The Occupation Shortage List (OSL), produced by Jobs and Skills Australia, provides a list of occupations in shortage in Australia and in each state and territory. The OSL is released annually and as a point-in-time assessment of the shortage status of occupations.

- An occupation has a shortage when employers are unable to fill, or have considerable difficulty filling, vacancies in that occupation at current levels of remuneration, current conditions of employment, and in reasonably accessible locations.

Machine-learning produced fill rates

The occupation vacancy fill rate (fill rate) is the percentage of vacancies that are filled.

A Gradient Boosting Machine (GBM) learning model initially predicts the proportion of vacancies filled (fill rates) for all occupations on OSL scope on 2013 ANZSCO basis. The OSL scope are ANZSCO-defined Skill Level 1 to 4 occupations. The GBM uses 24 data inputs from 13 different sources to predict the fill rates.

A modelling framework is used to blend the GBM-based fill rates and construct 80% confidence intervals for those fill rates. A set of rules are used on these blended fill rates and confidence intervals, along with other data sources to generate occupation ratings. Please refer to the 2024 OSL Methodology Report for more details.

Key steps in generating the occupation ratings using modelling framework

Key elements of the modelling framework
Estimating occupation fill rates using a GBM model
Blending the GBM-based fill rates with weighted moving average SERA fill rates
Producing 80% confidence intervals for the blended occupation fill rates
Converting the blended fill rates, confidence intervals and occupation employment sizes into occupation ratings using a set of decision rules

Converting GBM fill rates into occupation ratings

The OSL has four ratings to classify the shortage status on a specific occupation:

- Shortage (S): An occupation is in national shortage or overall shortage
- Metropolitan Shortage (M): An occupation is in shortage in a metropolitan area
- Regional Shortage (R): An occupation is in shortage in a regional area
- No Shortage (NS): An occupation is not in shortage.

National or overall shortages (S)

An occupation is rated as national Shortage (S) if:

- The occupation fill rate is below 67% in both metro and regional areas, or

- The occupation fill rate is below 67% in an area that has a majority employment share, which is defined as 80% or more.

Regional/Metropolitan shortages (R/M)

An occupation is rated as regional shortage (R) if:

- The occupation fill rate is below 67% in a regional area, and
- The employment share of the occupation in regional areas was between 20% and 80%.

A similar rule holds for Metropolitan shortage (M) if the above criteria are met for metropolitan areas.

Additional rules used to convert the occupation fill rates into occupation ratings.

1. For an S, M or R to apply, the employment size for an occupation must be above the bottom 25th percentile (approximately 1400 workers).

- If the employment size is between the 10th and 25th percentiles, and the occupation satisfies the above definitions and rules, an uncertainty flag (*) is applied.
- If the employment size is below the 10th percentile, an NS rating is applied as there is too much uncertainty to rate the occupation as S, M or R.

2. The occupation's fill rate is below 67% when 90% or more of an occupation's fill rates fall below 67% within an 80% confidence interval.
3. For the rating to differ from last year's OSL rating, there must be at least a 20-percentage point change in the percentage of fill rates that fall below 67% within an 80% confidence interval.

The 1st spreadsheet of the Workbook: 2024 OSL Aggregated Inputs, contains the results from the modelling framework, results of statistical tests and the additional data used to assist with applying the above rules to convert fill rates and confidence intervals into occupation ratings.

Mapping from 2013 ANZSCO to 2022 ANZSCO

One to one mapping. In the 2022 ANZSCO, about 834 occupations directly map to occupations in the 2013 ANZSCO. For these occupations, occupations ratings that are first generated on 2013 ANZSCO are applied to the 2022 ANZSCO versions.

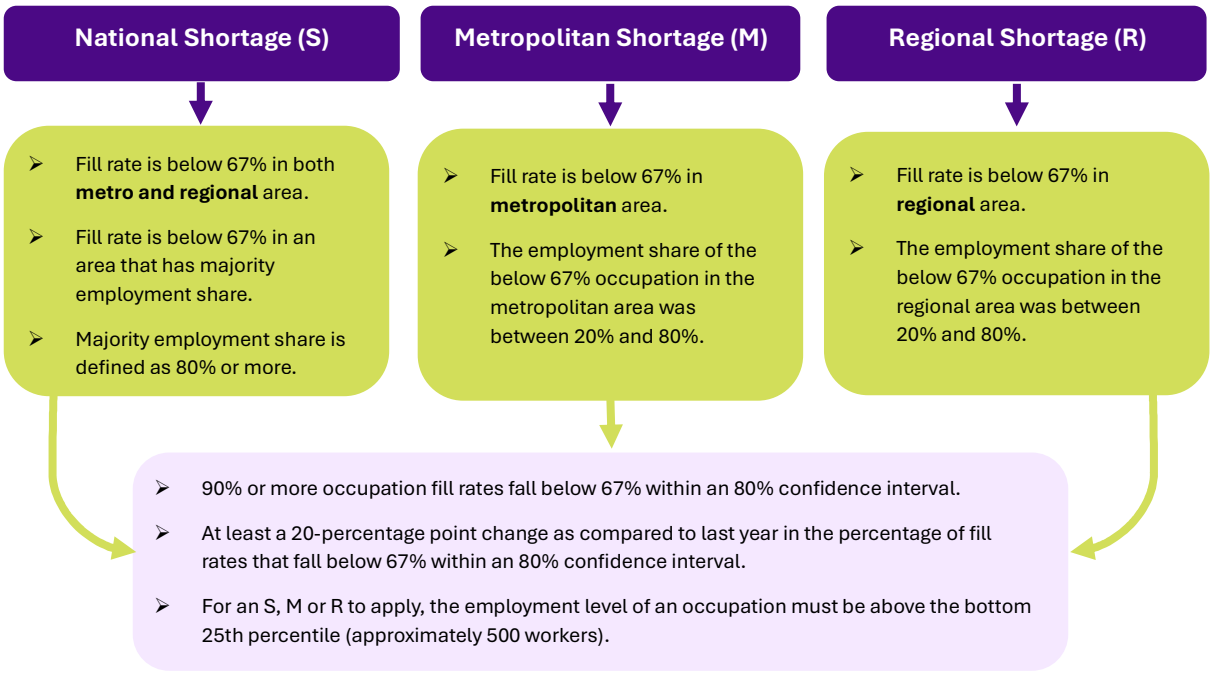
1. **Partial mapping.** There are 56 occupations where an occupation from 2013 ANZSCO splits into multiple occupations in the 2022 ANZSCO version. For these, the ratings for the 2022 ANZSCO occupations are equal to their parent ratings in 2013 ANZSCO.
2. **Mixed mapping.** There are 7 occupations where a combination of concordance and multiple occupations from the 2013 ANZSCO merge into one or more occupations in 2022 ANZSCO. When these amalgamated 2013 ANZSCO occupation have conflicting ratings, the 2022 ANZSCO ratings are based on the rating of the 2013 ANZSCO occupation with the largest employment size share (according to ABS Census 2021). In each of these 7 occupations, a single rating accounts for a clear majority.
3. **New Occupations.** These are 15 new occupations that appeared as part of the 2022 ANZSCO update and are not present in 2013 ANZSCO. These occupations lack data, including GBM fill rates. As such, the previous year ratings (2024 OSL) are applied as the

preliminary ratings for these occupations in the first instance, which are marked as double asterisk (**). These ratings will be verified further using additional evidence including feedback from stakeholders.

The 2nd spreadsheet of the Workbook: 2024 OSL Aggregated Inputs, highlights how the 2013 ANZSCO results are mapped to 2022 ANZSCO.

The below visual illustrates how the ratings are generated.

Figure 6: Occupation shortages scenarios



Source: Jobs and Skills Australia.

Manual verification of ratings

The model, data-driven based results are reviewed manually. This second stage is designed to ensure that the ratings generated by the modelling framework are sensible and reliable.

The data sources for the verification are:

4. Results from the *OSL Stakeholder Survey*
   3rd spreadsheet of the Workbook: 2024 OSL Aggregated Inputs
5. Analysis of *Internet Vacancy Index* (IVI) to unemployment ratios
   In Workbook: LD_LS_ratio_CI_2013
6. Analysis of Recruitment difficulty rates from Recruiting Employer Outlook Survey (REOS)
7. **In Workbook: REOS_RD_CI_2013**
8. Additional metrics from the Survey of Employers who have Recently Advertised (SERA):
   4th spreadsheet of the Workbook: 2024 OSL Aggregated Inputs
   - Qualified applicants per vacancy
   - Suitable applicants per vacancy
   - Vacancy duration (weeks unit, yearly from May 2023 to May 2024)
   - The gap between weighted qualified applicants per vacancy and weighted suitable applicants per vacancy.

Stakeholder survey scoring method

Responses from the stakeholder survey are used, along with other indicators, to verify the preliminary occupation ratings. To simplify the verification process, responses are collated, scored and aggregated to produce a signal (Shortage/No Shortage) and a signal strength (weak, moderate, strong) for each occupation reported on.

Both the signal and the signal strength for each occupation, in principle, are determined by the number of survey responses, the completeness of each response, the evidence supporting the response and the representativeness of the response.

Summary of the IVI to unemployment ratios

Metric	Description
Raw average ratio	The average across 3 financial years, 2021–22 to 2023–24.
The annual change in the raw ratio	The annual changes from: 9. 2021–22 to 2022–23; and 10. 2022–23 to 2023–2024.
The Z-scores of the ratio and their groupings according to their deviation from the mean[1]	The 3 year average ratio and 2023–24 ratio were converted into Z-scores and then grouped into whether they were 0.5 standard deviations below, within or above their respective mean.

Summary of recruitment difficulty rate metrics

Metric	Description
Raw average rate	The average across 3 financial years, 2021–22 to 2023–24.
The annual change in the raw rate	The annual changes from: 11. 2021–22 to 2022–23; and 12. 2022–23 to 2023–2024.
The Z-scores of the rate and their groupings according to their deviation from the mean[2]	The 3 year average rate and 2023–24 rate were converted into Z-scores and then grouped into whether they were 0.5 standard deviations below, within or above their respective mean.

Summary of additional SERA data metrics used

Metric	Description
Z-scores of the average	The average across 3 financial years, 2021-22 to 2023-24.
The annual change in the metrics	The annual changes from:

[1] The Z-scores are based on the log-transformed rates. Z-scores are a way to standardise raw figures. It is a statistical measure that quantifies the distance between a data point and the mean of a dataset. It's expressed in terms of standard deviations. It indicates how many standard deviations a data point is from the mean of the distribution. Z-scores above the mean are positive, while those below the mean are negative.
[2] As per footnote 12.

	2021–22 to 2022–23; and 2022–23 to 2023–2024.
The Z-scores of the metrics and their groupings according to their deviation from the mean	The 3-year average and 2023-24 metrics were standardised and then grouped into whether they were 0.5 standard deviations below, within or above their respective mean.

More detailed explanations are in the OSL methodology: 2024 Occupation Shortage List Methodology.

Additional definitions of metrics

The fill rate is the percentage of vacancies employers advertised that were filled. The metric is calculated by dividing the number of filled vacancies by total advertised vacancies.

The average number of qualified applicants per vacancy is calculated as the sum of the number of qualified applicants divided by the sum of advertised vacancies. Qualified applicants are the applicants who are assessed by employers as meeting the required qualification criteria of an advertised vacancy.

The average number of suitable applicants per vacancy is calculated as the sum of the number of suitable applicants divided by the sum of advertised vacancies. Suitable applicants are those who are deemed by employers to be suitable for the job advertised.

Metropolitan area refers to Capital City, while **Regional area** refers to Rest of State locations. Capital City and Rest of State areas are defined by the Australian Statistical Geography Standard (ASGS): Volume 1 – Main Structure and Greater Capital City Statistical Areas, July 2016.

Suitability gap is the difference between the average number of qualified applicants per vacancy and the average number of suitable applicants per vacancy as a proportion of the number of qualified applicants per vacancy.

A suitability gap greater than zero would mean that there are fewer suitable applicants than qualified applicants. The formula for calculating the suitability is:

$$\text{Suitability gap} = \left(1 - \frac{\frac{Number\ of\ Suitable\ applicants}{Number\ of\ vacancy}}{\frac{Number\ of\ Qualified\ applicants}{Number\ of\ vacancy\ Qualified\ applicants\ known}}\right) * 100\%$$

Additional information on derivation of metrics

Replacement rates and LD/LS ratios

The *average replacement rate* is the *weighted average* of the **replacement rate**, with the weights being the ***employment size*** but occupations which are out of scope are removed (for example, the weights for certain groups such as major group 8 will be smaller than if you just used all occupations).

The **vacancies** over **unemployed** has two measures:

- one based on **IVI** and,
- one based on the adjusted **IVI** results from the SAMM.

Both are *weighted* based on the *proportion of employment size in each group* that is in scope of the SPL.

Weights are apportioned like so

$$\omega_i = \frac{measure_i}{\sum_i measure_i}$$

Where measure is the variables selected as weights:

- For replacement rates, this is **employment size**.
- For vacancy/unemployed, this is **vacancies**.

Standard deviations are calculated as follows:

$$\sigma_x = \sqrt{E(x^2) - (E(x))^2}$$

Where $E(x)$ is the *weighted average* of x and $E(x^2)$ is the *weighted average* of x^2.

Because **vacancies** are the numerator, using it as a weight is not rational, so instead the standard deviation of unemployed over vacancies is derived and then related by the formula:

$$\sigma_{f(x)} = \left|\frac{df}{dx}\right|\sigma_x$$

$$\sigma_{1/x} = \frac{1}{x^2}\sigma_x = \left(\frac{1}{x}\right)^2 \sigma_x$$

$$\sigma_{vacancies/unemployed} = \left(\frac{vacancies}{unemployed}\right)^2 \sigma_{unemployed/vacancies}$$

For a full example, the standard deviation of the **replacement rate** is:

$$\sigma_{RR} = \sqrt{\sum_i {RR_i}^2 \frac{EMP_i}{\sum_i EMP_i} - \left(\sum_i RR_i \frac{EMP_i}{\sum_i EMP_i}\right)^2}$$

REOS

One of the questions in the REOS questionnaire is *"Overall, would you say that it was difficult to fill this vacancy?"*. Based on the output of this question, the *average recruitment difficulty* for REOS is calculated by creating a binary variable which returns:

- 1 when recruitment difficulty is *Yes*, and
- 0 when it is *No*.

For the binomial variable, the *confidence interval* can be calculated as follows:

$$RD\_CI = RD \pm z_{\alpha/2}\sqrt{\frac{RD\,(1-RD)}{n}}$$

Where

RD = recruitment difficulty

$z_{\alpha/2}$ is the z table score for $(1 - confidence)/2$. For 80% confidence it is approximately 1.28. Also, n is the total sample size