2009, NUMBER 12
2ND EDITION

PERFORMANCE MONITORING & EVALUATION

TIPS

DATA QUALITY STANDARDS

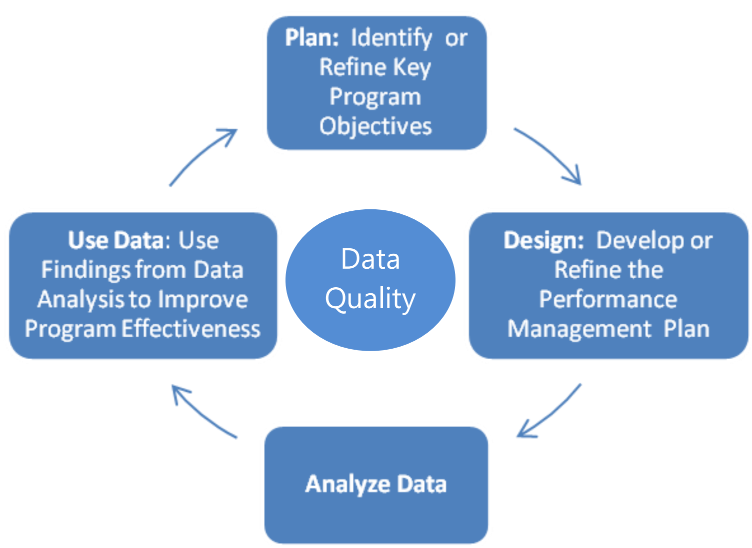
ABOUT TIPS

These TIPS provide practical advice and suggestions to USAID managers on issues related to performance monitoring and evaluation. This publication is a supplemental reference to the Automated Directive System (ADS) Chapter 203.

WHY IS DATA QUALITY IMPORTANT?

Results-focused development programming requires managers to design and implement programs based on evidence. Since data play a central role in establishing effective performance management systems, it is essential to ensure good data quality (see Figure 1). Without this, decision makers do not know whether to have confidence in the data, or worse, could make decisions based on misleading data. Attention to data quality assists in:

Figure 1. Data Quality Plays a Central Role in Developing Effective Performance Management Systems



- Ensuring that limited development resources are used as effectively as possible
- Ensuring that Agency program and budget decisions in Washington and the field are as well

- informed as practically possible
- Meeting the requirements of the Government Performance and Results Act (GPRA)
- Reporting the impact of USAID programs to external stakeholders, including senior management, OMB, the Congress, and the public with confidence

DATA QUALITY STANDARDS

Data quality is one element of a larger interrelated performance management system. Data quality flows from a well designed and logical strategic plan where Assistance Objectives (AOs) and Intermediate Results (IRs) are clearly identified. If a result is poorly defined, it is difficult to identify quality indicators, and further, without quality indicators, the resulting data will often have data quality problems.

One key challenge is to determine what level of data quality is acceptable (or "good enough") for management purposes. It is important to understand that we rarely require the same degree of rigor as needed in research or for laboratory experiments. Standards for data quality must be keyed to our intended use of the data. That is, the level of accuracy, currency, precision, and reliability of performance information should be consistent with the requirements of good management. Determining appropriate or adequate thresholds of indicator and data quality is not an exact science. This task is made even more difficult by the complicated and often data-poor development settings in which USAID operates.

The Five Data Quality Standards

1. Validity
2. Reliability
3. Precision
4. Integrity
5. Timeliness

As with performance indicators, we sometimes have to consider trade-offs, or make informed judgments, when applying the standards for data quality. This is especially true if, as is often the case, USAID relies on others to provide data for indicators. For example, if our only existing source of data for a critical economic growth indicator is the Ministry of Finance, and we know that the Ministry's data collection methods are less than perfect, we may have to weigh the alternatives of relying on less-than-ideal data, having no data at all, or conducting a potentially costly USAID-funded primary data collection effort. In this case, a decision must be made as to whether the Ministry's data would allow the Assistance Objective team to have confidence when assessing program performance or whether they are so flawed as to be useless, or perhaps misleading, in reporting and managing for results. The main point is that managers should not let the ideal drive out the good.

1. VALIDITY

Validity refers to the extent to which a measure actually represents what we intend to measure.[1]

Though simple in principle, validity can be difficult to assess in practice, particularly when measuring social phenomena. For example, how can we measure political power or sustainability? Is the poverty gap a good measure of the extent of a country's poverty? However, even valid indicators have little value, if the data collected do not correctly measure the variable or characteristic encompassed by the indicator. It is quite possible, in other words, to identify valid indicators but to then collect inaccurate, unrepresentative, or incomplete data. In such cases, the quality of the indicator is moot. It would be equally undesirable to collect

[1] This criterion is closely related to "directness" criteria for indicators.

good data for an invalid indicator.

There are a number of ways to organize or present concepts related to data validity. In the USAID context, we focus on three key dimensions of validity that are most often relevant to development programming, including: face validity, attribution, and measurement error.

FACE VALIDITY

Face validity means that an outsider or an expert in the field would agree that the data is a true measure of the result. For data to have high face validity, the data must be true representations of the indicator, and the indicator must be a valid measure of the result. For example:

> **Result:** Increased household income in a target district
>
> **Indicator:** Value of median household income in the target district

In this case, the indicator has a high degree of face validity when compared to the result. That is, an external observer is likely to agree that the data measure the intended objective. On the other hand, consider the following example:

> **Result:** Increased household income in a target district
>
> **Indicator:** Number of houses in the target community with tin roofs

This example does not appear to have a high degree of face validity as a measure of increased income, because it is not immediately clear how tin roofs are related to increased income. The indicator above is a proxy indicator for increased income. Proxy indicators measure results indirectly, and their validity hinges on the assumptions made to relate the indicator to the result. If we assume that 1) household income data are too costly to obtain and 2) research shows that when the poor have increased income, they are likely to spend it on tin roofs, then this indicator could be an appropriate proxy for increased income.

ATTRIBUTION

Attribution focuses on the extent to which a change in the data is related to USAID interventions. The concept of attribution is discussed in detail as a criterion for indicator selection, but reemerges when assessing validity. Attribution means that changes in the data can be plausibly associated with USAID interventions. For example, an indicator that measures changes at the national level is not usually appropriate for a program targeting a few areas or a particular segment of the population. Consider the following:

> **Result:** Increased revenues in targeted municipalities.
>
> **Indicator:** Number of municipalities where tax revenues have increased by 5%.

In this case, assume that increased revenues are measured among all municipalities nationwide, while the program only focuses on a targeted group of municipalities. This means that the data would not be a valid measure of performance because the overall result is not reasonably attributable to program activities.

MEASUREMENT ERROR

Measurement error results primarily from the poor design or management of data collection processes. Examples include leading questions, unrepresentative sampling, or inadequate training of data collectors. Even if data have high face validity, they still might be an inaccurate measure of our result due to bias or error in the measurement process.

Judgments about acceptable measurement error should reflect technical assessments about what level of reductions in measurement error are possible and practical. This can be assessed on the basis of cost as well as management judgments about *what level of* *accuracy is needed* for decisions.

Some degree of measurement error is inevitable, particularly when dealing with social and economic changes, but the level of measurement error associated with all performance data collected or used by operating units should not be so large as to 1) call into question either the direction or degree of change reflected by the data or 2) overwhelm the amount of anticipated change in an indicator (making it impossible for managers to determine whether progress. reflected in the data is a result of actual change or of measurement error). The two main sources of measurement error are *sampling and non-sampling error*.

Sampling Error (or representativeness)

Data are said to be representative if they *accurately reflect the population they are intended to describe*. The representativeness of data is a function of the process used to select a sample of the population from which data will be collected.

It is often not possible, or even desirable, to collect data from every individual, household, or community involved in a program due to resource or practical constraints. In these cases, data are collected from a sample to infer the status of the population as a whole. If we are interested in describing the characteristics of a country's primary schools, for example, we would not need to examine every school in the country. Depending on our focus, a sample of a hundred schools might be enough. However, when the sample used to collect data are not representative of the population as a whole, significant bias can be introduced into the data. For example, if we only use data from 100 schools in the capital area of the country, our data will not likely be representative of all primary schools in the country.

Drawing a sample that will allow managers to confidently generalize data/findings to the population requires that two basic criteria are met: 1) that all units of a population (e.g., households, schools, enterprises) have an equal chance of being selected for the sample and 2) that the sample is of adequate size. The sample size necessary to ensure that resulting data are representative to any specified degree can vary substantially, depending on the unit of analysis, the size of the population, the variance of the characteristics being tracked, and the number of characteristics that we need to analyze. Moreover, during data collection it is rarely possible to obtain data for every member of an initially chosen sample. Rather, there are established techniques for determining acceptable levels of non-response or for substituting new respondents.

If a sample is necessary, it is important for managers to consider the sample size and method relative to the data needs. While data validity should always be a concern, there may be situations where accuracy is a particular priority. In these cases, it may be useful to consult a sampling expert to ensure the data are representative.

Non-Sampling Error

Non-sampling error includes poor design of the data collection instrument, poorly trained or partisan enumerators, or the use of questions (often related to sensitive subjects) that elicit incomplete or untruthful answers from respondents. Consider the earlier example:

> **Result:** Increased household income in a target district
>
> **Indicator:** Value of median household income in the target district

While these data appear to have high face validity, there is the potential for significant measurement error through reporting bias. If households are asked about their income, they might be tempted to under-report income to demonstrate the need for additional assistance (or over-report to demonstrate success). A similar type of reporting bias may occur when data is collected in groups or with observers, as respondents may modify their responses to match group or observer norms. This can be a particular source of bias when collecting data on vulnerable groups. Likewise, survey or interview questions and sequencing should be developed in a way that minimizes the potential for the leading of respondents to predetermined responses. In order to minimize non-sampling measurement error, managers should carefully plan and vet the data collection process with a careful eye towards potential sources of bias.

Minimizing Measurement Error

Keep in mind that USAID is primarily concerned with learning, with reasonable confidence, that anticipated improvements have occurred, not with reducing error below some arbitrary level.[2] Since it is impossible to completely eliminate measurement error, and reducing error tends to become increasingly expensive or difficult, it is important to consider what an acceptable level of error would be. Unfortunately, there is no simple standard that can be applied across all of the data collected for USAID's varied programs and results. As performance management plans (PMPs) are developed, teams should:

- Identify the existing or potential sources of error for each indicator and document this in the PMP.
- Assess how this error compares with the magnitude of expected change. If the anticipated change is less than the measurement error, then the data are not valid.
- Decide whether alternative data sources (or indicators) need to be explored as better alternatives or to complement the data to improve data validity.

2. RELIABILITY

Data should reflect stable and consistent data collection processes and analysis methods over time.

Reliability is important so that changes in data can be recognized as true changes rather than reflections of poor or changed data collection methods. For example, if we use a thermometer to measure a child's temperature repeatedly and the results vary from 95 to 105 degrees, even though we know the child's temperature hasn't changed, the thermometer is not a reliable instrument for measuring fever. In other words, if a data collection process is unreliable due to changes in the data collection instrument, different implementation across data collectors, or poor question choice, it will be difficult for managers to determine if changes in data over the life of the project reflect true changes or random error in the data collection process. Consider the following examples:

> **Indicator:** Percent increase in income among target beneficiaries.

The first year, the project reports increased total income, including income as a result of off-farm resources. The second year a new manager is responsible for data collection, and only farm based income is reported. The third year, questions arise as to how "farm based income" is defined. In this case, the reliability of the data comes into question because managers are not sure whether changes in the data are due to real change or changes in definitions. The following is another example:

> **Indicator:** Increased volume of agricultural commodities sold by farmers.

A scale is used to measure volume of agricultural commodities sold in the

[2] For additional information, refer to Common Problems/Issues with Using Secondary Data in the CDIE Resource Book on Strategic Planning and Performance Monitoring, April 1997.

What's the Difference Between Validity and Reliability?

Validity refers to the extent to which a measure actually represents what we intend to measure. Reliability refers to the stability of the measurement process. That is, assuming there is no real change in the variable being measured, would the same measurement process provide the same result if the process were repeated over and over?

market. The scale is jostled around in the back of the truck. As a result, it is no longer properly calibrated at each stop. Because of this, the scale yields unreliable data, and it is difficult for managers to determine whether changes in the data truly reflect changes in volume sold.

3. PRECISION

Precise data have a sufficient level of detail to present a fair picture of performance and enable management decision-making.

The level of precision or detail reflected in the data should be ***smaller (or finer) than*** the margin of error, or the tool of measurement is considered too imprecise. For some indicators, for which the magnitude of expected change is large, even relatively large measurement errors may be perfectly tolerable; for other indicators, small amounts of change will be important and even moderate levels of measurement error will be unacceptable.

Example: The number of politically active non-governmental organizations (NGOs) is 900. Preliminary data shows that after a few years this had grown to 30,000 NGOs. In this case, a 10 percent measurement error (+/- 3,000 NGOs) would be essentially irrelevant. Similarly, it is not important to know precisely whether there are 29,999 or 30,001 NGOs. A less precise level of detail is still sufficient to be confident in the magnitude of change. Consider an alternative scenario. If the second data point is 1,000, a 10 percent measurement error (+/- 100) would be completely unacceptable because it would represent all of the apparent change in the data.

4. INTEGRITY

Integrity focuses on whether there is improper manipulation of data.

Data that are collected, analyzed and reported should have established mechanisms in place to reduce manipulation. There are generally two types of issues that affect data integrity. The first is transcription error. The second, and somewhat more complex issue, is whether there is any incentive on the part of the data source to manipulate the data for political or personal reasons.

Transcription Error

Transcription error refers to simple data entry errors made when transcribing data from one document (electronic or paper) or database to another. Transcription error is avoidable, and Missions should seek to eliminate any such error when producing internal or external reports and other documents. When the data presented in a document produced by an operating unit are different from the data (for the same indicator and time frame) presented in the original source simply because of data entry or copying mistakes, a transcription error has occurred. Such differences (unless due to rounding) can be easily avoided by careful cross-checking of data against the original source. Rounding may result in a slight difference from the source data but may be readily justified when the underlying data do not support such specificity, or when the use of the data does not benefit materially from the originally reported level of detail. (For example, when making cost or budget projections, we typically round numbers. When we make payments to vendors, we do not round the amount paid in the accounting ledger. Different purposes can accept different levels of specificity.)

Technology can help to reduce transcription error. Systems can be designed so that the data source can enter data directly into a database—reducing the need to send in a paper report that is then entered into the system. However, this requires access to computers and reliable internet services. Additionally, databases can be developed with internal consistency or range checks to minimize transcription errors.

The use of preliminary or partial data should not be confused with transcription error. There are times, where it makes sense to use partial data (clearly identified as preliminary or partial) to inform management decisions or to report on performance because these are the best data currently available. When preliminary or partial data are updated by the original source, USAID should quickly follow suit, and note that it has done so. Any discrepancy between preliminary data included in a dated USAID document and data that were subsequently updated in an original source does not constitute transcription error.

Manipulation

A somewhat more complex issue is whether data is manipulated. Manipulation should be considered 1) if there may be incentive on the part of those that report data to skew the data to benefit the project or program ***and*** managers suspect that this may be a problem, 2) if managers believe that numbers appear to be unusually favorable, or 3) if the data are of high value and managers want to ensure the integrity of the data.

There are a number of ways in which managers can address manipulation. First, simply understand the data collection process. A well organized and structured process is less likely to be subject to manipulation because each step in the process is clearly documented and handled in a standard way. Second, be aware of potential issues. If managers have reason to believe that data are manipulated, then they should further explore the issues. Managers can do this by periodically spot checking or verifying the data. This establishes a principle that the quality of the data is important and helps to determine whether manipulation is indeed a problem. If there is substantial concern about this issue, managers might conduct a Data Quality Assessment (DQA) for the AO, IR, or specific data in question.

Example: A project assists the Ministry of Water to reduce water loss for agricultural use. The Ministry reports key statistics on water loss to the project. These statistics are critical for the Ministry, the project and USAID to understand program performance. Because of the importance of the data, a study is commissioned to examine data quality and more specifically whether there is any tendency for the data to be inflated. The study finds that there is a very slight tendency to inflate the data, but it is within an acceptable range.

5. TIMELINESS

Data should be available and up to date enough to meet management needs.

There are two key aspects of timeliness. First, data must be available ***frequently*** enough to influence management decision making. For performance indicators for which annual data collection is not practical, operating units will collect data regularly, but at longer time intervals.

Second, data should be ***current*** or, in other words, sufficiently up to date to be useful in decision-making. As a general guideline, data should lag no more than three years. Certainly, decision-making should be informed by the most current data that are practically available. Frequently, though, data obtained from a secondary source, and at times even USAID-funded primary data collection, will reflect substantial time lags between initial data collection and final analysis and publication. Many of these time lags are unavoidable, even if considerable additional resources were to be expended. Sometimes preliminary estimates may be obtainable, but they should be clearly flagged as such and replaced as soon as possible as the final data become available from the source.

The following example demonstrates issues related to timeliness:

> **Result:** Primary school attrition in a targeted region reduced.
>
> **Indicator:** Rate of student attrition at targeted schools.

In August 2009, the Ministry of Education published full enrollment analysis for the 2007 school year.

In this case, currency is a problem because there is a 2 year time lag for these data.

While it is optimal to collect and report data based on the U.S. Government fiscal year, there are often a number of practical challenges in doing so. We recognize that data may come from preceding calendar or fiscal years. Moreover, data often measure results for the specific point in time that the data were collected, not from September to September, or December to December.

Often the realities of the recipient country context will dictate the appropriate timing of the data collection effort, rather than the U.S. fiscal year. For example, if agricultural yields are at their peak in July, then data collection efforts to measure yields should be conducted in July of each year. Moreover, to the extent that USAID relies on secondary data sources and partners for data collection, we may not be able to dictate exact timing

ASSESSING DATA QUALITY

Approaches and steps for how to assess data quality are discussed in more detail in TIPS 18: Conducting Data Quality Assessments. USAID policy requires managers to understand the strengths and weaknesses of the data they use on an on-going basis. In addition, a Data Quality Assessment (DQA) must be conducted at least once every 3 years for those data reported to Washington (ADS 203.3.5.2).

For more information:

TIPS publications are available online at [insert website]

Acknowledgements:

Our thanks to those whose experience and insights helped shape this publication including Gerry Britan and Subhi Mehdi of USAID's Office of Management Policy, Budget and Performance (MPBP). This publication was updated by Michelle Adams-Matson of Management Systems International (MSI).

Comments regarding this publication can be directed to:
Gerald Britan, Ph.D.
Tel: (202) 712-1158
gbritan@usaid.gov

Contracted under RAN-M-00-04-00049-A-FY0S-84
Integrated Managing for Results II