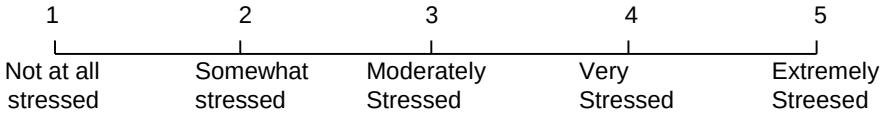
Constructing Rating Scales for Self-Rating and Observer Rating

You may wish to construct your own rating scales to gather data on certain program outcomes. Rating scales are much-abbreviated measurement tools, when compared to standardized tests. Still, they can be very useful for measuring change in client status or condition as a result of your family support services. They can be constructed so that the client rates himself or herself, or they can be constructed so that some other credible and knowledgeable person, such as a service provider, can rate the client.

A Likert-type scale is usually a 5-point to 9-point scale that has anchors at each end and definitions of some or all of the scale points that indicate increasing or decreasing qualities or properties of the scaled item. For example, you might construct a rating scale that focuses on stress, and employs a 5-point rating strategy. It might look like this:

"How stressed are you as a result of your child care demands?"



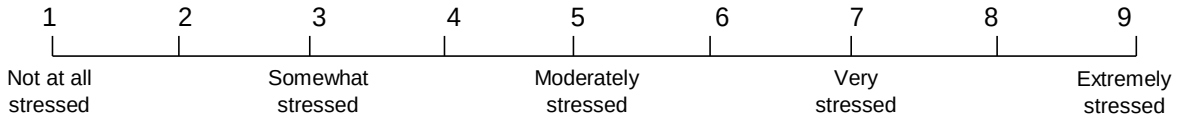
You are probably already quite familiar with this scaling technique. If you used a 7-point or 9point scale instead of a 5-point scale, you would increase the sensitivity of the scale. As used here, sensitivity refers to the likelihood that the scale will detect a true change in the dimension being measured, if one occurs. In other words, it takes a bigger change to move from a stress level of 3 to a stress level of 2 on a 5-point scale than it does to go from a 3 to a 2 on a 9-point scale because the intervals are smaller on the 9-point scale assuming that the overall scale includes the same upper and lower limits of stress regardless of how many points are on the scale.

You can see that the more points you have, the more sensitive the scale is to detect change. However, there is a trade-off with regard to increasing scale sensitivity. Practically speaking, as you increase the number of points on the scale, you have to come up with increasing numbers of scale point descriptions indicating the intended level or value of each rating possibility. Simply labeling a scale with a large number of points becomes problematic.

Furthermore, if you have too many points on the scale, then small differences become difficult to interpret and may only be due to measurement error. For example, imagine trying to interpret the meaning of a scale score change from 15 to 16 on a 21-point scale. Rating scales rely on subjectivity, and too much scale sensitivity can lead to error.

One technique you can employ to address this issue is to label every other point. We could increase the sensitivity of the stress scale above using this technique, and it might look like this—

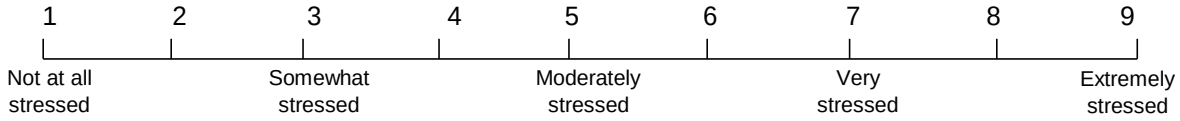
"How stressed are you as a result of your child care demands?"



You can see that scale values 2, 4, 6, and 8 do not have their own labels, but the values of those points are implicit because they fall equidistant between the labeled points on the scale. Generally speaking, 7-point or 9-point scales will give you sufficient sensitivity, and will not be too taxing to label.

As presented in the two preceding examples, the stress scale is constructed as a self-rating scale. The question asks, "How stressed are you…?" This same scale can be converted to an "observer rating scale" simply by changing the subject of the question and altering the grammar. For example, you might routinely ask your trained staff to rate the degree of stress perceived to be experienced by a client by using the scale below:

"How stressed does the primary caregiver appear to be as a result of his/her child care demands?"



The only difference between these two scale constructions is who does the rating.

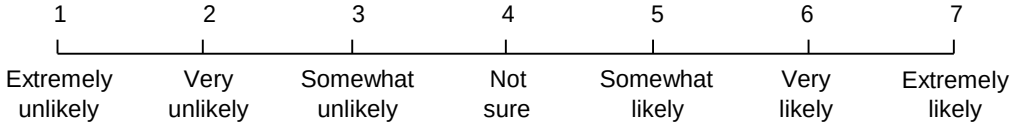
You would probably be wise to use both self-rating scales and observer-rating scales on (at least some of) the same dimensions for each client family. This allows you to corroborate ratings between different sources of data. If there are large differences in the ratings made by the natural caregiver and your staff caregiver, you may want to investigate the reasons for the differences, or perhaps use the information to suggest other services (such as counseling) that may be available for the client.

There may be logical reasons for differences in rating, depending on what is being measured. Natural family caregivers may be reluctant to admit that they have been considering foster care for their child, or they may be unwilling to rate their relationship with their spouse/partner as being very troubled. In these cases, observer ratings may be more reliable than self-ratings (or, a standardized test that asks a variety of questions relating to a topic like marital relations may be a better choice than a rating scale).

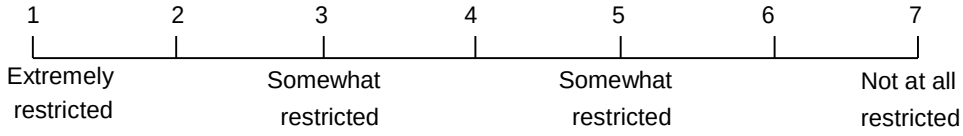
Some generic examples of rating scales are listed below. For the sake of simplicity, all employ a 7-point scale, but each could be changed to a 9-point scale to increase measurement sensitivity. Also, as presented, all represent self-rating forms of scale construction, but each could be changed to "observer-rating" construction with a simple change in grammar and replacing the implied natural caregiver with some other credible observer.

Generic Rating Scales

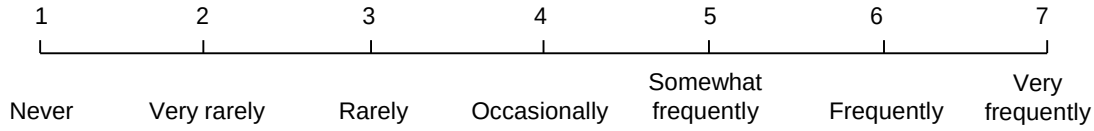
Q 1. Given your present child care demands how likely is it that you will (divorce, separate, place your child in foster care, hurt your child, neglect your child, etc.)?



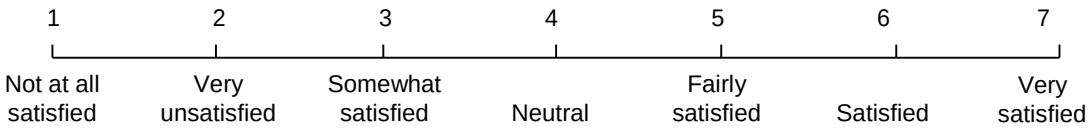
Q 2. At present, my ability to (pursue my education, keep up with daily chores, sleep soundly,visit aged parents, etc.) is



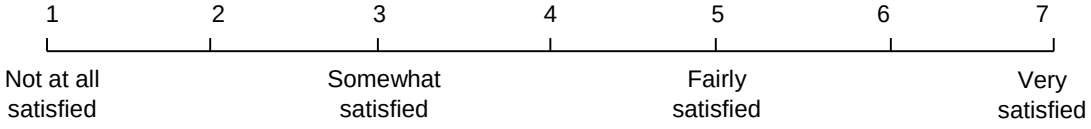
Q 3. How frequently do you (spend time with your children, have time to spend alone with your spouse/partner, have a chance to enjoy a hobby, go to church, feel depressed, etc.)?



Q 4. How satisfied are you with your (relationship with your child, ability to cope with your child's needs, living arrangements, child care, family relationships, service providers, hours of service availability, quality of service you receive, etc.)?



Q 5. How satisfied are you with your (relationship with your child, ability to cope with your child's needs, living arrangements, child care, family relationships, service providers, hours of service availability, quality of service you receive, etc.)?



You may have noticed some variations among the model scales (Q1—Q5), including variations in the manner in which the question is presented and also in the scaling techniques employed. For instance, Q1, Q3, Q4, and Q5 are all true questions, and the respondent answers the question from the scaled choices available. Q2, on the other hand, is a statement, not a question, and the respondent is asked to locate himself or herself along the continuum of "being restricted" used on the scale.

Note that in some cases all of the scale points are labeled and in some cases intermediate scale values are implied.

Q1 and Q4 employ a two dimensional scaling strategy that passes through a neutral midpoint. In the case of Q1, the scale values range from "extremely unlikely" to "extremely likely," passing through an equivocal "not sure" midpoint. Q4 ranges from "extremely unsatisfied" to "extremely satisfied" passing through a true neutral midpoint.

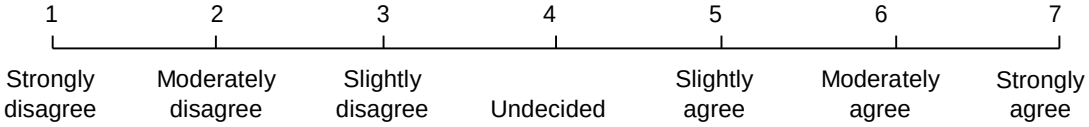
Q3 and Q5 employ unidimensional scales ranging from the absence of an attribute (i.e., "Never," and "Not at all satisfied") to a large quantity of the attribute (i.e., "Very frequently," and "Very satisfied"). Q4 and Q5 have the same root question, but employ different scaling techniques (two dimensional v. unidimensional).

Think about the clients you serve, and the information you want to know about them, particularly as it relates to

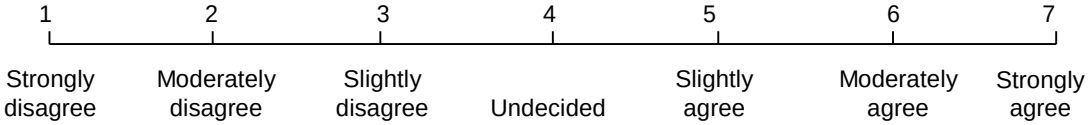
changes in their well-being. You may need to alter the scaling strategy to accommodate a particular type of information, and maximize the likelihood that your rating scale will detect change successfully.

Another variation on rating scale construction involves making a clear statement about the presence or absence of an attribute or occurrence and asking the respondent to agree or disagree with the statement along a multi-point scale. This type of scale is very common in satisfaction surveys, but can be used in other applications as well. Several examples are presented below:

Q 6. This program has reduced the stress in my family's life.



Q 7. The social contacts that my child has experienced during this program have improved his/her social skills.



Constructing rating scales takes a little practice, but the process is not terribly difficult. And, the rewards are well worth the trouble when it comes to evaluating your program. The ability to present changes in scores on scientifically validated standardized instruments, supported by changes in scale scores reported by both staff and clients, will go a long way to promoting your programs credibility. For a more complete discussion of scale construction, questionnaire construction and examples of other instruments you can use, see Bloom, Fischer and Orme (1995), and Royse and Thyer (1996).

Below are some examples of rating scale items that have been used in previous family support program evaluations, along with scale anchors that are appropriate for the content of the scales. In many cases, minor word changes make the same basic scale applicable in either type of program. For example, a client satisfaction item such as:

"I feel that I am treated with respect…," could be completed with, *"…by the staff at the center,"* for center based programs;

or,

"…by the staff that come to my home," for a home-based program.

Note that some items are worded to rate satisfaction, (e.g., " My child is well cared for at this facility"; and some are worded to rate the degree to which a client-centered outcome has been achieved (e.g., "This program has reduced the stress in my family's life").

In all cases the scaling strategy is "Strongly Agree" to "Strongly Disagree." Other scaling strategies can be employed, if desired (see previous discussion on scale construction).

Examples of Rating Scales for Client Satisfaction

(scaling strategy "Strongly Agree" to "Strongly Disagree")

My child is well cared for here.
My child enjoys his/her experience here.
I feel that I am treated with respect by staff.
Staff here listen to my concerns.
Staff here are knowledgeable about my child's needs.
This program meets my child's special needs.
I know what is happening to my child when I am not here.
I have sufficient choices about when to use this service.
Service was available when I needed it.
I have gotten support here from other parents like myself.
I was informed about how my child adjusted to the program.
I was informed about program policies and routines.
I will use this service again, if needed.
I will continue to use this service.

Examples of Rating Scales for Client Outcomes

I am spending more quality time with my husband/wife/partner.
I have a better relationship with my husband/wife/partner.
I have a better relationship with my child.
I feel closer to my children.
I feel much less stress than before.

I feel less likely to abuse (or neglect) my child.

I am happier in my role as parent since receiving services.

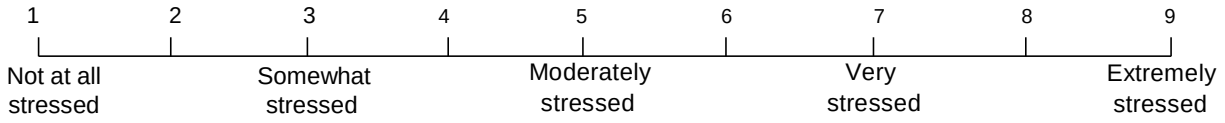
I am more competent as a parent.

I am better able to manage my child's care at home.

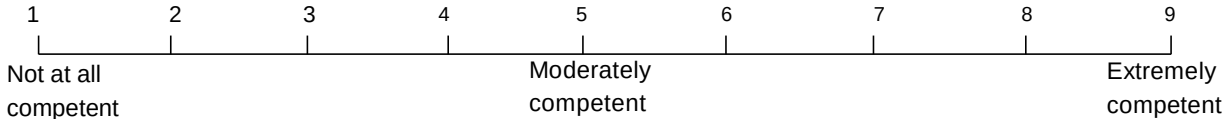
This program has made it possible for me to [pursue educational goals, go to church, seek housing, etc.]

The preceding lists are by no means exhaustive. They are simply examples of typical content items for rating satisfaction and competencies. Remember that you can increase the sensitivity of these scales for purposes of measuring change or achievement of outcomes by changing the scaling strategy. For example, in the preceding lists there are items relating to stress and parental competency. Rather than simply asking the respondent to agree or disagree that they have less stress or increased competence, you could scale the items to give much more detailed information, as follows:

"How stressed are you as a result of your child care demands?"



"How competent do you feel yourself to be as a parent of this child?"



The use of a 9-point scale increases the scales sensitivity to detect change. By phrasing the questions this way you can also use the scales as "pre/post" measures and look for change in the ratings, hopefully indicating improved outcomes as a result of your service.

Open-ended Questions

Unlike rating scales, open-ended questions do not provide a rating or scoring strategy for the client or rater to use when answering. As the name implies, open-ended questions simply ask a question and provide a space for the respondent to write down his or her answer (they can also be used during interviews, with the interviewer recording the answers given by the respondent). There are advantages and disadvantages to using open-ended

questions in your evaluation instruments. These advantages and disadvantages can best be understood by comparing their properties to those of the closed-ended or rating scale type questions.

You may have noted that closed-ended questions set boundaries around particular areas or interest. The scale items are specific, limited to a single issue and the respondent must answer using the scaling or scoring format

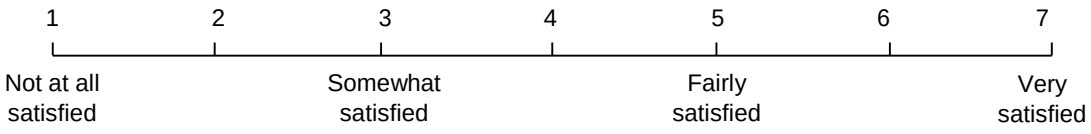
attached to that question. These boundaries are deliberately set because you are attempting to quantify the respondent's answers and obtain a scaled score. This allows you to assess individual clients and compare clients, as well as to measure change in clients over time. These features also make closed-ended questions efficient to administer and fairly easy to score.

However, closed-ended questions also limit the information that the respondent can give you. You may frustrate respondents by limiting their opportunity to give you information, and you may miss entirely important areas of inquiry if you use only closed-ended questions in your evaluation tools. Open-ended questions provide some options and partial solutions to these problems.

Open-ended questions are not limited by your language. They allow the respondents to use their own words to answer the question. Nor is the range of possible responses limited by openended questions. The respondent is free to be as descriptive as he or she desires. Open-ended questions invite respondents to tell you what is important; what they like/dislike; what could be changed or improved. They will often address issues in their responses that you did not anticipate, and this information can be very useful.

For example, you might be interested in assessing client satisfaction with transportation services you provide as part of your program. You could ask a closed-ended question like,

"How satisfied are you with our transportation services for the program?"



If the large majority of answers are in the "6 to 7" range, you are probably in good shape regarding the way you provide transportation. But if you have respondents that answer in the lower rating areas of the scale, you do not have any information about why they are dissatisfied. You could supplement your rating scale with the following open-ended question:

"If you are in any way dissatisfied with our transportation services, please tell us how the service could be improved?"

In response to this question, a client might write—

I am grateful for the transportation service that you provide, particularly in the winter when I have to allow extra time to commute to my job. One thing that I don't like, however, is that one of your van drivers is a heavy smoker, and the van is always filled with smoke when he arrives to pick up Ronny. Ronny really doesn't like the smoke and becomes agitated sometimes when getting on the van. I really don't like the idea of him riding for 45 minutes in the smoke filled van while the other riders are picked up. Also, maybe you could drive the "loop" in the same direction in the afternoon as you do in the morning. If you did that, my Ronny would not be the first person on the van in the morning and the last person off the van in the afternoon.

It should be obvious how much more and varied information is contained in the response to the open-ended question than in the closed-ended scale. Specifically, the three content areas discussed in the question (winter driving conditions, the smoking driver, and the direction of travel of the van during its routes) could never have been divined from a scale score rating of, say, "3," or "somewhat satisfied," on the closed-ended question.

You could use the information to adjust the pick-up schedule for all families on days when inclement weather may make commutes longer. This would be seen as a kind consideration by families that you serve. You might have just become aware of the need for a "no smoking" policy in the Center's van (or maybe you discovered a violation of an existing "no smoking" policy). The suggestion about which way the drivers go to/from the Center on their routes may be something that you could poll all family caregivers about. If the idea was generally viewed as a good one, you could implement this idea and further accommodate the needs of your client families. This is an idea you might never have had without open-ended inquiries on your satisfaction survey.

You can see that the open-ended questions tap areas that you might not have anticipated or that would be difficult to scale if you used only closed-ended questions. The disadvantage of open-ended questions is that they are time

consuming to administer and read. They may be labor intensive to interpret, if you serve a large number of clients. However, there is no better way to find out what clients think is important, or to get their ideas about how to improve services.

Below are some examples of open-ended questions that could be used in evaluations:

- What is the greatest benefit you have received from the program?
- What problems have arisen since you began receiving the service?
- How would you change the services that you receive?
- Have you changed the way you spend your time since you began using the service?
- What difference has the service made in your life, or in your family?
- How could we improve the service?
- If you are dissatisfied or unhappy about the service that we provide, how could we improve it to better meet your needs?