

ORGANIZING ASSESSMENT DATA FOR ANALYSIS

Office of Assessment of Teaching and Learning, Washington State University

Assessment data provide a means to look at student performance in order to offer evidence about student learning in the curriculum, provide information about program strengths and weaknesses, and guide decision-making. Analyzing the data -- in context -- gives meaning to the information collected and is essential in order to appropriately utilize and communicate the assessment results. Before any type of analysis can occur, the collected data needs to be compiled, organized, and cleaned. This resource is intended to help faculty consider best practices and plan for organizing and cleaning data collected about student learning as part of program or course assessment.

To organize and clean the assessment data, you will want to understand the context, purpose, and scope of the project. Assessment data may come from a wide range of assessment measures, including capstone papers, senior theses, dissertations, embedded assessments, observations of student performances, portfolios of student work, pre-test/post-test assessments, standardized tests, supervisor evaluations of interns, focus groups, interviews, surveys, and course evaluations. As a result, the information collected can be in a variety of formats, such as completed surveys, rubric scores of student work or performances, focus group notes, and completed tests or exam scores.

Types of Data

Generally, data collected for assessment fall into two categories: quantitative and qualitative.

Quantitative data analysis relies on numerical scores or ratings and is helpful in evaluation because it can provide quantifiable results that are easy to calculate and display.

Qualitative data consists primarily of words and observations, rather than numbers. Qualitative data can come from a variety of sources including open-ended survey questions, focus group notes, essay responses, and student portfolios, for example. Qualitative data are useful for answering “why” and “how” questions about student performance, approach, motivation, or experience.

Organizing and Formatting Data

Whether you have collected quantitative or qualitative data, it is important that the data are organized in a logical format that can be easily understood and analyzed. Microsoft Excel has many features that make it easy to manage and analyze data. To take full advantage of these features, it is important that the data are organized and formatted according to the following guidelines (an example of what the spreadsheet will look like follows). *Note: When organizing and formatting assessment data, it is good practice to archive a copy of the raw (i.e. original) data as a backup in case of any technical issues or errors.*

- Each column should be used to represent a unique piece of information collected about the students (i.e. survey or test response, rubric score, demographic characteristic).
- Each row should represent data for an individual student.
- Avoid blank rows and columns.

- Care must be taken to avoid including names or other identifying information in assessment results (or failing to, wherever possible, remove names or identifying information--including student, faculty, and staff--for assessment work):
 - Assign a unique identifier to each individual in the dataset and make sure that each student's responses or ratings are deidentified, meaning they are stripped of their name. *Note: it may be appropriate to include WSU ID numbers or maintain a key in a separate spreadsheet with student names matched to their unique ID (if demographic information is needed, for example; this decision will be determined by the assessment coordinator and project's questions and purpose).*
 - If you have rubric data scored by faculty, raters should also be coded by a unique ID, removing their names as well.
 - If you have survey responses or other qualitative data, names or other identifying information--including student, faculty, and staff--should be removed/redacted from the comments/responses wherever possible.
- Code text responses into numerical form, where possible, so that they are easier to analyze (e.g., 1=Yes, 0=No).
 - Enter data in a consistent format, such as always using a "1" to reflect female gender, rather than using various labels (e.g., "F," "female," "girl," etc.).
 - It can be useful to have one column with the text response and a second column with the coded numerical form (see Q4 – Satisfaction in the following example).
- Avoid inserting spaces at the beginning or end of a cell to indent data. These extra spaces can affect sorting, searching, and the format that is applied to a cell.

	A	B	C	D	E	F	G	H	I
	Student	Q1 - Education Label	Q1 - Education Value	Q2 - Credit Hours Value	Q3 - Summer Program Label	Q3 - Summer Program Value	Q4 - Satisfaction Label	Q4 - Satisfaction Value	Q5 - Strengths Text
2	1	High School	1	36	Y	1	Strongly Disagree	5	The Science program has a lot of flexibility; there were many electives.
3	2	College	2	40	Y	1	Strongly Disagree	5	The program is very flexible and allowed me to take the electives I wanted.
4	3	Advanced Degree	3	33	Y	1	Strongly Disagree	5	There are lots of options for electives.
5	4	College	2	36	Y	1	Neutral	3	I was able to take the exact courses I needed.
5	5	Advanced Degree	3	39	Y	1	Agree	2	I was able to take exactly the courses employers were looking for.
7	6	College	2	40	Y	1	Agree	2	The advisors are great!!

What if the data were collected electronically? In many cases, it is possible to download assessment data collected electronically (i.e. from a survey administered using Qualtrics or a fillable Adobe PDF form, an exam administered in Blackboard, etc.) directly into an Excel file that will be formatted following many of the previous suggestions. However, further formatting may be required to assign unique identifiers, redact names identifying information or code responses.

What if the data were collected by hand on paper? If the assessment data were collected on paper, then the data will need to be typed into Excel manually following the previous guidelines. If data must be entered manually, a data form in Excel (see following example) provides a convenient means to enter or display one complete row of information in a range or table without scrolling horizontally. You may find that using a data form can make data entry easier than moving from column to column when you have more columns of data than can be viewed on the screen.

Student	Q1 - Education Label	Q1 - Education Value	Q2 - Credit Hours Value	Q3 - Summer Program Label	Q3 - Summer Program Value	Q4 - Satisfaction Label	Q4 - Satisfaction Value	Q5 - Strengths Text
1	High School	1	36					
2	College	2	40					
3	Advanced Degree	3	33					
4	College	2	36					
5	Advanced Degree	3	39					
6	College	2	40					
7	Advanced Degree	3	32					
8	High School	1	47					

Dataset

Student: 20

Q1 - Education Label: High School

Q1 - Education Value: 1

Q2 - Credit Hours Value: 30

Q3 - Summer Program Label: Y

Q3 - Summer Program Value: 1

Q4 - Satisfaction Label: Strongly Disagree

Q4 - Satisfaction Value: 5

Q5 - Strengths Text: I was able to take exactly the courses I needed to pr

20 of 20

New

Delete

Restore

Find Prev

Find Next

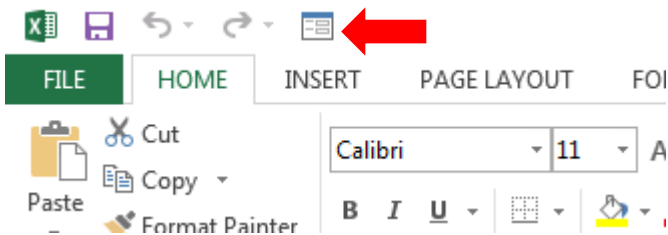
Criteria

Close

The data form automatically displays all column headers as labels in a single dialog box. Each label has an adjacent blank text box in which you can enter data for each column, up to a maximum of 32 columns.

Using a data form in Excel:

1. Add a column heading to each column in the range or table. Excel uses these column headers to create labels for each field on the form.
2. Click a cell in the range or table to which you want to add the form.
3. On the Quick Access Toolbar, click the Form button.



Note: If the Form button isn't on the Quick Access Toolbar, you can add it by clicking the arrow next to the Quick Access Toolbar and clicking "More Commands". In the "Choose commands from:" box, click "All Commands", and then select the "Form..." button in the list. Click "Add", and then click "OK".

4. To add a new row of data, click New and type the data for the new row. To move to the next field in the row, press TAB.
5. After you have finished typing data, press ENTER to add the row to the bottom of the range or table.

If you must enter data manually, decide what to do with obviously incomplete data (for example, surveys that are largely blank or rubric scores where the rater missed a whole section), if you're going to input them.

Cleaning Data

The usefulness and effectiveness of data depend on their being kept accurate and complete. After the raw data have been entered or downloaded into a spreadsheet, it is good practice to archive a copy of this raw data file to be kept as a backup.

No matter what type of data you have or what organizational method you use, be sure to go back and review data for errors. Checking for errors is commonly called “cleaning.” Cleaning the data is critical, as “dirty” data can influence your results. The most commonly used cleaning methods are spot-checking, eye-balling, and logic checks. The best practice is to use all three approaches to be sure you have caught all possible errors.

- Spot-checking involves comparing the raw data to the electronically entered data to check for data-entry and coding errors. To spot-check data, randomly select several participants and compare the raw data to the entered data. If you find an error in your first round of spot-checking, then you should randomly check another round of raw data. If you continue to find errors, then you may need to go over all of the raw data to ensure that the data was entered correctly.
- Eye-balling involves reviewing the data for errors that may have resulted from a data-entry or coding mistake. For example, “N” is assigned a value of 0, while “Y” is assigned a value of 1 from the previous example. Any number other than 0 or 1 in that column would be an obvious error.
- Logic check involves a careful review of the electronically entered data to make sure that the answers to the questions “make sense.” For example, if Student #1 indicated that he/she did not attend the summer program, then it would be illogical for this participant to have provided a satisfaction rating for the summer program.

This may also involve checking for duplicate data, missing data, or the prevalence of blank fields. In some cases, data from rating sessions can be messy – a rater may miss a lot or misunderstand the instructions, or there are problems pairing raters or second readers.

When cleaning data, it is best practice to record any changes made to the data set into some sort of log – this may be on another tab in the spreadsheet or in a separate text file. It is useful to record: the name of the person who entered the data, the name of the person who cleaned the data, the date(s) the cleaning occurred, and any specific changes that were made during the cleaning process. At this point, it is good practice to archive a copy of this “clean” data file along with the data cleaning log. When you are ready to begin analyzing the data it is a good practice to save a separate working copy of the clean data file – working from a copy of the clean data can help prevent data loss from unforeseen circumstances (i.e. if the file becomes corrupt).

Including Context about the Data

In addition to a data cleaning log, a short written description of the data collection processes, the number of participants, and a copy of any instrument used (i.e. rubric, survey, exam) should accompany the data file. This description should include basic data collection processes, including how the data were collected, who participated, and any known limitations of the data. Other factors to consider may include: How was the random sampling/size determined? What was the response rate? Were well-established, agreed-upon criteria (such as a rubric) used for assessing the evidence for each outcome? How were raters normed/calibrated? Did multiple raters review each piece of evidence? Has this measure been pilot tested and refined?

Assessment Data Stewardship: *It is important to remember that assessment data/results are valuable resources and must be carefully managed. Each individual with access to assessment data/results has the responsibility to use those data and any information derived from them appropriately. Non-public (i.e. internal or confidential) data/results should be labeled and only used to support assigned roles and duties. For more information, see ATL’s [Assessment Data Stewardship: Tips for Academic Programs](#).*