



SAFE DATA RELEASE FOR CAPACITY TO CONTRIBUTE

Direct measure of income refinement working group
paper

October 2020



SAFE DATA RELEASE FOR CAPACITY TO CONTRIBUTE

The Direct Measure of Income (DMI), recently implemented by the Department of Education, Skills and Employment (DESE), uses administrative data about parents and guardians of students at non-government schools to calculate Capacity to Contribute (CTC) scores.

The Australian Bureau of Statistics (ABS) contributes to the annual CTC process by:

- bringing data together;
- providing a secure environment for analysis; and
- ensuring confidentiality is maintained in the release of data.

ABS has worked with DESE to assess information needs and confidentiality risks associated with CTC data, to enable the release of data in a way that optimises its usefulness while protecting confidentiality.

Confidentiality

Confidentiality refers to protecting the secrecy and privacy of information collected from individuals and organisations. This means that information is only made available in a way that is not likely to allow individuals or organisations to be identified.

Australian Government agencies, including the ABS and DESE, are responsible for maintaining the confidentiality of those who provide information, in accordance with legislation such as the Privacy Act 1988.

CTC data is protected by the Census and Statistics Act 1905. This legislation strictly prohibits the release of information in a manner likely to enable the identification of a person.

CTC data - Confidentiality risks and treatments for safe release

Factors affecting confidentiality risk must be considered and appropriately treated before any CTC data can be released. Relevant factors include:

1. Population features: school level data may represent a small group of people, and may include people with unique characteristics which make them more likely to stand out.
2. The number of variables: as more variables are released, more information becomes available about a group of people, increasing confidentiality risk.
3. The level of detail in the data: a high level of detail reduces the size of groups represented in each statistic. For example, reporting income by decile breaks a population down into 10 groups, whereas using quartiles breaks it down into four. Cross-classifying variables provides additional information about group characteristics and can enable differencing by comparing the output of separate tables.
4. The frequency with which data are released: Releasing the same data frequently may enable information about a person or group to be inferred from changes in a statistic over time.

5. The type of statistic: Some statistics are inherently riskier than others, due to how they are calculated. Totals, means, and frequency tables are considered risky, because of the amount of information about a population they convey.
6. The structure and quality of the dataset: Datasets constructed from survey samples generally present lower risk than those containing an entire population, because with samples it is not known who is in the dataset. Aspects of quality, such as the timeliness of data, can also influence the level of risk.

Treatments commonly used to support the safe release of data include:

- Limiting the number of variables or statistics, or the level of detail provided.
- Ensuring the population or group represented is sufficiently large.
- Suppressing values which only represent a small group of people.
- Modifying values using perturbation (the making of small adjustments to data or tables).
- Top and bottom coding: for example, coding respondents over the age of 85 to an “85 and above” category, rather than using specific age groups.

In general, as the risk associated with releasing data increases, additional treatments are required, reducing the usefulness of the data.

Availability of CTC data

General CTC data analyses

High level analysis of CTC data has informed the understanding of the fitness-for-purpose of the DMI methodology and the development of the review process. This analysis is available on the [DESE website](#) in the following reports:

- A Data Quality Framework for the Australian Government’s Direct Measure of Income for Capacity to Contribute.
- The Technical Framework for the Capacity to Contribute Review process.

School level data

Releasing school level data not only enables the provision of CTC scores, but also enables schools and approved authorities to interpret CTC scores and participate in the review process. With safeguards to protect confidentiality in place, a limited amount of school level data can be released to DESE to support key information needs.

ABS and DESE have worked together to identify a set of variables that can be released to support stakeholders’ needs for school level data. The following table summarises this set of variables, the frequency with which they can safely be released, and factors which may affect their availability.

Data	Frequency	Purpose	Availability
CTC score	Annual	To inform Australian government funding under Australian Education Act 2013.	DMI based CTC scores not available for very small schools, and schools with very few families.
DMI score	Annual	To support interpretation of CTC scores.	As above.
Income coverage rate	Annual	To support interpretation of quality of CTC scores.	Availability depends on school size and factors associated with the annual release of data. Treatments may be applied.
Income deciles or quartiles	Annual	To support interpretation of CTC scores.	Availability depends on number of families at the school.
Linkage rate	As requested	To support review process.	Availability depends on school size and factors associated with the annual release of data and release of similar data (coverage rate). Treatments may be applied.
Average family size	As requested	To support review process.	Availability depends on school size. Treatments may be applied.
Quality metrics: missingness, accuracy & volatility	As requested	To support review process.	Availability & treatments will depend on the specific quality metric.

Definitions are provided in the Glossary at the end of this document.

Income distribution

Various factors regarding income distribution were considered, including:

- The safest approach is for one type of income distribution statistic to be made available for all schools (subject to safeguards being met). Providing one statistic for some schools and a different statistic for other schools would introduce an undesirable level of risk.
- Deciles or quartiles would be able to be made available for a large majority of schools, with quartiles available for more schools than deciles. Since deciles represent a higher level of detail, they can only safely be provided for larger schools. This is an example of the trade-off between a higher level of detail and greater treatment / reduced utility.
- Deciles or quartiles are useful for describing distribution around a median, which is the statistic on which DMI scores are based.
- Mean and standard deviation, and tables of population by income range, are not provided as options at this time, due to the reduced utility that would result from treatments required to make them safe.

The current assessment of confidentiality risk does not prejudice any future refinements to the CTC process. Future refinements to the CTC methodology or to processes affecting the quality of CTC scores must assess the corresponding impact on confidentiality risk. However, the release of CTC data, especially school level data, for a given reference period may prevent the release in future of additional data relating to the same reference period.



Median income associated with DMI scores

The median income associated with a DMI score is considered safe to release as it does not equate to a school's median income. ABS recommends the median incomes associated with a score be released in ranges, as this is easily interpreted as the range in which a school's median income falls, given its score.

Further information

For further information about confidentiality risks and risk mitigation, see the [ABS Confidentiality Series](#) (cat. no. 1160.0), August 2017.



Glossary

Average family size

Average (mean) number of dependants per family at a school, based on data from the 2016 Census of Population and Housing. Dependants are defined as “children under the age of 15 and members of the household under the age of 24 who are full-time students”¹.

Capacity to Contribute (CTC) score

A CTC score is a measure of the anticipated capacity of a non-government school community to contribute to the recurrent costs of the school. A school’s CTC score affects the amount of base recurrent funding the school receives from the Australian Government under the Australian Education Act 2013 (Cth). CTC scores may use a DMI or area-based methodology.

In 2020, a DMI-based CTC score will be the average of DMI scores for 2018 and 2019. From 2021, a DMI-based CTC score will be the rolling average of the previous three years’ DMI scores. 2020 and 2021 represent a transition period. During these years, the CTC score will be the most beneficial of: the rolling average DMI score; the 2016 SES score; or the 2011 SES score.

From 2022, the CTC score will be either the three-year rolling average DMI or the refined area-based score, if the DMI is not considered fit-for-purpose for quality or confidentiality reasons.

Direct Measure of Income (DMI) score

A methodology used to calculate capacity to contribute. The DMI score is based on the median Adjusted Taxable Income (ATI) of parents or guardians of students at a school. It is created by:

- calculating the total income for each student by summing the incomes of up to two parents or guardians;
- identifying the median family income for each school; and
- converting the median incomes for all schools into DMI scores via standardisation.

Income coverage rate

The income coverage rate for the school is the number of students included in the calculation of the school’s median income, divided by the number of students in the Address Collection.

Income deciles or quartiles

Income deciles are the income amounts which divide the sorted income distribution into ten groups of equal size. Income quartiles are the values which divide the sorted income distribution into four groups of equal size.

Linkage rate

The proportion of students in the Address Collection with at least one parent linked to the Multi-Agency Data Integration Project (MADIP).

¹ Source: Census of Population and Housing: Understanding the Census and Census data, 2016, (cat. no. 2900.0), see: [Count of dependent children in family](#).

Quality metrics: missingness, accuracy & volatility

A number of quality metrics are used by DESE in determining whether a school's DMI score is fit-for-purpose. These are described in Quality Gate 3, in [A Data Quality Framework for the Australian Government's Direct Measure of Income for Capacity to Contribute](#), available on the DESE website.