# It All Starts Here

The main focus of most articles on assessment is, obviously, assessment. Most publications including articles, books, and book chapters take a holistic approach to assessment. That is, they focus on all of assessment but not necessarily the components of assessment. This includes the logistics of an academic assessment program and the best practices of institutionalized academic program assessment. Granted, there are some evaluation articles that highlight certain assessment components such as evaluation instruments, results, and use of results. However, little is said of the data that is collected, coded, or analyzed. Yet, these types of data decisions form the foundation for assessment. Further, if assessment is to be at the heart of all academic programming, then the data should be revered as the arteries of assessment.

This begs the question, "What is good data?" Proffered are some best practices of data collection, maintenance, and use.

1.) **Singularity**: Good data should be singular. In other words, there should be a single purpose to any variable, field, or item. Conversely, a survey item should not serve several purposes. If a variable, field, or item has more than one purpose or dimension, it is impossible to separate the dimensions from each other and only make inferences on one of those dimensions.

   Hint: Construct survey items so that the respondents rate an item (program, service, facility) on one dimension.

   Bad Example: Rate the importance and satisfaction of the Bookstore (The respondent is asked to rate this item on two affective scales).

   Good Example: Rate the importance of the services of the Bookstore with respect to meeting your educational objectives.

2.) **Format**: Good data should have a consistent format. Specifically, all values within a field or variable should have the same "look."

Hint: Insist that all values within a given variable, field, or item have: (a) the same number of columns, (a) each column is similar with respect to valid values, and (c) the coding is consistent. In sum, the nomenclature should be consistent for every possible value for that field or variable. Why be so compulsive when constructing the initial formats? It will be easier and faster to detect possible errors and omissions, especially with large data sets.

Bad Example: MGT-103-01, MEDR-104-T3 (The course name has three or four letter prefixes, the last two columns have both numeric and alpha symbols, "03" and "T3" convert to the same description or value).

Good Example: MGMT-103-01, MEDR-104-03 (same number of columns, with numerals only in the last two columns).

3.) **Nothingness**: Good data construction includes identifying appropriate null values. In most instances, an item that is not answered should be coded as a blank. However, there are instances when a forced value is needed. Use a "NULL" value or, consistently, use the same value for non-responses or for variable values that are unknown. Also, include an item choice that lets the respondents opt out of responding to the item. The choices, "Does Not Apply" or "Neither" help to distinguish between a genuine rating and one that is forced because of the survey instrument. Avoid using the numerical symbol, zero, to substitute for non-response or the unknown especially in numerical fields.

Hint: Construct variables, fields, or items with a specified and uniform null value. Offer respondents a "Not Applicable" choice.

Bad Example: Use a 0.00 when *High School GPA* is unknown. Because it is possible to have a GPA of 0.00, there is no distinguishing between a very poor GPA and an unknown GPA.

Good Example: Use "NA" or keep the *High School GPA* field blank when GPA is not known.

4.) **Scale**: Good data construction includes attention to the scale. All values within a variable must be within the confines of the scale. Ergo, the values, regardless of their origin, have the same meaning. Also of concern are the minimum and maximum value of the scale. If negative values are possible, then include them. Also be mindful of detail. The precision of the field or variable should be determined to reasonably reflect differences between individual records or observations.

Hint: Construct or identify the scales of each variable, field, or item before data entry. For those variables that may have different scales, construct a conversion table. This will ensure that all values will conform to only one scale.

Bad Example: Enter "3" for a *High School GPA* on a scale of 0.00-4.00. Enter 3.0 for a second *High School GPA* that has a scale of 0.00-5.0. The difference in scale makes the comparison or aggregation of this field virtually impossible. This is a classic example of comparing apples to oranges. Also, the GPA is only reported as a whole number. Distinguishing "A" students from "B+" students is, likewise, impossible.

Good Example: Scenario one: Entering 3.12 for both aforementioned *High School GPA*s. This is not a disaster if the *High School GPA* field is linked to another field that identifies the scale. Scenario two: Provide a conversion table for *High School GPA* for high schools that have atypical grade point averages. Convert before data entry.

5.) **Consistency**: Good data is consistent on several dimensions. As mentioned, the format, null values, and scale should be consistently applied to all values within a particular variable, field, or item. Consistency also applies to data entry practices, field definitions, and retrieval. It goes without saying that all those involved in data entering should be using the same values for the same item choices. Less obvious is that the format for entering data should also be the same. In an alphanumeric field, a "0.00" could have a different meaning than "0." Similarly, field definitions and retrieval practices should also be consistently applied for data entry and for all administrations of the evaluation project. Finally, institute a double data entry verification process in which two persons enter the same data from the same survey. Comparisons on the consistency of entry can be made. Mistakes can be identified and possible processes to correct errors in data entry can be instituted.

Hint: Construct data entry practices, field definition, and retrieval standards concurrently with the actual variable, field, and item construction. Have every data entry professional review these practices and standards and verify that they understand them.

Bad Example: Two respondents live in Chicago. However, the city information is entered as "Chicago" for one respondent and "CHICAGO" for the other. This poses a potential data retrieval nightmare because the retrieval would have to include a query of all the possible combinations of lower and upper case letters that appear in Chicago.

Good Example: The title case, "Chicago," will be used for all respondents that reside in this fair city.

6.) **Data Entry Standards**: Although possibly subsumed in the consistency guideline, it cannot be overstated that good data construction includes data entry standards. Standards should be developed in both breadth and depth. Data standards should apply to all variables, fields, and

items. They also should apply to all values entered or to be entered. In terms of depth, data standards should be established regarding purpose, formatting, null value application, scale, and consistency. Likely, the last feature will be the most ventured. Most standards will speak to the consistency of the data.

Hint: Construct data standards that are reasonable to execute and easy to understand. Again, be consistent in the application of the data standards across all data entry professionals and administrations of the evaluation instrument.

Bad Data Standard: Data should be entered as soon as possible.

Good Data Standard: Expediency is never the goal of good data. Try instead, Standard 1: This field requires two decimal places. An '*' should be entered for unknown values.

7.) **Data Definitions**: Likewise, it cannot be overstated that good data construction includes data definitions. Data definitions explain the purpose of the field.

Hint: Construct data definitions that state the actual intent of the variable, field, or item. All feasible uses and values should be explicated when possible. Amass this information into a data dictionary. Make the data dictionary available to all campus users.

Bad Data Definition: The respondent's transfer institution type is entered in the *School Type* field.

Good Data Definition: The *School Type* field contains the type of institution that the respondent first transferred to after attending XYZ Community College. Only post-secondary institution types should be entered in this field. See the *School Type* Table for valid values.

8.) **Shadow Systems**: In short, good data can avoid shadow systems. Shadow systems are the databases, spreadsheets, and data files that reside on individual computers and servers that

are only accessible to individual users. When possible, keep it all on the mainframe, or main data system, main server, or main evaluation data clearinghouse. Again, a central location will help to identify any issues or problems regarding consistency, missing values, formatting, misinterpretation of data standards, use of data definitions, and the like. These problems can be mitigated by a systemic modification of process and procedure.

Hint: Be the Chief of the Shadow System Police. Encourage all users to enter their information on the central system. Also encourage curious onlookers such as academic Deans and Vice Presidents to use this central evaluation clearinghouse when retrieving data for academic program assessment. Offer tutoring and training for this central evaluation database on data entry and data retrieval. Importantly, be inclusive on granting permission to use this central evaluation database.

Bad Example: Mary maintains a database of a program assessment given by the Mathematics department on her personal computer. Jane does the same for the English Department.

Good Example: Have Mary and Jane upload this information onto a central evaluation clearinghouse, using similar format and data definitions. Resolve differences in scale, consistency, and data entry standards before this blessed event occurs.

9.) **Warehousing**: Good data is only as good as the extent of its usage. A data warehouse, or central evaluation clearinghouse, may not be an original objective but it should be a long-term goal. A data warehouse is a collection of data in which end-users have some control over how the data is aggregated. Watson (2000) explains that a data warehouse provides the user a source for accessing and analyzing data. A data warehouse should not be confused

with a data depository in which every piece of data is dumped into the warehouse. Instead, a data warehouse contains information that is organized logically.

Hint: Construct a data warehouse being attentive of the needs of the end-user. Develop appropriate mining tools aligned with end-users abilities.

Bad Example: A data warehouse has been placed on a server without concern to access, end-user tools, or purpose.

Good Example: Talk to potential users and contributors before developing data warehouse query options.

10.) **Best Practices**: As the data demons dissipate, exemplary data development should be noted. This includes specific attention to format, scale development, consistency across variables, fields, or items, data standards, and data definitions. Keep a log of how variables, fields, and items are developed, coded, used, and modified. Make a best practices list and post it on a public web site.

Hint: Have a best practice workshop or conference for staff and other end-users. Celebrate staff and departments that are adopting these best practices.

Bad Example: Ignore good practices.

Good Example: Recognize and reward good practices. Provide incentives for best practice behavior.

11.) **Support Groups**: As the college acclimates to good data development and coding, colleagues should be encouraged to meet. Discussion groups could involve problems and road blocks to data retrieval, development, and access. Best practices should also be shared and include item development, coding resolutions, and data warehousing construction. The

use of support groups is contagious. It encourages new end-users to jump into the arena of good data development. It also motivates existing data developers to continue their efforts.

Hint: Provide support groups by type of end-user application such as MS-Access, use of legacy system, or data warehouse. Also create support groups by the type of data being sought. For example, the Natural Science Department may have its support group because its data collection, coding, and retrieval issues may be different than those of other end-users.

Bad Example: One support group for the entire college will most likely be confusing to those who attend. The type of issues will be too broad in scope.

Good Example: An MS-Access Users group for the science and mathematics faculty will meet in the campus lounge next Tuesday.

This may not be an all-inclusive list. At minimum, these are the necessary tenets for procuring and maintaining good data. Consequently, good data will lead to accurate results and better results. In sum, good data practices will lead to good and useful assessment.

References:

Watson, J. (2000). Statement of Purpose. *Journal of Data Warehousing, 5*(4), 3-3.

About the Author: Dr. Mary Lou D'Allegro is the Director of Institutional Research at West Chester University that supports both the college's Academic Program Review and Assessment efforts.