

# A FIELD GUIDE TO HUMAN CAPITAL ASSESSMENTS

---

PRACTICAL, EVIDENCE-BASED INSIGHT ON HOW TO  
PROPERLY ASSESS AN ORGANIZATION'S GREATEST RESOURCE



**CLOVERLEAF**

**BY SCOTT DUST, PH.D.**

CHIEF RESEARCH OFFICER, CLOVERLEAF

# FOREWARD

We want to help bridge the gap between science and practice in the field of human capital. Assessments are a great way to do just that. Assessments are both widely studied in academia and widely administered in organizations around the world.

Along those lines, this field guide offers scientifically-informed, practically-relevant, and easy-to-read insights about human capital assessments. The goal is to give readers a resource for pinpointing answers to common human capital assessment questions. We hope you find it useful.

## ABOUT CLOVERLEAF

Cloverleaf is a team-based, automated coaching platform that integrates custom coaching into the tools employees use everyday to improve performance. Cloverleaf supports the most popular assessment tools and integrates with Google for Business, Microsoft 365 and Slack for integrated coaching. To learn more or start a free account visit [www.cloverleaf.me](http://www.cloverleaf.me)

## ABOUT SCOTT DUST

Dr. Scott Dust is the Chief Research Officer at Cloverleaf and the Dr. John F. Mee Endowed Assistant Professor of Management at Miami University. To get regular updates from Scott on topics of leadership, management, assessment and Industrial & Organizational Psychology visit [www.scottdust.com](http://www.scottdust.com).

# WHAT IS AN ASSESSMENT?

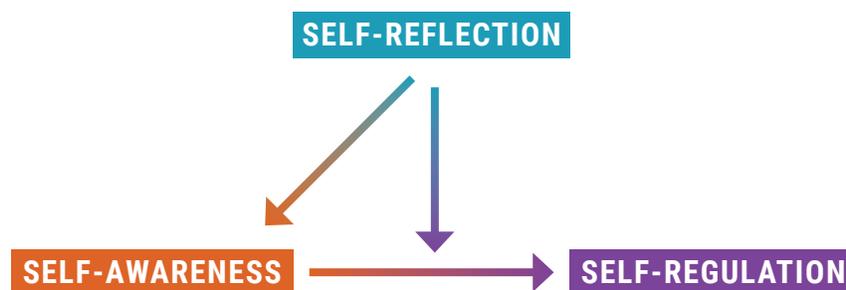
A concept is an idea that is generally agreed upon by others. All of the characteristics that we are interested in assessing, such as traits, behaviors, and emotions, are all concepts.

A concept becomes a construct when it is clearly defined and there is a method for quantifying the concept. The goal of assessments is to quantify a concept. In turn, assessments are what turn ideas (concepts) into measurable phenomena (constructs).

# HOW CAN ASSESSMENTS BE USED IN THE WORKPLACE?

## ASSESSMENTS AND THE SELF-LEADERSHIP PROCESS

A popular approach to using assessments is to learn more about one's self in order to grow and develop as an employee, colleague, manager, or leader. One way to conceptualize how this process unfolds is through self-leadership; a process that entails influencing one's self towards obtaining self-set goals (see the Figure below).



The first step in the self-leadership process is *self-reflection*, the act of systematically considering information about the self. Such self-reflection facilitates *self-awareness*, which is defined as an accurate understanding of one's characteristics (i.e., traits, behaviors, values, emotions, etc.).

Self-awareness is important because it influences the degree to which we engage in *self-regulation*, defined as the ability to choose appropriate behaviors in a given circumstance. This ability to self-regulate is critical for increasing positive (e.g., job satisfaction, job performance) and decreasing negative (e.g., emotional exhaustion, withdrawal) workplace outcomes.

Self-reflection also plays an important role in enhancing the relationship between self-awareness and self-regulation. The more one self-reflects, the more it reinforces the knowledge one garners about the self. This allows the individual to more effectively leverage this information for behavioral self-regulation. Thus, self-reflection has a dual role—it acts as a precursor to self-awareness and it enhances the degree to which self-awareness translates into self-regulation.

Self-awareness and self-regulation are cognitive in nature and hard to directly manipulate. Alternatively, self-reflection is a behavior that we can purposefully and proactively plan for and pursue. Assessments are the primary tool through which this self-reflection takes place.

## **ASSESSMENTS AND HUMAN CAPITAL MANAGEMENT**

Assessments can also be used to help organizations make strategic human capital management decisions. For example, organizations might employ assessments as a way to pinpoint ideal job candidates, uncover employees with leadership potential, or offer targeted feedback in a performance review. These assessment approaches are helpful in that they help organizations quantify subjective workplace characteristics in a way that allows them to properly manage their human capital resources.

# WHAT TYPES OF CHARACTERISTICS CAN BE ASSESSED?

When constructing and administering assessments it is important to first ensure that you are assessing the type of characteristics that you intend to assess. Below are some of the most common characteristics of interest in workplace settings.

## 1 TRAITS

Traits are characteristics that are consistent and stable across situations and time. Traits are unlikely to change much, if at all, regardless of whether the individual is in a different setting or as the person ages. The most common example of traits is the Big 5 personality framework, which includes conscientiousness, agreeableness, neuroticism (or its inverse, emotional stability), openness to experience, and agreeableness.

## 2 VALUES

Values are the things that are important to someone. Values can be long-term guideposts (i.e., terminal values) or in-the-moment personal expectations on how to live (i.e., instrumental values). While values are thought to be relatively consistent across time, they can and do change, especially as individuals transition through different life or career stages.

## 3 STRENGTHS

Strengths are another popular assessment characteristic. Individuals with specific strengths have a combination of knowledge, skills, and abilities that enable them to consistently provide superior performance in a specific task. Some examples of popular strengths include communication, empathy, determination, proactivity, flexibility, and creativity.

## 4 EMOTIONS

Emotions are instinctive states of mind that occur due to one's circumstances. Emotions are state-based, meaning that they ebb and flow from moment to moment. The most common models suggest that there are approximately eight basic emotions: joy, excitement, surprise, sadness, anger, disgust, contempt, and fear. Practitioners commonly use trait-based affect. For example, trait positive affect and trait negative affect, which entail the overall tendency to be in a good or bad mood, respectively, are quite common.

## 5 BEHAVIORS

Behaviors entail observable actions. Behaviors are traditionally evaluated as outcome variables. This makes sense, as we are ultimately interested in understanding how our characteristics influence what we do while at work. While the types of behaviors are wide-ranging, there are three popular categories. The first is performance (which can be broken down into task, proactive, adaptive, innovative, etc.) A second is positive workplace behaviors, such as organizational citizenship behaviors or voice behaviors. A third is negative workplace behaviors, such as counterproductive work behaviors or unethical behaviors.

## 6 ATTITUDES

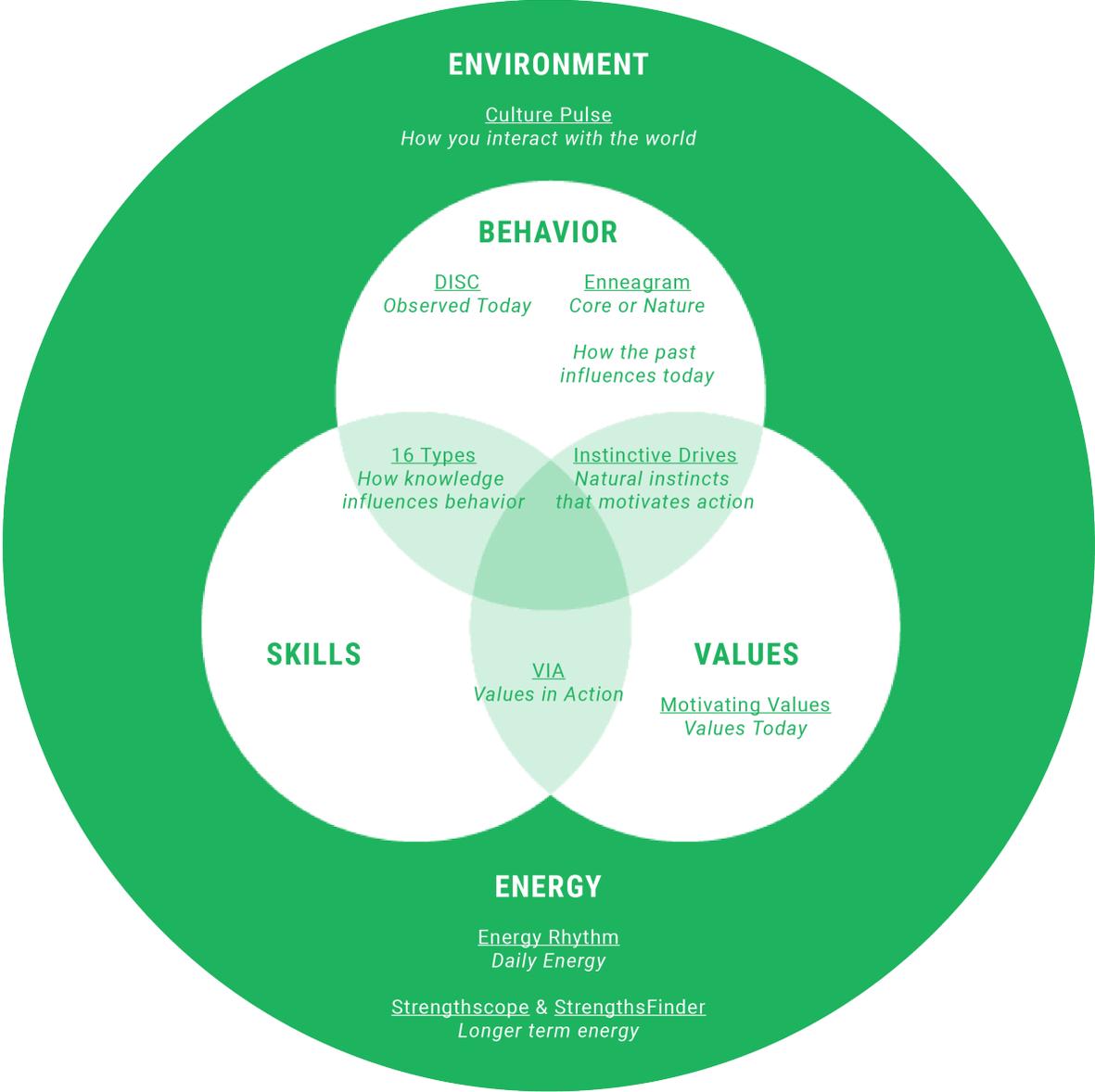
Attitudes are the way someone feels towards something. Specific to the workplace, this feeling could be directed towards their job, their organization, or a colleague. Attitudes are determined by an interaction of factors such as one's values, beliefs, and perceptions. Common workplace examples include job satisfaction, organizational commitment, and intent to turnover.

## 7 COGNITION

Cognition entails how one processes information. Although cognitive assessments are less common within the workplace literature, some of them are incredibly well-researched. The first is mindsets, which refers to the heightened accessibility of specific and distinct thought processes. The most popular in the workplace literature is a growth mindset (and its opposite, a fixed mindset); the belief that one's skills and qualities can be cultivated through effort and perseverance. The second is orientations, which entails your underlying motivations. One helpful example is the goal orientation framework which suggests that people can have a learning goal orientation or a performance orientation.

Cloverleaf provides a range of assessments to measure many of the characteristics for the people, teams and your organization. The following exhibit is a sampling of these assessments:

# ASSESSMENT OVERVIEW



# WHAT IS A CONCEPTUAL MODEL?

When we assess something, typically we're not just interested in that one concept, but also in understanding how that concept relates to other concepts. This is called the "conceptual model" (see the Figure below). To understand the relationships within a conceptual model, it can be helpful to think through three key questions:

## WHAT OUTCOME AM I INTERESTED IN?

**Start with the end in mind.** Are you interested in improving employee performance, employee job satisfaction, team psychological safety, or something else? It can be helpful to start with this outcome variable—called the *dependent variable*—and then work backward.

For example, let's say that you are a leader and you're interested in increasing employee performance. You want to know whether transformational leadership (i.e., idealized influence, inspirational motivation, intellectual stimulation, and individual consideration) increases employee performance. In this case, transformational leadership is the *independent variable* (also called the antecedent, predictor variable, or causal variable). It would then be appropriate to implement an assessment of transformational leadership as well as employee performance.



## WHY DOES THIS HAPPEN?

The next question is why an independent variable eventually leads to a dependent variable. As an example, in my research, I found that employee psychological empowerment—feeling meaningful, impactful, self-determined, and competent while at work—can explain the relationship between transformational leadership and employee performance. In this conceptual model, psychological empowerment is called a *mediating variable*. When leaders are transformational, employees react by feeling more empowered, which then increases their performance.

## HOW DOES THE CONTEXT IMPACT THIS RELATIONSHIP?

The next step is introducing *moderating variables*; constructs that change the nature of the relationship between variables. Continuing with the above example, I also evaluated whether the organization was organic (i.e., decentralized, informal procedures) or mechanistic (i.e., hierarchical, formal procedures). I found that in organic organizations, the impact of transformational leadership on employee psychological empowerment was stronger, but in mechanistic organizations, the impact disappeared. Moderating variables, therefore, are important because they help determine whether independent variables will have their intended impact.

# WHAT ARE THE DIFFERENT TYPES OF ASSESSMENTS?

## REFLECTIVE ASSESSMENTS

Reflective assessments ask participants to rate several items that “reflect” someone that would have the features of the construct. The questions are typically very similar, but get at the construct from varying angles or using different examples.

For example, if you were interested in psychological safety, you might ask questions such as “How comfortable are you speaking up when in your group?” or “How likely are you to admit a mistake when working in your team?” Both are questions that reflect the extent to which you feel psychologically safe.

## FORMATIVE ASSESSMENTS

Compared to reflective assessments, formative assessments are relatively rare in the human capital domain. Formative assessments include questions that represent each necessary component of the construct. It is typically used to measure unique outcomes.

For example, an organization might define team performance as a combination of three different metrics: the quality, quantity, and speed of team deliverables within a defined period of time. As this example illustrates, formative assessments can be very useful when you have several different indicators that are on different scales (i.e., quality, quantity, speed).

# WHAT IS ASSESSMENT RELIABILITY AND VALIDITY?

1

Develop a clear definition of a concept using preexisting theory.

2

Develop items that you think are reflective (or formative) of the concept.

3

Conduct content validity analysis, by asking participants to evaluate whether your items appropriately map onto your intended definition.

4

Conduct exploratory factor analyses to refine the potential items that should be retained and establish the dimensionality of the construct.

5

Conduct confirmatory factor analyses to confirm whether the items continue to have appropriate reliability and dimensionality.

6

Conduct convergent and divergent validity testing, which evaluates the correlations of the construct against similar and dissimilar constructs.

7

Conduct nomological validity testing, which entails testing the construct within a conceptual model.\*

## RELIABILITY

Assessment reliability is about the consistency of the measure. As an example, start by visualizing a dartboard. A reliable assessment will have the vast majority of the darts in the same location. One way that we assess reliability is by evaluating the inter-item correlations using Cronbach's alpha. The analysis entails evaluating the extent to which the items are correlated (i.e., do they vary in the same direction) with each other. The traditional threshold for acceptable reliability is 70.

Another form of reliability is test-retest reliability. This entails having a participant take the assessment a second time at a later date (preferably weeks or months later). Ideally, the assessment scores remain similar across time. Keep in mind that test-retest reliability isn't necessary for constructs that you would expect changes across time, such as state-based emotions.

## VALIDITY

Assessment validity is about the accuracy of the measure. Again, visualize a dartboard. A valid assessment will have the vast majority of the darts on the bullseye. But keep in mind that reliability and validity are both important. It doesn't help if the assessment is reliable but not valid—the darts are all in the same place, but not on the bullseye. It also doesn't help if the assessment is valid but not reliable—many of the darts are on the bullseye, but some are scattered around the board.

## THE STEPS OF CONSTRUCT VALIDITY

Construct validity is a multi-step process for establishing that new constructs are reliable and valid. Although there are many different methods, the approach generally goes as follows on the left.

\*In this stage, you are evaluating whether the construct behaves as you would expect in terms of its relationships with independent, mediating, dependent, and moderating variables.

# WHAT ARE ASSESSMENT QUESTION STEMS?

The question stem is the **text offered immediately before the assessment items**. There are two common question stem approaches.

## 1. OPEN-ENDED STATEMENT

One approach is to use a common, open-ended statement. For example, simply asking participants to rate the extent to which they agree or disagree with each question is straightforward and minimizes interpretation differences.

## 2. NUANCE

You could, if appropriate, add nuance to this common stem by focusing the user's attention on a specific referent or to a specific timeframe. For example, you might state "Consider what it's like while working with your team" or "Think about your team interactions during the last three months."

## 3. SCENARIOS

Another approach entails using scenarios. In this approach, you offer text that explains a specific situation or context and then ask the participant to respond to items based on this information. For example, you might ask the participant to read a short paragraph that asks them to assume that they are working on an important new project with a tight deadline. You might even go into details on the industry, the project components, or relationships with team members.

This scenario approach can be useful in two ways. First, some research suggests that priming a specific and/or emotion-laden context leads to more accurate responses. Second, the assessment can be more realistic as it relates to the specific scenarios employees are likely to encounter while working at their organization.

The downside, however, is that scenarios are very hard to write. The more details that are included, the higher the likelihood that participants will get confused or overthink their response. Further, if the assessment is being used broadly across an organization, the scenarios might not perfectly apply to all types of jobs or teams.

## WHAT TYPE OF LIKERT SCALE SHOULD BE USED?

Assessments are constructed as a list of items that are to be answered by the participant. These items are commonly listed as questions, phrases, or statements that can then be rated using a pre-determined Likert scale.

### AGREEMENT

The choice of Likert scale depends on the construct of interest. One of the most universal choices is to ask participants to rate the extent to which they disagree (low) to agree (high) on a multi-point scale (e.g., typically between 3 and 10 points). This tends to work well for traits, attitudes, and behaviors.

### EXPERIENCE

Another option is to ask users to rate how often they experience something. The scale ranges between never (low) to all the time (high). This tends to work well for emotions or cognitions.

### IMPORTANCE

Another popular option is to ask participants to rate how unimportant (low) to important (high) something is to them. This tends to work well for values or work preferences.

## HOW MANY LIKERT SCALE POINTS SHOULD THERE BE?

A handful of academics have made a career out of statistically proving that using a scale with either five (e.g., strongly disagree, disagree, neither disagree nor agree, agree, strongly agree) or seven (e.g., strongly disagree, disagree, somewhat disagree, neither disagree nor agree, somewhat agree, strongly agree) options is ideal. Using an odd number is important because it ensures that participants can select something in the middle when they genuinely feel neutral about the item.

Less than five items don't give participants enough options because they are being forced to respond with low, high, or neutral. More than seven items give users too many options and adds unnecessary variability based on one's tendency to rate items more conservatively or more liberally.



**Some assessments use a one to ten scale because it aligns with our numbering system. Research suggests, however, that it is better to use a five-point or seven-point scale and then rescale the findings on a 10-point or 100-point scale *after* the assessment is complete.**

# WHEN SHOULD I USE FORCED-CHOICE ASSESSMENTS?

A popular question is whether it's better to use assessments that use a Likert format (e.g., 1 = strongly disagree to 5 = strong agree) or a forced-choice format (e.g., pick the characteristics that best describes you). **In academia, Likert scale assessments are the norm.** Interestingly, some of the most widely used pay-to-play assessments (see the next section) are forced-choice assessments.

## COMMON CONCERNS:

### CONSTRUCT VALIDITY CONCERNS

With forced-choice format assessments the responses are dichotomous (i.e., you either do or do not select an item). This makes construct validity testing challenging because we don't know the degree to which one prefers one selection over alternative options. This gets even more challenging when forced-choice assessments have three or more options. Alternatively, Likert-based assessments evaluate one item at a time on a continuous scale (e.g., 1 to 5), making validity and reliability tests straightforward.

There are two common ways to evaluate the reliability and validity of forced-choice response assessments. First is the item response theory approach, which requires a large number of cases and is hard to run and interpret accurately. Second is the clustering approach, which is straightforward and easily administered but makes some big assumptions.

### COGNITIVE BIAS CONCERNS

Another reason people sometimes opt for forced-choice assessments is that it reduces inflation bias (i.e., a tendency to rate good things positively and bad things negatively) and social desirability bias (i.e., a tendency to rate things how you think others want you to rate them). Likert scales make it easy for assessment takers to evaluate whether items are positively or negatively worded. This presents the opportunity for biased responding. Theoretically speaking, forced-choice assessments can limit these biases, particularly when all choices within a question set are equally positive, negative, or neutral.

In most cases, a Likert-based assessment will still work. It may not be the best choice for self-rated performance, but it's unlikely to be a major issue for relatively neutral topics such as workplace preferences, values, strengths, or traits.

### INTELLECTUAL PROPERTY CONCERNS

Another reason people opt for forced-choice assessments is intellectual property protection. If you've developed the scale yourself and you're worried about someone stealing your intellectual property, forced-choice assessments are better. It's hard to reverse-engineer the scoring behind a forced-choice assessment, particularly when there are bogus items (i.e., items that don't count towards a dimension or characteristic) or questions that count towards multiple dimensions or characteristics.

# HOW CAN I USE ASSESSMENTS TO RATE OTHERS OR RATE TEAMS?

The most common example of using workplace assessments typically involves a participant completing an assessment about themselves. There are several other iterations of this approach that are also useful.

## OTHER-RATINGS

One approach is to ask an individual to assess the behavior of a focal colleague such as their direct manager. The individual could also rate each of the individuals on their team. These assessment scores can then be aggregated into a team-level score by averaging all of the individual scores. For example, you could average the scores of all of the team members' ratings of the team's direct manager. You could also aggregate the team member ratings about specific individuals on the team. This is common practice in 360-degree feedback.

## TEAM-LEVEL AGGREGATIONS

Another variation of this approach is to have the individuals of a team rate their team as a whole instead of specific individuals on the team. You are shifting the "referent" from the individual to the team. For example, it is common to ask all of the team members to rate the team as a whole for constructs like team psychological safety, team cohesion, or team information sharing, and then aggregate these ratings.

## LEVERAGING MULTI-SOURCE ASSESSMENTS

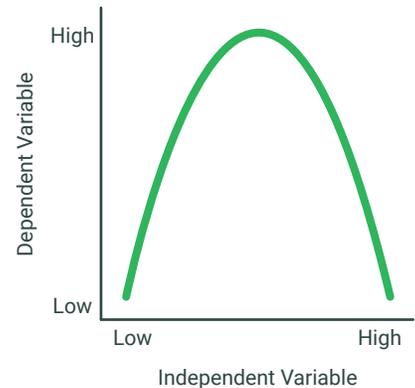
Leveraging multi-source ratings can be particularly helpful in uncovering differences in perceptions. For example, one could evaluate the extent to which an individual was self-aware by comparing the extent to which an individual's self-ratings were similar to or different from aggregated ratings from the team about that individual.

# CAN ASSESSMENTS UNCOVER NON-LINEAR EFFECTS?

We have a tendency to assume that the impact of a construct is linear. For example, being a humble leader (independent variable) should have a positive linear relationship with employee trust in the leader (dependent variable). This might generally be true, but it's still oversimplified. In many cases the relationship isn't linear, it's non-linear. The following are three common examples.

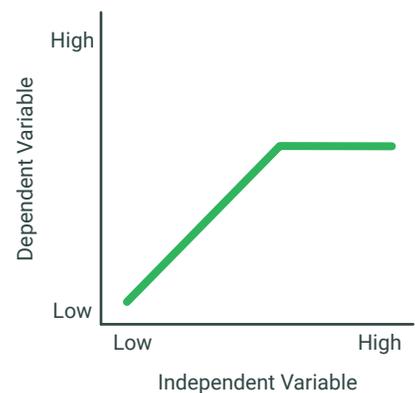
## THE TOO MUCH OF A GOOD THING EFFECT

As seen in the figure to the right, as we move from low to moderate levels of the independent variable, the dependent variable also increases. However, we eventually reach a tipping point, and when moving from moderate to high levels of the independent variable, the dependent variable begins to decrease. Continuing with the humble leadership example, this would suggest that being excessively humble might signal that you don't know what you are doing, which degrades employee trust.



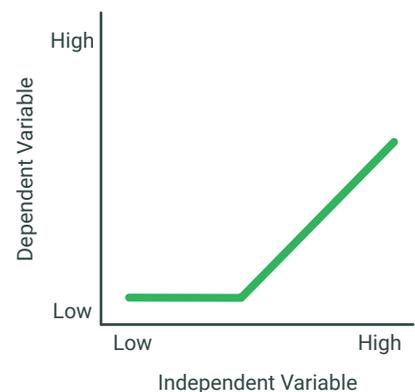
## THE GOOD ENOUGH EFFECT

A related scenario is when the impact of an independent variable levels off. As seen in the figure to the right, the independent variable improves the dependent variable up to a certain point, but after that, it doesn't add any additional value. This is helpful to understand because in many cases it is possible that doing reasonably well at something is good enough, and trying to improve it would be wasted effort.



## THE NOT ENOUGH EFFECT

Another scenario is when the impact of an independent variable only comes into play at a certain level. Specific to the humble leadership example, this would suggest that employees expect a moderate level of humility from their leaders. But once the leader exhibits a level of humility that is considered greater than the norm, its impact begins to have a positive effect on employee trust in the leader.



# CAN I EVALUATE MORE THAN ONE CHARACTERISTIC AT A TIME?

Assessments typically focus on one characteristic at a time. This is primarily because it's easier to interpret from a practical standpoint. However, individuals, teams, and organizations are complicated systems that involve several overlapping and interrelated characteristics. To fully appreciate this complexity entails considering a "profile"—a collection of characteristics.

Consider, for example, that we commonly evaluate where someone falls along the continuum of introversion to extraversion. Equipped with this information, we have a better idea of how to structure work processes that align with their communication preferences. But what about the other four Big 5 personality traits?

Let's assume that we're evaluating each of the Big 5 traits on a fivepoint scale (i.e., low = 1, high = 5). This equates to 3,125 potential profiles ( $5 \times 5 \times 5 \times 5 \times 5 = 3,125$ ). Make it a seven-point scale and there are 16,807 profiles ( $7 \times 7 \times 7 \times 7 \times 7$ ). To make things even more complicated, we could also evaluate several characteristics at a time; not just personality traits, but values, strengths, etc. This conversation on profiles highlight two important reminders.

- 1** First, it's important to balance specificity and accuracy on the one hand, and generalizability and usability on the other hand. Although it might be more specific and accurate to talk about hundreds of different profiles, this is unlikely to be helpful when trying to obtain practical insight. Alternatively, if we oversimplify assessments to the point where we assume that everyone fits within a very limited set of profiles, the implications and recommendations will never be fully representative of the assessment takers.
- 2** Second, it is important to consider whether there are multiple characteristics that might work together or counteract each other. Relatedly, it is important to not try to define individuals, teams, and organizations by one characteristic. While an individual, team, or organization might not have the ideal level of a specific characteristic, they might make up for it in other characteristics.

## HOW MANY ITEMS SHOULD AN ASSESSMENT HAVE?

For reflective assessments, the goal should be to ask as few questions as possible while maintaining high levels of reliability and validity. The minimum is typically three items because that allows you to properly assess reliability.

The more items you include, the higher the likelihood that you'll get an accurate assessment. However, too many items can also artificially inflate reliability. Additionally, if there are too many items in a survey users will exhibit survey fatigue, where they begin to randomly respond and/or pay less attention to the specifics of the questions.

Thus, the decision on the number of items is a balancing act. Across both academic and practitioner communities, the range of items is typically between three and 15, with simpler constructs being on the lower end and complex, multi-dimensional constructs being on the higher end.

## SHOULD REVERSE-CODED ITEMS BE USED IN ASSESSMENTS?

A reverse-coded item is a question that is worded in the opposite direction of the measure. For example, if you intend to measure team satisfaction with three items, you could ask users to rate the extent to which "they are satisfied with the work of their team members," "they enjoy working with their team," and "they dislike working with their team." The first two items are positively worded and the last item is negatively worded.

The benefit of using reverse-coded items is that it can reduce acquiescence bias; the tendency to always answer positively. Further, some items make more sense to the reader when worded negatively.

Many scholars advise against it, however, citing that it typically confuses the participant. Additionally, if you do employ reverse-coded items, you'll need to remember to recode the items in the appropriate direction before creating aggregated scores for the construct (i.e., taking the mean of all of the items).

## HOW CAN I MINIMIZE RANDOM RESPONDING?

Participants that randomly respond to assessments can reduce the overall quality of the human capital management initiative. To test for this, embed one item for every 30 items or so that states something like the following: “please select [insert scale choice here] to confirm that you are not randomly responding.” Analyses can then be run with and without the respondents that got the answer(s) incorrect.

If you are worried about survey participants being offended by this wording, consider using time-based analyses. For example, most survey hosting applications automatically collect the amount of time it takes for participants to complete surveys. You can pinpoint which surveys might be less accurate by first evaluating the median time to completion across users. Then, pinpoint users that take approximately less than half the median time to complete their surveys. Analyses can then be run with and without these surveys to see if there are any differences.

---

# REFERENCES

---

Aguinis, H., Edwards, J. R., & Bradley, K. J. (2017). Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods*, 20(4), 665-685.

Crawford, J. A., & Kelder, J. A. (2019). Do we measure leadership effectively? Articulating and evaluating scale development psychometrics for best practice. *The Leadership Quarterly*, 30(1), 133-144.

Gardner, R. G., Harris, T. B., Li, N., Kirkman, B. L., & Mathieu, J. E. (2017). Understanding “it depends” in organizational research: A theory-based taxonomy, review, and future research agenda concerning interactive and quadratic relationships. *Organizational Research Methods*, 20(4), 610-638.

Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, 13(4), 668-689.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating withingroup interrater reliability with and without response bias. *Journal of Applied Psychology*, 69(1), 85-98.

Landis, R. S., Beal, D. J., & Tesluk, P. E. (2000). A comparison of approaches to forming composite measures in structural equation models. *Organizational Research Methods*, 3(2), 186-207.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, 35(2), 293-334.

Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., & Lankau, M. J. (1993). Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management*, 19(2), 385-417.