

**IQRA NATIONAL UNIVERSITY**

**Subject**

Research Techniques In HRM

**Instructor**

M. Zeeshan Ibrahim

**Submitted By**

Hilal Malik

ID# 14376

**Assignment**

Final Term Exam

**Date**

23rd June 2020

# Q.No.1: What are scales in research kindly explain with reference to research papers?

Ans: Scales of measurement in research and statistics are the different ways in which variables are defined and grouped into different categories. Sometimes called the level of measurement, it describes the nature of the values assigned to the variables in a data set.

The term scale of measurement is derived from two keywords in statistics, namely; measurement and scale. Measurement is the process of recording observations collected as part of a research.

Scaling, on the other hand, is the assignment of objects to numbers or semantics. These two words merged together refers to the relationship among the assigned objects and the recorded observations.

A measurement scale is used to qualify or quantify data variables in statistics. It determines the kind of techniques to be used for statistical analysis.

There are different kinds of measurement scales, and the type of data being collected determines the kind of measurement scale to be used for statistical measurement. These measurement scales are four in number, namely; [nominal scale](https://formpl.us/blog/nominal-data), [ordinal scale](https://formpl.us/blog/ordinal-data), [interval scale](https://formpl.us/blog/interval-data), and ratio scale.

The measurement scales are used to measure [qualitative and quantitative data](https://www.formpl.us/blog/qualitative-quantitative-data). With nominal and ordinal scale being used to measure qualitative data while interval and ratio scales are used to measure [quantitative data](https://www.formpl.us/blog/quantitative-data).

**Characteristics of a Measurement Scale**

**Identity**

Identity refers to the assignment of numbers to the values of each variable in a data set. Consider a questionnaire that asks for a respondent's gender with the options Male and Female for instance. The values 1 and 2 can be assigned to Male and Female respectively.

Arithmetic operations can not be performed on these values because they are just for identification purposes. This is a characteristic of a nominal scale.

**Magnitude**

The magnitude is the size of a measurement scale, where numbers (the identity) have an inherent order from least to highest. They are usually represented on the scale in ascending or descending order. The position in a race, for example, is arranged from the 1st, 2nd, 3rd to the least.

This example is measured on an ordinal scale because it has both identity and magnitude.

**Equal intervals**

Equal Intervals means that the scale has a standardized order. I.e., the difference between each level on the scale is the same. This is not the case for the ordinal scale example highlighted above.

Each position does not have an equal interval difference. In a race, the 1st position may complete the race in 20 secs, 2nd position in 20.8 seconds while the 3rd in 30 seconds.

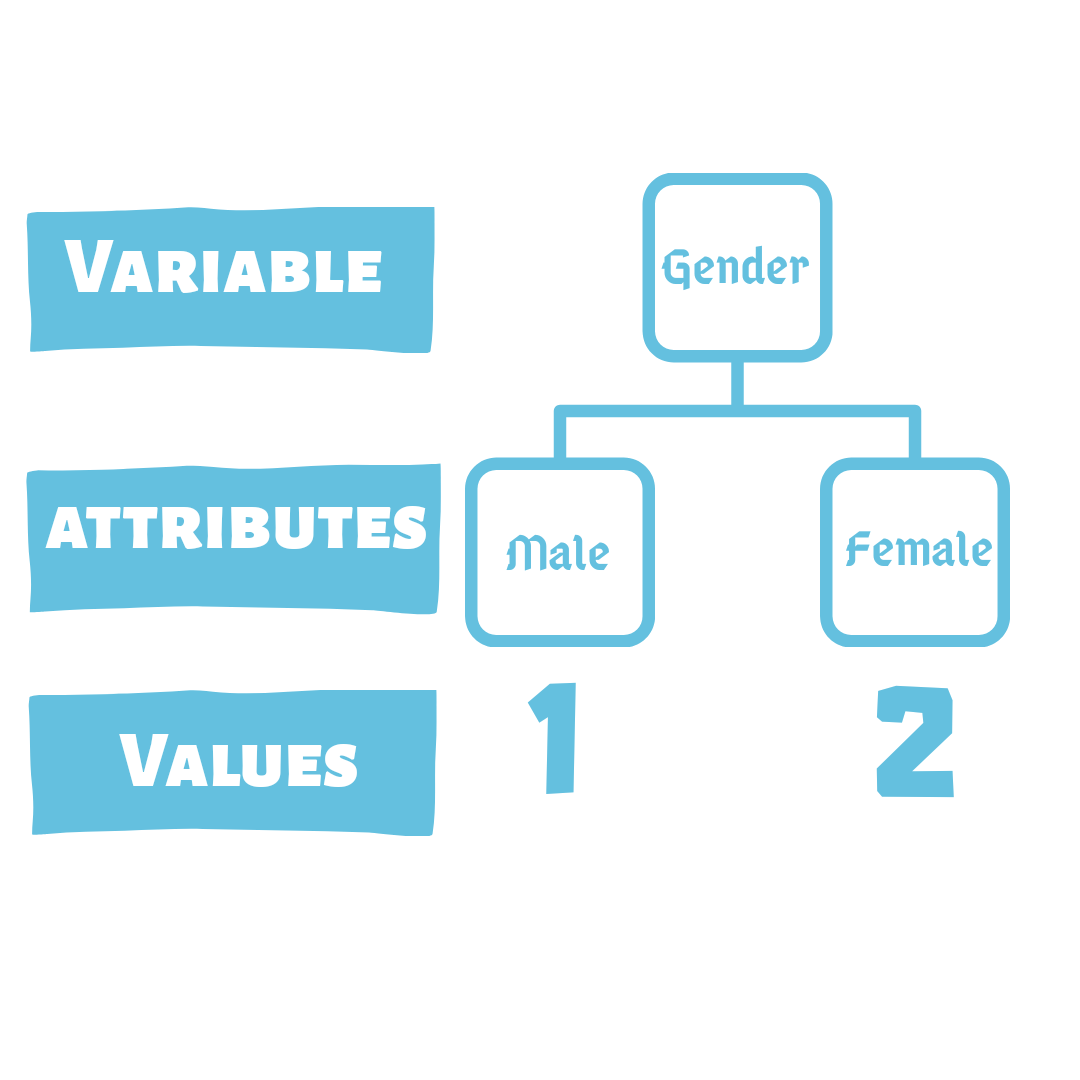
A variable that has an identity, magnitude, and the equal interval is measured on an interval scale.

**Absolute zero**

Absolue zero is a feature that is unique to a ratio scale. It means that there is an existence of zero on the scale, and is defined by the absence of the variable being measured (e.g. no qualification, no money, does not identify as any gender, etc.

**Levels of Data Measurement**

The level of measurement of a given data set is determined by the relationship between the values assigned to the attributes of a data variable. For example, the relationship between the values (1 and 2) assigned to the attributes (male and female) of the variable (Gender) is "identity". This via. a nominal scale example.



By knowing the different levels of data measurement, researchers are able to choose the best method for statistical analysis. The different levels of data measurement are: nominal, ordinal, interval and ratio scales

**Nominal Scale**

The[nominal scale](https://formpl.us/blog/nominal-data)is a scale of measurement that is used for identification purposes. It is the coldest and weakest level of data measurement among the four.

Sometimes known as categorical scale, it assigns numbers to attributes for easy identity. These numbers are however not qualitative in nature and only act as labels.

The only statistical analysis that can be performed on a nominal scale is the percentage or frequency count. It can be analyzed graphically using a bar chart and pie chart.

For example: In the example below, the measurement of the popularity of a political party is measured on a nominal scale.

Which political party are you affiliated with?

* Independent
* Republican
* Democrat

Labeling Independent as "1", Republican as "2" and Democrat as "3" does not in any way mean any of the attributes are better than the other. They are just used as an identity for easy data analysis.

**Ordinal Scale**

Ordinal Scale involves the ranking or ordering of the attributes depending on the variable being scaled. The items in this scale are classified according to the degree of occurrence of the variable in question.

The attributes on an [ordinal scale](https://www.formpl.us/blog/ordinal-data) are usually arranged in ascending or descending order. It measures the degree of occurrence of the variable.

Ordinal scale can be used in market research, advertising, and customer satisfaction surveys. It uses qualifiers like very, highly, more, less, etc. to depict a degree.

We can perform statistical analysis like median and mode using the ordinal scale, but not mean. However, there are other statistical alternatives to mean that can be measured using the ordinal scale.

For example: A software company may need to ask its users:

How would you rate our app?

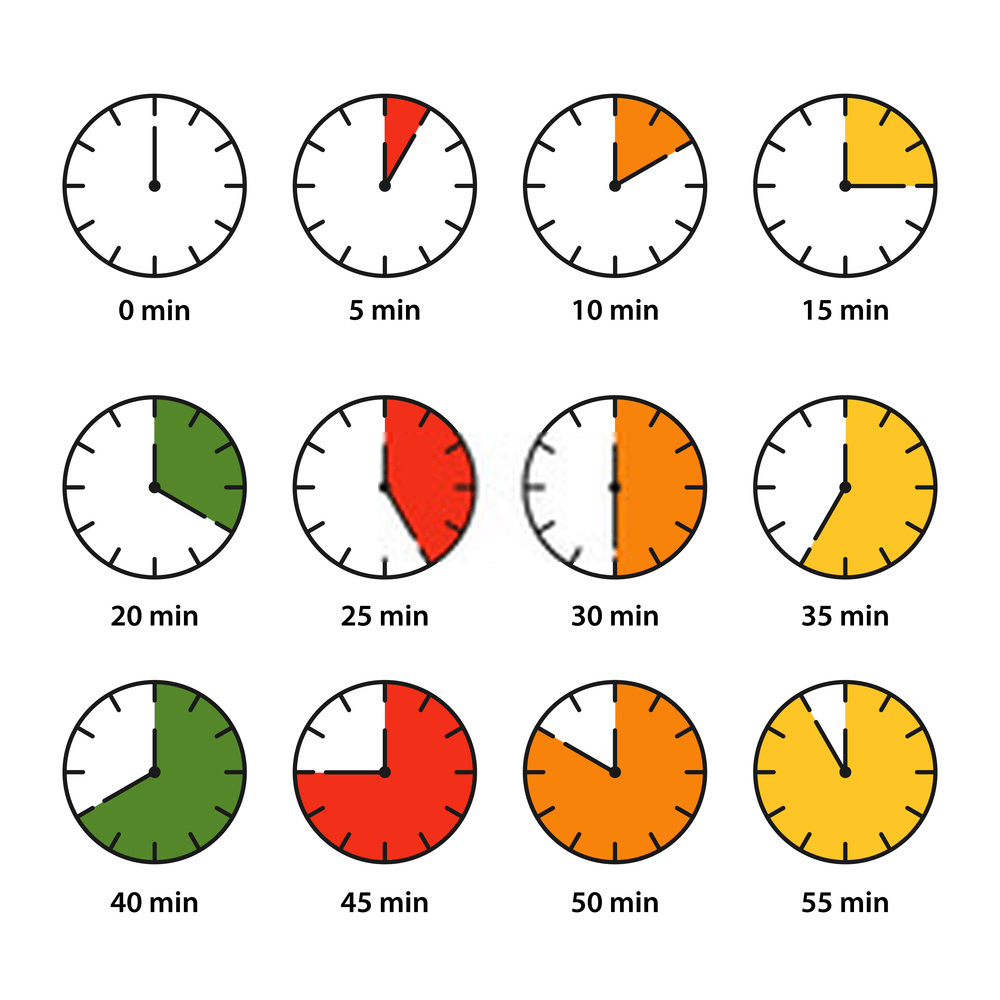
* Excellent
* Very Good
* Good
* Bad
* Poor

The attributes in this example are listed in descending order.

**Interval Scale**

The [interval scale of data measurement](https://www.formpl.us/blog/interval-data) is a scale in which the levels are ordered and each numerically equal distances on the scale have equal interval difference. If it is an extension of the ordinal scale, with the main difference being the existence of equal intervals.

With an interval scale, you not only know that a given attribute A is bigger than another attribute B, but also the extent at which A is larger than B. Also, unlike ordinal and nominal scale, arithmetic operations can be performed on an interval scale.



**A 5 Minutes Interval Time Scale**

It is used in various sectors like in education, medicine, engineering, etc. Some of these uses include calculating a student's CGPA, measuring a patient's temperature, etc.

A common example is measuring temperature on the Fahrenheit scale. It can be used in calculating mean, median, mode, range, and standard deviation.

**Ratio Scale**

Ratio Scale is the peak level of data measurement. It is an extension of the interval scale, therefore satisfying the four characteristics of measurement scale; identity, magnitude, equal interval, and the absolute zero property.

This level of data measurement allows the researcher to compare both the differences and the relative magnitude of numbers. Some examples of ratio scales include length, weight, time, etc.

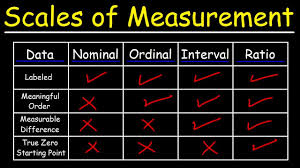
With respect to market research, the common ratio scale examples are price, number of customers, competitors, etc. It is extensively used in marketing, advertising, and business sales.

The ratio scale of data measurement is compatible with all statistical analysis methods like the measures of central tendency (mean, median, mode, etc.) and measures of dispersion (range, standard deviation, etc.).

For example: A survey that collects the weights of the respondents.

Which of the following category do you fall in? Weigh

* more than 100 kgs
* 81 - 100 kgs
* 61 - 80 kgs
* 40 - 60 kgs
* Less than 40 kgs



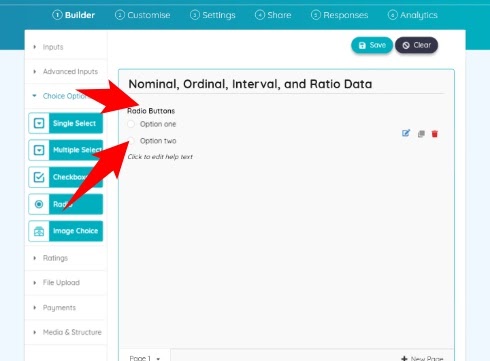
**How to Collect Nominal, Ordinal, Interval & Ratio Data with Formplus**

Formplus is the best tool for collecting nominal, ordinal, interval and ratio data. It is an easy to use form builder that allows you to collect data with ease. Follow the following steps to collect data on Formplus

**Step 1 - Select Feature**

We will be using the radio choice multiple-choice questions to collect data on Formplus form builder.

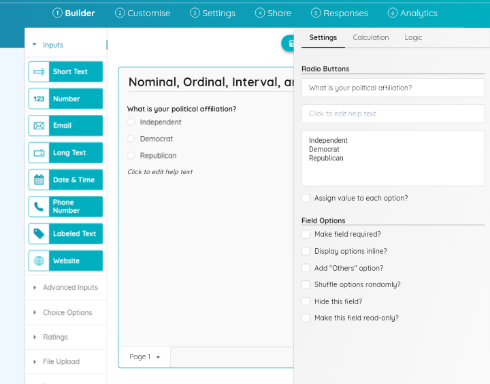
* Sign up or log in to your account on <https://www.formpl.us/>
* Click the "Choice Options" tab in the form builder menu.
* Click on the Radio button.



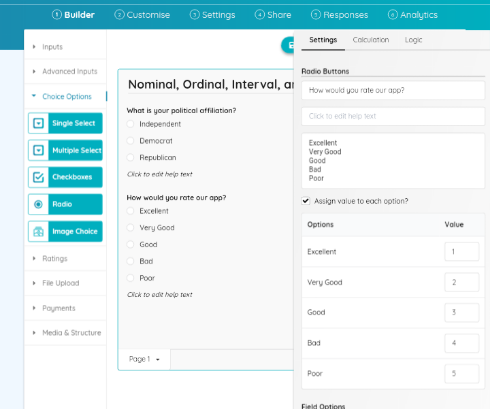
**Step 2 - Edit Form**

**Nominal Data**

* Click on the edit button to edit the form.
* Edit the question and choice options.
* Click the save button to save edits.

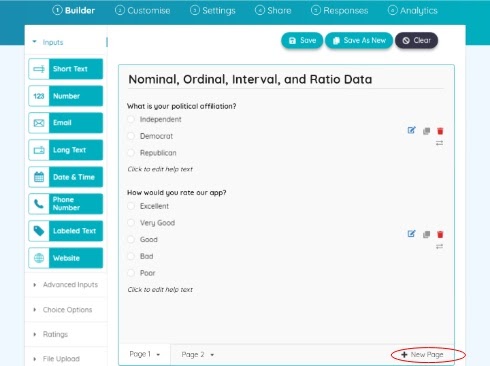


**Ordinal Data**

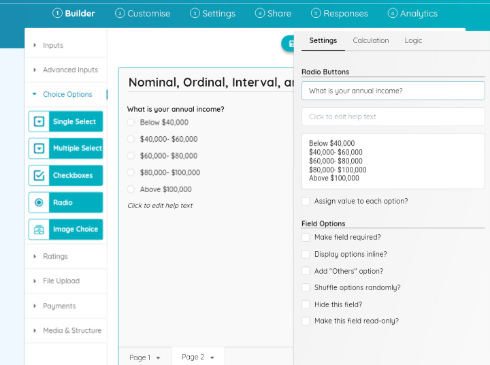
* Repeat **Step 1**.
* Click on the edit button to edit the form.
* Edit the question and choice options
* Assign values to the choice options.
* Click the save button to save edits.

**Interval Data**

* Click on the "+" sign at the bottom to add a new page.



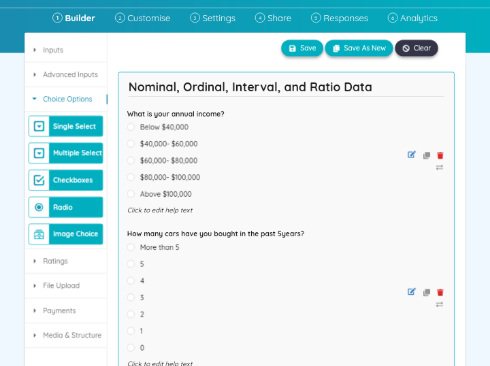
* Repeat **Step 1**.
* Click on the edit button to edit the form.
* Edit the question and choice options.
* Click the save button to save edits.



**Note**: The interval data options do not have a zero value.

**Ratio Data**

* Repeat **Step 1**.
* Click on the edit button to edit the form.
* Edit the question and choice options.
* Click the save button to save edits.
* Click the save button in the top right corner to save the form.



**Note:**that the ratio data example has a zero value, which differentiates it from the interval scale.

**Step 3 - Customize and Preview Form**

**Types of Measurement Scale**

There are two main types of measurement scales, namely; comparative scales and non-comparative scales.

**Comparative Scales**

In comparative scaling, respondents are asked to make a comparison between one object and the other. When used in market research, customers are asked to evaluate one product in direct comparison to the others. Comparative scales can be further divided into the pair comparison, rank order, constant sum and q-sort scales.

* **Paired Comparison Scale**

Paired Comparison scale is a scaling technique that presents the respondents with two objects at a time and asks them to choose one according to a predefined criterion. Product researchers use it in comparative product research by asking customers to choose the most preferred to them in between two closely related products.

**For example,** there are 3 new features in the last release of a software product. But the company is planning to remove 1 of these features in the new release. Therefore, the product researchers are performing a comparative analysis of the most and least preferred feature.

1. Which feature is most preferred to you between the following pairs?

* Filter - Voice recorder
* Filter - Video recorder
* Voice recorder - Video recorder
* **Rank Order Scale:**

In rank order scaling technique, respondents are simultaneously provided with multiple options and asked to rank them in order of priority based on a predefined criterion. It is mostly used in marketing to measure preference for a brand, product, or feature.

When used in competitive analysis, the respondent may be asked to rank a group of brands in terms of personal preference, product quality, customer service, etc. The results of this data collection are usually obtained in the conjoint analysis, as it forces customers to discriminate among options.

The rank order scale is a type of ordinal scale because it orders the attributes from the most preferred to the least preferred but does not have a specific distance between the attributes.

For example:

Rank the following brands from the most preferred to the least preferred.

* Coca-Cola
* Pepsi Cola
* Dr pepper
* Mountain Dew
* **Constant Sum Scale**

Constant Sum scale is a type of measurement scale where the respondents are asked to allocate a constant sum of units such as points, dollars, chips or chits among the stimulus objects according to some specified criterion. The constant sum scale assigns a fixed number of units to each attribute, reflecting the importance a respondent attaches to it.

This type of measurement scale can be used to determine what influences a customer's decision when choosing which product to buy. For example, you may wish to determine how important price, size, fragrance, and packaging is to a customer when choosing which brand of perfume to buy.

Some of the major setbacks of this technique are that respondents may be confused and end up allocating more or fewer points than those specified. The researchers are left to deal with a group of data that is not uniform and may be difficult to analyze.

Avoid this with the [logic feature on Formplus](https://www.formpl.us/features/conditional-logic). This feature allows you to add a restriction that prevents the respondent from adding more or fewer points than specified to your form.

* **Q-Sort Scale**

Q-Sort scale is a type of measurement scale that uses a rank order scaling technique to sort similar objects with respect to some criterion. The respondents sort the number of statements or attitudes into piles, usually of 11.

The Q-Sort Scaling helps in assigning ranks to different objects within the same group, and the differences among the groups (piles) are clearly visible. It is a fast way of facilitating discrimination among a relatively large set of attributes.

For example, a new restaurant that is just preparing its menu may want to collect some information about what potential customers like:

The document provided contains a list of 50 meals. Please choose 10 meals you like, 30 meals you are neutral about (neither like nor dislike) and 10 meals you dislike.



**Non-Comparative Scales**

In non-comparative scaling, customers are asked to only evaluate a single object. This evaluation is totally independent of the other objects under investigation. Sometimes called monadic or metric scale, Non-Comparative scale can be further divided into continuous and the itemized rating scales

* **Continuous Rating Scale**

In continuous rating scale, respondents are asked to rate the objects by placing a mark appropriately on a line running from one extreme of the criterion to the other variable criterion. Also called the graphic rating scale, it gives the respondent the freedom to place the mark anywhere based on personal preference.

Once the ratings are obtained, the researcher splits up the line into several categories and then assign the scores depending on the category in which the ratings fall. This rating can be visualized in both horizontal and vertical form.

Although easy to construct, the continuous rating scale has some major setbacks, giving it limited usage in market research.

* **Itemized Rating Scale**

The itemized rating scale is a type of ordinal scale that assigns numbers each attribute. Respondents are usually asked to select an attribute that best describes their feelings regarding a predefined criterion.

Itemized rating scale is further divided into 2, namely; Likert scale, Stapel scale, and semantic scale.

* **Likert Scale:**[**A Likert scale**](https://formpl.us/blog/point-likert-scale)**is an ordinal scale with five response categories, which is used to order a list of attributes from the best to the least. This scale uses adverbs of degree like very strongly, highly, etc. to indicate the different levels.**
* **Stapel Scale: This a scale with 10 categories, usually ranging from -5 to 5 with no zero point. It is a vertical scale with 3 columns, where the attributes are placed in the middle and the least (-5) and highest (5) is in the 1st and 3rd columns respectively.**
* **Semantic Differential Scale: This is a seven-point rating scale with endpoints associated with bipolar labels (e.g. good or bad, happy, etc.). It can be used for marketing, advertising and in different stages of product development.**

If there is more than one item being inherently investigated, it can be visualized on a table with more than 3 columns.

**Conclusion**

In a nutshell, scales of measurement refers to the various measures used in quantifying the variables researchers use In performing data analysis. They are an important aspect of research and statistics because the level of data measurement is what determines the data analysis technique to be used.

Understanding the concept of scales of measurements is a prerequisite to working with data and performing statistical analysis. The different measurement scales have some similar properties and are therefore important to properly analyze the data to determine its measurement scale before choosing a technique to use for analysis.

A number of scaling techniques are available for the measurement of the same measurement scale. Therefore, there is no unique way of selecting a scaling technique for research purposes.

Q.No.2: Explain the following terminologies with reference to research. (20)

a. Content Validity

When it comes to developing measurement tools such as intelligence tests, surveys, and self-report assessments, validity is important. A variety of types of validity exist, each designed to ensure that specific aspects of measurement tools are accurately measuring what they are intended to measure and that the results can be applied to real-world settings.

Before we move into discussing content validity, it is important to understand that validity is a broad concept that encompasses many aspects of assessment. For example, face validity describes the degree to which an assessment measures what it appears to measure, concurrent validity measures how well the results of one assessment correlate with other assessments designed to measure the same thing, and predictive validity measures how well the assessment results can predict a relationship between the construct of being measured and future behavior.

So, what about content validity? **Content validity** refers to how accurately an assessment or measurement tool taps into the various aspects of the specific construct in question. In other words, do the questions really assess the construct in question, or are the responses by the person answering the questions influenced by other factors? o how is content validity measured? How do researchers know if an assessment has content validity?

Content validity is most often measured by relying on the knowledge of people who are familiar with the construct being measured. These subject-matter experts are usually provided with access to the measurement tool and are asked to provide feedback on how well each question measures the construct in question. Their feedback is then analyzed, and informed decisions can be made about the effectiveness of each question.

b. Criterion Related Validity

In [psychometrics](https://en.wikipedia.org/wiki/Psychometrics), **criterion** or **concrete validity** is the extent to which a measure is related to an outcome.[[1]](https://en.wikipedia.org/wiki/Criterion_validity#cite_note-Cronbach&Meehl1955-1) Criterion validity is often divided into concurrent and predictive validity based on the timing of measurement for the "predictor" and outcome.[[1]](https://en.wikipedia.org/wiki/Criterion_validity#cite_note-Cronbach&Meehl1955-1):page 282 [Concurrent validity](https://en.wikipedia.org/wiki/Concurrent_validity) refers to a comparison between the measure in question and an outcome assessed at the same time. In [*Standards for Educational & Psychological Tests*](https://en.wikipedia.org/wiki/Standards_for_Educational_and_Psychological_Testing), it states, "concurrent validity reflects only the status quo at a particular time."[[2]](https://en.wikipedia.org/wiki/Criterion_validity#cite_note-:0-2) [Predictive validity](https://en.wikipedia.org/wiki/Predictive_validity), on the other hand, compares the measure in question with an outcome assessed at a later time. Although concurrent and predictive validity are similar, it is cautioned to keep the terms and findings separated. "Concurrent validity should not be used as a substitute for predictive validity without an appropriate supporting rationale."[[2]](https://en.wikipedia.org/wiki/Criterion_validity#cite_note-:0-2) Criterion validity is typically assessed by comparison with a gold standard test.[[3]](https://en.wikipedia.org/wiki/Criterion_validity#cite_note-3)

An example of concurrent validity is a comparison of the scores of the [CLEP](https://en.wikipedia.org/wiki/CLEP) College Algebra exam with course grades in college algebra to determine the degree to which scores on the CLEP are related to performance in a college algebra class.[[4]](https://en.wikipedia.org/wiki/Criterion_validity#cite_note-:1-4) An example of predictive validity is IQ tests, it was originally developed predict future school performance. Another example is a comparison of scores on the [SAT](https://en.wikipedia.org/wiki/SAT) with first semester [grade point average](https://en.wikipedia.org/wiki/Grade_point_average) (GPA) in college; this assesses the degree to which SAT scores are predictive of college performance.[[4]](https://en.wikipedia.org/wiki/Criterion_validity#cite_note-:1-4)

c. Construct Validity

Construct validity refers to the degree to which inferences can legitimately be made from the operationalizations in your study to the theoretical constructs on which those operationalizations were based. Like [external validity](https://conjointly.com/kb/external-validity/), construct validity is related to generalizing. But, where external validity involves generalizing from your study context to other people, places or times, construct validity involves generalizing from your program or measures to the concept of your program or measures. You might think of construct validity as a “labeling” issue. When you implement a program that you call a “Head Start” program, is your label an accurate one? When you measure what you term “self esteem” is that what you were really measuring?

I would like to tell two major stories here. The first is the more straightforward one. I’ll discuss several ways of thinking about the [idea of construct validity](https://conjointly.com/kb/construct-validity-idea/), several metaphors that might provide you with a foundation in the richness of this idea. Then, I’ll discuss the major [construct validity threats](https://conjointly.com/kb/construct-validity-threats/), the kinds of arguments your critics are likely to raise when you make a claim that your program or measure is valid. In most research methods texts, construct validity is presented in the section on measurement. And, it is typically presented as one of many different types of validity (e.g., face validity, predictive validity, concurrent validity) that you might want to be sure your measures have. I don’t see it that way at all. I see construct validity as the overarching quality with all of the other measurement validity labels falling beneath it. And, I don’t see construct validity as limited only to measurement. As I’ve already implied, I think it is as much a part of the independent variable – the program or treatment – as it is the dependent variable. So, I’ll try to make some sense of the various [measurement validity types](https://conjointly.com/kb/measurement-validity-types/) and try to move you to think instead of the validity of any operationalization as falling within the general category of construct validity, with a variety of subcategories and subtypes.

The second story I want to tell is more historical in nature. During World War II, the U.S. government involved hundreds (and perhaps thousands) of psychologists and psychology graduate students in the development of a wide array of measures that were relevant to the war effort. They needed personality screening tests for prospective fighter pilots, personnel measures that would enable sensible assignment of people to job skills, psychophysical measures to test reaction times, and so on. After the war, these psychologists needed to find gainful employment outside of the military context, and it’s not surprising that many of them moved into testing and measurement in a civilian context.

During the early 1950s, the American Psychological Association began to become increasingly concerned with the quality or validity of all of the new measures that were being generated and decided to convene an effort to set standards for psychological measures. The first formal articulation of the idea of construct validity came from this effort and was couched under the somewhat grandiose idea of the [nomological network](https://conjointly.com/kb/nomological-network/). The nomological network provided a theoretical basis for the idea of construct validity, but it didn’t provide practicing researchers with a way to actually establish whether their measures had construct validity. In 1959, an attempt was made to develop a method for assessing construct validity using what is called a [multitrait-multimethod matrix](https://conjointly.com/kb/multitrait-multimethod-matrix/), or MTMM for short. In order to argue that your measures had construct validity under the MTMM approach, you had to demonstrate that there was both convergent and discriminant validity in your measures. You demonstrated convergent validity when you showed that measures that are theoretically supposed to be highly interrelated are, in practice, highly interrelated. And, you showed discriminant validity when you demonstrated that measures that shouldn’t be related to each other in fact were not.

While the MTMM did provide a methodology for assessing construct validity, it was a difficult one to implement well, especially in applied social research contexts and, in fact, has seldom been formally attempted. When we examine carefully the thinking about construct validity that underlies both the nomological network and the MTMM, one of the key themes we can identify in both is the idea of “pattern.” When we claim that our programs or measures have construct validity, we are essentially claiming that we as researchers understand how our constructs or theories of the programs and measures operate in theory and we claim that we can provide evidence that they behave in practice the way we think they should. The researcher essentially has a theory of how the programs and measures related to each other (and other theoretical terms), a theoretical pattern if you will. And, the researcher provides evidence through observation that the programs or measures actually behave that way in reality, an observed pattern. When we claim construct validity, we’re essentially claiming that our observed pattern – how things operate in reality – corresponds with our theoretical pattern – how we think the world works. I call this process pattern matching, and I believe that it is the heart of construct validity. It is clearly an underlying theme in both the nomological network and the MTMM ideas. And, I think that we can develop concrete and feasible methods that enable practicing researchers to assess pattern matches – to assess the construct validity of their research. The section on [pattern matching](https://conjointly.com/kb/construct-validity-pattern-matching/) lays out my idea of how we might use this approach to assess construct validity.

d. Reliability

**Reliability** in [statistics](https://en.wikipedia.org/wiki/Statistics) and psychometrics is the overall consistency of a measure.[[1]](https://en.wikipedia.org/wiki/Reliability_(statistics)#cite_note-1) A measure is said to have a high reliability if it produces similar results under consistent conditions. "It is the characteristic of a set of test scores that relates to the amount of random error from the measurement process that might be embedded in the scores. Scores that are highly reliable are precise, reproducible, and consistent from one testing occasion to another. That is, if the testing process were repeated with a group of test takers, essentially the same results would be obtained. Various kinds of reliability coefficients, with values ranging between 0.00 (much error) and 1.00 (no error), are usually used to indicate the amount of error in the scores." [[2]](https://en.wikipedia.org/wiki/Reliability_(statistics)#cite_note-2) For example, measurements of people's height and weight are often extremely reliable. There are several general classes of reliability estimates:

* [**Inter-rater reliability**](https://en.wikipedia.org/wiki/Inter-rater_reliability) assesses the degree of agreement between two or more raters in their appraisals. For example, a person gets a stomach ache and different doctors all give the same diagnosis.[[5]](https://en.wikipedia.org/wiki/Reliability_(statistics)#cite_note-5):71
* [**Test-retest reliability**](https://en.wikipedia.org/wiki/Test-retest_reliability) assesses the degree to which test scores are consistent from one test administration to the next. Measurements are gathered from a single rater who uses the same methods or instruments and the same testing conditions.[[4]](https://en.wikipedia.org/wiki/Reliability_(statistics)#cite_note-themasb.org-4) This includes [intra-rater reliability](https://en.wikipedia.org/wiki/Intra-rater_reliability).
* **Inter-method reliability** assesses the degree to which test scores are consistent when there is a variation in the methods or instruments used. This allows inter-rater reliability to be ruled out. When dealing with [forms](https://en.wikipedia.org/wiki/Form_(document)), it may be termed **parallel-forms reliability**.[[6]](https://en.wikipedia.org/wiki/Reliability_(statistics)#cite_note-socialresearchmethods-6)
* [**Internal consistency**](https://en.wikipedia.org/wiki/Internal_consistency)**reliability**, assesses the consistency of results across items within a test.[[](https://en.wikipedia.org/wiki/Reliability_(statistics)#cite_note-socialresearchmethods-6)
* Reliability does not imply [validity](https://en.wikipedia.org/wiki/Validity_(statistics)). That is, a reliable measure that is measuring something consistently is not necessarily measuring what you want to be measured. For example, while there are many reliable tests of specific abilities, not all of them would be valid for predicting, say, job performance.
* While reliability does not imply [validity](https://en.wikipedia.org/wiki/Validity_(statistics)), reliability does place a limit on the overall validity of a test. A test that is not perfectly reliable cannot be perfectly valid, either as a means of measuring attributes of a person or as a means of predicting scores on a criterion. While a reliable test may provide useful valid information, a test that is not reliable cannot possibly be valid.[[7]](https://en.wikipedia.org/wiki/Reliability_(statistics)#cite_note-David-7)
* For example, if a set of [weighing scales](https://en.wikipedia.org/wiki/Weighing_scales) consistently measured the weight of an object as 500 grams over the true weight, then the scale would be very reliable, but it would not be valid (as the returned weight is not the true weight). For the scale to be valid, it should return the true weight of an object. This example demonstrates that a perfectly reliable measure is not necessarily valid, but that a valid measure necessarily must be reliable.

In practice, testing measures are never perfectly consistent. Theories of test reliability have been developed to estimate the effects of inconsistency on the accuracy of measurement. The basic starting point for almost all theories of test reliability is the idea that test scores reflect the influence of two sorts of factors:[[7]](https://en.wikipedia.org/wiki/Reliability_(statistics)#cite_note-David-7)

1. **Factors that contribute to consistency:** stable characteristics of the individual or the attribute that one is trying to measure

2. **Factors that contribute to inconsistency:** features of the individual or the situation that can affect test scores but have nothing to do with the attribute being measured.

These factors include:[[7]](https://en.wikipedia.org/wiki/Reliability_(statistics)#cite_note-David-7)

* Temporary but general characteristics of the individual: health, fatigue, motivation, emotional strain
* Temporary and specific characteristics of individual: comprehension of the specific test task, specific tricks or techniques of dealing with the particular test materials, fluctuations of memory, attention or accuracy
* Aspects of the testing situation: freedom from distractions, clarity of instructions, interaction of personality, sex, or race of examiner
* Chance factors: luck in selection of answers by sheer guessing, momentary distractions

The goal of estimating reliability is to determine how much of the variability in test scores is due to **errors in measurement** and how much is due to variability in **true scores**.[[7]](https://en.wikipedia.org/wiki/Reliability_(statistics)#cite_note-David-7)

A **true score** is the replicable feature of the concept being measured. It is the part of the observed score that would recur across different measurement occasions in the absence of error.

**Errors of measurement** are composed of both [random error](https://en.wikipedia.org/wiki/Random_error) and [systematic error](https://en.wikipedia.org/wiki/Systematic_error). It represents the discrepancies between scores obtained on tests and the corresponding true scores.

This conceptual breakdown is typically represented by the simple equation:

***Observed test score = true score + errors of measurement***

Q.No.3: Personal Interviewing method is not feasible for students what is your opinion about this statement. (15)

Ans: The two main types of interviews conducted in marketing research are structured and unstructured.

**Unstructured informal interview**

The unstructured informal interview is normally conducted as a preliminary step in the research process to generate ideas/hypotheses about the subject being investigated so that these might be tested later in the survey proper. Such interviews are entirely informal and are not controlled by a specific set of detailed questions. Rather the interviewer is guided by a pre-defined list of issues. These interviews amount to an informal conversation about the subject.

Informal interviewing is not concerned with discovering 'how many' respondents think in a particular way on an issue (this is what the final survey itself will discover). The aim is to find out how people think and how they react to issues, so that the ultimate survey questionnaire can be framed along the lines of thought that will be most natural to respondents.

The respondent is encouraged to talk freely about the subject, but is kept to the point on issues of interest to the researcher. The respondent is encouraged to reveal everything that he/she feels and thinks about these points. The interviewer must note (or tape-record) all remarks that may be relevant and pursue them until he/she is satisfied that there is no more to be gained by further probing. Properly conducted, informal interviews can give the researcher an accurate feel for the subject to be surveyed. Focus groups, discussed later in this chapter, make use of relatively unstructured interviews.

**Structured standardised interview**

With structured standardised interviews, the format is entirely different. A structured interview follows a specific questionnaire and this research instrument is usually used as the basis for most quantitative surveys. A standardised structured questionnaire is administered where specific questions are asked in a set order and in a set manner to ensure no variation between interviews.

Respondents' answers are recorded on a questionnaire form (usually with pre-specified response formats) during the interview process, and the completed questionnaires are most often analysed quantitatively. The structured interview usually denies the interviewer the opportunity to either add or remove questions, change their sequence or alter the wording of questions.

**Depth interviews**

Depth interviews are one-to-one encounters in which the interviewer makes use of an unstructured or semi-structured set of issues/topics to guide the discussion. The object of the exercises is to explore and uncover deep-seated emotions, motivations and attitudes. They are most often employed when dealing with sensitive matters and respondents are likely to give evasive or even misleading answers when directly questioned. Most of the techniques used in the conduct of depth interviews have been borrowed from the field of psychoanalysis. Depth interview are usually only successful when conducted by a well trained and highly skilled interviewer.

Other instances when depth interviewers can be particularly effective are: where the study involves an investigation of complex behaviour or decision-making processes; when the target respondents are difficult to gather together for group interviewers (e.g. farmers, veterinary surgeons, haulage contractors, government officials); and where the interviewee is prepared to become an informant only if he/she is able to preserve his/her anonymity.

Dillon et al1. believe that to be effective, the interviewer must adhere to six fundamental rules. These are:

 he/she must avoid appearing superior or condescending and make use of only familiar words

 he/she must put question indirectly and informatively

 he/she must remain detached and objective

 he/she must avoid questions and questions structure that encourage 'yes' or 'no' answers

 he/she must probe until all relevant details, emotions and attitudes are revealed

 he/she must provide an atmosphere that encourages the respondent to speak freely, yet keeping the conservation focused on the issue(s) being researched

Depth interviews involve a heavy time commitment, especially on the part of the marketing researcher. Interview transcripts have to be painstakingly recovered, if they are to be accurate, either from terse interview notes or from tape-recordings of the interviews. This can take many hours of often laborious work. The transcripts then have to be read and re-read, possibly several times, before the researcher is able to begin the taxing process of analysing and interpreting the data.

**Telephone Interviews**

Whilst telephone interviews among consumers, are very common in the developed world, these are conducted with far less frequency in the developing world. The reason is somewhat obvious, i.e. only a relatively small proportion of the total population has a telephone in the house. Moreover, telephone owners tend to be urban dwellers and have above average incomes and are therefore unrepresentative of the population as a whole.

To a greater extent, telephone interviewing has potential in surveys of businesses, government agencies and other organisations or institutions. Even then, it is still the case that telephone surveys are rarely without bias. Whilst it is true that many businesses have a telephone, small businesses and even medium-sized enterprises are far less likely to have access to telephones.

Telephone interviews afford a certain amount of flexibility. It is possible, for example, for interviewers to put complex questions over the telephone. The interviewers can probe, skip questions that prove irrelevant to the case of a particular respondent and change the sequence of questions in response to the flow of the discussion, and earlier replies can be revisited. The interaction between interviewer and interviewee that is possible over the telephone simply is not achievable through a mailed questionnaire. In comparison to personal interviews, telephone interviews do not appear to enjoy any margin of advantage. Perhaps the only advantages are those of speed and cost. Even then, manpower costs in developing countries tend to be very low and so only speed remains as a potential advantage over personal interviews.

In the developed world, the era of computer-assisted telephone interviewing (CATI) has begun. Researchers conduct the telephone interview whilst seated at a computer. Responses are entered directly into the computer, by the interviewer. The screen displays the questionnaire and any skipping of questions, due to earlier responses directing that some questions are not applicable in the case of the interviewee, is controlled automatically by the computer. Since the responses are entered directly into the computer the data is instantaneously processed. The computer can also be programmed to produce standardised marketing reports.

[**Figure 5.1 Types of personal interview**](http://www.fao.org/3/w3241e/w3241e06.jpg)

**Conducting the interviews**

It is essential, for both types of interview format, that the interviewer has a good grasp of the study's objectives, and of the information that is to be collected. This will enable 'probing' to elicit the right data required, and ensure all relevant issues are covered. Furthermore, some respondents may ask why a particular question was included in an interview, and it may be necessary for the interviewer to be able to 'justify' particular questions.

In rural areas it is customary before embarking on a formal interviewing survey to notify the relevant public authorities, e.g. village head, district union, etc. to ensure co-operation from respondents. Sometimes individuals may refuse to co-operate unless they are convinced that the interviewer has permission and approval to conduct the survey from the recognised local authorities.

Before commencing on interviews it is as well for the interviewer to prepare what he/she is going to say when he/she first meets a respondent. Decisions need to be made as to whether the respondent is to be told who is sponsoring the study, the purpose of the study, or how the data is to be used, and so on. These points need to be decided beforehand to ensure that a 'standardised' approach is used for each interview. Variations in approach style may lead to different types of response from respondents and therefore variations in results. If suitable introductions are prepared in advance, no time will be lost during the interview in lengthy explanations, and a good impression can be created from the start.

**Interview approach in the field:**It is important that the interviewer keeps as low a profile as possible in the rural setting. Interviewers should walk as much as possible and in small numbers - two in a team is often best. If the research team is large, it is advisable to divide the study area into a number of zones to avoid duplicating efforts or interviewing the same respondents.

Once an individual who appears to be worth interviewing is spotted in the field, it is best not to wander around indecisively creating suspicion. He/she should be approached directly. However, one should avoid startling potential respondents by running up to them and pulling out the questionnaire for interview. Blending into the local context as much as possible is obviously the best strategy. One should always be sensitive to the fact that most people may be suspicious of outsiders.

The timing of the interview can be very important. One should be aware of the daily schedule, seasonal activities, and work habits of potential respondents. For example, if a farmer is irrigating and receives water only once a week for an hour, he/she may not be interested in participating in an interview at that time.

**Interview introduction:**The introduction to an interview is crucial. A good introduction can effectively gain the respondent's co-operation and a good interview, but a bad introduction could result in refusal to co-operate or biased responses.

|  |  |
| --- | --- |
| **Greeting:** | This should be made according to local custom. |
| **Small talk:** | Being approached by a stranger will make the potential respondent feel uncomfortable. It is necessary to help him/her feel at ease by starting with polite small talk about the weather or crop conditions, (in the case of a farmer) or about the health of the family and the general economic climate in the case of non-farmers. |

**Overcoming apprehension:** The approach of an interviewer is still an unfamiliar experience to most people. Many people are suspicious of outsiders and particularly interviewers. Some may think the interviewer is an 'official' who has come to check up on them for taxes. Certainly many potential respondents will fear that the information they give will be used against them at a later date, or that the interviewer is trying to probe family secrets. To ensure cooperation it is important to:

 Keep the atmosphere relaxed and informal.

It can be helpful if the interviewer plays down the fact that he/she wishes to conduct a 'formal' interview. Respondents can be encouraged to think that the interviewer is interested in conversation rather than interrogation.

 Explain why the interview is necessary.

The respondent should be given a brief background as to the nature and purpose of the study. This will bring him/her into the interviewer's confidence.

 Stress the value/benefit of the study to the respondent

Respondents are more likely to co-operate if they think they will ultimately benefit from the study. If one can indicate that as a result of the study it will be possible to develop better and cheaper products for the respondent, then they should be encouraged to co-operate.

 Appeal to the instincts of pride and vanity of the respondent

The respondent needs to be made to feel important. He/she needs to be made to feel that the interviewer is particularly interested in his/her opinion because he/she is the 'expert' and 'informed'.

Additional points that may help to put the respondent at ease could include:

**Language:**It is advisable that marketing researchers should adopt the language of those from whom they hope to obtain information.

"... using local names for socio-economic characteristics, bio-physical characteristics, lands, customs, time, intervals and measures".

**Length of interview:**The respondent can be assured that the interview will be brief. It is unwise to be deceitful here, otherwise there is a danger that the interview may be stopped mid-way by an angry respondent.

**Confidentiality:**The respondent can be assured that the interviewer will not reveal the respondent's identity (and will use the data only in aggregate form) or give the results to official organisations.

**Closing interview:**After all relevant topics have been covered or the respondent's time exhausted, the conversation should be brought to an end. If the weather is unfavourable (too hot or too wet) or the respondent seems pressed for time it is best to prematurely stop the interview. The departure is best done gracefully, naturally and not too abruptly. The business-like 'Got to go' departure should be avoided. The respondent should be thanked for his/her time and given the appropriate customary farewell.

**Interview recording**

All the best interviewing is useless if it has not been adequately recorded, so it is important to ensure good recording conditions. In an open-ended interview it is difficult to make notes on everything during the interview. The best approach in team-work is to appoint a scribe, i.e. a person whose job it is to write everything down. How long one waits before writing up full field-notes depends on the setting, and the interviewer's personal style but it should be borne in mind that an interviewer's memory is limited. It is surprising how facts, ideas and important observations that one thinks one will never forget quickly slip away. Half of the details from an interview can be forgotten within 24 hours, three-quarters can be lost within 2 days and after this only skeletal notes can be salvaged. Jotted notes will help prompt memory later, but it is best to write up interview notes while they are still fresh in the interviewer's mind after the interview or at the end of the interviewing day.

**Use of tape-recorders:** A tape recorder can often be useful. It enables the interviewer to give THE respondent his/her full attention during the interview and avoid the need to be constantly scribbling notes. It can also enable data to be left until such time as analysis can be applied more rigorously and in a more leisurely way. It should be borne in mind, however, that not everyone likes to be tape-recorded. If taping is contemplated the respondents' permission should be sought first.

**Sources of error and bias**

In personal interviews there are many ways in which 'errors' can be made by both the respondent and the interviewer, and this can lead to 'bias' in the results. The objective of the interviewer should be to minimise the likelihood of such bias arising.

**Respondent induced bias**

**Faulty memory:**Some respondents may answer a question incorrectly simply because they have a poor memory. The key to avoiding this problem is to steer clear of questions requiring feats of memory. For example, questions such as, "Can you tell me what your crop yield was four years ago?" should be avoided. Other aspects of faulty memory that were mentioned in the previous chapter were telescoping and creation.

**Exaggeration and dishonesty:**There can be a tendency on the part of some respondents to exaggerate claims about their conditions and problems if they think it will further their cause and lead to improvement in their well-being. The interviewer must be alert to, and note any, inconsistencies arising. This is best achieved by checking key pieces of information with a variety of sources.

**Failure to answer questions correctly:**If rapport is not developed sufficiently, the respondent may be unwilling to respond or fail to give sufficient attention or consideration to the questions asked, and if the respondent does not understand a question properly he may give inappropriate answers. The interviewer needs to ensure that the respondent fully understands the questions being asked and is responding in the appropriate context.

**Misunderstanding purpose of interview:**Some respondents may perceive the purpose of the survey to be a long-winded form of 'selling', particularly if the interviewer is asking them what they think about a new product. Their comments, therefore, about such issues as 'propensity to purchase' need to be looked at within a context where they may be expecting to have to buy the product at some stage and are trying to strike a hard bargain. To avoid such problems arising it is important to carefully explain the objectives of the survey, the identity of the interviewer and sponsor, and what is required of the respondent, prior to the interview proper.

**Influence of groups at interview:**During interviews the presence of other individuals is almost inevitable. Most of the time other family members or neighbours will wish to join in the discussion. Such a situation has can have important implications for the type of data obtained. The respondent may be tempted to answer in a way that gives him/her credibility in the eyes of onlookers, rather than giving a truthful reply. In circumstances where the presence of third parties cannot be avoided, the interviewer must ensure as far as possible that the answers being given are the honest opinions of the individual being interviewed. The interviewer must again be alert to inconsistencies and closely observe and monitor the way in which the respondent is reacting and interacting with those around him.

**Courtesy bias:**In interview situations it is quite possible that one will come across the problem of courtesy bias, i.e. the tendency for respondents to give answers that they think the interviewer wants to hear, rather than what they really feel. The respondents may not wish to be impolite or to offend the interviewer, and may therefore endeavour to give 'polite' answers. Courtesy bias can be an obstacle to obtaining useful and reliable data and therefore needs to be minimised. Generally, however, the creation of a good interview environment and an appropriate relationship between the interviewer and the respondent can help avoid too much courtesy bias arising:

**Bias induced by interviewer**

It is also possible for the interviewer him or herself to introduce bias into an interview, and this must be avoided at all costs.

**Desire to help the respondent:**The interviewer may become too sympathetic to the problems and conditions of the respondent, and this can affect the conduct of, and results obtained from, the interview. Objectivity must be retained at all times.

**Failure to follow instructions in administering the questions:**It is often tempting for the interviewer to change the wording of a question or introduce inflections in questions. This can affect the respondent's understanding and can bias his/her replies. Particular problems may arise if the respondent does not understand the question as stated and the interviewer tries to simplify the question. The altered wording may constitute a different question. When questions are open-ended, this can involve the interviewer in formulating probing questions that go beyond the printed words. Unless the probes follow instructions faithfully the potential for bias is great.

**Reactions to responses:**When respondents give answers, the interviewer must be careful not to 'react.' A note of 'surprise' or 'disbelief may easily bias the respondent's subsequent answers. Interviewers must respond with a uniform polite interest only.

**Focus group interviews**

Focus group interviews are a survey research instrument which can be used in addition to, or instead of, a personal interview approach. It has particular advantages for use in qualitative research applications. The central feature of this method of obtaining information from groups of people is that the interviewer strives to keep the discussion led by a moderator focused upon the issue of concern. The moderator behaves almost like a psycho-therapist who directs the group towards the focus of the researcher. In doing so, the moderator speaks very little, and encourages the group to generate the information required by stimulating discussion through terse provocative statements.

**Characteristics of focus group interviews**

The groups of individuals (e.g. housewives, farmers, manufacturers, etc.) are invited to attend an informal discussion. Usually between 6 and 8 participants are involved and the discussion would last between 1 and 2 hours. Small groups tend to lose the mutual stimulation among participants, whilst large groups can be difficult to manage and may prevent some participants having the opportunity to get fully involved in the discussion.

The researcher raises issues for discussion, following a 'guide list of topics' rather than a structured questionnaire. The participants are encouraged to discuss the issues amongst themselves and with the researcher in an informal and relaxed environment. The researcher records comments made by the participants (usually utilising a tape or video recorder). Figure 5.2 shows how this list of topics is arrived at.

[**Figure 5.2 The process of developing a topic list for focus groups**](http://www.fao.org/3/w3241e/w3241e07.jpg)

In contrast to a personal interview survey, the number of interviews in a typical group interview survey is very small, usually between 3 and 4 would be sufficient for each type of respondent-sector (e.g. farmers or manufacturers). Generally from the first interview on an unfamiliar subject the researcher will learn a great deal. The second and third interviews will produce more information, but not all of it will not be new. By the fourth interview most of what is revealed will have been covered before, and the diminishing returns involved would generally not justify the cost of further groups.

The participants within a focus group are selected in such a way that they exhibit a high degree of homogeneity with respect to either background, behaviour or both. Consider, for example, a study carried out by a small African nation that is looking for a niche market for a new range of sparkling wines. It is decided that, as a first step, a series of focus groups be conducted. The researchers are keen to ensure that each group comprises people who are similar in age and behaviour with respect to wine consumption. Figure 5.3 depicts the kind of screening questionnaire that the marketing researcher would use.

Interviews may be either structured or unstructured. In the case of the former, the interviewer has well defined questions presented in a set sequence. Such questionnaires can only be used when a certain amount is already known about the topic/situation being studied. The unstructured interview, where the researcher has only a list of topics which can be added to or changed in the course of the interview, is employed when little is known about the subject of study.

The depth interview is a one-to-one encounter, concluded at length, that enables the researcher to tackle sensitive and/or potentially embarrassing topics with the individual. Depth interviews are also useful when the decisions or behaviour under study are complex.

Telephone interviews, though common in the developed world, are relatively rare in the developing world because of the low level of telephone ownership. In developing countries, telephone interviews are more often confined to industrial marketing research. The latest development in this field is computer - aided - telephone - interviewing.

Bias can originate from the respondent and/or the interviewer. Respondent bias arises from poor memory, exaggeration or dishonesty, a lack of rapport with the interviewer or a misunderstanding over the purpose of the interview. Interview bias is most often due to a lack of objectivity and/or failure to administer questions properly and consistently.

Focus group interviews involve small groups of 6-8 individuals who are encouraged to discuss a topic and are prevented from straying from that topic by a moderator. The moderator is armed only with a list of issues/topics to guide the discussion and, therefore, the focus group is an unstructured group interview and gives rise to qualitative data.