2014

A Decision Maker's Guide to Data

Collection, Analysis and Interpretation

A handbook for use by program managers and program teams as a concise introduction to the collection, analysis, and interpretation of data on programs performance.

2014

INTRODUCTION

We offer here a handbook for use by program managers and program teams as a concise introduction to the collection, analysis and interpretation of data on program performance. Although the booklet was designed to be used as part of a training workshop (and that may be how you are using it), it can also be used by programs on their own as they work on specific action research projects. This handbook is not the only reference you will need in any case. Expert advice and input should be sought for any program measurement project, but program teams should be informed consumers of that expert technical assistance. Far too many programs spend far too much of their limited funds for expert support of questionable quality.

The power of good action research done well and in a timely fashion is its ability to clarify otherwise murky issues and focus otherwise divergent debates. Evaluative information based on clear design, good instrumentation and informed analysis and interpretation of results can shed some light on some things. The strongest argument for expenditure on evaluation is an informed consumer. We hope that this helps you to become one.

Kate Toms, Ph.D.

Research Works, Inc.
www.researchworks.org
(877) 766-1370

YOU NEED INFORMATION BECAUSE:

You make decisions about efficiency of management and effectiveness of programs (as the achievement of results or outcomes).

Decisions are judgments, and your judgments are only as good as the information on which they are based.

These are two separate (yet related) sets of decision making responsibility, and they are informed using:

USE EFFICIENCY MEASURES FOR **MANAGEMENT**

USE EFFECTIVENESS MEASURES FOR **RESULTS OR IMPACTS**

<table>
<tr><td colspan="2">Efficiency measures are used to:

• provide decision makers with information on management of programs, or
• provide decision makers with information on system (coordinated program) infrastructure.</td><td colspan="2">Effectiveness measures are used to:

• provide decision makers with information on program outcome achievement in light of program goals, or
• provide decision makers with information on outcome appropriateness in light of system goals.</td></tr>
<tr><td colspan="2">Measurement can be :</td><td colspan="2">Measurement can be :</td></tr>
<tr><td>Ongoing</td><td>Periodic</td><td>Ongoing</td><td>Periodic</td></tr>
<tr><td>For example, extent of efficiency measured using day to day monitoring of expenditure, numbers of participants, levels of staffing, attrition, etc.</td><td>For example, nature of the overall management through a study of the analysis and interpretation of this ongoing data including dissemination of the findings.</td><td>For example, extent of effectiveness and magnitude of outcomes (outputs) being produced. Usually internal, TQM, etc.</td><td>Nature of overall and continuing need for the outcome(s) being produced, for example. Usually external.</td></tr>
</table>

DATA AS CONCRETE AND USEFUL

> Data is defined both as (1) things certainly known and from which conclusions may be drawn; and, (2) information prepared for and operated on a computer program.

How do we claim that things are "certainly known" so that we can draw conclusions from them? Whenever you operationally define and measure something, you must be concerned with validity and reliability.

Consider these two concepts:

- Validity: accuracy of the measurement process. How well a measuring instrument (survey, focus group question, checklist, etc.) measures the phenomenon under investigation. Major classes of validity include *construct validity, content validity, and criterion validity.*

- Reliability: ability to obtain the same value with repeated measures. The extent to which a measuring instrument (survey, focus group questions, checklist, etc.) would give the same value if used over and over providing the attribute measured did not change.

Examples:

Discuss each of these in terms of their Validity and Reliability.

- The state trooper did not have a radar printout, but testified that he judged the speed of the vehicle by the rate at which it passed the mileposts.

- The present IQ test is based on the Army Alpha Test developed just after World War I and bases intelligence scales on levels of sophistication of vocabulary.

- Driving tests are performance tests to measure driving proficiency. When I lived in Malaysia the test consisted entirely of reversing the car through traffic cones set up in a parking lot, in England I had to back the car around a blind corner into traffic, and in New York I failed the first time I took my test because I could not parallel park in one smooth operation.

DATA ANALYSIS AND INTERPRETATION

If your data is valid and reliable, what does it show you?

> Remember: All research is a search for pattern, and it is the pattern that you understand, not the data.

Problems arise when you jump to pattern identification conclusions and think you have understanding when you don't. What causes this jumping to conclusions?

⇒ Widely held assumptions, that may be false or flawed;

⇒ Collective wisdom about the source of the information - knowing what clients want without asking them or what they need without analyzing their concerns because you "know how they think"; and,

⇒ Practices that have become second nature and are repeated without question - we question everything else rather than deconstructing the practice.

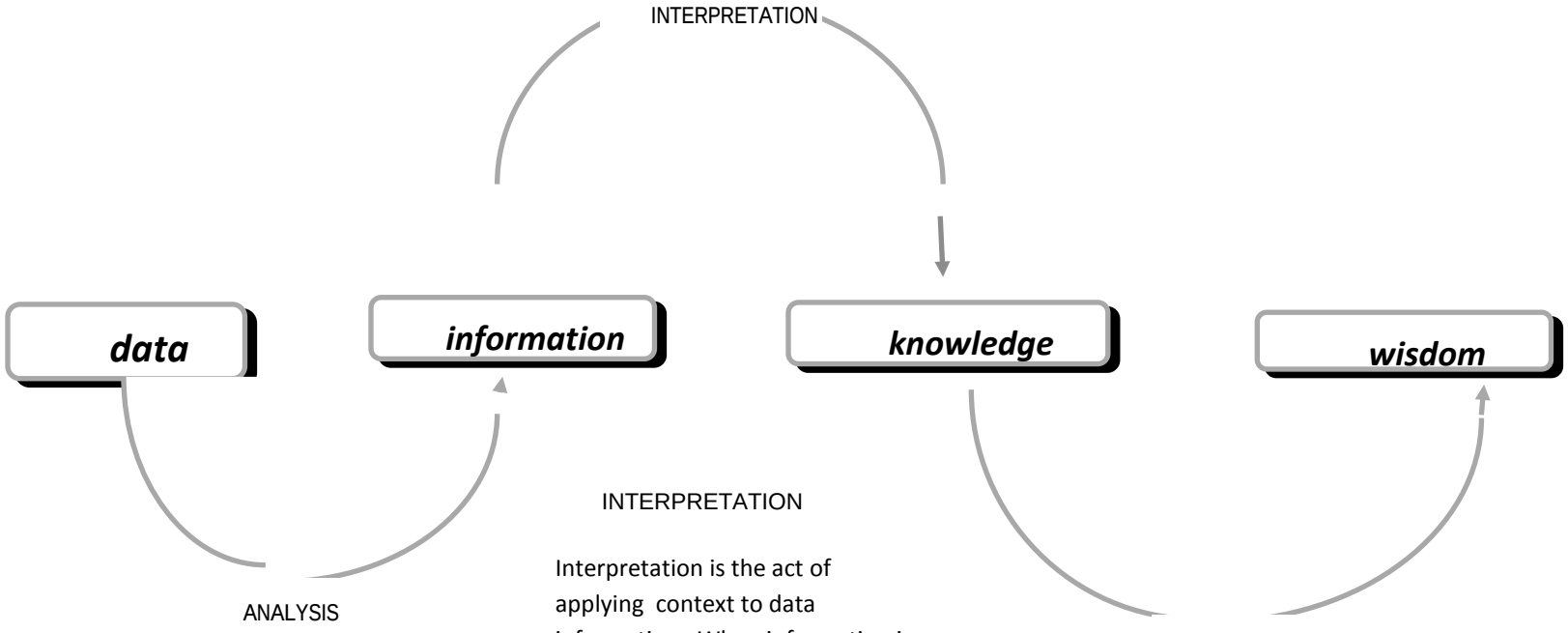
SEARCHING FOR PATTERNS

> One of the biggest mistakes made by those new to action oriented measurement of processes is that they collect way too much data and synthesize it into too little information.

Consider the following model for moving from **Data** to **Information** and from Information to **Knowledge**.

Data Analysis, Interpretation and Application of Information Model for Planners and Decision Makers

(Dr. K.T. Toms, Research Works, Inc., NY. Based on Harlan Cleveland: The Knowledge Executive)



INTERPRETATION

Interpretation is the act of applying context to data information. When information is interpreted by experienced professionals, knowledge-based decisions at all levels should follow. Ideally, interpretation of information is done through meetings, shared management and other open communication strategies thus inviting and taking into account multiple perspectives.

ANALYSIS

Data in and of itself is not useful. Analysis is the process whereby data becomes information. Quality data analyzed appropriately should provide quality information.

REFLECTION

Reflection, or the thinking through of the implications of knowledge based on experience, allows the system to improve. Issues of coordination of data collection rest here.

Data is important, but in and by itself, it is not useful.

Information gathering is at the heart of understanding observed phenomena, and actionable information is the keystone to efficient and effective management. Action research uses a wide variety of techniques that provide the *observations* that when analyzed yield actionable information. These techniques include questionnaire surveys, telephone surveys, content analysis of documents, direct observations of activities of interest, etc. This also includes observations collected for other reasons, such as attendance records, discipline records, testing records, report card information, and demographic information (gender, age, race, family structure, etc.). Observations collected through such processes are collectively called ***data***.

Data consists of measurements on the characteristics of interest. These measurements of the characteristics of interest are related, at least by the fact that you are interested in all of them at the same time, but that relationship is not overt.

Interpretation Leading to Information. You can't use a set of unrelated measurements (data) to inform decisions. In order to use data for effective management it has to be translated into information. This is done through a two stage process of analysis and interpretation. While analysis can be a lone occupation, interpretation of the analyzed data never has been.

> Knowledgeable decision-making is hard and years of research around decision theory have taught us at least that decisions in organizations are not purely rational, nor probably should they be. But you can use the information that you have to keep the deliberations prior to a management decision as clear and focused as possible.

HOW DOES ANYBODY KNOW ANYTHING?

> "Knowledge about the world is best obtained by carefully looking at the world not by looking at someone else's idea of the world."

RECOMMENDATION:

TO AS GREAT AN EXTENT AS YOU CAN, EMULATE THE SCIENTIFIC RELIANCE ON DEFENSIBLE EVIDENCE.

The remainder of this Guide is designed to help you to answer two questions:

What constitutes evidence?

How do we make sure that evidence is defensible?

SECTION ONE: WHAT CONSTITUTES EVIDENCE?

Things that we take for granted, but which are often *not proven* are called 'assumptions'.

You should not base decisions only on assumptions.

THE THREE COMMON WAYS OF KNOWING

There are three common beliefs about how things are known, that is, things on which we base our beliefs:

- Common Sense
- Advice of Colleagues
- Your Own Research

Following is a look at each of these as they pertain to the following situation.

SCENARIO

The ***I Know It All Program*** has had some problems in achieving their goals for client outreach. According to their catchment area information, the program should have 25% more information seeking contacts than they have been logging. At the moment, telephones are answered by all counseling staff, whoever is near a phone when one of the lines rings picks it up and deals with the caller and completes a log sheet. The manager suspects that some callers are not being dealt with properly, while others may not have their calls answered, and everyone knows that the logging system is cumbersome and less than efficient. Should the program (1) hire a receptionist to answer the phone; or (2) put in an electronic phone answering system.

Using the Three Common Ways of Knowing Our decision information might look like this.

Common Sense:

Working this out using common sense would lead to the following: A receptionist would be good because of the personal contact, and the ability of a human to deal with specific problems and direct calls appropriately. A receptionist would be a bad idea because a person can be ill, away from their desk, and cannot be at work all the time so that others in the office will lose their sense of shared responsibility about answering callers.

Applying common sense, complete these two statements.

A phone system would be good because _____

A phone system would be bad because _____

Problem with Common Sense: All alternatives seem equally viable.

Advice of Colleagues:

You can also try to make a reasoned decision based on the advice of colleagues whom you feel are expert in the area under consideration. These can be in-house or external experts. In-house often carries the same problem as using common sense, you are just using someone else's common sense. Asking experts from other programs can also bring you into a number of dilemmas. First, external experts can base their recommendations on what has worked for them in the past, disregarding your needs. Second, they have preferences that you may not share rapidity of answering and directing calls, for example, rather than caring contact with clients. Third, they may deal with a client population which is very different from yours, therefore necessitating different practices.

Problem with Expert Advice: What makes you an expert?

Below are two takes on the definition of expert knowledge and experts themselves. Our purpose is to caution you about accepting expert advice from those who are no more expert than you are. We do not mean to say there are no experts on whom you can call; indeed there are many knowledgeable and ethical experts to draw on. Please just consider that it is not always necessary to see defensible information from outside of your context.

> Spike Milligan on experts, "People are learning more and more about less and less until soon we will know absolutely everything about nothing at all."

> Scientific definition of expert: "X" equals an unknown quantity, and "spurt" is a drip, under pressure.

Collect Your Own Information

So how do you know who to ask? Do you ask the people who are getting through? Do you ask the people in the community that you think should be trying to get through but aren't? How do you find them? In what format do you ask them for information and believe that they will provide it?

Problem with Your Own Research: Error is inbuilt in the form of *bias* and *noise*.

<u>Every</u> method of knowing about the world is subject to some sort of error.

Our questions are:

What are the conditions necessary for us to consider something true?

Therefore, what are the characteristics of accurate information?

THE REMAINDER OF THIS GUIDE ADDRESSES THESE TWO QUESTIONS IN DETAIL.

SECTION TWO: HOW IS INFORMATION RENDERED DEFENSIBLE?

It is helpful to consider the criteria used by scientists to plan, control and carry out their investigations.

This is so because these are incorporated into Total Quality methods through:

- A reliance on evidence,
- A need for controlling error, and
- A standard of publicness.

These three methods should be used together; none of them can be used alone. Evidence should be balanced, that is, as unbiased as possible. Therefore, when collecting data you should be careful to collect it from a cross-section of sources. Controlling error also means looking carefully at the manner in which the data is collected, the collection mechanisms used, and the data elements collected. And finally, anything that you produce as defensible information should be able to withstand public scrutiny, even if that scrutiny has the potential to create challenges to your process.

RELIANCE ON EVIDENCE

> Experts can be wrong even if they all agree.
>
> *Bertrand Russell*

The reliance on evidence is not fool proof. In all of this, you have to be willing to accept that, at some point in the future, your interpretation of your error free data and its resulting publically supported conclusions and actions might be found to be wanting.

Here are some examples of widely held, expert opinion which has proven to be incorrect. Can you think of one more?

* The four food groups for balanced nutrition from the 1950's.
* The collective wisdom on the cause of ulcers.
* Quantum physics and all previous laws of physics.
* Plate tectonics.
* Assumption that the world is flat.
* _____.

NEED FOR CONTROLLING ERROR

Two types of error in any information gathering exercise: systematic error and random error, also known as **Bias** and **Noise**.

> The removal or reduction of sources of potential error is the responsibility of the action researcher.

Bias: is systematic, always in the same direction; always has a source. **This makes data hard to analyze**.

Sources of Bias by internal criteria: (1) your expectations, (2) the people you are studying, (3) your data plan.

1. Your expectations: because you define the problem in your terms you build in your own baggage. Baggage comes from your background, your experience, your deeply held beliefs or who you are (gender; race; age; religion, etc.)

2. Behavior of those you study: there are broadly four areas of this.

    a. The subjects don't want to be judged badly by the research: this can result in subject resistance to anyone studying their practice or behavior or opinion (non-return of surveys, stilted interviews, etc.).

    b. The subjects want to do what is expected: this at least indicates that the subjects know what is expected from them, although it can mislead as to whether they are following these practices.

    c. The subjects want to psyche out the study: most often because the study is viewed as a threat to some interest group or to the individual themselves.

d. The Hawthorne Effect is so named for the factory belonging to Western Electric that was the site of a major managerial study. In the study two groups of workers were given differing work conditions so that the researchers could study the effects of work condition on productivity. The group given better conditions had increased productivity, and the more the conditions were improved, the higher productivity went. However, as the work conditions of the other group were worsened their productivity levels went up!!! The conclusion, attention alone can increase productivity and managers should be wary of the Hawthorne Effect.

3. Your data plan is a tricky source of bias. As mentioned earlier, your data plan has to honestly reflect the subject that is the focus of your action research and include an honest cross-section of the data necessary for you to ***understand*** that subject.

Noise: is haphazard and random. **This makes information hard to interpret.**

Sources of Noise - Just about anything.

Characteristics of Noise:

1. Noise is pervasive and inevitable. Shannon and Weaver's model developed in Bell Laboratories in the late 1940's still holds: in order to maintain **the message** noise has to be reduced and allowed for. Remember: Noise can never be completely eradicated.

2. Because noise is unsystematic, it is "un-biasing". That means that left to itself noise will cancel itself out. In fact, noise if often introduced into research designs to reduce bias.

Conclusion: Mix the biases and live with the noise.

THEREFORE

1. **VIEW NOISY RESULTS AS <u>ESTIMATES</u>**

2. **LEARN TO SEE THROUGH THE NOISE BY BEING AWARE OF THE SOURCES OF NOISE:**

    a. **DIFFERENCES AMONG PEOPLE;**

    b. **FUZZY CRITERIA – IT IS IMPORTANT THAT STANDARDS BE SPECIFIC AND APPLIED CONSISTENTLY; AND,**

    c. **MIXING DATA BY SOURCE OR TYPE TO REDUCE BIAS.**

PUBLICNESS

Findings, at least in summary form, have to be included in any report.

Sometimes the results of a study are embarrassing, politically sensitive, or difficult to adjust professionally for individuals or groups of individuals. There are some things that actually make this easier:

1. Create a culture of data-based information in which all stakeholders have trust. This can only be done if the use of data-based information as actionable is balanced, and recognizable as undertaken for the good of the entire organization;

2. Ensure all employees are effective interpreters of data-based information, that understand the sources of this information, and the ethical application of it to organizational decisions; and,

3. Develop and support a shared decision-making culture wherein this use of actionable information is reassuring.

SECTION THREE:
TRUTH WITH A PINCH OF SALT

Actionable information is a close to truth as you can get it. But remember, it is impossible to be absolutely sure that you know everything you need to know – because you cannot completely wipe our bias and noise from your data, and because of our human nature.

> You always have to express and consider the chances that what you think the information is telling you is inaccurate or wrong.
>
> We are used to seeing this and taking in this sort of information: 95% chance of flood; polls reported 'within the margin of error'; number of wedding guests expected between 95 and 100, etc.

The analysis of your data that you do, or your evaluator/researcher does, provides measures of accuracy (standard deviation, margin of error, chi square statistic, etc.).

Here we point out a few things to keep in mind as you are designing your study, looking over the data and information from your study, and taking action using this information.

REMEMBER: EXPECTATIONS REACT WITH WHAT YOU ARE MEASURING

OBSERVATION: SEEING IS NOT BELIEVING

Observation in an action research study is not as easy as it seems to be in everyday life. Take this as an example:

> On Saturday, September 28, 1996 the New York Yankees defeated the Boston Red Sox. This was: □ a tragedy □ great □ of no interest to a civilized person □ who?

Point is: TRUTH is in the eye of the beholder.

To make sure you have vision as clear as possible – pay attention to measurement: both the things you use to measure and the clarity of the information that the measure provides.

⇒ The things you use to measure are referred to as *instruments.* Instruments include, but are not limited to:

- Surveys or questionnaires: written as paper and pencil or electronic sets of questions posed to identified informants. All members of an informant group are given the same questions, but questions on the same topic can be modified for different informant groups (say parents and teachers, for example).
- Interview Protocols: sets of questions for informant groups to be asked directly either in person or by telephone (newer ones also operate online in real time). These can be general category questions with specific questions left to the interviewer, or 'scripted' interviews that are more verbal surveys.

        - Group interviews: for people who know each other or focus groups for people who do not. Protocols necessary.

    - Observation Protocols: when observation of activities or practice is required, a set observation focus is necessary to maintain cross informant data. Many observation protocols are available so the one that fits your purpose most closely can be chosen and then modified to fit perfectly.

    - Rubrics: are sets of expectations of the quality of a task completion against which the task can then be measured. Rubrics are used in a number of instances, for example, to produce a quality assessment scale for a practice or activity; when doing a document review either for a single document type or for multiple document types to record levels of information against expected; etc.

⇒ If you use data collected for another purpose or by someone else, it is called use of secondary data. Typical data of this type are things such as: student assessment data, attendance data, discipline data, report card data, demographic data, etc. You are then limited to the format of the data as it was originally collected.

Beware of reaching conclusions based on a single measure. In action research the carpenters' rule holds, only here it is:

Measure twice – conclude once.

That repeat measurement may be done using the same tool (or instrument) or it can be done using a different instrument.

So, what we are looking at is really not that complicated. Remember, back on page 2 we asked you to consider these two concepts:

- Validity: the accuracy of the measurement process. How well a measuring instrument (survey, focus group question, checklist, etc.) measures the phenomenon under investigation. Major classes of validity include *construct validity, content validity, and criterion validity.*

- Reliability: the ability to obtain the same value with repeated measures. The extent to which a measuring instrument (survey, focus group questions, checklist, etc.) would give the same value if used over and over providing the attribute being measured did not change.

The practice of measuring the same thing in multiple ways or multiple things using the same instrument type is called ***triangulation*** (taken from orienteering where you take multiple readings off a single datum point to "orient" yourself - so you won't get lost in the woods).

And always remember that people remain people, despite the data, instruments, analysis and interpretation.

Because you interpret what you see, you often see what you expect:

- There is a folk story from India popular with the authors of books on research methods. The story tells of three blind men being told to feel different parts of an elephant - the tail, the trunk and a foot. Each man described what he felt and the animal he believed he was touching, and each gave a description of a very different animal.

  Remember: Action research only tells you about what you are studying, and tells you nothing about anything else.

- You can, however, learn to see, reducing the effect of your prejudice if you face the fact that everyone is prejudiced in some way or another.

- Active social science research assumes that "normal" is contextual, meaning that phenomena must be studied in its natural context.

- And beware of a positive orientation, which means you see only positive things, often called "the halo effect". Remember the kid at school who never got into trouble no matter how awful they were, or the person a work that seems to be unable to do anything wrong?

YOU CANNOT SEE EVERYTHING

Strangely enough, one of the major problems with people new to action research/evaluation and the use of evaluative information is their tendency to collect way too much data and therefore produce way too little information.

You do not have to measure everyone, you do not have to ask everyone's opinion, and you do not have to seek input from each individual person.

You do have to measure the perspective of each relevant stakeholder group, ask for opinions in non-threatening ways, and seek input that is representative of each individual person.

THUS, YOU SAMPLE

For a survey, a reasonable sample is 20% of the population. You may also stratify the sample to ensure that it is representative. To do this you simply compute the % of the population that has one or more characteristics that you think will influence their perspective (gender, for example, is often one) and maintain those percents as you "pull" the 20% sample. Thus, if you have 100 participants and 10% of your participants are female and 90% are male, the sample would be 20 people, two women and eighteen men.

YOU PRODUCE SOME OF WHAT YOU SEE.

Computer nerds have a great saying:

GIGO Garbage In, Garbage Out

In action research you have to ensure that the instruments that you use are appropriate, have the correct scales and will be responded to. Sounds pretty simple to do.

Rule: Use good instruments when **GOOD means:**

⇒ An appropriate type (don't plan telephone interviews of the homeless, for example).

⇒ That use the correct scales (if you want to know 'yes' or 'no' don't give them a five point choice).

⇒ That will be completed and returned (the best information in someone's head won't help you).

You can ask subjects things directly using interview, questionnaires and comment (sometimes called suggestion) cards. You can gather information indirectly using "non-reactive" measures such as traffic counts, document review (files, case files, memos, letters, etc.), and context observation (how busy is the waiting room, how clean are the bathrooms, do people bring their children or leave them home). Test information is another indirect measure, usually considered "reactive".

Researchers use various techniques for collecting data which will be analyzed using other techniques to answer specific research questions. There is no such thing as a 'universal analysis technique'. You can ruin a perfectly good study by analyzing the data it generates with the wrong analysis technique.

USING QUESTIONS TO GET THE INFORMATION YOU NEED

Most of what you are doing in Action Research is asking questions. Be careful, asking the question restricts what you can find out, so you have to be careful about how you ask the questions.

In writing on questionnaires and reaction cards and verbally in interviews (individual or group), questioning is the most common way in which we try to elicit information. In English, there are five ways to ask questions. The way you ask the question restricts the way that people will answer you.

In English, there are **Five English Question Forms:**

1. Inflection Questions: making a statement and raising my voice at the end, inflection questions obviously only work verbally. Closed question: answer = yes or no.

   Your name is Bob?

   You broke the vase?

2. Inversion Questions: moves the verb to the front of the sentence if it is a form of 'to be' and inserts the 'empty do form' for any other verb. Closed question: answer = yes or no.

   Is your name Bob?

   Did you break the vase?

3. Specific Questions: using question words (Who, What, Where, When) to seek specific information. Elicits specific information with greater accuracy than yes/no questioning.

   What is your name?

   What did you break? (if I heard a crash), or

   What did you do? (if you look guilty and I'm not sure why)

4. Open Questions: using question words which elicit cause effect or reflective responses. Open questions, seeking explanation and reflection

   Why won't you tell me your name?

   Why did you think balancing the vase on your head was a good idea?

5. Tag Questions: use informative sentence form with negative tag indicating question. (nearly impossible for ESL clients to understand).

   Your name is John, isn't it?

Yes answers align with the sentence stem (Your name is John?), no answers align with the tag (isn't it?). Thus, Yes, my name is John or No, my name isn't John.

You broke the vase didn't you? – other than forms of 'to be' again use the 'empty do' verb forms, this time in the tag.

Tag Questions are especially confusing (even to native English speakers) if you negate within the stem "Your name isn't John, is it?" Even though the rule still applies: i.e., Yes, my name isn't John or No, my name is John. Trouble is double negatives!

These are important to remember because:

⇒ When you review the design of data instruments, you should have some idea of whether the questions you are asking will get you the information you are looking for,

⇒ Group interviews, focus groups, and individual interviews are time consuming and expensive data collection techniques if the information is unclear or misleading,

⇒ If you do not get a response it may mean that the subject did not understand the question not that s/he has no answer.

SURVEYS & INTERVIEWS: SCALES, CHECKLISTS & OPEN QUESTIONS

There is really not the time here to go into survey research in detail here, but a few points are important enough to mention.

Likert-type scales are very popular and often misused. We call them 'Likert-Type' because the present use of these scales does not match the parameters laid out by their inventor, Rensis Likert.

Likert-type scales offer a choice of one of a series of numbers and label the numbers with response - type vocabulary words (for example, 1= strongly agree to 4=strongly disagree).

Use these scales with caution and remember:

⇒ Label all the options, not just the highest and lowest, or you won't know what the respondents meant.

⇒ The stem should be a statement, not a question and responses should be affective (agreement, comfort) or temporal (a lot, a little, never).

⇒ The 'scale' is not a numeric scale (the difference between strongly agree and agree is not the same as between agree and disagree) so some numeric analyses are not appropriate.

⇒ Researchers differ in the provision of a mid-point that would indicate no real opinion - done by using an odd numbered scale. I favor even numbered scales to force and opinion.

⇒ If you honestly think that "don't know" is a viable answer, or "not applicable", then add to the scale separate from the core scale.

⇒ Analyze using frequency, standard deviation and chi-square.

The **scale** can be purpose built to get the best and clearest information. Levels of understanding of various domains can be elicited by providing a scenario scale. "Mark 3 if you feel you could explain this to someone else and they would understand what to do" "Mark 2 if you feel you can use this but could not explain it to someone else" "Mark 1 if you would like someone to explain this to you again" and so on.

Sliding scales are also popular with each end labeled and respondents asked to mark where they would place themselves on the continuum. Good eyeball info, but has to be analyzed as binary (from midpoint up = 1 and from midpoint down = 2).

For example, put an X on the line where you would put yourself after hearing about the Yankees winning the World Series.

Happy Sad

Checklists work well if you offer the "other (please specify)" option. The lists themselves should be generated using other data you have collected (to triangulate) or isolated through an extensive literature review.

Open-ended questions are good if you want to know Why? or What do you think?, but people will not answer too many or find the time to write an essay. You can provide guidance (for example statements) but should be careful not to lead the witness.

And as always PILOT any instrument

FIRST AND FOREMOST: BE CAREFUL !!!

It is often difficult to establish if Cause ⇉ Effect relationships exist. Once you have established that there is some relationship (called a correlation) two details about the relationship need to be established.

All correlations are not related through cause and effect. For example, shoe size and reading age in elementary students is statistically correlated, but one does not cause the other. Many correlations are important, even those with no cause and effect link.

When looking for a cause – effect relationship, it is often hard to see which is the cause and which the effect, called the 'direction' of the relationship.

In the second, the relationship might be established, but the overall relationship is not guaranteed to be important.

Remember: some causal relationships are spurious and some are ridiculous. Few are real and important.

Three questions as guidelines for determining causal relationships:

⇒ Can you find a weakness in the claim that a relationship exists?

⇒ Can you turn the relationship around, or argue that the variable that is supposed to be the cause did not occur before the other variables?

⇒ Can you identify plausible third variables which might be the cause or part of the cause?

DATA ANALYSIS AND INTERPRETATION

OVERVIEW

This section of the Guide gives you a brief, and I mean brief, overview of the basics of the research part of Action Research. This is not a complete course, it is meant to give you the basic information you need to follow a conversation about a study, the data it is supposed to produce, and what the study director wants to do 'to and with' that data. This discussion is framed around the basic choice you have: qualitative data only, quantitative data only, or a use of both data types in combination.

Choices – qualitative and/or quantitative

Analyzing qualitative data

- Content analysis – extracting themes
- Affinity diagrams
- Controlling for bias

Analyzing quantitative data

- Descriptive statistics
- Inferential statistics

OVERVIEW: QUALITATIVE METHODS

A qualitative description of program, process and experiences both flexibility and depth of information. Depth and detail will depend on the purposes of the study and stakeholder requirements. Regardless of the depth and detail, however, qualitative methods will help you to understand the context of the situation and to provide you with insight into main actor perceptions.

Qualitative Data Analysis

Inductive analysis: Going from specific information to general categories

In inductive analysis the study team review all the material collected in order to identify the patterns, themes and categories. Qualitative researchers often refer to the fact that in qualitative analysis the data speaks: that is, the patterns and themes emerge from the data.

Semi-inductive analysis: Basically is 'switching horses mid-stream', transitioning from pure inductive analysis and using initial findings to frame a deductive completion.

In semi-inductive analysis, the study team reviews some of the material collected and identifies the patterns and themes they find in it. Then they use these patterns as they review remaining material in terms of those patterns and themes.

Characteristics of Qualitative Methods

1. Study team is part of the study process
2. Data collection, data analysis and interpretation overlap and are interdependent
3. The study itself evolves as new questions emerge

4. The study design is flexible

5. And results provide an up close and personal view point

Qualitative Data Collection

Use qualitative techniques with the following data sources:

⇒ Narrative documents

⇒ Open-ended interviews

⇒ Responses to open-ended questions on questionnaires

⇒ Focus groups

⇒ Diaries

⇒ Observations

Using Case Analysis as a Analytic Tool

Cases can be specific parts of the program, groups, critical events, or people.

When Case is at the Program Level: use all the records, interviews with staff and participants, observation of the program in action.

When the Case is at the Individual Level: use specific informant interviews, observations, diaries, and narrative records.

Compile Case Records

Keep records to ensure you have all the information for each case

Review during data collection to ensure they are complete, synthesized, and that all parts fit together

Each case can then be compared, which is the point.

Triangulation

You can use triangulation when you have any or all of the following:

- Multiple sources of the same information; and/or
- Multiple sources of different information; and/or
- You are using different data collectors, i.e., different interviewers or observers

If the same findings are ascertained through different data collection methods, it adds more credibility to the results

Triangulation allows you to combine the depth of one method with the breadth of another method. For example, drawing corroborating information from interviews, focus groups and questionnaires, or from questionnaires, available data, and expert panels. Another typical combination is of observations, program records, interviews or interviews, diaries, and available data.

Early Steps

While collecting data:

1. Keep good records
2. Write up interviews, impressions, notes from focus groups
3. Make constant comparisons as you progress
4. Meet with the team regularly to compare notes and make adjustments

Keeping Records: The Contact Summary Report

This is a one page summary after each major interview or focus group. It includes:

- Main issues
- Major information obtained
- What was most interesting, illuminating or important?
- What new questions need to be explored?

Keeping Records: The Subjectivity File

- Your own reactions during the study
- Your feelings, hunches and reactions

Keeping Records: A File of ideas

- Your ideas that emerge as your proceed
- Title and sub-titles of report

Keeping Records: The Quotation File

A record of quotable quotes

Data Analysis: Get Ready

Review study questions, purpose, and focus:

- Purposes and issues of the study
- Stakeholders' concerns

Review insights developed during data collection by reviewing summaries and files

Make sure all the information is in one place.

- Is all the information complete?
- Are there any holes that need to be filled?

Make copies and place originals in a central file for both electronic and print materials.

Use copies to write or make notes on, cut and paste as needed, etc.

Developing Categories

Recurrent themes, ideas, words, phrases

- Do they fit together?
- Are there differences?

Categories should be large enough to capture a range of views but not so large as to be meaningless. Categories should be distinct from each other or the analysis becomes too confusing.

Coding Data

Develop coding scheme

1. Train coders
2. Do pre-test with small sample of qualitative data
3. Check inter-rater reliability.
4. If there are problems, fix them.
5. Pre-test again.

Analyzing Data

Bringing order to the data

- Data may be placed in sortable conditions as available – this includes in print (on index cards, for example) or data may be placed in a spread sheet
- Computer software is available to analyze qualitative data – RWI uses NVIVO. (The list changes often, an 'as of today' list is included here as Appendix III).

Your purpose is to sort data to reveal patterns and themes

Interpreting the Data

Making sense of the data, finding meaning and significance

Linking themes and categories to processes of the program and/or to the outcomes.

– Are some themes more prevalent when discussing process issues?

– Are some themes more prevalent when discussing outcome issues?

– Look for rival explanations and alternative ways of understanding the data.

Iterative Dialogue:

– Share information early and often with key informants.

– Have others review early draft with the intention of eliciting information, questions, other ways of interpreting the data, other possible sources of data.

– Meet with stakeholders with preliminary findings to ensure the analysis is on target.

Affinity Diagrams

Affinity Diagrams are a group process for analyzing qualitative data: The process includes members of the study team and relevant stakeholders, if appropriate.

The process:

– Begins in silence

– Participants write down ideas and themes

– These are posted on a wall or other position to be seen

– The group then works to sort these into similar categories

The group then discusses the result, once basic categories are identified.

– Usually this results in a need to refine the original category sort.

Categories become the framework for the report

Affinity Diagrams Advantages

– Allows everyone to say what they want.

– All ideas are equal.

– Limits dominance by an individual.

– Fast way to generate many ideas.

Content Analysis

Used to analyze written or spoken words or visual images.

It is a way to convert qualitative information into quantitative information.

Possible Uses of Content Analysis

– Review of narrative portions of applications for work or school.

– Analysis of open-ended responses to questionnaires.

– Analysis of focus group transcripts.

Content Analysis: Steps

1. Select the material to be analyzed.
2. Key issue is to develop the coding categories.
    a. These are based on your study questions and measures.
    b. Categories must be exhaustive and unambiguous.
3. Develop decision rules for coding.
4. Essential to success:
    a. Pre-testing the analysis to ensure that the data is where you think it should be.
    b. Training all team members in what is expected, and how to categorize the data correctly.
5. Teams of at least two people review and categorize the material – inter-rater reliability is critical.
    a. Their ratings are compared. If they are different their process is reviewed and corrected.

It is important to minimize content analysis challenges. Reliability is essential

Report Writing

Report major themes:

- "Three major themes emerged from the data."
- Do not report numbers or percent.
- "Some participants said…."
- "Others said…."

Highlight interesting perspectives even if only said by one or two people.

No cookbook here.

Stay focused. With so much data, it is easy to get lost.

Ask yourself:

– Does this information answer the study questions?

– Does this information add value to the stakeholders?

– If not, leave it on the cutting room floor.

COMMON QUANTITATIVE ANALYSES

Types of Qualitative and Quantitative Variables

Data differs by how its constituent data elements are constructed. This is because the analysis techniques used with quantitative data have been designed for use with particular data element types. A variable is a characteristic that can vary in value among subjects in a sample or population.

Qualitative Variables

For qualitative variables, distinct categories differ in quality, not quantity or magnitude. The set of categories for a qualitative variable is called a ***nominal scale***. Here one answer does not represent a greater value than another, such as in questions of religion, gender, or country where you were born. **Examples: categories such as gender, religion, country of origin, etc.**

A Little Bit Qualitative – A Little Bit Quantitative

Between the qualitative variable type and the quantitative variable type is the ***ordinal scale.*** This scale consists of categorical scales which have a natural ordering of values, but undefined interval distances between the values, such as information on social class, political philosophy, or answers to questions on opinion. These scales are not nominal because their categories are naturally ordered. **Examples: order but not numerical such as opinion scales, or age and income categories, etc.**

> Both the nominal and ordinal scales consist of a set of categories. Each observation falls into one, and only one, category. Variables having categorical scales are called ***categorical variables.***

Quantitative Variables

When the possible values of a variable do differ in magnitude, the variable is called quantitative. These variables have a numerical scale, called an ***interval scale***. Interval scales have a specific numerical distance or 'interval' between each pair of levels. **Examples: real numbers such as age, income, weight, height, etc.**

Careful: In many respects, ordinal scales more closely resemble interval scales and should be analyzed using quantitative methods. This can require a recoding of the data in question, but the benefits in the variety of methods available for data analysis, particularly for data sets with many variables, can make the extra data cleaning worth it.

Or Call Variables Discrete or Continuous

Discrete variables (also called categorical variables) differ in kind rather than in degree, amount or quantity. Examples are eye color, religion, occupation, and most kinds of treatments or educational methods. They are discrete because the categories are fixed and variables fall into one category only.

Continuous variables (also called measured variables) vary in degree, level, or quantity, rather than in categories. Differences in *degree* results in 'rank ordering' and then in placement on a continuum (putting people, for example, in some rank-ordered sequence, without identifying how far apart the rankings are). Putting in an interval of measurement imposes the difference in ranks, and lifts the level of measurement from ordinal to interval. Thus the variable is now continuous – we have a continuum, with intervals, showing less and more of the characteristic.

Quantitative Data Analysis

Quantitative data analysis uses set procedures for data collection, coding and analysis. Quantitative data is often incorrectly associated with numeric data only. In many cases, data collected in what is called 'alpha' format (i.e., data from either written or spoken language) can be systematically transposed into numeric format and thus analyzed using quantitative procedures and techniques.

Quantitative methods do rely on predictor and effect variables, can only operate on numeric data (either originally numeric or transposed for alpha sources), and generally procedures of statistical analysis. That analysis is prescriptive, meaning that the choice of statistical analysis approach used is restricted by the manner in which the data is formatted during collection.

This section of the Guide will use these topics to cover quantitative analysis:

- The concept of *frequency* and its use in both count-based and percent distributions
- Knowing the distribution - measures of central tendency of a distribution: Means, medians, and modes
- Dispersion as Standard Deviation or Variance
- Comparison of means as an analysis
- Cross-tab analysis as the Contingency Table

Count and Percent Frequency Distributions

Frequency is the rate at which something occurs or is repeated over a particular period of time or in a given sample. In Action Research studies, it is often helpful to collect and report on information as a comparison. For example number of program participants overall is a good indicator of basic program success, but the same information by gender, by grade level, or by days of the week might be more informative for planning and management decisions.

A simple frequency distribution is a set of the same measured elements arranged in order of magnitude along the *x-axis* and the frequency (as a count) of each measure is represented along the *y-axis,* such as test scores.

If the measured elements are further analyzed into groupings that make up a portion of the total number of elements calculated as a percent of the total, this is a percent distribution. Frequency distributions can tell, at a glance, something about the shape of the distribution, and this can be important in determining subsequent steps in the analysis. They also help the action research/study team to stay close to the data, important as there is a great benefit to getting a 'hands-on feel of the data'.

For example, how many male and how many female students are in the program?

Distribution of participants by gender:

	Male	**Female**	**Total**
Count	100	200	300
Percent	33%	67%	100%

Measures of Central Tendency

Known as 'the 3-Ms', the mode, the median and the mean are the three most commonly used measures of central tendency.

- The Mean is the arithmetic average of all data in the distribution, and the measure of central tendency with which most people are the most familiar. The mean is the point in a distribution about which the sum of the squared deviations is at a minimum. The second is that the mean is a very effective statistic where scores within a distribution do not vary too much, but it is not so effective when there is wide variance in the data set. Of the three measures of central tendency the mean is the only one that does not necessarily appear in the distribution.

- The Median value in a set of numbers is that value which divides the set into equal halves when all the numbers have been ordered from lowest to highest. The median is particularly appropriate when the distribution of numbers is skewed, unlike the mean.

- The Mode is the most frequently occurring number in a set of scores. It is possible to have two numbers each with the same frequency in a data set. A distribution with two modes is called 'bimodal'. In cases of the use of Likert-type scales, the mode of response (the most often chosen) is a better choice than the mean, for example.

Which One to Use?

Depends on the type of data that you have, but as a rule of thumb:

- ❖ Nominal Data: mode
- ❖ Ordinal Data: mode and median
- ❖ Interval: mode, median and mean – mean if the distribution is normal, median if it is not.

Variation: Standard Deviation and Variance

The standard deviation is the most common measure of variability. These are distances of individual measurements from the mean of the distribution. They are then standardized to give us the standard deviation. In one number it summarizes the variability of a set of data. The more spread out the scores, the larger the standard deviation.

The variance is the square of the standard deviation (or the standard deviation is the square root of the variance).

While standard deviation is most often used in descriptive statistics, the variance is more commonly used in statistical inference, you can still calculate one from the other.

Working with Three Types of Variable

Description versus Explanation

Description and explanation represent two levels of understanding.

⇒ The **Descriptive Study** is undertaken to **<u>understand</u>** what happened, or how things are proceeding, or what something or someone is like. This has a more restricted purpose than explanation as you can describe without explaining. A Descriptive Study focuses on what is the case. Description can be a first step to explanation.

⇒ The **Explanatory Study** is undertaken to **account for** what happened, or for how things are proceeding, or for what something or someone is like. It involves finding the reasons for things, events and situations, showing why and how they have come to be what they are. Explanatory studies focus on why something is the case.

Knowing why or how something happens, we are in the position to predict what will happen, and perhaps be able to control, or at the very least influence, what will happen in the future.

Independent, Control, and Dependent Variables

Independent Variable: Variable which you believe explains a change in the dependent variable, the action or effect producing variable.

Control Variable: a variable whose effect we want to remove or control because we suspect that it might confound comparisons we want to make or relationships we want to study. The term is to 'remove its effects' or 'partial it out' of the analysis.

Dependent Variable: Variable you want to explain, the outcome, effect, or result.

Crosstabs aka Contingency Tables

A contingency table is a two-way table that is a useful tool for examining relationships between categorical variables. This is most commonly used as a descriptive tool. The entries in the cells of a two way table can be frequency counts or relative frequencies.

For example:

Gender	Dance	Sports	TV	Total
Men	2	10	8	**20**
Women	16	6	8	**30**
Total	18	16	16	**50**

Reporting results of survey that asked 20 women and 30 men what is their favorite after work activity.

Used when working with nominal and ordinal data, also can be used with interval/ratio data that has been categorized.

Inferential Statistics

Welcome to the world of inferential statistics!

Inferential statistics are used to analyze data from randomly selected samples.

But there is the risk of error because your sample might be different from the population as a whole.

To make an inference, you first need to estimate the probability of that error.

Statistical Significance

Statisticians have provided us with the tools to estimate how likely our results are in error.

These are called tests of statistical significance.

They allow you to estimate how likely it is that you have gotten the results you see in your analysis of sample data as a result of chance.

It has a benchmark of 5% – .05 Alpha level or 'P value'

It means we are 95 certain that our sample results are not due to chance or the results are statistically significant at the .05 level.

Hypothesis Testing

The **research hypothesis** is your best guess as to the relationship between variables.

The Hypothesis is always expressed as a positive assertion: There is a difference between the per capita incomes of men and women in South Africa.

The **null hypothesis** is always a statement that "there is no difference" or "no impact" between our variables: There is no difference between the per capita incomes of men and women in South Africa.

Testing for statistical significance:

What is the probability of getting a Rand 4,600 difference if we assume there is a no difference in the population from which this sample was drawn?

If the probability is “small” for getting a Rand 4,600 difference, then we reject the null hypothesis.

Small is defined here as less than 5%.

Remember:

A significance test is nothing more than an estimate of the probability of getting the results by chance if there really is no difference in the population.

Tests for Statistical Significance

Most Common Tests:

Chi Square: nominal and ordinal data

T-tests: Dependent Variable: ratio data; Independent Variable: 2 categories

ANOVA: Dependent Variable: ratio data; Independent Variable: 3+ categories

F-tests: Interval data

Education & Income

You want to know: Is there a difference in income based on highest degree obtained?

– How do you word the Research Hypothesis?

– How do you word the Null Hypothesis?

APPENDIX I:

Are the following statements about the thinking through of questions and their answers *true* or *false*? (from David Hackett Fisher)

a) A logical argument which is flawed in some respect is therefore flawed in all respects.
- ☐ TRUE
- ☐ FALSE

b) A logical argument which is flawed in some respect, or even in every respect, has therefore substantively false conclusions.
- ☐ TRUE
- ☐ FALSE

c) A flaw in a logical argument is an external sign of its author's depravity.
- ☐ TRUE
- ☐ FALSE

d) Sound thinking is merely thinking which is not misleading or based on error.
- ☐ TRUE
- ☐ FALSE

e) False or mistaken beliefs exist independent of particular purpose and assumptions.
- ☐ TRUE
- ☐ FALSE

APPENDIX II: TESTING

Testing

from The Answer Sheet, a newsletter distributed to ETS's test makers.

"We've a splendid testing system. If you'd like it I shall list 'em,"
Said the city superintendent with a holy little smile.
"We measure kids and test kids to see what things infest kids,
And then repeat the process every little while.

"We give grammar tests and hammer tests and also Katzenjammer tests,
And German tests and vermin tests, the best we can compile,
Appreciation, condensation, information, lucubration,
To say nothing of vocation - Oh, a tall, tall pile.

"Our tests are often mental, but they may be merely dental
Or sometimes environmental (about the domicile).
Versatility and ability, then utility, then debility -
With indefatigability we choose the latest style.

"Constitution, restitution, home pollution, destitution,
Go-to-college, moral knowledge - just wait a little while;
Aptitudes and attitudes but seldom the beatitudes
For measurement of platitudes serves only to beguile.

"Physiology, sociology, entomology, and geology,
For present-day psychology says these things we should compile;
Metaphorical and clerical, historical, hysterical,
Our tests are quite numerical, and very much worthwhile.

"Spelling tests and yelling tests - no, I'm not selling tests,
But schools that seldom use them are very, very vile.
We give our tests, record our tests (I wish we could afford more tests)

AND I KEEP THEM - KEEP THEM - IN A GREAT, LARGE PILE."

APPENDIX III: QUALITATIVE ANALYSIS PROGRAMS

Computer Help

Software to analyze qualitative data:

- NVIVO
- Atlas-Ti
- Dedoose

From the University of Illinois at Urbana Champlain: Listed below are some good examples of high performance CAQDAS platforms that are offered free of charge. Some have graphical user interfaces (GUI) and others do not. Some investigation will be required to determine what will best suite your needs. Note that two of these free programs, Aquad and RQDA, make use of the powerful statisical analysis package R. While there is only a plug-in for Aquad, you will need to install R to use RQDA. R is free to download and supported by a vast user community.

- Aquad: Aquad is a platform developed in Germany (you will have to translate the website, which is also in German) that supports text of any kind, audio, video, and image files. There is a plugin available to use with R, the open source statistical analysis software.

- Coding Analysis Toolkit (CAT): CAT was designed to utilize keystrokes rather than the mouse as a coding assistant. CAT can import an ATLAS.ti project to parse quantitatively, though it has a coding mechanism built into itself as well. CAT won the Best Research Software award from the Information Technology & Politics in the American Political Science Association in 2008. CAT is Web based.

- Compendium: Compendium models itself as an application to visualize the connections between information and ideas. It is very visually based and offers similar functionality as most CAQDAS platforms.

- QDA Miner Lite: A "lite" version of the proprietary software, QDA Miner Lite has a slightly limited functionality. You can still perform the basic functions of the full version but don't have some of the more advanced options. You can see all the differences side by side

comparison: http://provalisresearch.com/products/qualitative-data-analysis-software/freeware/compare-versions/

- RQDA: RQDA is another package from the exciting R Project, a statistical analysis platform that has developed a constantly growing community of users and developers that is designing all sorts of software for use in data analysis. What a plugin to such a power statistics program means is that implementing quantitative data to your project will be seamless and effective.
- Weft QDA: Weft was a project by a student who was fed up with the pricing of many of the proprietary software providers. Weft provides solid coding capabilities of text in addition to full text search using Boolean operators (AND, OR, NOT). There is no visualization capability. Unfortunately, Weft has been abandoned, so there is no support from the developers. Probably best used by researchers only looking

Computer Help

Some use text search

Some use flagging

Some use codes

Varies in features