

Submission on Redevelopment and Audit of the Higher Education Data Collection

Ittima Cherastidtham

Higher Education Fellow, Grattan Institute

ittima.cherastidtham@grattan.edu.au; (03) 9035 8653

Andrew Norton

Higher Education Program Director, Grattan Institute

andrew.norton@grattan.edu.au; (03) 8344 0060

Grattan Institute

8 Malvina Place Carlton VIC 3053

1. Introduction

This submission focuses on sections two, four and five of the *Redevelopment and Audit of the Higher Education Data Collection discussion paper*. It supports extending the USI to all higher education students. Parental educational attainment provides useful insights beyond aggregate socio-economic measures and should be retained. Likewise, we recommend that the Department of Education and Training continue collecting ATAR data for the applications and offers collection, and require ATAR data from students whose basis of admission is not secondary education when available. This would improve data reliability and the accuracy of results drawn from ATAR analysis.

2. Identifying students using USI

Most domestic students are assigned a Commonwealth Higher Education Student Support Number (CHESSN) and a tax file number (TFN). These students' outcomes can be tracked reasonably well over time. But many students are not assigned with one of these identifiers. A CHESSN is only assigned to students who receive government support through direct tuition support (CGS) or the loan scheme (HELP).

For students who do not receive government support, however, there is no way to accurately track their activities over time. Every student receives a student identification code from their higher education provider. But this code is not consistent across institutions. When students move to another provider the ability to track them is limited. Probabilistic matching based on personal information such as name and date of birth can be used, but the results are not as accurate as using a universal identifier like CHESSN or a Unique Student Identifier (USI), which is already used for vocational education and training.

Extending the USI to higher education students would improve the government's and the sector's ability to accurately identify and track students irrespective of their education level or citizenship status.

Using USI for all students would also allow us to better understand students' vocational education experience. More than one in ten commencing domestic bachelor degree students are admitted based on their vocational education experience.¹ In 2016, 9 per cent of vocational enrolments reported a bachelor degree or above as their highest previous education level.² Understanding transitions between sectors will help both sectors and the government to better understand student pathways.

Since many students have vocational education experience, many are also likely to have an existing USI. To avoid duplication, the government should consolidate the current list of CHESSNs and USIs prior to allocating any new USIs to students. All students should be asked for an existing USI before issuing a new number. Perhaps the Department could use the benefit of having a national comprehensive record of their education and training online to encourage students to provide their existing USI.

A USI should be assigned when higher education students apply rather than when they enrol. Currently nearly 100 per cent of people who apply through tertiary admission centres but only about 85 per cent of those who apply directly to university have a CHESSN. Collecting or assigning a USI at application would help track the approximately 8 per cent of students who

¹ About 13 per cent of commencing enrolments at table A institutions in 2015, Department of Education and Training (various years).

² Total VET activity program enrolments, NCVER (2018)

defer.³ Their interrupted pathway can provide useful insights, especially as low-SES students are more likely to defer than high-SES students.⁴

Recommendation: Extend the USI to all higher education students and applicants with a consolidation process to ensure USI can uniquely identify students.

3. Relevance of parental educational attainment

Parental educational attainment data was first collected in 2010 to supplement student socio-economic background calculations, due to concerns that the geographic measures then used misclassified individuals. During its first year of collection just under 70 per cent of commencing enrolments had valid parental education attainment information, reducing its usability as a SES measure. This has since improved, with a response rate of nearly 90 per cent in 2016.⁵

The Department now uses a more nuanced geographic measure, shifting SES calculations from Census Collection Districts (CD) to the smaller Statistical Area1 (SA1) in 2014. A student's permanent home address is converted into an SA1 code. Each code is mapped onto an index of education and occupation (IEO) using the ABS Socio-economic Indexes for Areas (SEIFA) framework. Switching from SA1 to CD improved the accuracy of SES measures through greater granularity of the SES calculation.

While parental education data and IEO scores based on SA1 data are correlated, discrepancies between them remain. Parental education data represents individual-level information, while on average an SA1 area has 400 people and can include up to 800 people. Figure 1 shows the discrepancies between the two measures for 2016 commencing domestic bachelor degree enrolments. One in five students classified into the bottom IEO decile have a parent with a bachelor degree or above qualification, which would typically indicate higher-SES, and a similar share of students classified into the top decile have a parent with no post-school qualification.⁶

³ The share is 9 per cent if unknown offer responses are excluded. Data from 2013-2016, Department of Education and Training (various years)

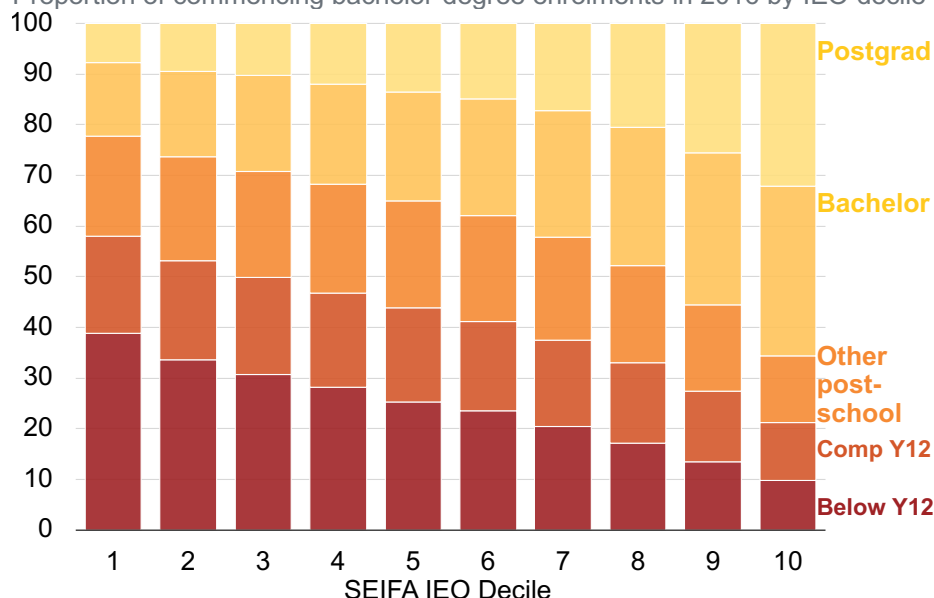
⁴ For example, 11 per cent of low-SES students defer compared to 9 per cent of high-SES students between 2013 and 2016, *ibid.*

⁵ Bachelor-degree domestic commencing enrolments at Table A institutions. Based on parent 1 data (element 573). Parent 2 variable (element 574) follows a similar trend but with a larger share of missing data, *ibid.*

⁶ Based on parent 1 data.

Figure 1: A SES measure using the SEIFA index of education and occupation cannot replace parental education data

Proportion of commencing bachelor-degree enrolments in 2016 by IEO decile



Notes: Commencing domestic bachelor students from the 2016 cohort. Highest level of education taken from students' "parent 1". SEIFA Index of Employment and Occupation (ABS) applied to students' permanent home residence SA1. Source: ABS (2013a); Department of Education and Training (various years)

Parental educational attainment has a link to students' performance independently of IEO. Grattan analysed the first-year fail rate of 2014 domestic students. We find that while IEO has a link to the first-year fail rate, so does a mother's higher education qualification. Having a mother with higher education is associated with a marginally lower first-year fail rate; the effect is statistically significant after controlling for other personal characteristics, including the student's IEO decile.⁷

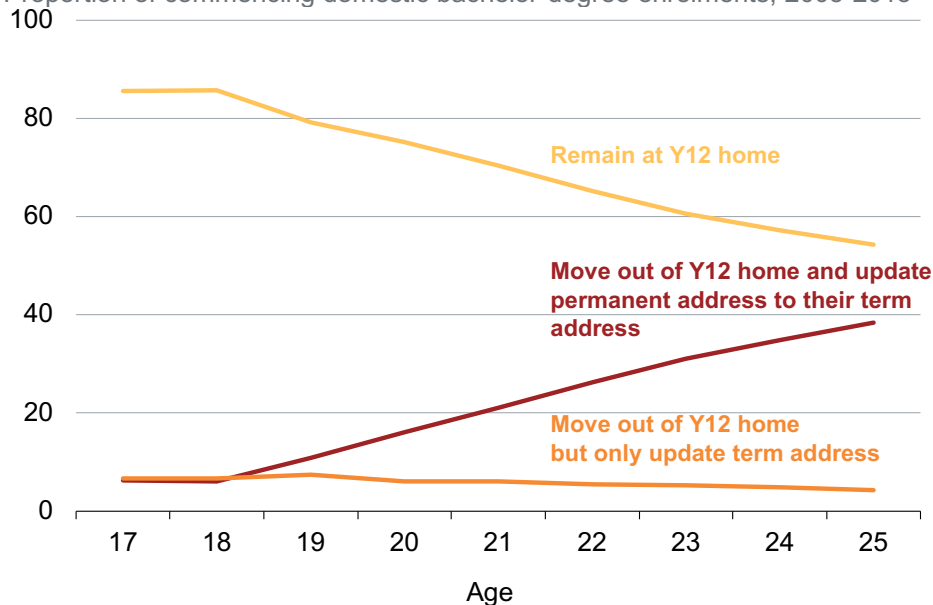
The underlying address data used in the IEO calculation may also compromise its accuracy. IEO is derived from what students report as their permanent address. For older students, this is likely to be a good approximation of their long-term location and socio-economic status. For young students, however, the address may not represent their background circumstances. At least 20 per cent of higher education students aged under 25 are likely to have moved out of home in the 2016 Census data.⁸ Some of these students give a permanent address like their parents' home that is a long-term location. But as Figure 2 shows (red), a significant share of students give their permanent address as their term address, where they have moved to study. The tendency to use term address for permanent address increases as students get older.

⁷ Excluding double degree, students who commenced in the second semester, non-table A students, previously enrolled in a bachelor degree between 2005 and 2013. Using OLS regression controlling for institution, field of education and prior academic history such as ATAR, *ibid.*

⁸ Students at a university or other tertiary institution who live alone or in a group house, ABS (2017)

Figure 2: A significant share of students update their permanent address to their term address when they move out

Proportion of commencing domestic bachelor-degree enrolments, 2005-2015



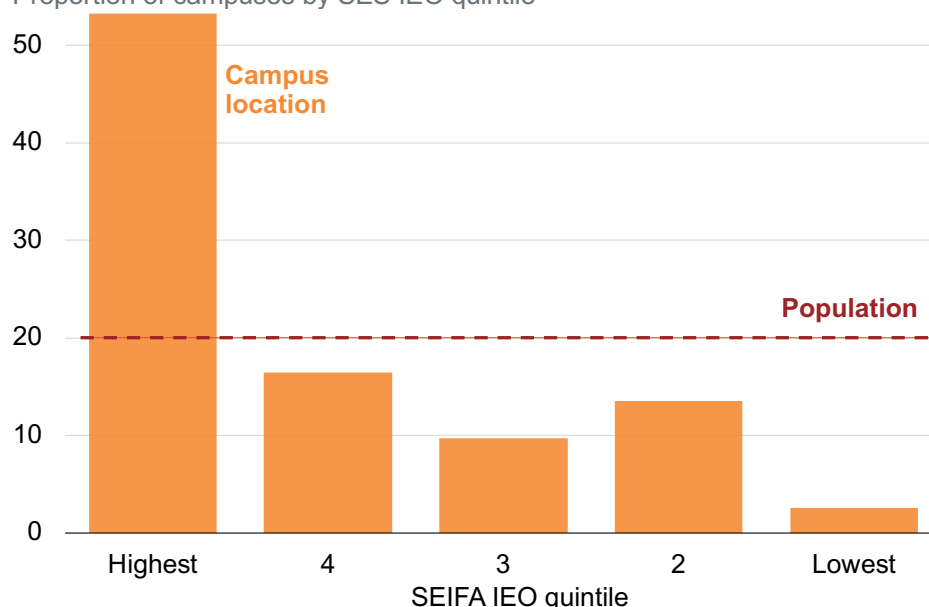
Notes: Domestic bachelor students commencing between 2005 and 2015. Only those with a valid term residence postcode, permanent home residence postcode and year 12 home postcode are included. Addresses are considered the same if they are within the same postcode, the smallest location measure common to the three variables.
Source: Department of Education and Training (various years)

A student's term address does not necessarily represent their background circumstances. IEO scores increase with a greater proportion of students in tertiary education and of people working in professional and managerial jobs. University campuses tend to attract high IEO scores because of student accommodation and their highly skilled workforce living nearby. More than half of higher education campuses are located in the top IEO quintile (Figure 3). For students who report their term location as their permanent address, their IEO score can represent the socio-economic status of their university campus rather than their own, which is likely to be different. This problem could be substantially reduced by using year 12 home address for younger students. But the IEO cannot replace the benefits and the granularity provided by parental education data.

Given the potential issue with IEO and its discrepancies with parental educational attainment, the Department could supplement its equity performance reporting with results by parental education. The additional measure would reduce pressure on IEO results and make its equity reporting more comprehensive.

Figure 3: More than half of campus locations of higher education providers are in the highest SES quintile based on the index of education and occupation

Proportion of campuses by SES IEO quintile



Notes: Campuses of all higher education providers in 2016. Campus postcode is mapped to an IEO score using Postal Area (POA) IEO tables. There were 547 institution campuses that had postcodes matched to an IEO quintile. Eighty could not be found in the POA table and were excluded.

Sources: ABS (2013b); Department of Education and Training (various years)

Recommendation: Retain the collection of parental educational attainment data

4. Improving data quality

Course name

Institutions report course names differently.⁹ Because the variable has a maximum character length of 72 characters, some institutions report their course names with abbreviations that are difficult to interpret. In addition, most but not all course names include qualification level data. Some universities use an abbreviation of B. for bachelor degree.

The Department could consider dropping qualification level data from the course name as it is already reported in a separate element. This would reduce the pressure to abbreviate the course name without losing much information.

One complexity that would be triggered by this change is for double degrees where one course is with honours and the other is not. To avoid confusion, the Department could extend the course type code into two elements; the first to represent the main course and the second for the secondary course.

Recommendation: Exclude qualification level data from the course name

Highest participation prior to commencement

We agree that providers report the highest educational participation data inconsistently, which greatly reduces its usability. The main issue is an unclear hierarchy of qualifications. Reviewing

⁹ There are also variations within institution, Department of Education and Training (various years)

and clarifying coding notes would ameliorate the problem. But we recommend that the Department consider an alternative framework for collecting participation data.

Participation data could be collected using a set of elements, one for each qualification level. Each element can take on one of three scenarios: never commenced, commenced but incomplete with year last enrolled, and completed with completion year, as described in Table 1.

The change increases the number of elements but is unlikely to be difficult to implement, and would provide additional useful information. This framework follows existing requirements in the applications and offers data collection (elements 730-737). Many institutions already report this information for their students. For them, changing the reporting system is unlikely to create much additional cost. The new structure would provide information about all qualifications rather than only the highest.

With the new structure, one required clarification is for when a student commenced two qualifications at the same level but only completed one. In this case, a completed course could take precedent and the completion year should be reported.

The Department could also consider removing the elements once a USI is in operation which allows for comprehensive tracking between the two datasets (this requires USI to be assigned to all applicants). This is contingent on mandatory reporting of elements 730-737, which we also recommend (section 5).

Table 1: Potential coding structure for highest participation prior to commencement

Code	Meaning
10000	Has never commenced such a course
2YYYY	Had commenced but did not complete all the requirements of such a course and the last year of enrolment was YYYY
3YYYY	Has completed all the requirements of such a course and the last year of enrolment was YYYY

*Note: Based on element 731 from
Source: 2017 Data requirements, HEIMSHelp*

Recommendation: Adopt a reporting structure of prior qualification for highest participation prior to commencement

ATAR

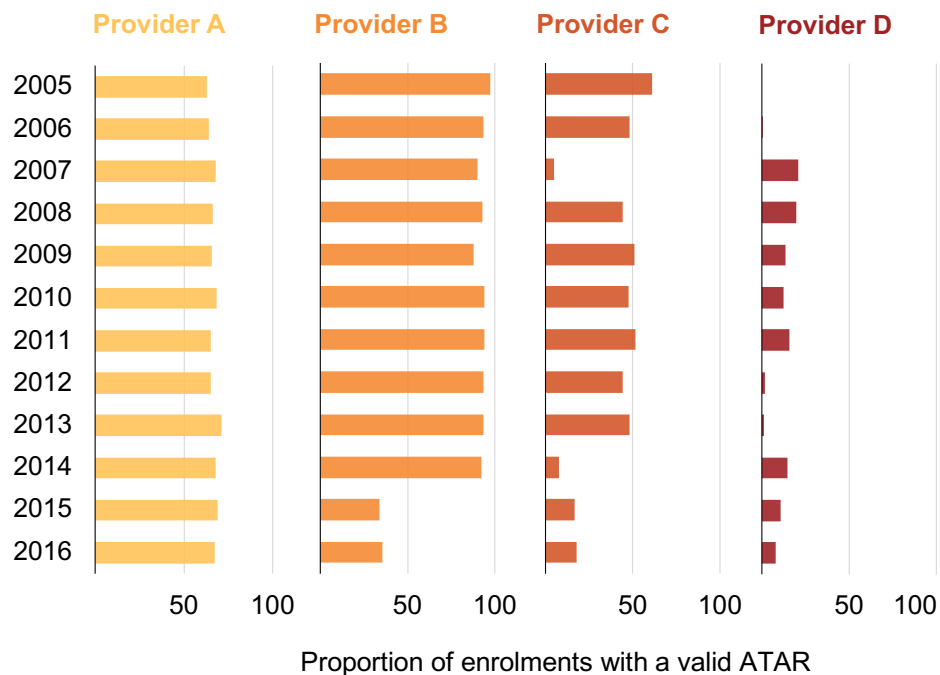
We agree with the Department that ATAR reporting is inconsistent, which reduces its accuracy. Because ATAR is widely used by the Department, TEQSA and researchers, improving its quality is crucial.

Removing the inconsistencies around bonus scores would improve the accuracy of ATAR data. A new element should be created for bonus scores. Given that universities need to report these scores as part of the Higher Education Standards Panel’s recommendations in its report, *Improving the Transparency of Higher Education Admissions*, any additional reporting burden is likely to be small.

Inconsistencies in ATAR reporting extend beyond bonus scores. The proportion of students with valid reported ATARs can vary significantly across institutions and over time. Figure 4 shows four sample universities. Provider A shows consistent reporting of ATAR while provider B is likely to have made a policy change in reporting starting from 2015. The ATAR data for

providers C and D shows irregular reporting practices. These inconsistencies exist in many other institutions, which affect the overall accuracy of ATAR data.

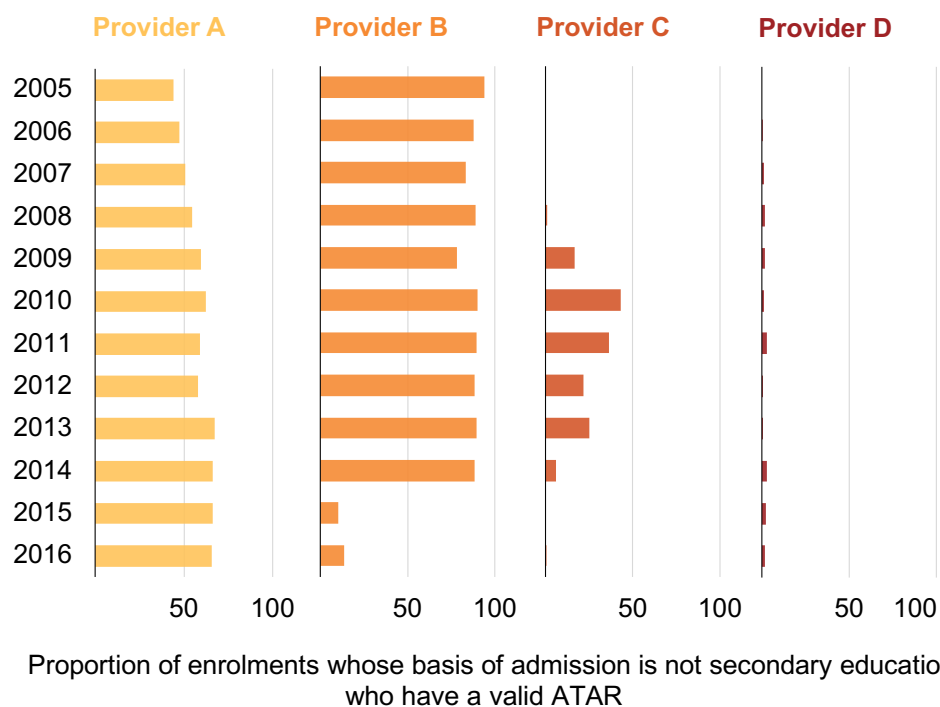
Figure 4: ATAR reporting practices vary across institutions



Notes: Commencing domestic bachelor students 2005-2016. A valid ATAR is one between 30 and 100, inclusive.
 Source: Department of Education and Training (various years)

The main cause of this inconsistency is that universities vary in what they report for students whose basis of admission is not secondary school. For these students, universities are not required to report their ATAR. Figure 5 shows the same universities as in Figure 4. Provider A remains relatively consistent between 2005 and 2016. Provider B changed their ATAR reporting policy from 2015, which corresponds with the proportion of students with a reported ATAR dropping from 92 per cent in 2014 to 34 per cent in 2015 in Figure 4. Provider C, once again, shows irregular behaviour, probably because of multiple changes in reporting policy. Provider D generally does not report ATAR for students whose basis of admission is not secondary school.

Figure 5: The proportion of valid ATARs largely corresponds to university policies on whether to report ATAR for students whose basis of admission is not secondary education



Notes: Commencing domestic bachelor students 2005-2016. A valid ATAR is one between 30 and 100, inclusive.
Source: Department of Education and Training (various years)

Irrespective of basis of admission, ATAR provides a useful insight into students' likely performance at university. A study of Monash University students compares first-year performance of those who were admitted based on secondary education and those admitted through a TAFE pathway program. The paper finds ATAR to have a significant impact on first-year weighted average mark in both entry groups.¹⁰

To reduce reporting inconsistencies between institutions, we recommend that the government require universities to report ATAR irrespective of admission basis. Since many students do not have an ATAR, it is not realistic to require one from every student. The government could instead require universities to request an ATAR and report when students provide one. This would improve ATAR data availability.

We also recommend that the Department enhance the ATAR data collection by creating separate categories for ATARs less than 30 and students who did not complete Year 12. Currently they are classified into one group which also includes non-domestic students.

The Department should maintain ATAR in the applications and offers data collection. It can use the ATAR data in the applications and offers collection to better understand students from equity groups such as low-SES. For example, overall low-SES students are marginally less likely to receive an offer.¹¹ But after controlling for ATAR, the opposite is true. As Figure 6 shows, low-SES students are more likely to receive an offer than high-SES students at a similar ATAR. So lower ATARs explain the lower offer rate.

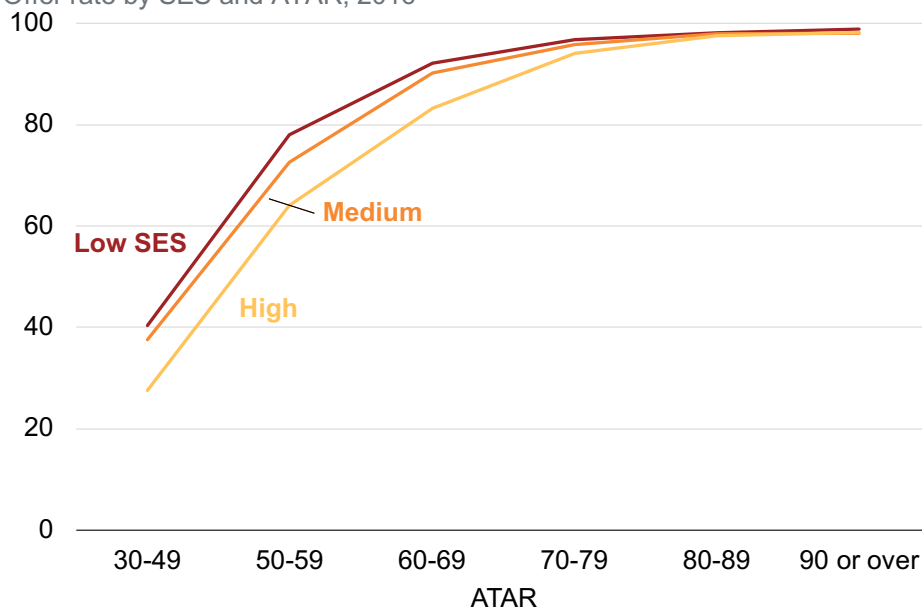
¹⁰ The effect of ATAR is linear for TAFE and increasing as ATAR increases for school leavers, Diamond and O'Brien-Malone (2018), p.10.

¹¹ About 83 per cent offer rate for low-SES versus 84 per cent for high-SES students in 2016, Department of Education and Training (2016) p. 30.

TEQSA could use ATAR offers data as part of assessing whether higher education providers are complying with the admissions standards. Both the Department and TEQSA could use the ATAR data to monitor student applications, offers and enrolments, especially for any unusual patterns and risks to the higher education system.

Figure 6: Low-SES students are more likely to receive an offer after controlling for ATAR

Offer rate by SES and ATAR, 2016



Notes: Domestic bachelor degree applicants and offers from 2016 who applied through tertiary admission centres with a valid CHESSN and an ATAR. Based on the index of education and occupation (SEIFA 2011). Low-SES represents the bottom 25 per cent. High-SES represents the top 25 per cent.
Source: Department of Education and Training (various years)

Recommendations:

- Require bonus scores to be reported separately from the ATAR data
- Require universities to request and report ATAR scores irrespective of admission basis when available
- Continue to collect ATAR data in the applications and offers data collection

5. Additional issue of mandatory elements in the applications and offers data collection

Applications and offers are reported directly by providers as well as by tertiary admission centres (TACs). For TACs, many data elements are not required to be reported but some should be mandatory.

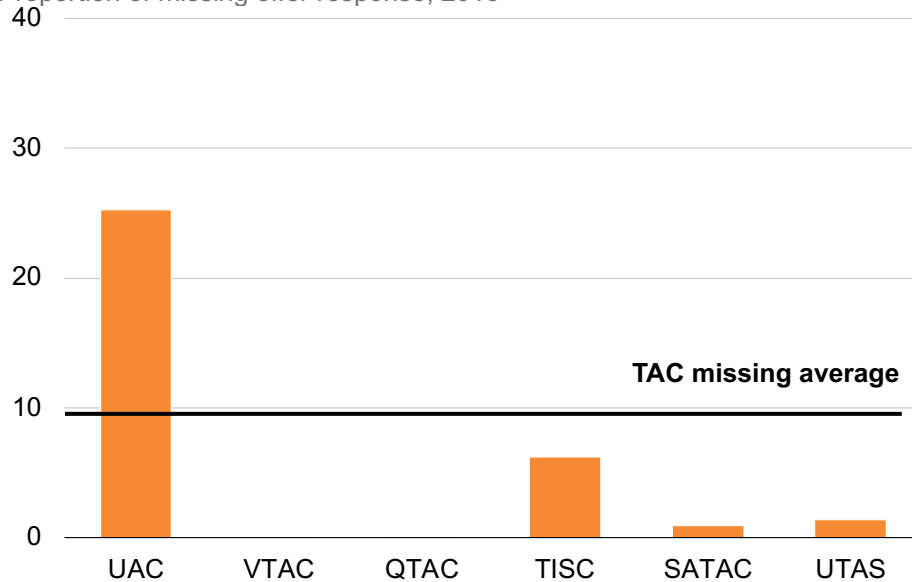
The response code in the applications and offers data identifies students' response to an offer. The data is used in the Department's publications including the annual *Undergraduate Applications, Offers and Acceptances report*. But because reporting the code is not mandatory, nearly 10 per cent of the data is missing.¹² The problem varies between TACs. As Figure 7 shows, a quarter of offer responses at UAC are missing. A large share of missing information is

¹² Response unknown or not yet recorded

likely to affect data accuracy, which limits the value of any analysis and conclusions drawn from the data.

Figure 7: A quarter of offer responses at UAC are missing

Proportion of missing offer response, 2016



*Note: Missing refers to response unknown or not yet recorded.
Source: Department of Education and Training (various years)*

We recommend that the Department make more data elements mandatory, especially those that are used in its publications. This would include offer response code, ATAR, and previous education information.

Recommendation: Make important data elements mandatory including ATAR, offer response code, and previous qualification

References:

- ABS (2013a) *Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2011, Cat. 2033.0.55.001*, Australian Bureau of Statistics from <http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012011?OpenDocument>
- ABS (2013b) *Technical paper: socio-economic indexes for areas (SEIFA) 2011, Cat. 2033.0.55.001*, Australian Bureau of Statistics
- ABS (2017) *Census of population and housing, 2016, TableBuilder Pro, Cat. 2073.0*, Australian Bureau of Statistics
- Department of Education and Training (2016) *Undergraduate applications, offers and acceptances 2016 report*, Department of Education and Training
- Department of Education and Training (various years) *Higher education statistics collection*, from Department of Education and Training
- Diamond, M. R. and O'Brien-Malone, A. (2018) 'Pathways to performance: an examination of entry pathway and first-year university results', *Asia Pacific Journal of Education*, p 1-19
- NCVER (2018) *VOCSTATS: TVA program enrolments 2014-2016*, National Centre for Vocational Education Research from <http://vocstats.ncver.edu.au/webapi/jsf/tableView/tableView.xhtml>