

Assessment is a common but high stakes practice in organisations used for;

- Hiring the right people
- Shaping organizational culture via core competency assessment
- Assessing and developing technical competency
- Assessing and developing leadership competencies
- Ensuring Compliance – for example health & safety behaviours
- Improving productivity via Performance Appraisal
- Evaluating Staff engagement levels
- Assessing organisational climate and culture



So, getting it right is critical.

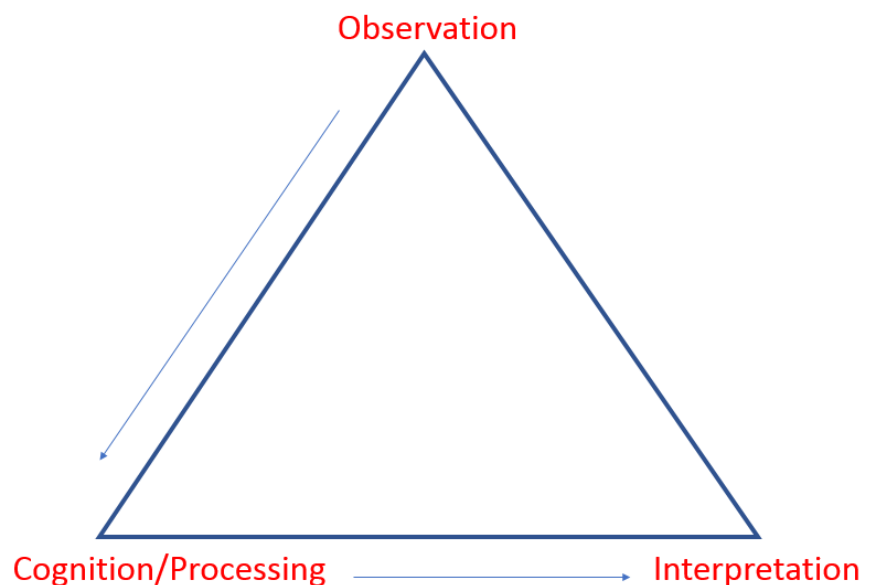
Unfortunately, it is fair to say that the vast majority of organisations do not get it right. Assessment is a process fraught with bias of many kinds.

Is bias inevitable?

When we think of assessments, we think of giving a rating or score that represents the demonstration of a work task, a responsibility, a behaviour or other performance standard.

But there is a sequence of events that underlies this judgement that we are not consciously aware of.

First there is (hopefully) an observation of the person or a self-report. Then a processing phase where a comparison is made with a standard – that should be clearly set out in the assessment tool. Finally, this is interpreted into a score and sometimes a written comment as well.





1. Observation

The observation process is about how assessors pay attention to and actively select information about the assessee and their activities. Unless the assessment is an on-job assessment this means drawing upon memory.

Assessors must identify relevant information that they will use as a basis for their judgment. That is they must recognise what is relevant and actually perceive it, then store it in memory.

The observational process is influenced by many conscious, unconscious, situational and personality factors. Research has shown that assessors shown the same video of an assessee will pay attention to different aspects of their activity or behaviour.

2. Cognition/Processing

This is the phase in which assessors retrieve the information they have gathered and use contextual information and their own prior knowledge to make sense of it.

They make use of a categorisation mechanism – a comparison with some sort of standard – implicit or explicit.

An explicit standard would be a very specific indicator or statement in the assessment tool.

However, in most cases, the standard is not specific. In these cases, the assessor will compare what they have observed with an implicit standard. A standard derived from their own beliefs and experience; their own idea of competency/performance and some specific examples they can recall.



Examples used include people assessed in the past, recall of their own level of performance and skills in a similar context, and of staff and colleagues with different levels of experience and expertise.



3. Integration/interpretation

In the final phase assessors combine different sources of information to form an overall judgement.

The information from the observation and processing phases is reviewed. It is weighted and put together into a cohesive mental picture. From here assessors make judgements which have to be translated into a format used by the assessment tools – typically a rating scale and comments.

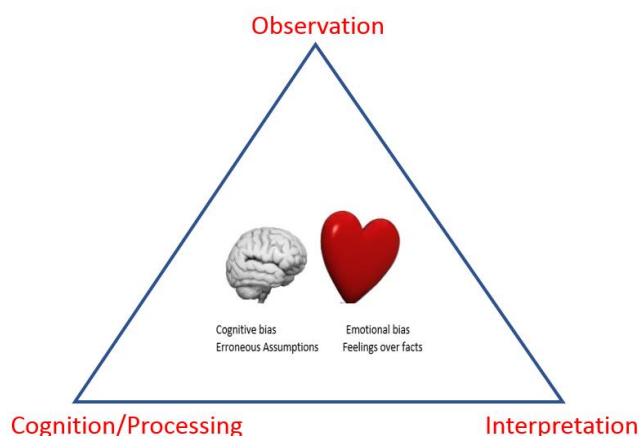
The strength and clarity of their mental picture will affect the assessor's confidence in their judgment.

Where does it go wrong?

The assessment process is prone to bias at all stages.

Bias is defined as *a systematic error, or deviation from the truth, in results or inferences.*

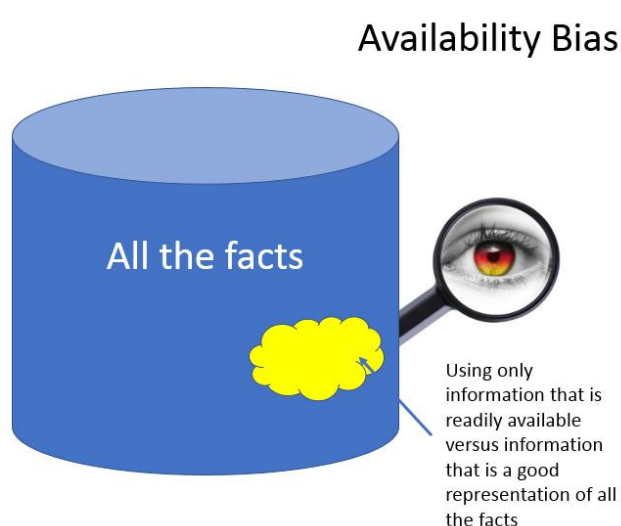
Biases can operate in either direction: different biases can lead to underestimation or overestimation of the true situation. They vary in magnitude: some are small, even trivial, but some are substantial (so that a judgment may be entirely due to bias).



Bias stems from our thoughts and our feelings.

Cognitive bias

Cognitive bias is faulty thinking. There are many types. Some are particularly applicable to the assessment process.



Availability bias is the use of information that is easily available rather than making the effort to look at the whole range of information that exists.

Various types of recall bias are the main contributors.

Recall bias is a systematic error that occurs when people don't remember events or experiences accurately or some details are not recalled. What we remember is influenced by subsequent events and experiences.

- We remember recent events best
- We remember the first and last of a sequence of events better than those in the middle
- We remember negative experiences more than positive ones
- Dramatic events over the routine.
- We are not good at remembering things that happened in a different context than the present.

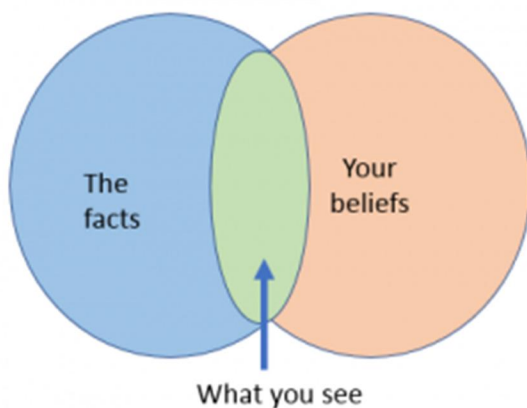
Anchor effects also contribute to availability bias. An example is information on a previous assessment. Assessors use that available information as a basis for their new assessment, rather than making a completely independent evaluation.

Observer bias

The collection of information for assessments is the phase most prone to bias. Selective perception is a common bias where expectations about people and situations affect what is seen and heard. The stereotypes we have about categories and groups of people shape these expectations subconsciously.



Confirmation Bias



Then there is a tendency to see and hear those things that confirm existing beliefs and to filter out things that don't agree.

The **format of the assessment tools** affects accuracy. Where assessment items are not specific research shows large differences in interpretation. The same applies to rating scales that don't have extended descriptions for each rating point.

The layout and order of items in the assessment can encourage raters to mark many items with the same rating without proper consideration.

Emotional Bias

Feelings trump facts. Our feelings influence how we think.



The halo effect

We tend to carry the positive or negative traits of a person from one area to another in our perception of them. The most common example is physical attractiveness. Those who are more attractive tend to be rated more positively on any dimension. In fact, research has shown that in elections it is often the most attractive person who prevails, regardless of their policies.

Research shows that very often, instead of direct observation, the overall impression of a person is applied to assess specific attributes. For example, in a 2019 study of 360 feedback on leadership competencies analysis of all the interactions between raters, rater type and the assessment items showed that the ratings reflected the perception of overall personality, rather than the specific competencies.

Many research studies have established that performance appraisals are primarily a measure of the staff/supervisor relationship, not the staff member's actual perform.

Wanting to be liked

Related to this is that assessments are typically more favourable than the assessor's true opinion. This may be due to a need to be positively perceived in the workplace. In particular so they can avoid the discomfort of awkward conversations with offended assessees with marginal results.

What about self-assessments?

Self-assessments prone to bias for many of the same reasons. In addition, the assessee complete the assessment to fit their perception of supervisor expectations.

Overall people tend to rate themselves optimistically, most people believe they are 'above average'. Finally, people overestimate their insight into their own motives and actions.

We also overestimate our ability to know what others think, feel and believe.

What can be done to minimise bias?

What we can do about bias depends on the perspective we take on assessment. There are 3 prevalent perspectives.



1. Assessors are trainable

- Use calibration sessions to clarify the anchors on rating scales to minimise the variability in interpretation.
- Ensure they focus on the specific observable aspects to be assessed.
- Discourage assessors from using their own performance/competence or characteristics as comparative reference points
- Remind them to be fair, have an independent mind and take an egalitarian approach. (Research has shown that calling attention to stereotypes actually accentuates their use).
- Ensure they view marginal assessments as an opportunity to help staff in their development.
- Encourage assessors to take the time to consider each item separately.
- The assessment uses the terminology specific to your organisation



2. Assessors are fallible

In this perspective errors will occur despite training because people are easily influenced. This means the assessment tool must have safeguards to minimise potential bias.

- Headline the assessment with clear instructions on its content and how to make the judgments.
- Make the questions as specific and unambiguous as possible
- Use suitable rating scales
 - for very specific standards just a yes/no option
 - for development assessments a frequency scale is appropriate
 - for core and leadership behaviours an agreement scale works well.



- Make sure the scoring system matches the scale. For example, a neutral mid-point such as neither agree nor disagree is equivalent to no assessment. It should have a NULL score.
- Ensure rating scales have suitable descriptive labels for each point.
- Ensure that assessors space out their assessments so as to minimise the contrast effect (the tendency to make the assessment different than the preceding one(s)).
- Each assessment should be independent. Make sure the assessor does not have access to previous assessments for the individual that may influence their judgment.
- Make use of automated or suggested scoring – so that the scoring of each factor is properly aligned to the scoring of its performance indicators or indicative behaviours.
- Encourage the use of shareable journals to reduce the reliance on memory.

The assessor has a unique view



The variation in assessments is due to the forming of relevant and valid differences in opinion, partly as a result of differences in the observation context. This is especially relevant where behaviours are not directly observable – for example ‘professionalism’, ‘reliability’, ‘integrity’.

Consequently, assessors may spot different aspects of an individual’s performance and form different interpretations of them. Variations in assessor judgements may very well represent variations in the way performance can be understood, experienced and interpreted. The inconsistencies among assessors’ interpretations might be complementary and equally valid.

This is particularly applicable to core and leadership competency assessment.

- Ensure the assessment tool has as much explanation as possible around broad terms such as ‘Working in a Team’
- Provide specific examples that are worked through and rated.
- Require comments to provide the context for the ratings given.

References

- Gomez-Garibello, C. , Young, M. (2018) Emotions and assessment: considerations for rater-based judgements of entrustment. *Medical Education* 52: 254–262
- Gingerich A, Kogan J, Yeates P, Govaerts M, Holmboe E. (2014) Seeing the ‘black box’ differently: assessor cognition from three research perspectives. *Medical Education* 48 (11):1055–68.
- Gauthier G, St-Onge C, Tavares W. Rater cognition: review and integration of research findings. (2016) *Medical Education* 16;50 (5):511–22.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kazdin, A.E. (1977) Artifact, bias, and complexity of assessment. *The ABC’s of reliability. Journal of Applied Behaviour Analysis* 10. 141-150