

Setting the assessment standards and rating approach

On this page

Standards for assessment

Using rating scales and benchmarks

Cognitive ability tests

Clear standards and a consistent rating approach lead to a reliable assessment process and fair and defensible selection decisions.

Standards for assessment

It is good practice to be clear upfront on the standards you will use to determine whether candidates meet the requirements of a role.

The role description and **job advertisement** help to inform the standards you need for assessing capabilities, knowledge and experience.

Assessing capabilities

The **role description** shows the level required for each capability (from foundational to highly advanced). In the recruitment context, you need to assess candidates against the capability levels in the role description.

In the **NSW public sector capability framework** each capability is supported by behavioural indicators at the different levels. These are indicative behaviours allowing scope for agencies to contextualise them according to the particular role and agency setting.

Assessing knowledge and experience

Knowledge and experience need to be assessed when they are included in the role description as they form part of the standards for the role. The Role Description Development Guideline contains advice about when to include knowledge and experience in role descriptions.

The assessment does not need to be separate from the assessment of capabilities as knowledge and experience as these are often demonstrated along with the capabilities. You can design your assessments in a way that draws this information out. For example, an Executive Assistant role may have an experience requirement to 'provide wide-ranging support to a senior leader' (i.e. CEO or head of an agency). You could design your interview questions, for example, to allow candidates to explain their experience in one or more questions.

Make sure you set clear standards for knowledge and experience so you can record your findings against these when consolidating results.

To determine standards for assessing either subject matter or functional knowledge and experience, consider the depth, breadth and context needed for the role. Then decide how this should be demonstrated through the assessment.

For a legal role, the level of complexity of cases previously managed and breadth of legal practice may determine or inform your standards.

Assessing essential requirements

Essential requirements are usually yes/no answers. You can often ask whether candidates meet the essential requirement in the application form. Responses of successful candidates should be validated at some stage during the recruitment and selection process, often through pre-employment checks.

Assessing other attributes

Other attributes that can be assessed include motivation to do the particular type of work (e.g. child support worker), to work in your organisation or the NSW public sector more generally, or willingness to undertake certain requirements of the role such as travel. It is up to you to decide the standards but this should be based on the role context and needs of the organisation.

Using rating scales and benchmarks

A rating scale or benchmark is used to evaluate candidates' responses and behaviours in assessment activities.

A systematic approach to assessing candidates involves:

- applying the same approach to all candidates
- all assessors using the same approach

- applying the method consistently.

Rating scales or benchmarks should be used to evaluate the capabilities or other essential requirements being assessed, not skill on the assessment itself (e.g. interview technique).

The consistent use of a standardised rating scale:

- allows for comparison between the ratings of different assessors
- enables meaningful comparisons between candidates
- supports the integration of assessment results for each candidate at the end of the process.

Rating scales and benchmarks allow you to determine which candidates meet the standards for the role. Once you have identified those candidates, you need to decide who is well suited to the role requirements. This may not necessarily be the person who scored the highest rating.

Rating approaches

Different rating scales and benchmarks are suited to different situations or stages in a process. The following table shows the different rating approaches to consider depending on the expected number of candidates.

	Benchmark (met / not met)	3-point rating scale	5-point rating scale
Low-volume recruitment (e.g. for a single role)	Yes	Yes	
High-volume recruitment (e.g. bulk recruitment, graduate intake, establish a talent pool)		Yes	Yes

Rating scales

A rating scale allows for differentiation between candidates when there could be a number of candidates who meet the standard required. Descriptive rating scales generally result in more consistent scores between assessors and help to minimise subjective biases because all assessors understand what they are looking for. Descriptions can be kept fairly broad to reflect general performance against each capability.

While rating scales help to differentiate between candidates, you also need to consider other factors such as motivation and fit for the role when making your selection decision. Your aim is to find the person who meets the standards and is well suited to the role requirements, not necessarily the highest scorer.

See [examples of 3-point and 5-point descriptive rating scales](#).

Benchmark approach

A benchmark approach (e.g. met / not met) is useful when it is less important to make fine distinctions between candidates, for example if you receive few applications.

Combining approaches

You may also consider combining approaches by using benchmarking during pre-screening and a rating scale for the capability-based assessments.

Example of mixed model for Customer Service Representative (9 roles)

Svetlana decides to use the benchmark approach to assess two targeted questions in applications for the nine Customer Service Representative roles she is recruiting for. This allows for the initial selecting out of those candidates that do not meet the focus capabilities addressed in the targeted questions.

For the capability-based assessments (in this case, a cognitive ability test, group exercise, role-play and behavioural interview) Svetlana will use a 5-point rating scale. This will help her to make finer distinctions between the remaining candidates and to progressively reduce the candidate pool to a more manageable size.

Cognitive ability tests

Setting cut-offs and ranges

Cognitive ability tests are scored and interpreted by qualified professionals. The score is the raw number of correct answers. To give meaning to the score, test results are compared to an appropriate **comparison group** who have completed the test.

You can either consult your in-house experts or seek advice from suppliers on the **Talent Acquisition Scheme**.

Comparison or 'norm' group

The comparison group should be chosen based on the requirements for the role.

Comparison groups are often available by country (e.g. Australian adults), sector (e.g. public sector), industry (e.g. engineers, science and technology) or role level (e.g. manager and professional).

It is important to consider the appropriate comparison group each time you use a test to ensure your benchmark is set appropriately.

Applying rating scales to cognitive ability test results

Your rating scale can be applied to cognitive test results to help integrate them with other assessment results.

This table shows an example where a hiring manager is assessing the Think and Solve Problems capability using a cognitive ability test. The candidate’s raw score is 28. In relation to the comparison group, her score is in the 86th percentile. Using either the 5-point or 3-point rating scale, she achieves 4 or ‘exceeds’ on Think and Solve Problems.

Cognitive score ranges	Percentiles	5-point rating	3-point rating
Well above average	91-99	5	Exceeds
Above average	71-90	4	Exceeds
Average	31-70	3	Meets
Below average	11-30	2	Development required
Well below average	1-10	1	Development required