data across jurisdictions. In some cases it was possible that data from the regulator may be able to be used where that data is not available from the employer. In particular this was used in jurisdictions where Catholic or Independent sector data was not available, and where the regulator has comfort in its sector data, based on a regular audit programme.

The summary of this analysis was presented to the Working Group for decision on which data to formally request.

In some instances, regulatory data was used to supplement gaps in employer data. For example, in Victoria, Queensland and the ACT, regulator data, identifying the sector was used to tag these teachers. For Victoria and Queensland, where Catholic and Independent sector data was unavailable for the employer, this provided the most benefit in terms of populating a number of demographic fields. These states were chosen due to advice from the regulator on the extent and coverage of audit activities performed to check that only registered teachers are actually working in schools.

The following two pages provide a summary perspective on the availability of each data item for each jurisdiction and each authority in that sector.
<table>
<thead>
<tr>
<th>#</th>
<th>Data Item</th>
<th>ACT</th>
<th>NSW</th>
<th>NT</th>
<th>QLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Year of Birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ATSI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Country of Birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Year of arrival in Australia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Previous occupation group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Qualifications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-12</td>
<td>Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-14</td>
<td>Institution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-15</td>
<td>Year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-16</td>
<td>Specialisations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Professional Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Area</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher Registration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Registering authority</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Conferral year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Teaching Restrictions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Specialisations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Registration Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Current Employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Year started with current employer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Time fraction employed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Type of employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Salary range</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Salary increment level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Current School or Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>School Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>School Sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33A</td>
<td>School Location - Remoteness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33B</td>
<td>School Location - Statistical Area 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Current Teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Subject area(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Year level(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Other duties</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exit From Teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Year of leaving</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Reason for leaving</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Destination</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher Supply</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Teacher education graduates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Teacher pools</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Net migration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher Need</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>Adequacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Current requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>Future requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### National Teaching Workforce Dataset

<table>
<thead>
<tr>
<th>SA</th>
<th>TAS</th>
<th>VIC</th>
<th>WA</th>
<th>National</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reg</td>
<td>Gov</td>
<td>Cath</td>
<td>Ind</td>
<td>Reg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Legend:**
- **Green:** Government schools
- **Brown:** Catholic schools
- **Light Green:** Independent schools

---

19
Once the WG was comfortable understanding the availability of data, there was discussion on how to move towards collection of the data and how it could be used. This included how to integrate the data collected between employers and regulators to get a more complete perspective of each individual teacher. This also needed to be undertaken within an environment that minimised the risk of any one individual teacher being identified.

An additional area raised by regulators was whether the Act under which they were established enabled the release of data. This also aligned to specific concerns about personally identifiable information. Regulators were also not a signatory to the Teacher Quality National Partnership, and as such had not received any funding support for either the extraction of data or to make legal queries on whether the data could be released. Arrangements were made with each regulator to enable them to receive advice and/or extract necessary data.

The NSW Institute of Teaching, in particular, had statutory limitations in respect to being unable to release data without changes to the Institute of Teachers Act.

**Data Management Framework**

A Data Management Framework was defined to provide all potential custodians with an understanding of the WG’s intention in relation to use of the data. This framework, developed after consultation with jurisdictional privacy commissioners around the country, also outlined the processes and controls in place, within EY, to minimise the risk of identification of any one individual teacher within the NTWD. The framework sought to minimise the risk of identification, both internally to the data collected and held by EY, as well as to release of data from the NTWD.

Prior to release of any data, each data custodian signed a confidentiality agreement with EY, which was supported by a letter from the Australian Government Department of Education. This outlined the processes from the Data Management Framework and stated that nothing could be released outside what was agreed to in the confidentiality agreement. This enshrined the approach documented in the Data Management Framework into a formal agreement.

It is worth noting that the Data Management Framework did not provide detailed directions on the processes or approvals required to be undertaken within each data custodian to provide data. The Data Management Framework concerned itself with the process of removing direct identifiers of teachers, while providing a way to integrate data from employers and regulators, and all subsequent activities to report on, and release data.

Data requests for specific data items were issued to individual data custodians

Data was extracted by custodians. In some cases, EY provided support with the extraction in terms of specific fields that were sought for capture. In addition to the data items requested, each custodian was requested to extract, if available, three specific data items for each teacher:

- Name and Date of Birth
- Employer Number
- Registration Number

The below steps in collection of the NTWD summarise the approach documented in the Data Management Framework.
**Key Creation**

EY provided an identical executable application to each custodian. This took the three data items (Name and Date of Birth, Employer Number, Registration Number) and used the SHA-512 hashing algorithm to provide a unique value for each of these items.

The hashing algorithm is such that for the same input, the same output would be produced. This then would enable later integrating of data across custodians without the need to reveal the underlying data. The application had some control built in to minimise risk of erroneous, including requiring data to be provided in a specific format, removal or special characters and numbers, and conversion to lower case, prior to hashing.

**Data Extraction**

Data custodians indicated that they needed to write scripts or use existing reporting systems to extract data from their systems. In many cases for any one individual, there were multiple pieces of data associated with one data item in the NTWD (e.g. teachers with more than one undergraduate degree).

Following the extraction custodians asked to assign the hash value to the extracted data. Custodians did not need to join tables as this was performed by EY.

**Management of the Quarantine Environment**

The quarantine environment held individual level data received from data custodians but no Personally Identifiable Information (PII).

Custodians provided their data to EY through a number of delivery mechanisms, including physical hand-over, secure file transfer, or secure courier. Once data was received by EY in the quarantine environment, it was loaded onto EY servers.

The data was checked to confirm that the upload was successful and basic profiling was performed to ensure data received aligns to that requested. Once data was confirmed as meeting expectations, data custodians were contacted and advised.

The file size of data received from custodians was not significant and disk storage space was not a challenge for this collection.
**Data Classification**

One of the key strategic decisions made by the Working Group was that data would be normalised to a national standard within the quarantine environment, rather than requiring custodians to perform this. This relieved some burden on the custodians and also reduced the risk of different approaches to national consistency. Data received from custodians was classified to, where possible, existing national data standards. In some instances, this leveraged data standards established by the ABS, in other this used standards outlined by ACER in the Framework Report, and in others EY created a standard, based on data captured. The classification proved particularly challenging for items captured in free text by custodians. In many cases, numerous forms of essentially the same data item were provided, with qualifications presenting the greatest amount of variation across the country.

The classification standards applied for each data item, for which teacher level data was captured is opposite. As can be seen from the table, there were a number of classifications for which an appropriate national standard was not available. This was not unexpected by the project team as many of the values permitted for some of these fields collected are very broad in nature. For example, employee classification within the NTWD only holds four values (principal, deputy/assistant principal, executive teacher, classroom teacher) and all employers allowed more variation than this. In some cases where an external data standard was used, this was then supplemented by other values to fit the captured data. An example, is qualification institution, where TEQSA standards are used but additional values for “overseas” qualifications was used in the NTWD.

<table>
<thead>
<tr>
<th>Data item</th>
<th>Data standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>ACARA Data Standards Manual 2012</td>
</tr>
<tr>
<td>Age</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Aboriginal and Torres Strait Islander status</td>
<td>ACARA Data Standards Manual 2012</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>ABS 1269.0 Standard Australian Classification of Countries (SACC) 2011</td>
</tr>
<tr>
<td>Qualification Level</td>
<td>ABS 1272.0 Australian Standard Classification of Education (ASCED) 2001</td>
</tr>
<tr>
<td>Qualification Field</td>
<td>ABS 1272.0 Australian Standard Classification of Education (ASCED) 2001</td>
</tr>
<tr>
<td>Qualification Institution</td>
<td>Based on Tertiary Education Quality and Standards Agency (TEQSA) National Register Code</td>
</tr>
<tr>
<td>Year of Graduation</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Regulatory Authority</td>
<td>Customised for NTWD</td>
</tr>
<tr>
<td>Years since Conferral</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Teaching Restrictions</td>
<td>Customised for NTWD</td>
</tr>
<tr>
<td>Registration Level</td>
<td>Customised for NTWD</td>
</tr>
<tr>
<td>Employment Status</td>
<td>Customised for NTWD</td>
</tr>
<tr>
<td>Years with Current Employer</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Time Fraction Employer</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Type of Employment</td>
<td>Customised for NTWD</td>
</tr>
<tr>
<td>Employee Classification</td>
<td>Customised for NTWD</td>
</tr>
<tr>
<td>Salary</td>
<td>Customised for NTWD</td>
</tr>
<tr>
<td>School Type</td>
<td>Customised for NTWD</td>
</tr>
<tr>
<td>School Sector</td>
<td>Customised for NTWD</td>
</tr>
<tr>
<td>School Remoteness</td>
<td>1270.0.55.005 Australian Statistical Geography Standard: Volume 5 – Remoteness Structure 2011</td>
</tr>
</tbody>
</table>
Classification Details

This section is not intended to be exhaustive in outlining all steps to classify the captured data, but rather to provide an understanding of the work undertaken in some key fields.

Aboriginal and Torres Strait Islander status

Custodians collected Aboriginal and Torres Strait Islander status at different levels. Where some custodians have data available to define teachers as either Aboriginal but not Torres Strait Islander, not Aboriginal but Torres Strait Islander, Aboriginal and Torres Strait Islander; other custodians only captured either Aboriginal or Torres Strait Islander. This does reduce the utility in the data and classification was needed to the later level.

Country of Birth

This field was generally supplied as a free text entry. Free text will be a common theme, in terms of challenges to classification, but in the case of country of birth, it was relatively trivial to normalise to the ABS standard. In some cases, Wales, Scotland or England was provided as the birth country, while in other cases this was the United Kingdom (UK). In these cases, all were normalised to the UK.

Qualification – general comment

The free text in the qualification field was often difficult to classify, due to free text and extensive use of acronyms in the data provided. Specific challenges with each individual item are presented next.

Qualification Level and Field

Qualification data items represented the most significant data classification challenge. Many of these were provided as free text, and level and field were often in the same entry (e.g. Bachelor of Education). Text mining tools were used to seek to extract the relevant information from data provided, but there were in excess of 250,000 unique qualification items provided, across all custodians.

Qualification Institution

This field posed similar challenges to the qualification level and field data. In addition to those, there were many qualifications that were conferred at institutions that do not exist, with the same name, as what they are now. To maintain currency, historic institutions were integrated to align to their present day name.

Employer data – general comment

In some cases, employers provided multiple records for an individual teacher where that teacher was working in different schools. The school where the teacher had the highest time fraction employed was used for that teacher in the NTWD. This established the location, school type and employment classification. Time fraction employed was summed across all records.

Employee salary

Where multiple records existed for a teacher, salary was calculated based on the time fraction employed. Some custodians supplied the annual salary in each record, which required checking to ensure salary was not over calculated.
**Data Linkage**

With the hashed value captured for each teacher, matching was possible across different custodians. With an intention to get as rich as record as possible about each teacher, the integration of employer data with regulator data presented a method to achieve this. Linkage was done across the three hashed values as a method to identify records in each dataset.

The figure below is sample, as represented in the Data Management Framework, of how the linkage across different custodians would be performed.

For any one teacher in an employer dataset, being matched against teachers supplied by the regulator, there are a number of permutations to determine the likelihood that these records represent the same person. The table below presents the possible matches that could have been found. Where an employer and regulatory record did not match the criteria, then this was passed to the next check. If no matches could be identified for a particular record, then it was labelled an orphan; either an employer record orphan or a regulator record orphan.

Analysis showed that that matching approach adopted was very effective in matching employer and regulator data to provide an enriched record. The data report highlights the extent of where data for a teacher was provided by a regulator but where no corresponding employee record could be identified (labelled additional registrants). The data below shows the matching rate, per jurisdiction, of employer data and where a regulator record could be identified. Excluding the issues with incomplete NSW data, and noting likelihood of transcription differences, the overall matching rate for the other states is 97.7%.

The suggests that the linkage approach taken was largely effective. Further investigation into unmatched known employed teachers was not possible due to the hashing process but with visibility of the differences, it may be possible to further refine the matching approach.
**Data Enrichment**

In addition to core data items, further data was added to allow investigation into focus areas as per direction of the WG.

**School location**
Many custodians supplied the school address within the data. These were geo-coded and then the appropriate ABS location was determined for the location of the school, this included SA1, SA2, SA3, SA4 and remoteness.

**Socio-Economic Status**
Using the school location, each school was mapped against the SEIFA (Socio-Economic Indexes for Areas) measures established by the ABS. Schools were then classified based on their SEIFA measure into one of five categories, from high to low SES.

ICSEA (Index of Community Socio-Educational Advantage) established by ACARA and reported on the My School website was not used.

**Additional data collected for the NTWD**
In addition, the quarantine environment also included data received from the Australian Government Department of Education, ACER, DIAC and the ABS. Public data from the ABS population census was also included.

Data was analysed for the purposes of this report. Analysis was provided of individual data items collected. In addition, the Working Group outlined a series of focus areas for further detailed analysis to identify trends and insights.

**Data Release**
A key objective of the NTWD was to provide data back to custodians to support their own analysis and benchmarking. To that effect, four releases of data have been agreed as below:

- **Release 1A**: A national release that may include geographic identification to SA4 with no sector information. This release is intended for employer custodians and the Australian Government Department of Education. A map of the 106 SA4s is shown below.

- **Release 1B**: A national release that provides no geographic data other than remoteness, and has a sector identifier. This release is intended for employer custodians and the Australian Government Department of Education. A map of the 5 remoteness areas is shown below.

- **Release 2**: An employer specific release with data only for that employer enriched with regulator data.

- **Release 3**: A to-be-defined release of data for ATRA members. This release is not included in the current confidentiality agreement but is committed to by the Australian Government Department of Education and EY.
Part III: Improving the National Teaching Workforce Dataset
Value of teaching workforce data

Data collection and usage currently pervades nearly all areas of society and is a key driver of innovation. Organisations, both within the public and private sectors, are looking at how they can use both the internal data they collect, and external data available to them, to improve decision making. Though sometimes challenged, the use of data and analytics is generally perceived to be a more objective method to understanding a specific subject area. Within the public sector, diligent data collection and analysis can yield valuable insight into policy and program performance and direction. The use and presentation of data also supports greater transparency and accountability for decision making. Valid and reliable data, while providing an understanding of historical and current trends, also lays groundwork for more advanced analytic applications, including modelling and forecasting future states.

Within the teaching workforce, a complete and unit level dataset would prove valuable to understand the current workforce and determine national priorities for achieving desired outcomes related to improved teacher quality.

Data and its application will not always provide the desired answer, particularly in an area as complex as understanding the Australian teaching workforce. As such, the value of data on the teaching workforce is not just to provide answers but also allow for the generation of more meaningful questions, that can then be used to direct additional and focussed research activity.

In the timeframe of the initial NTWD, much has been achieved in understanding availability and then collecting data to support an ambitious national collection of the teaching workforce. The data collected, as evidenced in the data report, is generally of good quality and, with appropriate use, allows for inference and insight to be developed.

The methodology adopted for the initial NTWD collection assessed data from custodians on a number of dimensions. In this section, we consider the possibilities should data be able to meet these qualities. We also discuss some of the most significant barriers to obtaining data and present options to address these. The data qualities that we focus on are:

- Completeness
- Accuracy
- Depth
- Access
- Consistency
Summarising the way forward

This page highlights a simplified roadmap to enhance the data captured on the teaching workforce, and subsequently the ability to interrogate the data to understand issues affecting teacher quality. Each area presented is intended to address issues to improve overall data quality as underpinned by the five dimensions. The coloured circles correspond to the data quality objective that will be enhanced by adopting that step in line with the legend opposite.

- **Evaluate usefulness of current data and determine core data items required to meet national and stakeholder objectives**
- **Establish data standards to support nationally consistent data and definitions**
- **Evaluate the opportunities that exist to rationalise and enhance data collections**
- **Consider alignment of teaching workforce data collection to existing collections (specifically NSSC and registration processes)**
- **Determine appropriate long term custodian for data collected**
- **Establish strategy for appropriate release of data across multiple interest groups, including interactive experiences**
- **Formalise longitudinal requirements, including approach to de-identification and integration across employer and regulator data**
- **Data custodians to assess impact of system and process change required to support national standards**
- **Impact assessment on VIC, NSW and QLD Catholic sector of providing unit level teacher workforce data**
- **Impact assessment of NSW BOSTES capturing data on all NSW teachers**

**Data Quality objective**
- **Improves completeness**
- **Improves accuracy**
- **Improves consistency**
- **Improves access**
- **Improves depth**
Options for future collection

There are multiple potential ways in which a future teaching workforce dataset could be undertaken. As a general principle, there is a trade off between improving the data quality objectives mentioned earlier, against the financial and time cost of implementing such change. The following table highlights three pathways forward across the dataset from the current “as-is” state through to an “improved” dataset to a “comprehensive” dataset. Broadly, “as-is” could be undertaken immediately, “improved” outcomes could be achieved within 2 years while comprehensive outcomes have a longer horizon for realisation. Option decision making though does not sit solely within a vertical column and solutions can be mixed to achieve required outcomes.

<table>
<thead>
<tr>
<th>Completeness</th>
<th>As-is</th>
<th>Improved</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Strong coverage across Government</td>
<td>• QLD, NSW, VIC Catholic sector engagement</td>
<td>• Integration with NSCC or registration enabling full capture incorporating Independent sector</td>
</tr>
<tr>
<td></td>
<td>• Full regulator coverage, except NSW</td>
<td>• Catholic and Independent sector gaps</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>As-is</th>
<th>Improved</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Reliance on custodian to capture data accurately with checks by EY</td>
<td>• Custodian liaison to update source data based on NTWD values</td>
<td>• All information verified</td>
</tr>
<tr>
<td></td>
<td>• Accuracy impacted by lack of national consistency</td>
<td>• Data standard driven and limiting of free text collection</td>
<td>• Custodian systems control data entry and permissible values</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Aligned to national and local needs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Depth</th>
<th>As-is</th>
<th>Improved</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Uses the hash algorithm approach which has proved very effective as an interim measure</td>
<td>• Employer requirement to capture registration number</td>
<td>• National teacher identifier from entry into study and through career</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Access</th>
<th>As-is</th>
<th>Improved</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Data releases are provided by EY in strict alignment to confidentiality agreements</td>
<td>• Long term storage identified</td>
<td>• Data released to open/linked data standards</td>
</tr>
<tr>
<td></td>
<td>• No additional release scheduled</td>
<td>• Development of risk framework for release</td>
<td>• Interactive portal with analytics and visualisation capabilities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Query access through portal</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consistency</th>
<th>As-is</th>
<th>Improved</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Uses a data custodian independent classification approach to achieve national consistency</td>
<td>• Establish and publish national standards for required data items</td>
<td>• Integration of national standards into data custodian systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Established processes for standards refresh</td>
</tr>
</tbody>
</table>
Improving the Completeness of data

While sampling methods can enable extrapolation of data across wider populations, a key benefit of a complete dataset is inherent in its ability to provide comfort in the conclusions drawn from it. While the data can age due to the dynamic nature of the teaching workforce, this impact is observed across all data collections.

The benefit of a unit level data collection is the ability of analysts and researchers to query the data to test specific hypotheses. Acknowledging appropriate privacy mechanisms, this granularity means that users are not constrained by pre-prepared summaries of data. This granularity also enables opportunity for appropriate integration of data across disparate datasets. This is discussed in more depth later in this report.

Examples of the type of questions a complete dataset could address include the following. Some of these have been able to be partially addressed through the current collection.

- How do teachers in lower SES areas differ from those in higher SES areas, and should there be programs put in place to address any differences?
- How many self reported Aboriginal and Torres Strait Islander teachers are there in Australia, and are they working to support indigenous communities, or are they elsewhere?
- What key skills or subject matter expertise will Australia lose over the next 5 years as baby boomers approach and reach retirement?
- Where do I have subject gaps, and should programs be implemented to attract new teachers to these subjects or try and attract qualified teachers back from other professions?
- What might the teaching workforce impact be of disruptive change in education – including change to education delivery, changes in funding etc. Is the teaching workforce adequately prepared for these changes?
- How might Australia seek to predict future teacher need?
- If I need to run a survey to address a specific issue, based on the question I have, how can I target this for the best response rate?

Regulator participation
All regulators actively participated in, and supported, the NTWD. Within NSW, the Board of Studies, Teaching and Educational Standards (BOSTES) could only supply data for all accredited teachers i.e. for those teachers entering teaching employment from 1 October 2004 onwards. This remains a gap for the non-government sector but which can be met by full provision of data by NSW non-government employers.

Employer participation
As noted, not all employers were able to participate in providing data for the collection. Fundamentally, this was driven by the more decentralised manner in which schools in the Independent sector, and some Catholic jurisdictions, are governed and managed. This impacts how data is stored and can be released. A statement from the National Catholic Education Commission (NCEC) and Independent Schools Council of Australia (ISCA) is provided.

Data burden
Data burden is a significant issue impacting the ability to collect, and analyse, the data needed to understand the teaching workforce. This issue was raised by a number of custodians, as part of this project. This is an issue that is wider than this collection but generally data burden is raised when the cost and impact of extracting the data exceeds the perceived value of providing the data. This lends itself to the adoption of a principle where data is captured once and used for multiple purposes. A unit level collection on teachers, such as the NTWD, lends itself well to be a key source for other aggregated teacher data provision and reporting. Appropriate design of data requirements, as highlighted earlier, both for the NTWD and extended to address other needs, should work favourably to reduce the data burden, though at some probable cost in utility to the data consumer.

Option:
An inventory of data collections, both national and jurisdictional, could be compiled to identify duplication and the opportunity for rationalisation. The creation of a single point of truth, that is appropriately accessible for multiple purposes should be, while idealistic, a guiding principle.
Non-Government sector perspectives on data collection

Statement from the National Catholic Education Commission

The Catholic sector has been active in its participation through the AEEYSOC National Teacher Workforce Dataset (NTWD) project. The sector supports the need for national data that will enable monitoring and reporting on workforce trends to inform workforce planning nationally and within and across education jurisdictions and sectors.

The amount of data, the number of surveys and their timing are issues that impact on schooling authorities with meagre resources and time to respond. An aim of the NTWD project was to reduce the amount of data being collected by reducing the duplication across surveys presently happening. It is recommended that all data collections should be collected at the one time, only once per year, using a provided online data site similar to the Australian Government Department of Education Schools Service Point.

The collection of data from the Catholic sector is complicated by the diverse governance structures in each jurisdiction. In some jurisdictions, the Commission does not hold nor collect the data being sought. The data is owned by the schools and their schooling authorities (22 in Queensland alone) and what data is available and its form is also just as diverse as the schooling authorities.

There are difficulties in meeting the demands of a model that requires data collected at unit record level. Data collected by some jurisdictions is in aggregated form and simply not available in “unit” form, in an electronic format nor collected by all authorities. This is notwithstanding the issues around complexity, sensitivity, confidentiality, ownership and storage and deletion of data.

To progress this project with the Catholic sector, particularly in Victoria, New South Wales and Queensland, the Government would need to negotiate the purpose, nature, availability and timing of a meaningful contribution from the Catholic sector and hold discussions on the content, form, collection method, security, confidentiality and frequency as well as costs inherent in any software, hardware and personnel requirements.

Statement from the Independent Schools Council of Australia

The Independent sector has been active in its participation through the AEEYSOC National Teacher Workforce Dataset (NTWD) project. The Independent sector represents about 15% of school students, and by inference will have 15% of the teaching workforce.

Given the very decentralised nature of the Independent sector, there will always be challenges with a data collection, especially one not mandated by legislation or other requirements. Schools are already saturated with data requests and while alignment to the annual financial questionnaire could be considered, the mixing of this with teacher workforce data may be confusing. The financial questionnaire is a challenge already and changes are often made, so there are already challenges around these collections that schools are coming to grips with.

Many individual schools may not be able to easily provide the requested teacher level data, though aggregates may be possible (at a school level) if appropriate support for the idea and concept can be engendered across the schools community. The concept is likely to be acceptable to schools but minimising the impact on schools needs managing. In this respect larger schools are likely to be better placed to support any data request, than smaller schools.

The articulation of benefits of the collection needs to be clearly made and what advantage they will gain from their involvement. AIS’s are generally not involved in the market or in workforce issues and ISCA would be expected to have a similar view. The nature of sector means there is no unified system of reporting and the Australian Government Department of Education could consider establishment of a portal but the IT and people requirements could be significant.

The collection of data at the point of initial registration and subsequent registration is a possibility as it moves the burden off schools but regulators would need to aware of the impact.
Aligning the NTWD to other collections

Some custodians would like consideration of integration of collection for the National Dataset with other collections. The following collection points are considered:

- **National Schools Statistics Collection (NSSC)**
  According to the ABS website, the NSSC is “census, conducted annually as a collaborative arrangement between State, Territory and Commonwealth education authorities and the ABS. Data is collected from the relevant authorities on a range of issues relating to schools, students and staff in primary and secondary schools throughout Australia, from both the government and non-government sectors. Schools, Australia (ABS Cat. No. 4221.0) publishes information on the number, age, sex, year/level, category of school, apparent retention rate and participation rates of students at both the State and national levels. As well, information on school staff is published at State and national level” (source: http://www.abs.gov.au/ausstats/abs@.nsf/DOSSbyTopic/6F7111FCBD0121C0CA256BD00027255B?OpenDocument).

The appeal of the NSSC is that it represents an established process and captures all schools in Australia. However, there will be challenges for schools should collection of data for the teacher dataset be integrated with this collection. In addition, the integrity and completeness of the teachers dataset may be at risk.

Schools are already feeling pressure to complete the NSSC in the determined time frames, and this is being further amplified by additions to the NSSC collection from 2014, including the Students with Disability. The NSSC is primarily also a school level, rather than teacher level, collection, and would require a significant cultural change programme to be implemented. A further challenge is within education authorities, generally the Government sector, who seek to provide some quality assurance of the data prior to submission. Any such activities undertaken on the teacher dataset would be significant, particularly for larger schools.

In respect to data integrity and completeness, the NSSC will only provide information on currently employed teachers in a school. This explains why there are some observed gaps between the NSSC and the NTWD collection numbers. There would also be no data captured for registered but not teaching individuals, and teachers who work in multiple schools may be captured on multiple occasions.

- **Registration**
  Another option exists to collect data at the point of registration or registration renewal. This will have a significant, and generally unresourced, impact on regulators and the implementation could be expected to take an extended period of time. This may require legislative changes as well to enable this to occur. NSW also has the challenge of an incomplete register, which is not a trivial issue and was highlighted earlier. There are also current challenges with the renewal period, which differs across jurisdictions. There will also be a reliance on self-reporting of much of the information requested, and subsequent data quality risks. This though is an existing issue in much of the data at present.

The benefits of this approach are that it removes the burden on schools to collect data, by moving this to the regulator. All teachers, including those not active, will be captured across all jurisdictions and sectors and a teacher will only be identified, and so reported, once (per jurisdiction that they register in).

**Option:**
A diagnostic activity could be undertaken that assesses, in more depth, the relative costs and benefits of leveraging a data collection through the NSSC or the regulators. This needs to consider legislative reform, as well as privacy considerations, system, process, communication and change management impacts. A more in depth comparison could then be undertaken to determine integration of the collection of teaching workforce data with an existing collection, as opposed to being undertaken stand alone.
Improving the Accuracy of data

Data collected by the NTWD was reliant on data quality processes within the custodian to ensure accuracy. Within the project team, fundamental issues of data quality based on what was provided were checked. These included:

• Values of data provided aligned to previously communicated expectations on permissible values (e.g. the gender field held only the expected three values).
• Referential integrity existed across different tables of data as supplied by custodians (e.g. an individual with qualifications also must have demographic information supplied, even if the values are null or unknown).
• Duplicate data being provided where the same apparent teacher had multiple different records.
• Outlying values beyond the expected (e.g. year of birth being 1850; or FTE > 1.0) or in the wrong field (e.g. a salary value in the country of birth).
• Inability to link records in provided tables for a single teacher without further supporting information (e.g. a single teacher with two salary entries in one table, and two FTE value in another but not shown which FTE value corresponds to which salary value).
• Metadata inconsistencies (e.g. advise from the custodian indicates we should receive 3,000 teachers but supplied data has 6,500).
• Missing reference table look ups (e.g. we are supplied with a number in a field but no way to determine what literal value the number corresponds to).
• Unknown classification used by custodians (e.g. teachers’ country of birth is supplied as a SACC code but some codes provided are not in the standard definitions).
• Low match between employer and regulator data. Typically this was the result of data quality issues in the capture of information (e.g. registration number), resulting in a low percentage match on the derived hashed values.

Limitations in extraction capability

An issue identified through the process was the varying ability of data custodians to support the extraction and analysis activities associated with the provision of data. For many, especially smaller, organisations, this was not a core role and is likely to be a causal factor for some of the issues identified above.

Many of the issues identified that impacted data accuracy can be attributed to the process of the NTWD being a new one to all custodians. It is probable that should future collections seek similar system based data extractions, then these errors would be reduced.

Other issues though are inherent within the system that capture and hold this information and these are beyond the scope of the NTWD project to address. However, a number of custodians commented that the visibility to their data provided through engagement with the NTWD project team yielded benefits for improved data quality. This included actions system owners took to directly remediate obviously incorrect data. Other custodians more proactively started to adopt frameworks to minimise the use of free text and enabled menu driven selections for some values. In some cases, this extended beyond the data required by the NTWD to other data. Some custodians looked to adopt the data standards of the NTWD as direction in terms of permissible values for their own systems. The above outcomes are not unexpected as a fundamental premise to improve data quality is to first expose the quality of the data.

Further foundational steps are required in data custodians to improve the accuracy of data. In depth discussion of these is beyond the scope of this report but directionally, these could be expected to include the establishment of data governance policies and procedures across the life-cycle of data management.

Option:

Aligning data standards to an accepted and agreed national standard gives custodians some clarity in respect to how they collect and organise their own data. Many of the data standards were customised for this collection. Should this data collection be continued, formalising the data standards could be undertaken to allow custodians to adapt their systems in capture data in this manner. This should then reduce the complexity of future data extraction and classification.
Achieving National Consistency

For the first time, data has been collected and integrated across jurisdictions, employing authorities and regulators. The classification processes that were undertaken, in consultation with data custodians, have enabled a nationally consistent dataset to be developed, that accounts for the differences in jurisdictions and authorities.

The process of classification to a nationally consistent data format means that information is lost as unique custodian values are collapsed into a more limited number of national categories. For a number of data items, historical antecedents have led to the need for different values to be captured in different jurisdictions. This is expected to have been driven by different policy decisions or different approaches to policy implementation both across and within the jurisdictions. For some custodians, attempting to classify their data to a national classification was sometimes problematic as there was a concern that it failed to capture the subtlety of how education is delivered or teachers are regulated.

The loss of information is the cost applied to be able to gain a national understanding and enable reasonable comparisons of the teaching workforce from across Australia. For a workforce, as large and as critical to Australia’s future prosperity as teachers, being able to develop a national perspective is a critical step to develop an objective understanding that can, and should drive, policy to close gaps.

In considering what information is important, it is critical to understand the extent of information loss that is acceptable. Primarily this centres in identifying an appropriate balance on maximising the utility of the data collected against the resource requirements to standardise the data to a national format. At one extreme, there is no, or very limited, change made to data received from custodians which will limit the extent to which national comparisons will even be possible. In this interim NTWD, this would be the data captured and held in the quarantine environment. At the other extreme, all data is transformed to a very limited number of nationally consistent values.

This extreme renders the data invaluable for analysis and the ability to make informed data decisions is lost. An easy to appreciate, if unlikely, example is if all ages were grouped into a single bucket of 15 years to 95 years. This provides no value as all teachers look the same.

To achieve this balance requires a combination of two key ingredients:
- Clarity on the fundamental purpose of the dataset and the understanding that is being sought.
- Knowledge about policy implementation across different jurisdictions and how this represents itself in data stored by custodians. This project has captured knowledge of the data for the in-scope items and has a high level appreciation of the policy that underpins the data provided.

The initial step to determine what understanding is required is essential to determine which data items should even be sought to be collected. Across the life of this project, this issue was raised intermittently with stakeholders expressing similar interests from the data, including:
- Teacher workforce profile
- Teacher supply and demand
- Understanding registered teachers not in classroom or school leadership positions

The above areas of interest share similar themes but different data might be required to fulfil all these requirements. The initial NTWD collection followed the ACER Framework report rigorously in the request for data and the formats that this was to be translated to. This leads to our first recommendation for future data collections.
Option:
A summative evaluation could be undertaken with data custodians to understand the value they received from the NTWD process and deliverables. This can then support moving to a position of clarity on the purpose, or purposes, of the dataset. This should inform and direct future requests for data and provide custodians clarity in respect to what their data will be used for, and how it contributes to a national understanding.

Once that clarity is provided, bringing this together with the policy implementation and data knowledge will enable clarity on feasibility of currently held data to be able to meet needs. This step also includes outlining the level of granularity needed to be maintained in the data. This is represented by the different classifications that will be required to achieve the objectives.

As mentioned previously, the current collection followed the ACER Framework report and did not seek to challenge the classifications that this report recommended. The Framework report classifications were only adapted when it was apparent that they were not suitable for the data that was being received, They were not changed based on a changing reporting or analysis need.

Refinement of the data required and how it will be classified will require a process of facilitation. For some data items, such as demographic areas, this activity is relatively trivial. For others, such as those related to registration level, this is more complex.

It is unlikely that needs will be able to met “out of the box” by custodians, and two key challenges, would be expected to emerge that would threaten being able to obtain the required data to meet the stated objectives. Both of these were seen in this collection and the remedial activities taken to address them are presented.

• Data custodian holds the data but not in the format
  Of all the challenges, this is probably the easiest to address and this project did this through the classification methodology outlined earlier. A fundamental decision is required about whether this should classification should be performed centrally or at the data custodian level. This project performed this task centrally as it reduced the risk of different classification across different custodians. Some custodians, as a result of this process, are considering adapting their own data collection processes and systems to align to this collection, which will reduce the need for centralised classification. This should be encouraged but only at a time when the earlier steps of clarity of purpose and required data have been endorsed.

• Data custodian does not hold data or not in an easily useable format
  This represents a more significant challenge as this requires changes in the processes or systems, at the data custodian, to enable this data to be provided. The extent of change to existing practice could have significant resourcing implications. A simple, to understand, example is the Aboriginal and Torres Strait Islander data item. As this is self reported, custodians generally capture this in three ways: Reported Aboriginal and Torres Strait Islander, Reported Not Aboriginal and Torres Strait Islander and Not Reported. This approach enables comparison between those that have reported. Some custodians though only capture this as Reported Aboriginal and Torres Strait Islander and Not Reported, enabling a different type of comparison to be performed but not necessarily one that enables Aboriginal and Torres Strait Islander vs. Not Aboriginal and Torres Strait Islander analysis. A change would be required to enable this to occur but has highlighted earlier, this needs to be driven by the objectives sought from the data.

In other instances, it may be that data is either incompletely collected or not collected at all. An example relates to limitations around qualifications data. While some jurisdictions, such as the ACT, enjoy the benefits of being smaller and more recently established, and so being able to define data requirements up front, this is not the case nationally. Perhaps the most obvious current example relates to the NSW regulator. Significant resourcing, and potentially legislative and regulatory change, may be required to enable data to be requested. This would then be followed by project management effort to then capture the data.
Collecting deeper data with a longitudinal perspective

As much as there is benefit from being able to profile and understand the teaching workforce at a point in time, being able to have an understanding of how the teacher workforce changes over time, will be a benefit from future collections. This can be enabled in two broad directions:

Macro changes in the teaching workforce

The first level of analysis over time would be to understand how key macro aggregates and averages change across subsequent years of the data collection. This is a common approach across sample based collections that are taken at different times. The Staff in Australian Schools survey is an example of this.

The granular level of the data though may better enable an assessment of the impact of policy implementation in specific geographic, socio-economic or teaching areas. For example, the data captured provides some perspectives on Aboriginal and Torres Strait Islander teachers and how they differ from non Aboriginal and Torres Strait Islander teachers. A future collection could provide some perspectives on the extent to which outcomes from the More Aboriginal and Torres Strait Islander Teachers Initiative (MATSITI) are being achieved.

Critically, continued collection could provide a feedback loop that supports continuous performance improvement. By identifying where programs have, or have not, met their objectives, further investigation could be undertaken into what factors have influenced success. This further investigation may be able to be supported, in some part, by the data but more likely other activities will be required to gain this understanding.

Individual changes in the teaching workforce

The methodology undertaken created, in the absence of an existing national identifier and to manage privacy, a series of hashed key values to uniquely identify each teacher. Should the process be repeated, and the hash values again used, then this will enable, at an individual level, changes over time to be analysed and understood, and then aggregated over a national population. The granularity of the data enables more insight to be gained into these changes.

This is possibly best demonstrated by the sample questions that could be addressed, in the call out box, below.

With future data collections, macro questions that could be addressed include:

- Is the teaching workforce more rapidly that the population and in what areas is this most obvious?
- Has the demographic profile of teachers in remote Australia changed and in what ways?
- What additional qualifications do early stage or graduate teachers have, and how has the mix changed?

Option on a universal identifier

The use of hashed values proved an effective, interim and low cost approach to integration of data between regulators and employers. It is though a work around developed for a specific issue at the time and better long term identifiers exist.

The first option is the establishment of a national teacher identifier. This would allow an individual to remain identified in the face of cross sector and cross jurisdictional employment movement but the administrative requirements to implement this, particularly in the absence of any other purpose make this problematic. The second approach would require employers to capture, in an accessible format, the registration number (or numbers of registered in multiple jurisdictions) of their teachers. This is already done by some employers and would provide a simpler mechanism, than name, to integrate data with employers.
Access to data

In the context of the current collection, access to data was focussed on the ease with which custodians could extract the requested data. Most data was stored electronically, and where this was held in a paper based form, this was not requested to be provided. The challenges with digitisation of data are significant and it is not clear that a compelling case for change exists to warrant this be undertaken.

In addition to capture of data, access also, and further in this section, refers to making the data available to other users.

The methodology detailed the scheduled releases of data. These support those stakeholders most closely aligned with the NTWD project but do not provide scope to provide data for education interest groups, researchers or the general public.

This yields two related questions:

- What is an appropriate organisation to store and host the initial NTWD?
- What level of access should be granted, beyond current data releases, to make the collected data available for interrogation by other users?

To address these questions requires a recognition of two competing priorities:

- Recognition of the importance of privacy and minimising the risk of identification of any individual
- Utility of data to researchers. This report has previously commented on the utility of the data through adoption of classification algorithms. In this instance, utility is used more widely to consider the utility of a dataset provided based on the application of anonymisation techniques used to protect it


Within this context the data releases provided a simple model for the release of data. Protections put in place to minimise the risk of identification of an individual included:

- Size of any aggregate must include at least 10 teachers
- Release of only two tables of data to minimise any risk or cross tabulation across released datasets

It cannot be underestimated how critical data custodian support is to determining an appropriate balance on data release. This recommendation assumes appropriate engagement across employer and regulator data custodians.

Consideration of long term storage custodian

An appropriate long term storage custodian, beyond EY’s involvement, is required to be identified. A number of organisations already exist that are expected to have appropriate data security and portals for access to data. This includes the ABS and the Australian Data Archive, though others may exist.

The WG should establish criteria to determine suitability for this dataset and then evaluate candidate organisations against this criteria. Criteria, not subject to any weighting, could include:

- Data protection
- Access to data, including online/portal driven
- Where required, procedures to authenticate, authorise and approve access to the data for new users
- Cost to maintain
- Ongoing involvement support required by subject matter experts

It is worthwhile to reinforce that amended confidentiality agreements are required before any data can be moved from EY to a long term storage custodian.
Future release of data

Options on data to be maintained
The methodology adopted, from data collection to analysis report, resulted in the data being subject to significant transformation and it is held in multiple iterations of that transformation.

The following versions of the data are the most essential from a reproducibility perspective as well as querying capability. These should be considered to be maintained:

- Raw data as provided directly from custodians. This data is that has not been subject to any change or modification as part of the classification process. This data could be supplied to researchers who wish to adopt a different classification method to the data; for example to address more specific question for which the classification used is too general.

- Modified and linked data. This data has been subject to classification and linking and is the basis from which the analysis report has been prepared. This would be expected to form the basis for most future queries of the data by users.

Options on future data release
To support wider possible release of data, an option is to develop a risk based methodology. This involves establishing criteria to determine the risk of identification of an individual, while still delivering a dataset that holds value to the user.

Risk based methodologies for data release are attracting increasing attention for seeking to quantify the risk of identification of an individual based on the utility of the dataset that is proposed to be provided. The balance can then be adapted based on the risk appetite of each organisation. For example, a release of data to a trusted third party, bound under contractual obligations for securing that data, may have a lower risk profile than a public release of data. As such, the granularity of data released may differ.

The subject of data anonymisation (masking and deidentification) is beyond the scope of this report. The National Statistical Service, led by the ABS, provides guidelines for (http://www.nss.gov.au/nss/home.nsf/NSS/BA2EF335C739A1CECA25763F000B81A4?opendocument#11.2).

Comment on Open Data
Open government is often underpinned by open data and providing appropriate access to meet the requirements of different stakeholders. The Office of the Australian Government CTO within the Department of Finance and Deregulation, currently maintains the data.gov.au website where users can find, access and reuse public datasets from Government. Determining an appropriate strategy for public release of data is a task that requires further investigation.

Options on future data interaction
Today’s volumes of data mean that many organisations are seeking to provide compelling and visual experiences to enable non-technical audiences to engage with complex data. Modern browsers, as the general entry point, provide opportunities for users to explore and enhance the understanding of underlying data. Should a public release be determined then embracing a visual aspect should be considered.
Part IV: Extracting value from the National Teaching Workforce Dataset
Using data to improve Teacher Quality

The NTWD in its own right has value in improving the understanding of the teaching workforce. However any future NTWD can extend the value of the collected data.

This section presents two future directions that the NTWD is already designed to support:

- Integration with other datasets to drive increased value. This was evidenced in the collaboration with the MATSITI project and the greater depth to which the Aboriginal and Torres Strait Islander teachers were able to be explored
- Workforce supply and demand modelling. Different options for modelling, and respective advantages and disadvantages are discussed
Teacher supply and demand

Workforce planning is essential for the delivery of effective education outcomes, whether this be at a national, jurisdictional, authority or school level. At a national level, workforce planning would seek to align supply and demand to ensure equitable delivery of education across Australia. There can be many factors that influence both supply and demand and forecasting is a complex area, with a number of influences. In addition, within a national model, there would be expected to be a number of variances at a more granular level, including geographic areas (remote and regional areas), subject areas (maths and sciences) and school type (primary vs. secondary) as examples of commonly stated supply shortages.

The data collected by the NTWD is supply side and not fully adequate to develop an appropriately rigorous model. As such, in this section, we present two different workforce planning frameworks and make commentary on their potential application and development. This section is not intended to be a comprehensive discussion of different approaches but rather to consider some of more popular approaches within the context of school teaching in Australia, based on learnings gained through the NTWD project. Our focus is more specifically on the supply side of the model.

While both approaches presented adopt the same framework (as presented opposite), the fundamental difference between these is the granularity of data required to act as input for the model.
A framework for modelling teacher supply and demand

Determining supply and demand requires gathering of available information that may influence either of these. Some examples are below, but this should not be considered to be exhaustive. Although there may be many influencing factors, models can quickly become overcomplicated and difficult to manage. Appropriate selection of variables to include generally relies on domain expertise and experience.

1 – External Drivers
Demand for school teachers is expected to be driven by a number of external factors. Factors that are expected to be relevant to understand to build an understanding of external drivers includes:

• Population growth rate including birth and immigration rates of school aged children
• Changes in profile of school leavers (including leaving later)
• Changing student demand for different subjects
• Impact of disruptive technology on preferences for education delivery

2 – Education strategy
Education policy as determined at both Commonwealth and jurisdictional levels may impact demand. This may include programmes or activities that:

• increase school enrolment and/or attendance
• change the student to teacher ratio

3 – Labour market
The availability of individuals to seek and become teachers drives the supply side. Key considerations in this area include:

• Enrolments and completions of initial teacher education and postgraduate teacher courses
• Perception of pay levels amongst different professions
• General economic outlook and availability of work
• Retirements and exit from teaching
• Extent and availability of registered teachers not in a classroom or school leadership role

4 – Individual preferences
Individual choices also impact the supply of teachers. Information to support understanding this includes:

• Increased request for part time or casual work arrangements
• Desire for flexible work arrangements
• Extent of extended leave for personal (e.g. maternity/paternity leave) or professional (e.g. sabbatical) reasons

5 – Develop supply and demand assumptions
With available data collected in the earlier steps, an historical perspective can be established. This leads to development of reasonable assumptions in terms of the direction of a key input measure.

6 – Forecast supply and demand
The next task involves the application of the collected data and the assumptions to model future years supply and demand.

7 – Forecast supply and demand gap
With expected, supply and demand numbers the determination of the gap or excess between supply and demand is a relatively simple task.

8 – Workforce strategy
Determining the workforce strategy requires establishment of a strategic path to close any workforce gaps that have been modelled. This will lead to policies and programmes that, generally, seek to address two key areas:

• Workforce capacity: do we have enough teachers for the demand we expect
• Workforce capability: do we have the right kind of teachers for the demand we expect
Options for Implementation

A range of approaches should be considered and there are pros and cons which would need to be examined for different possible approaches. Two approaches are presented here.

**Approach 1: Granular data**

Using this approach requires the collection of data at the individual teacher level or at small levels of aggregation.

In this respect, as the NTWD project has shown the availability and quality of data collected at this level is problematic. Options for improving the completeness and accuracy of data have been previously discussed, but these focussed on the specific data items of the NTWD. In fact, the NTWD data would be one input into the supply and demand model and does not request items relating to individual preferences. SiAS does hold data in this regard but it is not, under current arrangements, possible to integrate this data with other data within the NTWD.

The greater granularity of data, as mentioned earlier, enables greater ability to query and understand the teaching workforce and where specific hot spots may exist. There is much anecdotal evidence of shortages of STEM (Science, Technology, Engineering and Mathematics) teachers; shortages of secondary teachers; and an excess of primary teachers.

**Approach 2: Aggregate data and performance metrics**

This approach uses the collection of aggregate and available data sourced from multiple custodians.

This approach is characterised by drawing different datasets together to draw necessary supply and demand conclusions. In fact, it may be that a model that seeks to unify disparate data is not even used but rather an approach that monitors key measures for longitudinal trends. Should a measure move outside its historical norm then this would lead to further investigation as to whether this impacts supply or demand, or whether there are other mitigating factors present.

This approach is adopted by the Australian Government’s Department of Employment’s Leading Indicator of Employment monthly reports.

Despite its apparent simplicity, this approach is intuitive to users and low cost. Key annual measures that could be tracked include the following, many of which are presented on the framework:

- Education course enrolments, in-place students and completions
- Current and new teacher registrants as well as teachers leaving the register
- Current retirement rates and exit from teaching
- Employment availability
- Number of school aged children
- Teacher to student ratios
Utilising the NTWD with other datasets

While datasets alone hold intrinsic value, increased understanding and insight is being gained when disparate datasets are joined. The integration of different datasets is one of the core drivers of the global appetite and enthusiasm for Big Data solutions. As much as the context has generally been applied in the private sector, the opportunity equally exists in the public sector and education, and the application of principles remains consistent.

This project itself was a data integration task by bringing together twenty one different teacher level datasets and four aggregated datasets. The data was enriched with ABS population census data and sense checked against NSSC and SIAS data. Further integration is technically possible. Below are some options on how to further value can be extracted the NTWD. This considers only the possibilities rather than the practicalities (technical linkages) and processes required to enable release of data (where not publically made available).

While some examples below are based as a school level comparison, it remains equally possible to aggregate any measure to a higher level of geographic abstraction for either release of data or for reporting purposes.

ACARA MySchool data
MySchool holds and presents a lot of data at the school level. Some of what could be integrated with the NTWD is below.

The analysis of socio-economic status that is currently used is as provided by the ABS. The ICSEA (Index of Community Socio-Economic Advantage) was supported by members of the project Working Group as a more appropriate and familiar measure to use. Use of ICSEA would allow a greater depth of investigation into the level of the school’s advantage and the teacher’s working at that school. Current analysis is based on the school location only.

NAPLAN scores could be correlated against the profile of teachers working within a school to provide some insight into this, as a driver, for a NAPLAN score. This score could be either the actual absolute score, or the extent of improvement or deterioration over different years.

ABS population census
The extent of this has only been touched on in the current data analysis report. Further work could be undertaken using publically released data by the ABS or with approved selected release of specific data.

An earlier example was understanding the extent to which Aboriginal and Torres Strait Islander teachers are working in low SES areas.

University datasets
There is an increased attention and focus on initial teacher education at present, and its ability to produce the right kind of teachers that are needed in schools. Only aggregated graduate information was provided for the NTWD but additional data, especially when used longitudinally, would be of increased value.

ATAR (Australian Tertiary Admission Rank) scores often attract attention in education with observation of reduced requirements for education qualifications. Stakeholders have commented on a number of potential reasons for this trend, including lower fees, and education being the “default” degree that the liberal arts used to be. Regardless of cause, understanding both the entry ATAR and then that of those that complete their education, register with their regulator, and then take a teaching role, may provide a more complete story.

Students failing to complete their education course is a concern both for the training institutions as well as having a potential impact on overall supply. Being able to incorporate information across the full teacher life cycle from education to employment and then exit, would provide an holistic perspective about which areas should be targeted to retain teachers, including prospective ones. This will also support the provision of key market information to a number of stakeholders including prospective students, parents, universities and employers.
Part V: Indicative Supply and Demand Measures

Determining teaching workforce supply and demand requires an understanding of the complex relationships between factors both internal and external to the school environment.

One of the primary purposes of the NTWD was to monitor and report workforce trends and inform workforce planning issues including workforce supply and demand. The NTWD provides data relevant to teacher supply and gives a point in time snapshot of the teaching workforce, but does not hold historical data from which supply trends can be modelled. The indicative measures in this section focus on the employed teaching workforce.

The granularity of demand data does not match the granularity at which teacher level data has been collected. As such, the following pages present a high level, national perspective of supply and demand, to provide indications of how data could be used. The data presented should be considered as indicative only and is drawn from both the NTWD as well as other publically available sources. It should also take into account the limitations set out earlier in this report.

The key findings from analysis, in this section, that warrant further investigation are:

- The ratio between teaching workforce and student numbers is currently 9.5. This is lower than NSSC ratios as the NTWD attempts to determine all those in a teaching position over an extended period of time.

- Currently, the most influential factor in determining the teaching workforce to student numbers ratio is the rate at which teachers exit the profession.
Supply and Demand Measures

An initial, indicative baseline of measures that support an understanding of teaching workforce supply and demand, and their data source are used in the following pages, to model some scenarios. These are modelled to understand the current, and changes to, the ratio of students to the teaching workforce.

Supply

The supply measures shown below are intended to represent key stages of an individual from commencing an initial teacher education course, completing this, registering, entry into the workforce and subsequent exit. This is supplemented with information on overseas teachers entering Australia. The baseline starting point and assumptions for future projections are outlined.

Initial Teacher Education Commencements

The Higher Education Statistics data sourced from the Australian Government Department of Education is used as source data for the number of teachers commencing initial teacher education. This data indicates that commencements in initial teaching education are increasing 3% year on year (CAGR from 2001 to 2012).

Initial Teacher Education Completions

The Higher Education Statistics is also used as source data for the number of teachers completing initial teacher education. Comparing the completions against commencements from both three and four years before the completion, provides an indicative completion rate of 70%.

Registered

The NTWD suggests that the proportion of teachers who will register with their jurisdictional regulator, relative to the completion rate is around 80%. Some graduates will not register as they take up other roles, including working in the early childhood sector.

Registered to Teaching Workforce

The NTWD suggests that approximately 80% of teachers who are registered are working in a teaching role. Combining this data point with the percentage who register indicates that around 64% of students who complete studies will have a position in the workforce. This differs to the Graduate Destination Survey reported in AITSL’s Initial Teacher Education Data Report 2011, which showed this percentage to be closer to 75%. This survey differs to the NTWD in that it is taken four months after completion, whereas the NTWD adopts a snapshot perspective over a five year period.

Inbound migration

DIAC data captured for data item 42 suggests 1,900 arrivals of teachers each year. The NTWD also provides some indication that there are additional teachers who have qualified at overseas institutions, who will not be captured in the Australian completions and commencements data.

Teachers exiting the profession

The NTWD data estimates that teachers leaving the profession is 5.7% based on the number of teachers who allowed their registration to lapse in the previous 12 months. It is not clear how many were in teaching positions prior to this lapsing. The Victorian Department of Education and Early Childhood Development in its 2011 Supply and Demand estimates attrition at 5%. As some of these may move to casual employment, our baseline estimate of exit rate is 4.5%. This is used for all exit reasons including, as per the SiAS report, resignation from teaching (8% - 10%), retirement (23% - 26%) and overseas migration (3% - 4%).

The NTWD data has shown that the teaching workforce has an uneven age distribution and the exit rate, based on an increased number of retirements, could be expected to increase.
Total Teachers
The NTWD estimates the number of additional registrants at approximately 62,000 which suggests that the number of teachers working in schools through a year is around 380,000. This estimate is similar to an extrapolation of the NSSC 2012 head count to the NTWD for the government sector, where the NTWD is larger by a factor 1.33 (estimating casual VIC and NT teachers). Using this factor across the NSSC total count yields 385,000 teachers.

Demand
Detailed demand measures and assumptions were not used. For a national perspective, the simplified demand measure of the number of students was used.

Student numbers
Schools Australia provides a count of the number of full time and part time students in schools. Further breakdown of this across other factors, such as geographic need, subject requirements, and split between primary and secondary school type, were not considered.

Schools Australia shows that the number of students is increasing 0.9% year on year (CAGR from 2006 to 2013).

Student to teaching workforce ratio
The above measures enable an estimate of the current numbers of teachers in the workforce to the number of students. The teaching workforce measure is intended to measure headcount rather than FTE. Not every teacher in the workforce is in the classroom on any particular day and the NSSC head count provides a better indicator of teachers needed on any one day, and this is why this ratio is lower than other reported ratios. This value represents a baseline value against which possible changes in supply or demand measures are estimated.

Further investigation
Current data does identify areas for future consideration. These include:

- Understanding the drivers of the proportion of students who commence but do not complete initial teaching education courses may provide insight into how to increase completion rates.

- Understanding students who complete but do not register as a teacher, is another area that may warrant further investigation. Anecdotal evidence indicates that the lower entry requirements and cheaper course costs means that a number of students see this as the path to a tertiary qualification, if preferred options are closed.

- Determining changes to the rate at which teachers exit the workforce and how this is impacted by internal and external factors, will be critical to understand – especially as the teaching workforce ages.
Impact of Changes to Measures

An in depth supply and demand model was out of scope of the NTWD work. Using the trend data from the previous page enables scenarios to be determined based on changes in the assumptions presented.

The chart opposite each scenario maps the annual increase in student numbers (lighter colour and right vertical axis) and net increase in the teaching workforce count (darker colour and left vertical axis).

Scenario 1: No Change to Assumptions
With no change in assumptions, the teaching workforce continues to grow in total numbers and the ratio of students per teacher reduces.

The modelled assumptions suggest that, today, more teachers are leaving the profession than joining it. This trend though is short lived and the increased commencement rates impact almost immediately. Over time, the proportional growth in teachers exceeds the proportional growth in student numbers, and the estimated teaching workforce to student numbers falls to 9.4.

Scenario 2: Increased rates of completion of education courses and movement to register
This effect of lowering the teaching workforce to student numbers ratio is more pronounced in the second scenario if completion rate and move to registration rates increase. This is evidenced in the chart by the teaching workforce growth demonstrating a steeper gradient than when compared to student numbers.

Scenario 3: Increased rates of completion and registration but increased rates of exiting the profession and increased growth in student numbers
The last scenario demonstrates that even if the completion and registration rates increase, an increase in exit rate coupled with further growth in student numbers, will result in the ratio increasing. If all assumptions held, it would take ten years before more teachers were entering the workforce than leaving. It takes nearly twenty years before the proportional growth of new teachers entering the workforce exceeds the proportional growth of students.
A blue arrow indicates a change in a measure that will increase the supply of teachers and a red arrow indicates a change to the measure that reduces supply of teachers or increased demand.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1. No change in assumptions</th>
<th>2. Improve completions and registration</th>
<th>3. Exit rate increases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Teacher Education Commencements</td>
<td>No change</td>
<td>No change</td>
<td>No change</td>
</tr>
<tr>
<td>(initial assumption: growth at 3% p.a.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Teacher Education Completions</td>
<td>No change</td>
<td>75%</td>
<td>75%</td>
</tr>
<tr>
<td>(initial assumption: 70% of commencements)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registered</td>
<td>No change</td>
<td>85%</td>
<td>85%</td>
</tr>
<tr>
<td>(initial assumption: 80% of completions)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overseas Teachers</td>
<td>No change</td>
<td>No change</td>
<td>No change</td>
</tr>
<tr>
<td>(initial assumption: 0% growth p.a.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers exiting the profession</td>
<td>No change</td>
<td>No change</td>
<td>7%</td>
</tr>
<tr>
<td>(initial assumption: 4.5% reduction p.a.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Numbers</td>
<td>No change</td>
<td>No change</td>
<td>1.0%</td>
</tr>
<tr>
<td>(initial assumption: 0.9% growth p.a.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ 5 Years Ratio of Student to Teaching Workforce</td>
<td>9.8</td>
<td>9.5</td>
<td>10.7</td>
</tr>
<tr>
<td>(total teaching workforce count)</td>
<td>(385k)</td>
<td>(393k)</td>
<td>(352k)</td>
</tr>
<tr>
<td>+ 10 Years Ratio of Student to Teaching Workforce</td>
<td>9.8</td>
<td>9.4</td>
<td>11.5</td>
</tr>
<tr>
<td>(total teaching workforce count)</td>
<td>(400k)</td>
<td>(418k)</td>
<td>(344k)</td>
</tr>
<tr>
<td>+ 20 Years Ratio of Student to Teaching Workforce</td>
<td>9.4</td>
<td>8.7</td>
<td>11.7</td>
</tr>
<tr>
<td>(total teaching workforce count)</td>
<td>(457k)</td>
<td>(493k)</td>
<td>(374k)</td>
</tr>
</tbody>
</table>
## Impact of Changes to Measures

There are many scenarios that could be modelled. The table below shows the change to the current ratio of student numbers to teaching workforce if that measure is changed and all other measures remain unchanged. For example, if the number of completions of initial teacher education increase to 80% (as compared to the baseline completion rate of 70%) then the ratio of students to teaching workforce will, in five years, reduce from its baseline of 9.8 (per opposite page), by 2.4% to now be 9.5. Only the percentage movement is shown, not the new ratio value.

With no baseline growth numbers for overseas teachers, these were not considered.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Change in 5 years</th>
<th>Change in 10 years</th>
<th>Change in 20 years</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial Teacher Education Commencements</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline 3% growth year on year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commencements growth to 4%</td>
<td>-0.2%</td>
<td>-1.3%</td>
<td>-6.3%</td>
<td></td>
</tr>
<tr>
<td>Commencements growth to 2%</td>
<td>+0.2%</td>
<td>+1.3%</td>
<td>+5.5%</td>
<td></td>
</tr>
<tr>
<td><strong>Initial Teacher Education completions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline 70% of commencements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completions grows to 80%</td>
<td>-2.4%</td>
<td>-4.6%</td>
<td>-7.9%</td>
<td></td>
</tr>
<tr>
<td>Completions fall to 60%</td>
<td>+2.4%</td>
<td>+4.6%</td>
<td>+7.9%</td>
<td></td>
</tr>
<tr>
<td><strong>Completions to Registration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline 80% of completions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registration grows to 85% of completions</td>
<td>-1.1%</td>
<td>-2.0%</td>
<td>-3.5%</td>
<td></td>
</tr>
<tr>
<td>Registrations fall to 75% of completions</td>
<td>+1.1%</td>
<td>+2.0%</td>
<td>+3.5%</td>
<td></td>
</tr>
<tr>
<td><strong>Registration to Teaching Workforce</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline 80% of registrations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registration grows to 85% of completions</td>
<td>-1.1%</td>
<td>-2.0%</td>
<td>-3.5%</td>
<td></td>
</tr>
<tr>
<td>Registrations fall to 75% of completions</td>
<td>+1.1%</td>
<td>+2.0%</td>
<td>+3.5%</td>
<td></td>
</tr>
<tr>
<td><strong>Teachers exiting the profession</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline 4.5% of current workforce</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher exit rate increases to 7%</td>
<td>+10.8%</td>
<td>+17.8%</td>
<td>+24.7%</td>
<td></td>
</tr>
<tr>
<td>Teacher exit rate increases to 10%</td>
<td>+22.3%</td>
<td>+34.7%</td>
<td>+44.1%</td>
<td></td>
</tr>
<tr>
<td><strong>Student Numbers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline 0.9% growth year on year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student numbers growth to 1.1%</td>
<td>+1.0%</td>
<td>+2.0%</td>
<td>+3.9%</td>
<td></td>
</tr>
<tr>
<td>Student numbers growth to 1.5%</td>
<td>+2.9%</td>
<td>+5.8%</td>
<td>+11.2%</td>
<td></td>
</tr>
</tbody>
</table>