

360-Degree Feedback Survey Response Scales

When a new 360 survey is developed, one of the tasks is to determine which response scale is most appropriate. This *Insight* white paper describes some of the issues and criteria that should be considered, and provides several examples of possible scales.

Background and Criteria

Industrial-Organizational psychologists have researched the pros and cons of alternative response scales for many years, but there is no consensus on the "one best scale." What works best depends on the purpose of the survey and on the types of items included. However, there is consensus on the basic criteria for what constitutes a good response scale:

- The scale is simple and easy to understand.
- To the extent possible, raters have a common understanding about what the scale means – they interpret the scale in a similar fashion.
- The words used to define the levels are clearly ordered (e.g., it may not be clear which is higher – "Excellent" or "Outstanding").
- The scale is suitable for the types of items in the survey (e.g., a scale where one of the levels is "Always" may or may not be appropriate, depending of the types of items).
- The scale leads to variance in ratings (competencies and participants are not all rated the same – there is a good spread in ratings).
- The scale is aligned with the main purpose of feedback (e.g., different scales would be appropriate when the purpose is developmental versus evaluating performance against expectations/requirements).

Three other general conclusions can be drawn from the research:

1. Increasing the number of rating levels will, up to a point, lead to a greater spread in ratings (which is desirable). In most situations, having at least five levels is preferable. However, scales with more than seven levels can lead to more error being introduced.
2. The labels or anchor terms tend to confuse raters when there are more than six levels, and each level is labeled.

3. In many cases, regardless of the themes or number of levels used, raters tend to inflate ratings, and there is less variance in scores than one would desire.

General Types of Response Scales

Listed below are the general types of response scales, based on the main theme or aspect on which the ratings are made:

- **Simple qualitative** – how would you rate the person in this area? (e.g., from "Poor" to "Excellent").
- **Extent** – to what extent does the statement describe the person? (e.g., from "Very little extent" to "Very great extent").
- **Frequency** – how often does the person demonstrate effective use of the behavior/ skill? (e.g., from "Almost never" to "Almost always").
- **Developmental** (strength/development need) – how would you rate the person's capabilities in this area? (e.g., from "Significant development need" to "Exemplary—a role model").
- **Comparison with others** – compared to others with whom I have worked, this person is _____ in this area (e.g., from "Significantly below average" to "Best I've ever worked with").
- **Performance** – how would you rate his/her performance in this area? (e.g., from "Does not meet expectations" to "Far exceeds expectations"). Note that this theme would make sense for performance appraisal purposes, but probably wouldn't for development purposes where peers and subordinates also provide ratings.

It is possible to construct a response scale that has more than one theme. For example, an anchor point of "Excellent—best I've worked with" combines the qualitative and comparison themes. However, unless this is carefully done, there is a danger of the response scale becoming too complex, and thereby confusing.

As a final comment on general types of response scales, it is important that each level in a scale relate to the same theme(s). It is not appropriate, for example, that some levels of a scale relate to frequency, and other levels relate to capability (developmental). This is a common failing of many "home grown" response scales.

Anchor Points

Response scales vary in terms of two main factors – the number of levels (5-point, 6-point, etc.), and the theme or themes captured by the anchor terms and scale definitions. There are obviously a very large number of possible scales. Normally, each level in the scale has a descriptive anchor when there are four or fewer levels, but all the levels do not necessarily have to be anchored when there are five or more levels. The example below shows Censeo's standard scale for established surveys:

- 5 – Exceptional strength
- 4 –
- 3 – Competent
- 2 –
- 1 – Weak

The theme for the 5-point scale above is developmental, and only three of the levels are anchored. It is highly questionable whether adding anchors for the second and fourth levels, using intermediate terms, would lead to more accurate and valid ratings. Raters implicitly form a mental image of the scale and its meaning, and are able to easily understand that a rating of 4 is halfway between "Competent" and "Exceptional strength."

The response scale below is an example of a 5-point frequency scale where only three points are anchored:

Never		Some- times		Always
1	2	3	4	5

The Appendix on the following page shows several additional examples of response scales.

Getting Started

While choosing the best response scale (or developing a new scale) for a 360 survey is an important task, it is not something over which one should obsess. The reality is that, regardless of the care that goes into scale development, a good portion of raters will form their own mental map, with the lowest value being "bad" and the highest value being "good." This does not diminish the value of the 360 survey process, but does mean that the people being rated should not interpret the scale too literally. Information on relative scores (which competencies are highest/ lowest), and normative comparisons (how do I stack up against others?), should be the focus rather than the absolute values of the ratings.

When a new 360 survey is being developed, it's usually advisable to pilot a draft of the survey in a focus group with typical raters. In addition to input solicited on the items themselves, key questions regarding the response scale include:

- Do people interpret the scale the same way?
- Does it "feel right" to raters for the kinds of items in the survey?
- Does the scale work well for all rater roles that will be included (managers, peers, etc.)?
- Will the scale provide adequate discrimination versus everyone being rated the same?

Censeo Corporation is a human resource consulting firm offering consulting services and Internet-based assessment products. The company's mission is to help clients maximize their investment in human capital by collecting and then strategically using accurate and valid information about competencies and work preferences. For more information, visit www.censeocorp.com, email info-request@censeocorp.com or call 407.645.1600 x150.

Appendix: Examples of Response Scales

Qualitative Scales

4 – Excellent	5 – Outstanding	6 – Extraordinary
3 – Good	4 – Very good	5 – Superior
2 – Fair	3 – Good	4 – Very good
1 – Poor	2 – Fair	3 – Good
	1 – Poor	2 – Fair
		1 – Poor

Extent Scales

4 – Exactly descriptive	5 – Very great extent	6 – Completely true description
3 – Very descriptive	4 – Great extent	5 – Largely true
2 – Somewhat descriptive	3 – Some extent	4 – Somewhat true
1 – Not descriptive	2 – Little extent	3 – Somewhat false
	1 – Very little extent	2 – Largely false
		1 – Completely false description

Frequency Scales

4 – Almost always	5 – Almost always	6 – 100% of the time
3 – Usually	4 – Most of the time	5 – 90+% of the time
2 – Sometimes	3 – Often	4 – 80+% of the time
1 – Seldom	2 – Sometimes	3 – 70+% of the time
	1 – Seldom	2 – 60+% of the time
		1 – Less than 60% of the time

Developmental Scales

4 – Towering strength	5 – Exemplary, best possible	6 – No room for improvement
3 – Strength	4 – Real strength	5 – Significant strength
2 – Competent	3 – Fully competent	4 – Strength
1 – Development need	2 – Development need	3 – Competent
	1 – Weakness	2 – Development need
		1 – Significant development need

Comparison Scales

4 – One of the very best	5 – Far above average	6 – Top 5%
3 – Better than most	4 – Above average	5 – 80-95 th percentile
2 – Better than some	3 – Average	4 – 50-80 th percentile
1 – Not as good as most	2 – Below average	3 – 20-50 th percentile
	1 – Far below average	2 – 5-20 th percentile
		1 – Bottom 5%

Performance Scales

4 – Commendable	5 – Far above requirements	6 – Exceeds all standards
3 – Exceeds	4 – Above requirements	5 – Exceeds most, meets others
2 – Meets	3 – Meets requirements	4 – Meets most, exceeds others
1 – Does not meet	2 – Below requirements	3 – Meets all
	1 – Far below requirements	2 – Meets most, below on some
		1 – Below on many standards