Workbook too

Observing and evaluating in a personnel diagnostic setting

Practical benefits:

*You will learn to **conduct behavioral observation and assessment in personnel diagnostic settings in a structured and objective manner** to make well-founded and comparable suitability assessments.*

Introduction and overview

Observing and Evaluating in the Context of Personnel Diagnostics

In everyday life, people **observe and evaluate** almost constantly, often without consciously realizing it. These processes help individuals **navigate their environment**, but they are strongly influenced by **personal experiences and socialization**. As a result, **perception biases** can arise, leading to **misjudgments**.

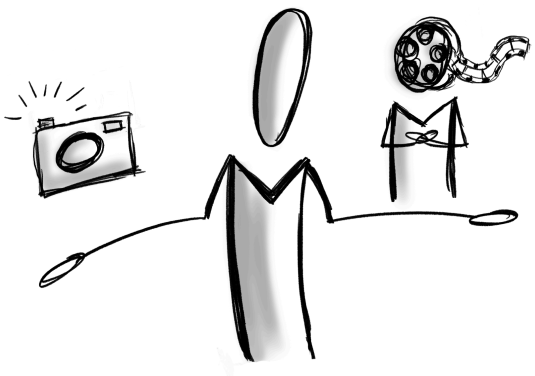
To **counteract these biases**, **standardized procedures** are used in **aptitude assessment**, such as **group discussions, role-playing exercises, presentation tasks, or interviews**. The requirements for such procedures, as well as for **systematic behavioral observation and evaluation**, are outlined in **DIN 33430**.

The goal of **structured behavioral observation** is to **systematically capture observed behavior** and analyze it in relation to **predefined aptitude criteria**. This involves not only assessing the **behaviors themselves** but also considering the **conditions and situations** in which they occur.

To ensure this, **precise observation systems** are necessary, including:

- ➢ Operationalization of aptitude criteria
- ➢ Development of suitable observation forms
- ➢ Thoughtful combination of different diagnostic procedures

The **target audience** for training in these requirements includes **personnel diagnostics professionals and observers** who are part of the **personnel assessment process**.



The observation and evaluation process

Structured behavioral observation follows a **clearly defined process** to ensure an **objective and transparent assessment**. In this process, **aptitude criteria** are **precisely operationalized, systematically observed, and evaluated** based on **established criteria**. The following **steps** are essential for a **well-founded behavioral observation and evaluation**:

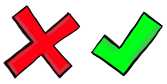
Creating an observation system

Before **targeted behavioral observation** can take place, **aptitude criteria** must be **clearly operationalized**. This means describing them in a way that makes them **observable in behavior**. A **structured observation system** and an **appropriate observation sheet** help to systematically capture **relevant behaviors**. The **selection and combination of suitable diagnostic methods** ensure a **comprehensive assessment foundation**.

Targeted observation

During the observation, the focus is on identifying **specific behavioral anchors** and **accurately documenting** the observed behavior. The observation should be conducted as **objectively as possible**, without interpretation or premature conclusions. Both the **actions displayed** and the **situational conditions** in which they occur are taken into account.

Assessment based on suitability characteristics



After the observation, the recorded notes are assigned to the **relevant competency criteria** and checked for accuracy. The **evaluation is based on predefined assessment criteria**, ensuring that the judgment remains **consistent, comparable, and transparent**.

Integration of the individual results

ITo obtain a well-founded overall assessment, **individual observations are consolidated**. This can be done through an **observer conference**, where different perspectives are reconciled, or through **mean score calculations**, which provide a mathematical summary of the evaluations. The goal is to achieve an **objective and valid assessment** of the observed individual.

What steps have you already been involved in? What experiences have you had with it?

Observing

Observation is a **core element of aptitude diagnostics**, as it allows for **systematic recording and objective evaluation of behaviors**. To minimize biases, observation follows **defined criteria and standardized procedures**, ensuring a **reliable assessment**. This chapter explains the **fundamentals, steps, and techniques** of structured behavioral observation to enable a **well-founded and practical application**.

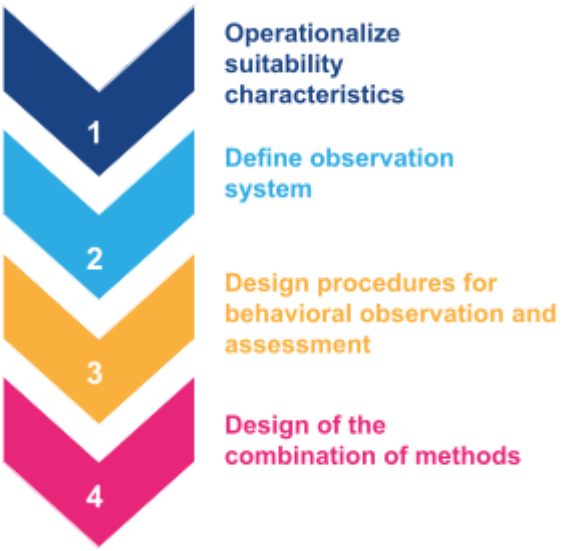
Characteristics of Structured Behavioral Observation

Unlike intuitive everyday observation, **diagnostic behavioral observation** is characterized by the following features:

- **Systematically bound** → Observation follows **defined rules, specific criteria, and controlled conditions** regarding location, time, and duration. The evaluation process is also clearly structured.
- **External** → The focus is not on one's own behavior but on that of another person. This is conducted by **trained aptitude diagnosticians and observers**.
- **Laboratory** → Observations take place in a **deliberately arranged setting** designed to make relevant behaviors **visible**.
- **Direct** → Behavioral observation occurs **in real time**, rather than relying on later recollections.
- **Non-Participatory** → Observers generally remain **passive**, meaning they do not actively influence the situation.

Structured preparation of the observation

For **structured behavioral observation** to be **reliable and meaningful**, careful preparation is essential. This involves several **interconnected steps**:



1. Operationalize suitability characteristics → Before targeted observation takes place, relevant suitability characteristics must be clearly defined and translated into observable behaviors.

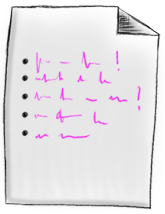
2. Define observation system → A structured system ensures the systematic documentation of behavior.

3. Design procedures for behavioral observation and assessment → The correct selection of methods minimizes distortions and ensures comparability.

4. Design of the combination of methods → Often several procedures are combined to get a more comprehensive picture.

The following sections look at these steps in more detail.

1. Operationalize suitability characteristics



Through a **requirement analysis**, the relevant **suitability criteria** are identified, determining the key attributes necessary for success in a **professional role, training program, or academic study**. These criteria serve as a **central reference** for observations, ensuring that only **job-relevant behaviors** are recorded and assessed.

To ensure that assessments remain **objective and consistent**, **suitability criteria** must be **precisely operationalized**. This involves formulating **behavior-based examples** that describe typical **actions or statements**. By doing so, all **assessors and observers** share a **common understanding** of the criteria.

To structure the **operationalization process**, three levels of **abstraction** are typically distinguished:

The abstraction levels	Description	Examples of observation notes
Abstract description of characteristics	General definition of suitability characteristic	*Dealing with change: Uses emerging changes constructively and deals with them flexibly.*
Behavioral operationalization	Specific behaviors that make the trait recognizable	*+ React squickly to changes* *+ Changes to plans are well received* *+ Quick conversion to new solutions* *- Persists in the original solution, even if it no longer fits the new situation* *- No change in perspective* *- Change leads to stagnation*
Concrete description of behavior	Relevant examples for observation	*Participating person accepts plan changes well:* *+ Accepts new conditions without resistance or excessive criticism* *+ Revises outdated problem solutions*

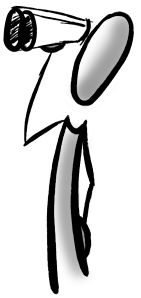
2. Define observation system

The next step is to establish the basic framework of the observation system. This provides a **clear structure for data collection** and ensures that the observation is **targeted** and **comparable**. The following key questions serve as a guideline:

- ☐ **When does the observation take place?** → *Is it conducted once or over a longer period? Does it take place during a specific task or across multiple phases?*
- ☐ **Where does the observation take place?** → *Is it conducted in a natural setting or under standardized laboratory conditions? Are there external influencing factors that need to be considered?*
- ☐ **Which behaviors should be observed?** → *Which suitability characteristics are the focus? Which specific behavioral indicators should be recorded?*
- ☐ **How are the results documented?** → *Is the documentation open-ended or standardized, e.g., using coding systems, rating scales, or written observation protocols?*

A well-defined observation system simplifies later evaluation and significantly contributes to the objectivity and traceability of the assessment.

3. Design procedures for behavioral observation and assessment



Various methods are available for observing suitability characteristics, including presentations, role plays, group discussions, or case studies. To ensure these methods provide reliable results, they must be carefully designed and standardized.

DIN 33430 specifies that all candidates must be observed and assessed under the same conditions. To achieve this, essential measures include **observer training** and **precise instructions** for both observers and participants.

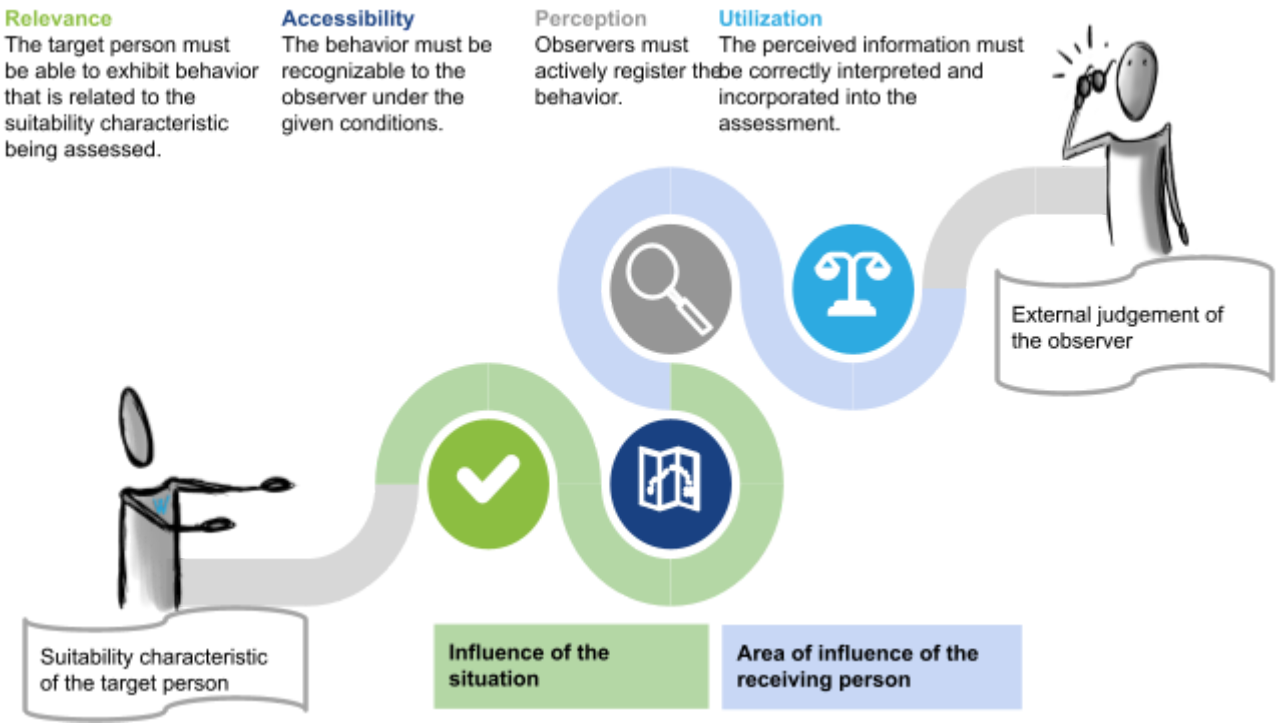
Additionally, exercises should be designed to be **low-distraction** and appropriately challenging. Observations must take place in situations that genuinely elicit the behavior being evaluated. For example, if creativity is being assessed, the exercise must provide opportunities for demonstrating creative solutions.

According to the **Trait Activation Theory (TAT)** by Tett & Gutermann (2000), two key factors should be considered in method design:

- **Situational Stimuli** → The exercise must provide sufficient triggers to elicit the relevant behavior.
- **Freedom of Action** → Candidates need enough room to behave authentically.

This prevents artificial test settings from leading to unnatural or forced behavior.

Another useful model for precise behavior observation is the **Realistic Accuracy Model (RAM)** by Funder (1995), which outlines four key areas that contribute to an accurate assessment of suitability characteristics.



To **ensure the quality of observation**, exercises should be tested in advance and optimized if necessary. In particular, **role-play participants** need detailed instructions to display comparable behavior across all participants.

4. Design of the combination of methods

Once the **suitability characteristics** and **exercises** have been defined, it must be determined **which characteristics will be observed in which exercises**. Ideally, **multiple exercises** should be used for each characteristic to create a broader basis for assessment. The following table provides an example of **which suitability characteristics could be assessed in different exercises**.

Sustainability / Exercise	Role play	Group discussion	Presentation	Case Study
Creativity			✖	✖
Solving ability		✖		✖
Conflict capability	✖	✖		
Leadership skill	✖			✖
Expressiveness		✖	✖	

Additionally, it should be planned **who will observe each exercise** to ensure an **unbiased assessment**. **Rotating observers** can help **reduce biases**. To avoid **overloading observers** and to ensure the **clarity of the assessment**, a maximum of **five suitability characteristics (preferably three)** should be considered per exercise. **Rotating observers** can further help **minimize biases**.

Observations of **suitability characteristics** should be recorded in an **appropriate format** – preferably using a **standardized observation sheet**.

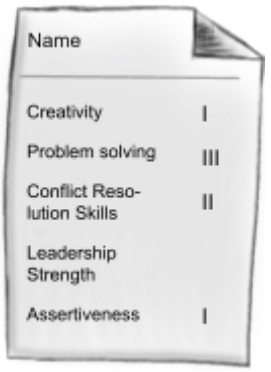
Observation techniques

After the **observation** has been systematically prepared, the question arises as to which **methods** are suitable for capturing behavior. In **aptitude diagnostics**, three **main types of observation techniques** are used. The **choice of method** influences both the **observation process** and the **subsequent assessment** of the candidates.

Variant A: Free observation form

Variant B: Categorized observation form

Variant C: Checklist



Variant A: Free observation sheet

An empty observation sheet is used, with only a narrow margin on the side to assign the respective **suitability characteristics** after the observation phase. This variant can be classified under the **behavioral observation approach** in scientific research.

Advantages	Disadvantages
+ Maximum flexibility in recording behavior + Unexpected or spontaneous behavior patterns can be documented + Particularly suitable for exploratory analyses	- High effort for later evaluation - Lower comparability between observers - Greater subjectivity of assessment

Variant B: Categorized observation sheet

This sheet already provides fields for the relevant **suitability characteristics**. Observers assign their notes directly to the predefined characteristics during the observation. This simplifies and accelerates the later assessment. This variant can be classified under the **frame-of-reference approach**.

Advantages	Disadvantages
+ Structured and systematic recording of behavior + Improved comparability between observers + Less effort required for evaluation because the assignment is already made during the observation	- Less flexibility because observations have to be classified into predetermined categories - Potential bias if relevant behaviors are not included in the categories

Variant C: Checklist

In this case, the **indicators** for the observed **suitability characteristic** are predefined, and only a tally sheet needs to be maintained. In the end, a **frequency list** is available – however, without any indication of **duration** or **intensity**.

Advantages	Disadvantages
+ Objective and quick documentation + High comparability of results + Simple and time-saving handling	- No recording of the context, duration or intensity of the behaviors - Limited differentiation of complex or situation-dependent characteristics

Evaluate

After observing the behavior of the participants, the actual **assessment** follows. The central challenge is to derive a **fair, objective, and comprehensible evaluation** from the collected impressions.

Three key steps are crucial:

1. **Rules for individual assessment** → Which criteria are included in the evaluation, and how are they weighted?
2. **Methods of assessment** → Which approaches enable a systematic evaluation, e.g., the **operationalization approach** or the **behaviorally anchored rating scale**?
3. **Redundant observation and judgment formation** → How can assessments from multiple observers be combined to avoid biases?

This chapter explains how **assessments** can be structured and validated to provide a **reliable foundation** for suitability diagnostics.

Rules for individual assessment

In **suitability diagnostics**, numerical **assessment** has largely become standard. In advance, it is determined which **suitability characteristics** are particularly relevant. These can be given greater weight and thus have a larger impact on the overall evaluation.

The **observation notes** are then compared with the **assessment scale**. In the **behavioral observation approach**, protocol notes are assigned to the respective suitability characteristics. In the **frame-of-reference approach**, the focus is on whether the notes correctly correspond to the predefined categories.

Various **criteria** can be used for assessment:

- **Frequency:** How often was the behavior displayed?
  Example: "How many times did the participant behave cooperatively?"
- **Duration:** Over what period of time was the behavior observed?
  Example: "Was cooperative behavior demonstrated throughout the entire exercise?"
- **Intensity:** How pronounced was the behavior shown?
  Example: "How strong were the cooperative behaviors demonstrated?"

"The more, the better" doesn't automatically apply. Some **suitability characteristics** are more appropriate in moderation than in excessive expression. For instance, a person who occasionally offers help may be perceived positively, while excessive helpfulness could be seen as intrusive.

Additionally, **organizations** can define further **assessment rules** tailored to their **specific requirements**.

Methods for assessment

There are various methodological approaches to ensure a structured and comprehensible assessment. These help to make suitability characteristics tangible and to put evaluations on a reliable basis. Two common methods are: **Operationalization approach**, in which a feature is broken down into partial aspects, and the **behaviorally anchored rating scale**, which supports the assessments with concrete examples of behavior.

Operationalization approach

A common approach to designing **assessment scales** is the **operationalization approach**. Here, a **suitability characteristic** is divided into **different subcategories** to enable a more **differentiated evaluation**.

This method helps to make **abstract concepts more tangible** and reduces **subjective judgments**.

Example: creativity			
Positive aspects		**Negative aspects**	
Contributes innovative ideas to finding solutions		Just take the traditional solution route	
Does not act strictly according to plan and also allows for other solutions		Difficult to adapt to new situations	
Can tolerate ambiguity		Don't deviate from one point of view	
Develops new perspectives within a short period of time		Deals inflexibly with changes	
Stands out for the original way in which tasks are handled		Insists on his own point of view and leaves no room for other creative solutions	
Can reuse and improvise familiar objects			

Based on these **characteristics**, the **suitability-related assessment** is conducted on a **scale from 1 to 6**, where **higher values indicate a stronger expression** of the characteristic.

This method offers **high differentiation** and **better comparability** between different participants. It also reduces the risk that assessments are based on an **intuitive overall impression** rather than **specific behaviors**.

However, there are also challenges. The classification into **positive and negative aspects** may create the impression that all characteristics are equally weighted, even though some may be **more important** for the respective position. Additionally, in **complex situations**, making clear assignments can be **challenging**.

The **operationalization approach** is particularly useful when a **systematic analysis of suitability** is required, and **various facets of a characteristic** need to be **clearly defined**. It is well suited for situations where assessors require **concrete reference points** for their evaluation.

Behaviorally anchored rating scale

The **behaviorally anchored rating scale (BARS)** goes beyond a **pure numerical assessment** by supplementing each **rating level** with **specific behavioral examples**. This makes the **evaluation more transparent** and **comprehensible**, as it is **clearly defined** which **behavior corresponds to which level**.

This approach helps **reduce subjective interpretations** and achieve **higher observer agreement**, as all assessors have a **consistent point of reference**. Additionally, it allows for **more detailed feedback**, enabling participants to **specifically identify** which **behaviors should be improved**.

Scale	**Behavioral example on the topic of creativity** (6 = completely fulfilled until 1 = not fulfilled at all)
6	Frequently contributes ideas and innovative solutions, responds flexibly to spontaneous changes and can deviate from original perspectives.
5	Designs presentation and goes beyond the specified standard.
4	Does not bring new ideas, but is open to other people's creative solutions.
3	He or she is irritated by open-ended work instructions.
2	Fill out the introduction form without any particular creativity.
1	Insists on a conservative solution, even if it is not optimal. Wants to assert one's own point of view and is not open to other people's creative solutions.

The **strength** of this approach lies in the **clear connection between theory and practice**. Assessors do not have to evaluate **abstract categories** but rather **observe and assign specific behaviors**. This increases **objectivity** and **comparability** between participants.

However, there are also **challenges**: Developing a **behaviorally anchored scale** is **time-consuming**, as **specific behavioral examples** must be defined for each **suitability criterion**. Additionally, in **real observation situations**, the **exact described examples** may not always occur, requiring assessors to **exercise interpretation**.

Conclusion

Both the **operationalization approach** and the **behaviorally anchored rating scale** meet the **requirements of DIN 33430** for a **behavior-based specification** of **suitability criteria**. While the **operationalization approach** enables an **analytical breakdown** into **sub-aspects**, the **behaviorally anchored scale** provides a **practical reference** based on **concrete behaviors**.

Which of the two methods do you like more? And why?

Redundant observation and judgment

After the **evaluation has been carried out** using a **structured approach**, the question arises as to **how multiple individual judgments** can be combined. Especially when **multiple observers** are involved, it is crucial to **systematically integrate the judgments** without causing **distortions**.

To **maintain objectivity**, observers should first **submit their assessments independently**. A later **harmonization** can be useful but also **carries the risk** of **evaluation errors due to mutual influence**.

Therefore, **clear rules** should apply when **merging judgments**:

- ➢ If **coordination** on the **evaluation** takes place, it should be **documented**.
- ➢ **Guidelines** on how to handle **evaluation discrepancies** help to **avoid distortions**.
- ➢ **Frequent disagreements** may indicate that the **personnel diagnostic process** should be **revised**.

The **integration of individual judgments** can take place in **three ways**:

1. Joint decision-making through an observer conference

The **observers meet** and **discuss their individual assessments** to **reach a consensus judgment**.

Advantages	Disadvantages
+ Promotes the exchange of different perspectives + Disagreements can be resolved together + Qualitative aspects can be better taken into account	- Danger of **harmonization**: Individual observations are lost as observers adapt - Dominant people can influence the decision - Time consuming, especially with lots of reviews

2. Judgment formation by offsetting

The **individual evaluations** are **mathematically calculated**, usually by **determining an average**.

Advantages	Disadvantages
+ **Objective and reproducible**, as personal influence is minimized + **Fast** and **simply** feasible + Reduces distortions caused by group dynamics	- **No qualitative differentiation**: Deviating observations are leveled out - Particularly high or low values can distort the mean - No way, **individual justifications** to include

3. Hybrid Form: Average as a Starting Point + Discussion

First, an **average is calculated**, then the result is **discussed in an observer conference** and **adjusted if necessary**.

Advantages	Disadvantages
+ **Combines objectivity with reflection**	- Requires additional time for discussion
+ Extreme values can be discussed and justified	- Increases the complexity of the assessment
+ Reduces the influence of single dominant people	- There is still some risk of harmonization

The choice of method depends on the **framework conditions** of **suitability diagnostics**. In **highly standardized procedures**, calculation is often **more efficient**, while in **more complex situations**, an **observer conference** or a **hybrid form** may be **more appropriate**. A **well-structured process** helps ensure that **evaluation remains fair, transparent, and comparable**.

Review all steps of **behavioral observation and assessment** once again. If you are **already involved** in the **personnel diagnostic process**, **reflect** on **your approach** and **consider where adjustments might be necessary**. If you **have not yet gone through or planned a process**, you can now **develop your own approach**!