# Make Your Teaching Innovations Count: A primer on quantitative approaches in educational research

Allyson Skene
Sarah King
Janice Patterson
Centre for Teaching and Learning

and

Tim Cho Kin Cheng
Department of Psychology

*University of Toronto Scarborough*

*This project was supported by a 2011 Educational Developers Caucus (EDC) Grant of the Society for Teaching and Learning in Higher Education (STLHE).*

*Thank you to all UTSC faculty who contributed ideas and questions, participated in discussions, and looked at drafts of the document, in particular Clare Hasenkampf for her ongoing support and Elaine Khoo for the examples and ideas she brought to the early stages of this document.*

# Contents

## Introduction

If you teach or run academic support programs at a university or college, chances are that you regularly experiment with new programs, teaching, and assessment methods. Chances are that you also gather data about these innovations (whether or not you think of it as “data”), through attendance, students’ oral and written feedback, perception surveys, informal conversations, and your own or other instructors’ perceptions. You then use this data to inform your decisions about where to go next. In other words, you are already engaging in a rough form of educational research.

Five good reasons to systematize your research on teaching:

1. Improve teaching effectiveness
2. Document growth and success as a teacher
3. Meet expectations for accountability and argue for increased support
4. Identify gaps and student needs
5. Communicate findings with colleagues

It is less likely, however, that you are researching your teaching innovations systematically—i.e. developing a focused research question and a controlled research design, compiling, analyzing and interpreting data about student learning and then presenting your data to colleagues at conferences and through publications. In other words, you are probably not making the most of the data, in terms of your own learning and the benefits of sharing it.

Why do most of us, even as we take teaching and educational programming seriously, not actively engage in systematic research on our teaching? The multifaceted responsibilities and pressures of working at university or college often make it difficult. Lack of time, lack of training in educational research and statistics techniques, lack of some key data component, ethical concerns—there are many reasons why even those of us who would like to have better data, to understand how and whether our students are learning, and learning as we intend them to, are not able to do it. Maybe you have an idea for a small research project but are not sure if it is necessary to submit it to your institution’s research ethics board; maybe you have started a project but are not confident in the statistical procedure used to analyze the data; maybe you just have not found the time to think about researching your teaching.

The goal of this toolkit is not to stimulate education-based research—if you are teaching or running student programs, you are already well on your way. Rather we hope to enable educators to make more effective use of already existing data. This document suggests ways to do structure data into possible research questions, recommends practical strategies for managing data, and finally provides a clear introduction to the statistical analyses most commonly used in educational research. In addition, we offer strategies you can use to face challenges and resolve problems you may encounter along the way, and we briefly introduce approaches to important ethical considerations.

Our focus is on quantitative research, not because it is better or faster than qualitative research, but because we know, based on our own experience, that it can be intimidating to begin. In addition, in an era of competition for limited budgets, numbers matter. Presenting your program or teaching using data may be a means of showcasing your work and successes more effectively, and in terms that may be useful in obtaining support. Finally, as we’ve already mentioned, the increasing use of electronic devices in teaching and educational programming has significantly increased the amount of teaching-related data available to instructors. It has also increased the ease with which we can collect and analyze large quantities of data. We believe that quantitative research on teaching and learning is both satisfying and rewarding. It can confirm your intuitions, clarify your results, support informed choices rather than intuitive ones and lead to better teaching.

Additional benefits of quantitative analysis:

1. Quantitative research allows you to quantify the impact you have made with your innovation. It also allows you to compare the impact across different innovations.
2. Quantitative research makes it easier to manage data from and draw meaningful inferences about a large population.

What will you, as a reader, get out of this document? You will:

- Work step by step to quantitatively assess and reflect on your own programs and initiatives, addressing your own teaching as a scholarly endeavour ;
- Learn to organize data that you might have collected;
- Learn to utilize available data to address a research question;
- Learn basic statistical approaches to educational research;
- Learn to interpret statistical results; acquire an understanding of the limitations of research findings, and learn to address the limitations in future research;
- See a range of examples of educational action research.

Most guides to research design start at the beginning, e.g. developing a research question. We suggest that you begin in the middle, using data you already have. From there you can assess what research questions might be answerable using that data. In effect, this gives you practice thinking about the data. Then the toolkit walks you through the steps of analyzing and interpreting data, explaining common statistical tests in accessible language, using examples drawn from a higher education context.

Quantitative research is an iterative, cyclical process. Each research question you ask and answer leads to more questions, and tells you more about what you don't know. Perhaps our most important goal with this document is to get you started on quantitative research by taking you around the cycle once, so that the process is familiar and has lost any intimidation factor it might once have had.

You may not be able to answer key questions using data that you have collected without a question in mind. However, once you have been through the process we recommend here, we predict you will emerge with a clear idea for a meaningful question you could ask, and how to use gather, organize, and analyze the data you need to answer it. In addition, one advantage of quantitative research is that it is cumulative. As you cycle through the process, both questions and answers become richer as new data is added to existing data.

The authors of this document, are not (with one exception) experts in quantitative analysis. Rather we are university instructors and program coordinators who have been developing our interest and ability to assess our educational programs. Taking our own difficulties and lack of knowledge as a starting point, we inferred that there must be many others in similar situations at Canadian universities. With the help of a small grant from EDC, we hired a graduate student who **is** an expert, and have worked together to develop these materials. This document is a work-in-progress, and we welcome questions, feedback, and suggestions.

Allyson Skene
Sarah King
Tim Cho Kin Cheng
Janice Patterson

University of Toronto Scarborough
October 2012

## Section 1: Research Design

**Starting with Data**
Most guides, whether to statistical analysis or educational research, recommend starting with the research question, then gathering data, analyzing and interpreting it. While this approach conforms to best practices in conducting research generally, we believe the emphasis on starting with design and planning is one of the reasons that a lot of educational research never gets off the ground. The process takes time and can be intimidating, particularly if you've never done it before. Instead, we start in the middle, by suggesting taking a look at the data you already have. Then you can work backwards and see what questions this data, maybe with a little supplementing, might be able to answer.

**Exercise 1: Brainstorming**
Take a few minutes to quickly jot down answers to the following questions. (If unsure how to respond, see Table 1 for ideas):

1. What innovations have you introduced into your course(s) or program(s)? List as many innovations, large and small, that you can think of.

2. What data do you have or could you collect relatively easily? (e.g. grades, evaluations, attendance, etc.)

3. In this data, what might you be able to measure or count?

The following table gives some examples of data related to teaching innovations. It is intentionally not yet complete. Feel free to adapt it to your own teaching by adding elements

**Table 1: Thinking about Data: Some examples**

| Teaching or Programming Innovation | Data sources | Measurable data |
|---|---|---|
| <u>New technology</u><br>• discussion boards<br>• blogs or learning journals<br>• e-portfolios<br>• Twitter<br>• tablet<br>• Turnitin<br>• iClicker<br>• e-mail, skype, or virtual tutoring<br><br><u>New pedagogical approach</u><br>• different approach to concept<br>• re-organization of course material<br>• posting lecture notes<br>• in-class active learning exercise<br>• peer review (in-class, online)<br>• extracurricular workshops<br>• departmental help centre<br>• in-class writing workshop<br>• peer study group or group tutoring<br><br><u>Assignment Re-design</u><br>• new assignment type<br>• scaffolding<br>• group projects<br>• rubric | <u>Data automatically collected</u><br>• Student grades on all assignments, quizzes, tests, exams, participation<br>• Course evaluations<br>• Electronic copies of assignments, possibly including comments and grades<br>• iClicker results<br>• Learning or Content Management Systems (e.g. Blackboard, Moodle, Joomla, Drupal)<br>• Turnitin data<br>• Notes recorded on individual tutoring or counselling sessions<br><br><u>Data you could collect</u><br>• Attendance<br>• Sample assignments<br>• Additional perception surveys (note that course evaluations are perception surveys)<br>• Pre- / post- test results (e.g. administer same test or survey at beginning and end of program, course, module or session) | <u>Attendance/participation</u><br>• Number of times attended<br>• Number of times participated (e.g. discussion board post)<br>• length of post (# words, keywords)<br>• hits on particular site/document; time on site<br><br><u>Grades</u><br>• Correct answers; responses to multiple choice questions<br>• Frequency of errors in specific areas; areas where greatest progress was made<br>• Change in results<br>• Change in grades on similar assignments<br><br><u>Survey responses</u><br>• Course evaluation scores<br>• Perception survey responses - responses, response rate or frequency |

**Categories of Data in Educational Research: Participation, Performance, Perception**

Now that you have a better sense of the number of teaching innovations and the data that you have, the next step is to draft some research questions that the existing data might be able to answer. In order to do this, it's useful to have a sense of the typical types of data and questions used in educational research.

Whether for assessing a faculty development workshop, an academic support program, or a classroom initiative, educational research is, broadly speaking, attempting to determine whether our efforts are leading to students' meeting our (learning) objectives. Have students, or workshop participants, gained knowledge, skills, or values? Will they change their behaviour (e.g. alter their teaching approach, change their study habits, approach writing assignments differently) as a result of our support? To determine answers to these questions we need ways of measuring these changes, whether directly, by observing actual behaviour and responses in a test or real world situation, or indirectly, by asking students and participants whether they perceive these changes in themselves. Another way of putting this is, does participating lead to change in either performance or perception?

As a result, then, educational research tends to measure one or two of the following three things: Participation, Performance, and Perception. Generally, a research question will take two of the categories and ask a question about the relationship between them. For example:

- Do students who **participate** in Writing Centre tutorials **perform** better on writing assignments than students who **do not participate**?
- Did **participating** in an activity of developing possible multiple-choice questions improve **performance** on final exam?
- Did **participating** in a series of four vocabulary-building workshops affect students' **perceptions** of their academic vocabulary?
- Does the number of times students **participated** in an online discussion forum relate to their **performance** in the course?
- Do student **perceptions** of a particular teaching innovation affect their **performance** in a course?

Research questions, particularly when they refer to the specifics of measurable data, don't necessarily use the terms, but they are implicit. For example:

- Did adding two in-class workshops on Using Sources Effectively (student **participation**) reduce number of essays flagged by Turnitin as having significant textual similarities to existing documents? (student **performance**)?
- Do student grades (**performance**) in a course correlate with high ranking of the instructor in course evaluations **(perceptions**)?

Look at the data you have and categorize it in terms of participation, performance, and perception.

### Approaches to Educational Research: Experimental and Correlational

There are two possible approaches to educational research: the correlational approach and the experimental approach. Broadly speaking, the correlational approach measures whether there is a relationship between two (or more) variables, For example, is there a correlation between attendance at study groups and course grades? The experimental approach requires both a test group and a control group, then measures either change or difference between the two groups. For example, if you had an optional assignment in your class, you could ask whether completing the assignment made a difference to final performance. The students who did not do the assignment function as the control group, while those who did are the test group. (See the Ethics section of this document for more information about using the experimental approach in your research.)

Tables 2 and 3 summarize the elements of each of these approaches and show some sample research questions. Which approach is best for you may depend on your data, or on your research question. Some data lend themselves to either approach. Figure 1 shows an example where the data can be analyzed either way.

**Table 2: Common Experimental Approaches in Educational Research**

<table>
<tr><th colspan="2">What are you measuring?</th><th>Required data</th><th>Sample research question</th><th>Test group</th><th>Control group</th></tr>
<tr><td rowspan="3">PERFORMANCE</td><td>Change in performance</td><td>Paired scores (e.g. grades from exams or tests) both before and after intervention</td><td>Did activity (after midterm) where students developed and assessed multiple choice questions improve performance on final exam?</td><td>Final Exam scores for students who participated in activity</td><td>Midterm exam scores for students who later participated in activity</td></tr>
<tr><td rowspan="2">Differences in performance between participants and non-participants</td><td rowspan="2">Ability to identify participants, combined with scores from tests, course grades or surveys of those that did and those that did not participate</td><td>For first-year students whose average in high school calculus was below 80, did participating in a math transition program lead to better grades in first-year calculus compared to those students who did not participate in the program?</td><td>Students taking first-year calculus who participated in transition program</td><td>Students taking first-year calculus who did not participate in transition program</td></tr>
<tr><td>Did students who participated in optional course-based study groups perform better in the course than those who did not?</td><td>Students in course who attended study groups</td><td>Students in course who did not attend study groups</td></tr>
</table>

| | | | | | |
|---|---|---|---|---|---|
| **PERCEPTION** | Change in perceptions or values before and after | **Paired** ratings from pre and post tests | Did attending a series of vocabulary-building workshops increase participants' confidence with academic vocabulary? | Post-test after completion of multiple workshops | Pre-test before students begin first vocabulary workshop |
| | Differences in perception between participants and non-participants | Ability to identify participants, combined with scores from participants and non-participants | Did new Annotated Bibliography assignment (scaffolding of research proposal) change student assessment of the value of the assignments? | Current class responses to question 11 on course evaluation forms that asks about value of assignments | Previous class responses to same question on course evals |

**Table 3: Common Correlation Approaches in Educational Research**

| Required Data | What are you measuring? | Sample Research Questions | Data required to answer sample question* |
|---|---|---|---|
| Data on a scale for each category:<br>• **Participation data** could include frequency of attendance, frequency of use of tool or resource<br>• **Perception data** could include survey results on a ranked numerical scale<br>• **Performance data** could include grades, number or percentage of correct answers or errors | Relationship between **participation** and **performance** | Does attending more writing centre appointments correlate with higher grades on the final essay? | Attendance at writing centre AND Final essay grades |
| | | Does # of times students participated in online forum relate to their performance on the course? | Number of times student participated in forum AND Final grade |
| | Relationship between **perception** and **performance** | Is there a correlation between student ratings of high test anxiety and performance in test and non-test situations? | Student ratings of their test anxiety AND grade differential between test and non-test performance |
| | Relationship between **participation** and **perception** | Is there a correlation between perceived difficulty of the course and number of times students attend study groups? | Student ranking of course difficulty AND attendance at study groups |

* See Section 2 for ideas on how to collect data.

### What Approach(es) Might Best Fit Your Data? Mapping Research Design Options

Data can often lend itself to different research questions. Figures 1and 2 give some examples showing how data can lend itself to several different analyses. After looking at these, select some of your own data for a teaching innovation and map your options.

**Figure 1: Optional course-based study group.**

Background: A course has optional study groups associated with it, organized by the learning skills centre. The coordinator of the study group program has access to data on student attendance at the groups, student grades in the course, and two separate perception surveys, one completed by students who attended the group and one by students who did not.

EXPERIMENTAL APPROACH
- attendance
- course grades

RQ: Do students who attend study groups achieve better grades than students who do not attend?
OR
RQ: Are students with lower grades more likely to attend study groups?

CORRELATION APPROACH
- attendance
- grades

RQ: Is there a correlation between the number of times students attended study groups and their course grades?

**Available data**
- **Attendance (participation)**
- **Course grades (performance)**
- **perception surveys**

EXPERIMENTAL APPROACH
- perception surveys
- attendance

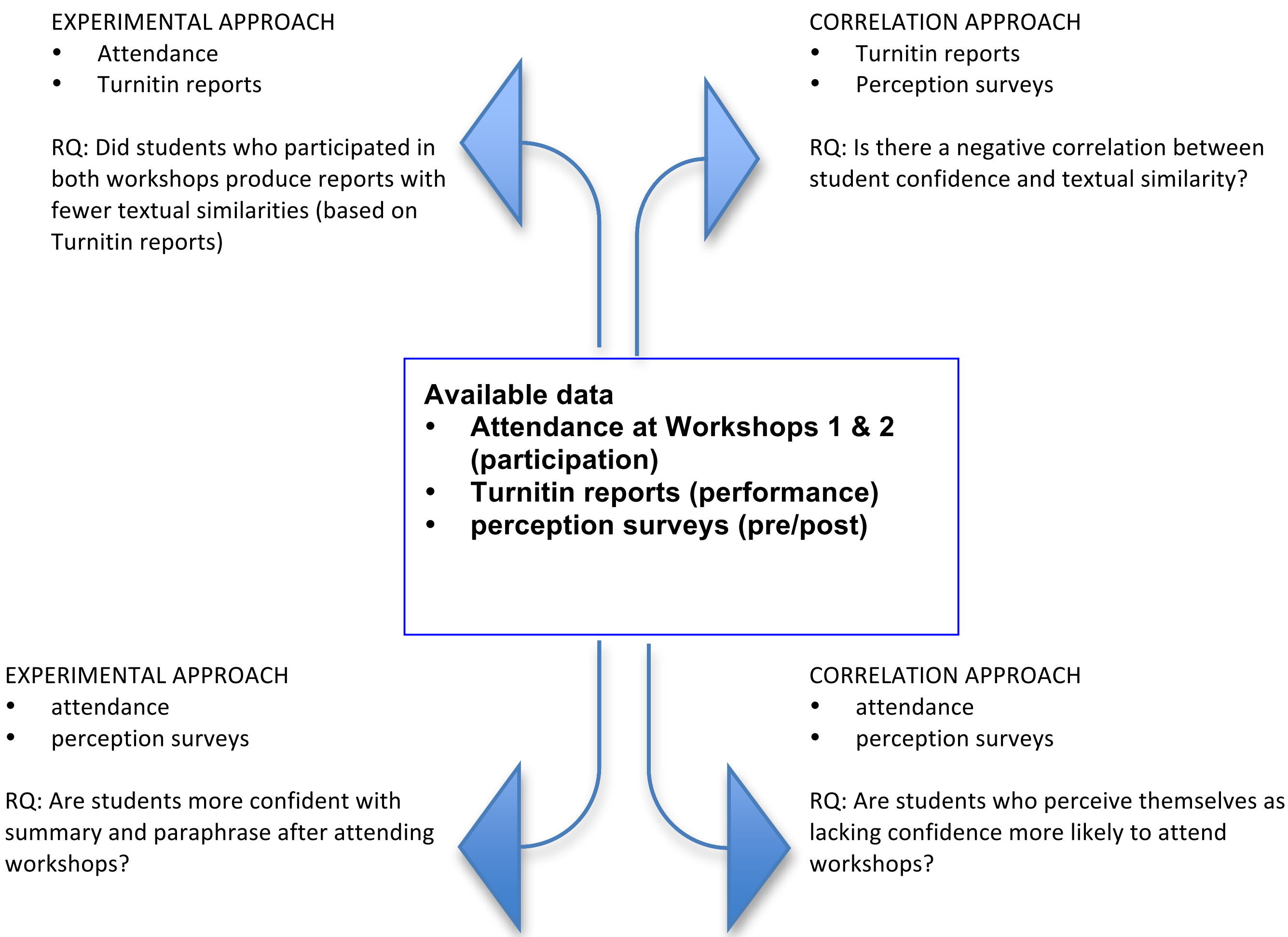
RQ: Are students who perceive the course as difficult more likely to attend study groups than students who perceive the course as easy?
OR
RQ: Do student who attend study groups rank the course more positively overall than students who do not?
OR
RQ: Are students more confident with the materials after attending study groups?

CORRELATION APPROACH
- perception surveys
- attendance

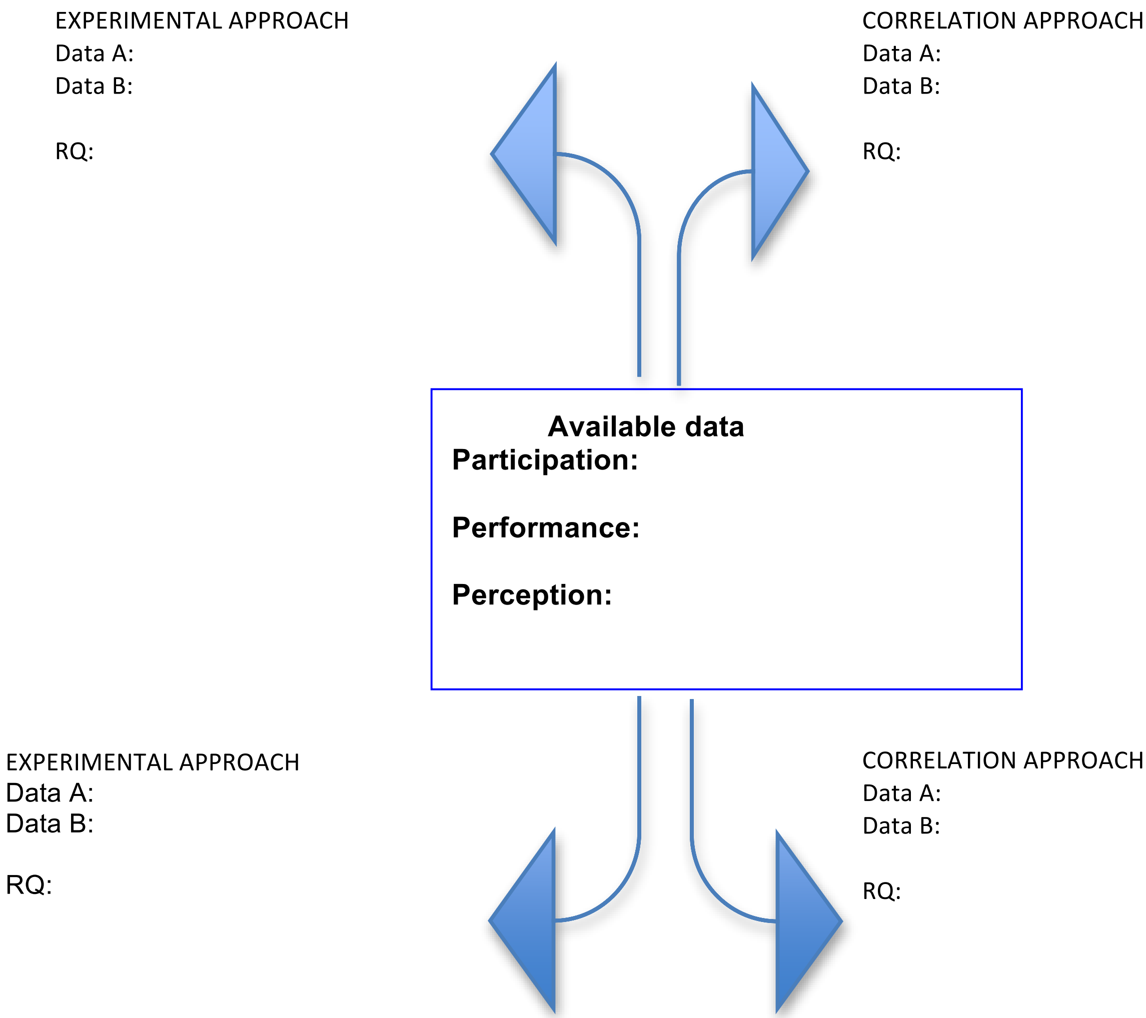
RQ: Is there a correlation between perceived difficulty of the course and number of times students attend study groups?

**Figure 2: Teaching students to summarize and paraphrase sources**

Background: A course instructor invited the Writing Centre to offer two short in-class workshops on summarizing and paraphrasing sources. Students submit papers through Turnitin, so the instructor has textual similarity reports, as well as grades and workshop attendance. The instructor also has the reports on a similar assignment from last year, in a class where there were no workshops. Finally, the instructor distributed a short survey to students about their confidence level in using sources before the first workshop and after the second workshop.



**Exercise 2: Mapping Research Design Options**
Choose data you have from a teaching innovation and map out what experimental and correlational approaches might be possible, and what research questions you might develop:



**Troubleshooting Research Design Options**

As you consider possible directions and research questions that you might be able to answer with existing data, you may encounter one or more of the following challenges. In the following table, we have suggested some options that might lead you to solutions.

| **Challenge** | **Possible solution/option** |
|---|---|
| Missing control group | Measure against a fixed standard, such as average course grade; look at previous sections of the course before the innovation was implemented; omit innovation or make it optional when teaching the course next time |
| Missing data –e.g. attendance | Reframe your research question to deal with data you do have.<br><br>Consider other data types that you may be able to access:<br>• Obvious/common sources: attendance; demographics (year, first generation in university/college, program of study, language, etc); grades; survey or focus group responses; pre and post tests;<br>• Less commonly used: scores from assessment instruments ( e.g CLASSE); statistics derived from LMS re usage (e.g participation in discussion forums, which can be used to compare with final grades or to comment on level of engagement); samples (analysis of student output)<br><br>Consider obtaining data from a different source:<br>• partnerships with others, including departments, colleagues,<br>• institutional data (registrar; NSSE).<br>• Campus organizations, such as Student Life, AccessAbility, etc. that may have research data<br><br>Determine how to efficiently collect more next time. |
| Concerns about whether the impact of the program or service will be measurable | Set up your measurement to an appropriate scale. For example, don't expect that one workshop will measurably alter students' final grades. Instead try the following:<br>• focus on what the program is intended to improve. For example, a language workshop may improve students' vocabulary, but it might not significantly improve their English course grade even though vocabulary is one of the components in an English course.<br>• use a pre- post- test based on your key learning objectives<br>• focus on assessing participants' perceptions of what they've gained from the experience<br>• conduct a follow-up survey with participants to see if their initial perceptions still hold<br>• combine data across semesters, allowing you to have a larger sample—the larger the sample size, the greater the likelihood that even a very small change will be statistically significant |

**Exercise 3: Putting it all together**
Lay out one or two sample research designs using the following table

| **Teaching innovation** | **Research Question** | **What will I be measuring? (participation/perception/performance)** | **What approach am I using? (correlational/ experimental)** | **What data will I need?** |
|---|---|---|---|---|
| *Added optional facilitated study groups (FSGs)* | *Did students who attended FSGs perform better than students who did not?* | *Performance related to participation: Difference between grades of students who attended vs students who did not attend* | *Experimental – Test group: students who attended at least one FSG Control group: students who did not attend* | *Attendance data for all students at all FSGs*<br><br>*Final grades of all students* |
| | | | | |
| | | | | |

**Alignment**

One of the key principles of designing an effective study is ensuring that all elements line up. Two key questions to ask yourself about your research design:

1. Is my Research Question consistent with the goals or learning objectives of the program, course, assignment, or activity?

2. Am I measuring what I think I am measuring? i.e. is an improvement in grades a sign of improvement in learning?

Finally, it is worth considering whether you have any reason to believe that your results are or will be unusual? Are there any unusual characteristics of the group of students or the learning context, or the data gathering process that might affect the results?

## Section 2: Working with Data

### Organizing Data

Once you have collected your data, whether on paper or digitally, you will need to collate and structure it into a usable form. The following are a few principles of data organization to help you with this procedure:

1. *Stay raw*. If you pre-compile data, you limit its usefulness for future questions and analyses.

   Consider the following example: In a writing centre, students complete paper perception surveys of one-on-one tutoring on an ongoing basis. Because these have primarily been used for feedback on the performance of individual writing instructors, and in order to save time on data entry, data is compiled and entered for each instructor. This provides valuable information about instructor performance and enables the centre to compare instructors, but does not allow the data to be mined for other information. If instead data were entered in its raw form, it could more easily be re-sorted (e.g. by student) in order to ask different questions: e.g. is there a correlation between the number of appointments a student has had and their rating of their learning about writing? Or, is there a correlation between the discipline and writing topics covered in appointments?

2. *Think granular*. To maximize flexibility when analyzing and interpreting data, enter information into discrete categories wherever possible. Avoid putting two pieces of information together (e.g. first name and last name) when you can keep them separate.

   For example, the term in which a specific placement occurred was entered as part of the title of that placement for a service learning program. This helped ensure that recurring placements were counted accurately (each offering of the same placement appeared as unique because of the term indicator), but at the same time this approach made it more difficult to tabulate how often a given placement had been offered over the years, as well as how many and which students had participated in that particular placement.

   The solution to this problem may be to separate out placement title and term, but note that it may be worthwhile to have redundant information in multiple columns (or rows). To extend the example above, one way to resolve the problem would be to keep the column with the unique name (placement + term), as well as adding another two columns, one with just the placement title and the other indicating the term. This redundancy would allow for the same data to be sorted according to different criteria and for different purposes.

3. *Be linear*. Align data in columns and rows systematically, ensuring clear categories for each aspect of the data, with no repetition of headings or interruption. Either one of the following options can work well; choose the one that makes it easiest for you to navigate your data.

Example A

| Participant | Question# | Response |
|---|---|---|
| | | |
| | | |
| | | |

OR

Example B

| Placement Title | | | |
|---|---|---|---|
| Term | | | |
| Student ID | | | |
| Student ID | | | |

The goal here is to make your data easy to tabulate, not easy to read. When making your decision, consider how much and which direction you will need to scroll to view your data and whether you are working with a finite data set – or will be adding new information over time. In general, Example A will be easier to work with, especially for data with a large number of participants, because it is easy to scroll vertically through a large list of participants. However, Example B may work better if you will be adding in information about new participants at a later date.

Note that you can also transpose columns and rows if you find that your initial set up isn't working well for you. To facilitate this, it is helpful to keep your raw data separate from your analyses and calculations. (See also Pt. 3 in Managing Data, below, for a caution about transposing data.)

4. Remember: there is always strength in numbers
   Numbers are much easier to tabulate than text. So, for example, instead of using "male" and female" as cell entries indicating the sex of participants, convert this data to numerals such as 0 and 1. This will make later calculations much easier to formulate (but do be sure to keep an index, so that you can remember what the numbers represent).

**Exercise 4: Organizing Data**
In the two blank tables below, test out a couple of different possible structures for your data set. Which alignment do you think will work best for you? Will you need to further manipulate your data to make it easy to analyze?

| | | | | |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

| | | | | |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

**Managing Data**

A few extra tips to help you manage your data effectively:

1. *Always keep a copy of raw data.*
   If you need to merge (or add) columns or rows to facilitate calculations, be sure to keep an untouched version of your raw data so that you can go back to it if need be.

2. *Document your changes.*
   Keep notes or indexes of the changes or calculations that you have made. This can be done in headings or as annotations on a separate sheet – whatever works best for you and your situation. The main requirement is to keep the information somewhere where it will be easy to find. Otherwise, you may return to your data several months later and not be able to make sense of it.

3. *Be aware that changing columns and rows may affect your calculations.*
   When using a spreadsheet, many formulae will refer to other cells. However, if you find later that you need to add a new column or row, or otherwise shift your data, you can accidentally erase all of your hard work. Key to avoiding this problem is to be aware of *absolute* and *relative references*. With the former, your reference will target the specified cell, regardless of what other changes you make to your file. With the latter, on the other hand, your cell references will automatically update when you add in new columns or rows. If you choose wisely in the initial set-up, you can preserve absolute references when you need them, and update other columns or rows automatically when you add in new calculations or information.

4. *Consider abstract references*.
   Ongoing data organization and management can be facilitated by referring to cells abstractly, that is by identifying the row and column numbers, rather than the content of the cell. Doing this allows you to create a template that you can easily plug additional or different data into.

**Visualizing Data:**
Once your data is organized, the next step is to take a brief snapshot to see if the results are as you suspect. Inspecting data visually can give you an overall summary and help you determine whether it is worthwhile to spend the effort to conduct more detailed statistical analysis.

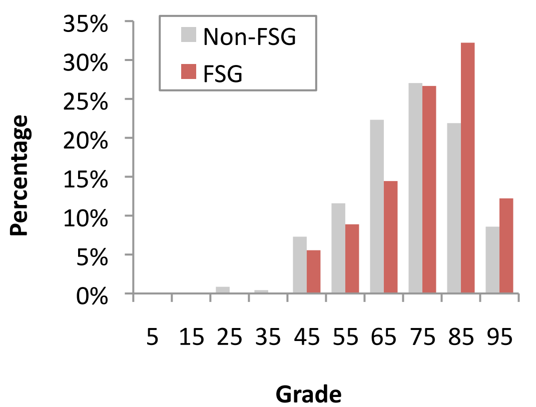
The most common ways to visualize a data set are histograms and scatterplots. The former work well for the experimental comparison of control and test group, whereas the latter are ideal for correlational data:

**Experimental Approach**

Histograms

A histogram is a bar (or line) chart that displays the distribution of your data, shows whether the data is normal or skewed, and clarifies the frequencies of specific intervals. Overall you gain a picture of what, if any, differences there are between your control group and test group.

**Figure 3: Relative frequency histogram comparing final grades of students who participated in FSGs with those who did not.**



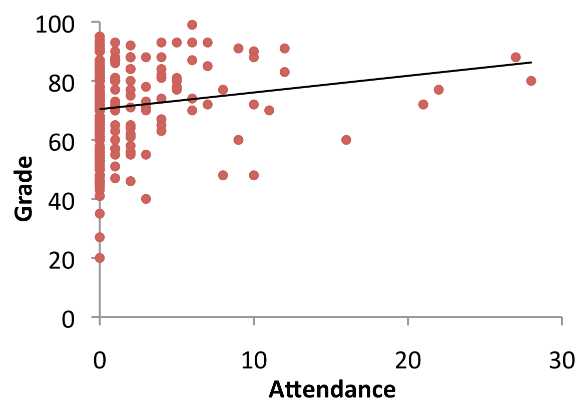

This graph quickly reveals that more of the FSG group earned an A or B, while more of the non-FSG group earned lower grades. Visual inspection also reveals that the peak (the highest frequency count) of each group is also different. In the FSG group, the peak is in the B range; whereas for the non-FSG group, it is in the C range. This indicates that there is indeed a difference between the two groups, and that further analysis could likely be fruitful. (See also Fig. 6 for another method of quickly visualizing differences between control and test group.)

**Correlational Approach**

Scatterplot

Scatterplots are used for correlational data to see if two variables are related. Each dot represents an observation that is located according to the x- and y- axes. In the example below each student is located by the number of times they attended FSGs and the final grade they received.

**Figure 4: Scatterplot correlating final grades of students who participated in FSGs with the number of times they attended study groups.**




This graph quickly reveals that there is no correlation between the grades participants earned and the number of times they attended because the observations are not clustered around the regression line*. Students who attended only once, earned grades as low as a D and as high as an A+, as did students who attended up to 5 times. This tells us in one glance that further statistical analysis is not likely to be worthwhile. (But see p. 26 for suggestions of other possible ways to examine this data.)

**The line that best fits the data because it comes as close as possible to all of the points.*

Here is some information on creating your own histogram or scatterplot.

***How to create a Histogram***

1. Identify the range of data that represents your control group and test group. A spreadsheet of continuous data will need to be sorted by whichever column (or row) contains the unique identifiers for each group.

2. Define *bin limits*: Bin limits are categories that delimit the increments that will be compared. In Figure 1, the final grades on the x-axis are divided into increments of 10. This number was chosen because it reflects our common understanding of grades in categories from A - F. The axis labels then reflect the midpoint for each bin.

   If it turns out that your frequencies are concentrated in a couple of bins, you might want to decrease the bin size in order to increase the level of detail that you can see. Conversely, if you have very few counts in all of the bins, you might want to increase the bin size, to better enable generalizations.

3. Create frequency columns for both control and test group. This column should reflect the number of observations that fall within each bin. For example, in Figure 1, no observations were found for FSG participants in the 10 - 20 bin, but a few students from the control group were. [Note that bin limits and frequency columns should be adjacent or in a separate table for easy graph construction.]

4. Generate the chart. This is typically an automatic procedure in most spreadsheet and statistics software.

***How to create a Scatterplot***

As scatterplots plot every observation in the data, they are very simple to create. In Figure 2, each dot represents one participant in the FSG program, located by the number of times that student attended (x-axis) and the final grade (s)he received (y-axis). Here are basic instructions:

1. Identify the range in the data that you will be correlating. In Figure 2, the range included two columns: the number of times each participant attended and the final grades each one received.

2. Define x and y axes. Each of the variables you will be correlating will be on one or the other of these axes. The convention is to graph the predictor on the x-axis, and the criterion on the y-axis. In Figure 2, we are using attendance to predict grade, so we put attendance on x-axis and grades on y-axis.

3. Insert regression line. This is typically an option available in most spreadsheet and statistics software.

4. Create chart.

Try generating a histogram or scatterplot with your own data. If you are having difficulties figuring out exactly how to do so, you can use the terms and procedures outlined here to enable your search through the spreadsheet or software program's help files.

# Section 3: Analyzing and Interpreting Data

**Descriptive Statistics**
Descriptive statistics are essentially quantified summaries. Taken together, they comprise the total of your observations and provide a brief snapshot of key aspects in a numerical format. Depending on your specific purposes, however, you may not need to use all of the possible options:

- ***Averages*** include the most common descriptive statistics: mean, median, and mode. Together these three types of average reveal the central tendency of the data -- that is, what the most common data point is. Comparing these three numbers ensures greater clarity and accuracy when reporting results from anomalous data. For example, a mean of 65 could be the result of a normal bell curve distribution, but it may also be a product of a group of students who did extremely well and another that did poorly. Reporting median and mode can help clarify which is truly the case.

  Note that while it is common practice to calculate averages from Likert-style surveys, many argue that this is not best practice because it relies on the assumption that the intervals in the scale are equal, when often they are not. The difference between “very good” and “outstanding”, for example, is not necessarily equal to the difference between “good” and “very good” in the same way that the difference between 4 and 5 is the same as the difference between 2 and 3. As such, averages do not necessarily accurately reflect the data collected.

- ***Frequencies*** measure the number of times specific data points occur. For Likert-style perception surveys, it is considered best practice to tabulate in terms of frequencies or percentages.

  For example, after attending one-on-one writing centre tutorials, students completed a perception survey in which they ranked the quality of instruction on a seven-point scale, with 7 being outstanding, 6 being very good, 5 good, etc.

  Rather than presenting a weighted average, as is commonly done (e.g. *average student rating of instruction was 6.36 on a seven-point scale*), it is more accurate – and sometimes more effective -- to state it in terms of percentages: e.g., *88% of students attending one-on-one writing centre tutorials rated the quality of instruction as “outstanding” or “very good.” No one ranked instruction below “good.”*

- ***Standard Deviation*** allows you to see the variance in your data. This calculation determines the range of data points, both above and below the mean. See Figure 5 for an illustration.

- ***Skew*** describes a type of distribution that deviates from the norm. Whereas normal distribution appears as a symmetrical bell curve, skew indicates asymmetry, and will appear as a tail

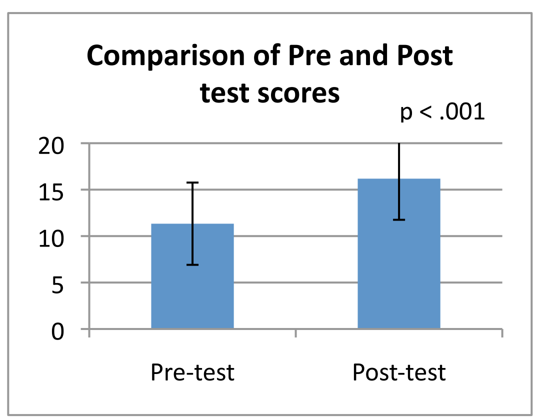



**Figure 5: A bar chart comparing means of pre and post test scores from a summer learning institute program. Vertical error bars indicate the standard deviation, and indicate that overall post-test scores are higher.**

extending from either end of the peak. Negative skew is elongated on the left side; positive skew is elongated on the right side. While visual inspection can indicate skew, actual calculation of skew allows you to determine whether the assumptions necessary for further statistical testing hold. (This calculation requires both mean and standard deviation).

- ***Correlation*** determines the degree of relationship between two variables. This relationship goes beyond easily observable comparisons such as whether one score is higher than another, in that it measures whether there is a direct relationship in scattered data.

  To measure correlation, create a scatter plot and insert a regression line in order to visually inspect your data. At this point, you will be able to judge whether a correlation is likely because the data points will be clustered around the regression line. (See Figure 4, above, for an example of uncorrelated data.) The more closely the points are clustered around the regression line, the stronger the correlation is likely to be because the regression line is itself based on the trend of the data. This visual information can then be confirmed by calculating the actual correlation. The result will be somewhere between -1 and 1. The former indicates a perfect negative correlation, while the latter indicates a perfect positive correlation. A result of zero indicates that there is no correlation between the two variables.

  In many cases you will find that there is some amount of correlation, however small, between the two variables, and so the next step would be to determine whether the variables are actually related to each other. Testing for the significance of the correlation will answer this question. (See *Inferential Statistics* below for more details.)

> **Tip**: The calculations described throughout this document do not necessarily require you to know the math. Most can be done automatically with Excel or other statistics software.

**Inferential Statistics**

Inferential statistics are what most people think about when they think of "statistics." Going beyond mere summary, inferential statistics allow you to conduct further analysis and draw meaningful conclusions about the effectiveness and impact of your innovations on a wider population.

The starting point of inferential statistics is always the null hypothesis. Any differences seen through descriptive statistics are initially assumed to be a product of chance – as this is the most probable explanation. In the simple terms of odds, the very foundation of inferential statistics, a difference between groups is much more likely to be coincidental than meaningful. Consider, for example, the differences illustrated in Figures 3 and 5 above. In both cases, there are clearly observable differences between the two groups. But what exactly, explains those differences? While we certainly hope that it is our interventions and innovations, there are a plethora of alternative explanations, including the possibility that the results are simply coincidental.

One reason for this is that the groups being compared are always a sample of a much larger population – and oftentimes a very small sample at that. As such, the sample may not be sufficiently representative of the larger population to draw any meaningful conclusions. But even in ideal circumstances with impeccable research design and large samples, inferential statistics are always drawing conclusions from observed differences – and without rigorous procedures to constrain our inferences, we open the door to a wide array of possible errors and spurious conclusions. Statistical calculations provide the requisite mathematical rigour to help us determine whether the null hypothesis holds (i.e differences are likely due to chance) or can reasonably be rejected (i.e. differences are quite likely to be an effect of our intervention).

Understanding the Null Hypothesis

| | Null Hypothesis | Alternate Hypothesis |
|---|---|---|
| Purpose | Acknowledge that any difference, effect, or correlation may be entirely due to chance | Identify a possible explanation for a difference, effect, or correlation, with the hope of rejecting the null |
| Effect | Independent variable has no effect on dependent variable | Independent variable does have an effect |
| Resulting predictions | Observed differences are not significant | Observed differences are significant |

- **Significance**

Significance is the most common measure to determine whether the differences seen from the descriptive statistics are a result of the program or simply due to chance. There are a number of ways to test for significance:

- T-test is most commonly used for interval or ratio – i.e. continuous--data (e.g grades). Two assumptions required for t-test include that the initial data is normally distributed, and that the data in both groups has equal variance.
- Chi square is most commonly used for nominal or ordinal – i.e. categorical -- data (e.g. relationships between demographics and choice of discipline) or ordinal data (e.g. Likert-scale surveys). While it is not uncommon for researchers to treat Likert-scale ordinal data as interval data, technically the Chi square is more appropriate test.
- Analysis of Variance (ANOVA) can be used for both types of data, and is very useful for comparing differences between three or more groups.

Each of these tests will give a value which can then be used to calculate the percentage likelihood that results are due to chance (the p-value). For example, if the p-value is .05, this is interpreted as a 5% possibility that results are due to chance.

> **Tip**: Most spreadsheet programs and statistical software will provide you with the p-value, or percentage likelihood that results are due to chance, in a single step.

The p-value is central to determining significance because it gives us reason to reject (or fail to reject) the null hypothesis. If the p-value is high (e.g. 0.3), it is very likely that results are due to chance (30%), which means that the null hypothesis is likely correct. The lower the p-value, the less likely it is that the event occurred by chance – giving us greater reason to reject that null hypothesis.

Where exactly the cut-off for determining whether a p-value is or is not significant varies depending on the type of data one is working with and other factors, including the associated risks encountered if there are errors. For example, for a medical treatment, the threshold for significance may be as low as <.001 (or less than 0.1% possibility that results were due to chance) whereas the standard in educational research is typically <.05 (or 5% probability that results are due to chance).

- **Effect size**

Effect size allows you to measure the magnitude of the observed differences between your test group and control group. There are a number of possible formulae to calculate effect size; which one you choose will depend on the significance test you are using. For the T-test, effect size is calculated using a statistical test called *Cohen's d*; for correlation, effect size is calculated with a test called *Pearson's* $r^2$. Effect size calculations complement significance testing by quantifying the size of the impact. For example, if a calculation reveals an effect size of 0.8, this can be interpreted as the intervention having a fairly large impact on the group. Coupled with a significance test that reveals the p-value to be 0.05 or less gives us good reason to conclude that our program or service was a success.

**Exercise 5: What statistics would you use in the following scenarios and why?**

***Scenario 1***: You've developed a new approach to teaching a co-curricular writing workshop that involves more active engagement from the students. Students signed a participation sheet at the beginning of the workshop and filled out a survey evaluation form at the end. What statistical tests might you employ to analyze the information you have currently? What additional data might you obtain, and how might this change your analysis? (See Appendix A for an answer key.)

***Scenario 2***:

You have introduced a new assignment into your class, requiring students to participate in online discussions of weekly readings to help support their learning. Although students did not receive marks for these assignments directly, you have tracked participation through the learning management system. You know how many times each student has posted, and how long each post is. All grades for the course are in. How would you go about analyzing this data statistically? (See Appendix A for an answer key.)

***Scenario 3***: Look at your own research question and data. What tests are most likely to be of value to you? And why?

**Data interpretation**

Because statistics quantifies the possibility that any results are entirely due to chance, there are a number of issues that can arise when drawing conclusions.

***How significant is significance?***

While a useful measure for determining whether the results are due to chance, the value of significance should not be overestimated. In a large sample, for example, very small differences can register as significant. Yet, the differences might not have any practical importance. For example, in a large class of 1500, you might find one group earned better grades than the other, but by only, say, a percentage point difference in scores. While this may be statistically significant, we normally would not consider it very important that one group earned on average a 65, while the other earned a 66.

With small samples, on the other hand, statistically significant differences may still be a product of chance because the odds that a coincidental connection will be found in a small sample are very high. For example, if you notice that a sample group of 10 participants has a higher average than another sample of non-participants, it could still easily be coincidental that those 10 students had higher grades. In cases like these, calculating effect size can be helpful, as it provides a sense of the magnitude of the difference without the influence of sample size.

Even with an ideal sample sizes, however, there are always other possible explanations for significant results.

**Intervening and Confounding Variables**

Even when data show that a teaching innovation has had an effect, the possibility remains that the innovation itself did not actually cause that effect. Instead, it is quite possible that some other factor has interceded. On one hand, the effect could have been mediated by a third variable: the *intervening variable*. On the other hand, the effect might have been caused by another, altogether unaccounted for, variable: the *confounding variable*.

*Intervening variables* are those in which the relationship between the intervention and the result is not direct, but supported by another, related variable. For example, if every student who participated in a research skills workshop showed significantly higher grades than those who did not, it would certainly be tempting to conclude that this was because of their participation in the workshop. But other factors may be at play: perhaps the students who participated had some other feature in common, e.g. were all highly motivated, or above-average intelligence. Or perhaps they all coincidentally took part in another course or program that led to the observed gains.

In cases like these, the conclusion may very well be that the *intervening variable* (e.g. student motivation or intelligence) is more directly responsible for the observed results, rather than the *independent variable* that is the focus of the analysis (successful completion of the workshop).

*Confounding variables* are those that lead to an entirely different explanation for the observed differences that were not addressed in the analysis. For example, if one instructor introduces a new pedagogical approach, such as discussion boards or reading blogs into one section of a course, while another instructor doesn't in another section of the course, it could very well be that any difference in average (assuming there is one) is a result of the difference in the instructor, rather than the specific pedagogical approach.

The possibility of intervening and confounding variables will inevitably temper any certainty that we may have about our conclusions.

***What can I do to overcome the problem of intervening or confounding variables?***
It isn't possible to control for everything, especially with educational research where any number of factors may affect a student's learning or perceptions. And this problem is magnified by the approach we are taking in this workbook, making the most of existing data, rather than crafting a full-scale research study. Even so, there are a number of possible solutions to counteract this problem:

1. *Triangulation*
   Look for ways to synthesize multiple data. One way to do this is to approach data from different perspectives. Statistical significance alone, for example, suggests that the results are not a product of chance, but there still may be an intervening variable. Participants, for example, may all have been highly motivated. But if you can then also show a high positive correlation between number of times attending and performance (e.g. on a post-test or final grade), then the intervening variable is mitigated as there is now greater reason to conclude that motivation alone is not sufficient to explain the increases in performance.

   Triangulation can also be accomplished by compiling information from different sources. Perception surveys can complement performance measures; institutional or national data can support individual research projects; results on standardized assessment instruments can provide baselines for comparison, and so on.

2. *Refer to external literature*
   For common programs, services, or teaching strategies, there is likely already a significant literature documenting the effectiveness of various aspects. Similarly, most disciplines have at least one journal dedicated to teaching in that area. Even for unique or individualized workshops or truly innovative approaches, you may be able to find related research by searching key words from your research question or the specific activities or tools that you've used.

3. *Pool data from across terms or from different centres or sources*
   This solution is especially helpful for overcoming problems related to a small sample size, but can also be used effectively to find additional measures of performance or perception.

4. *Administer an independent assessment*
   This is rarely a practicable solution for working backwards and assessing existing data, but can be extremely useful for ruling out intervening and confounding variables. Independent assessment instruments (e.g. the Teaching Concept Inventory, the Student Opinion Scale, the Academic Proficiency Scale) can provide a valuable baseline by which to compare and assess participants and non-participants. The National Centre for Postsecondary Improvement at Stanford offers a list of instruments that may come in handy for this purpose: http://www.stanford.edu/group/ncpi/unspecified /assessment_states/instruments.html#basicSkills

**Troubleshooting Data Analysis and Interpretation**

| | |
|---|---|
| My data was not normally distributed, and the assumptions of typical tests did not hold. | Unless data deviates substantially from normal distribution, it is common practice to continue to use the statistical tests described here.<br><br>However, in cases where the distribution is substantially skewed, different types of statistical analysis (that are beyond the scope of this introductory workbook) will be required. For more information see the section on Looking Forward. |
| I do not have a comparator or control group. | To create a substitute control group, you can compare your results with a fixed standard. Suppose, for example, a learning skills centre is able to track the grades of students who have participated its service(s), but has no access to information from students who didn't. Instead of an authentic control group, you could compare the performance of your students with a predetermined value, such as a course, departmental, or institutional mean grade, whichever matches most closely with the sample you are analyzing. Remember, though, the more removed the comparison, the less likely it is valid.<br><br>To conduct this test, you will need to create an artificial control group, in which each member is pre-determined to have earned the standard mean. (That is, ultimately this approach is following the procedures of a two-sample t-test, with only one sample.) |
| | *One-sample t-test*:<br><br>Another procedure would be to generate a measure of improvement for each student over time (e.g. through pre- and post- tests, exam scores, writing grades, or other), then calculate the mean of this improvement plus standard deviation. From this, the one sample t-test will help determine if the improvements are significant). Effect size calculations can also supplement one-sample t-test results. |
| My sample sizes are too small | Small samples can be problematic as even statistically significant results could still be due to chance. Solutions: examine effect size to see if it is large; pool data over terms or years; or combine data with other programs. |

| | |
|---|---|
| There doesn't seem to be any correlation between my variables. Does that mean my program or service doesn't have an impact? | It can be very difficult to find strong correlations in educational research, as there are so many factors that can come into play. However, you may be able achieve different results if you re-analyze your data with different parameters.<br><br>For example, suppose that you have found that the number of times a student participated in an activity does not correlate with performance in a linear fashion. (That is, those who participated twice did not always do better than those who participated once; those who participated 8 times didn't always do better than those who participated 6 times, and so on). In this case, participation data could be re-categorized and re-analyzed to see if, e.g., those who participated 6 -10 times did better than those who attended 1 - 3 times.) |
| I'm worried my data will be meaningless because of intervening and confounding variables | You can reduce the likelihood of intervening variables by reducing individual variance, that is by measuring results in the same individuals over time, both before and after the intervention. When results are paired in this way, the problems generated by different levels of motivation, intelligence, or skill are mitigated. The appropriate significance test in this scenario is the two-sample paired T-test.<br><br>For example in a course focusing on English Language Development, students are given a vocabulary test in the middle and at the end of term. Because it is the same individuals that provide both the control and the test group, the problem of individual variation does not arise. |
| | Unfortunately, the above approach is not always possible, especially when working backwards. Intervening variables can also be addressed by administering independent assessments to measure things like motivation, basic skills, and so on. These provide a baseline for controlling for the different variables. |
| | A third possibility is to mine other data sources. Perception surveys, for example, can get at students' evaluation of their own abilities. While this data may not be as accurate, it can at least give a clearer picture of who is participating in your teaching innovations. For example, is it only top students participating or the full range? Are there more students with low confidence about their abilities or more students with high levels of confidence? |
| I'm not sure if my innovation will have a big enough impact to even measure. | Attend to the scale of your intervention when deciding what to use as your outcome indicator. For example, when assessing the effect of a single workshop on performance, consider that it is much more likely to have an impact on a post-test or related assignment, than on an unrelated assignment, final grade in a course, or overall GPA. |

## Section 4: Ethical Issues

Up until now, we have focused on the analytical side of educational research projects. But there is another, equally important, aspect of educational projects that has to be considered – ethics.

Ethics is a key consideration of any educational research project and our document would not be complete without an acknowledgment of its importance. There are many ethical considerations that go into the design and conduct of educational research projects. Singleton notes these can be roughly organized into *"….three broad areas of ethical concern in scientific research: the ethics of data collection and analysis, the ethics of treatment of participants, and the ethics of responsibility to society."* (p. 48). Below is our brief outline of some key issues; there are numerous good and important resources that you can use to become well-informed.

For educational research purposes, a critical document to be aware of is Canada's *Tri-Council Policy Statement – Ethical Conduct for Research Involving Humans* (TCPS 2). This document is a joint effort of Canada's three largest research funding organizations -- Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and the Social Sciences and Humanities Research Council of Canada. This policy includes guidelines for the work of Research Ethics Boards (REBs) which are the local institutional units that would provide an ethics review of your research project, if one is required.

Two key areas of concern for an education researcher are: a) whether your project needs an ethics review, and b) clarity of the potential ethical issues regarding the human participants – your students – in your project.

**a. Ethics review of your course/program research project**
In educational institutions, instructors are not required to seek ethics review for the normal evaluation and assessment in their own courses (Article 2.5, TCPS).

If your research study goes beyond 'normal educational requirements', it still might not require an ethics review. It depends on exactly what your research plan proposes to do. Having said that, even if you determine that ethics review is not necessary, but there is a chance you might wish to publish your research at some point, it might be prudent to seek an ethics review up front, saving you from the complications of seeking it post-research. The REB might be willing to approve your project after the fact, but you might face issues such as obtaining student consent.

For the ethics review process itself, Cresswell (p. 158) outlines four general steps:

a. start by finding out about the review process used by the …..review board on your campus.
b. determine what information the review board needs about your project.
c. develop an informed consent form for participants to sign before they participate in the study.
d. submit a description of your proposed study to the institutional review board. *NOTE: in Canada, REBs will require specific, detailed information not just a description of the project.*

When planning your project, be sure to allocate sufficient time for the various steps, so that it can be accomplished within your timeline.

**b. Human participants – your students**
A foundational tenet of all research involving human participants is that the research must do no harm to those participants.

In educational research, there are possible situations or conditions that can cause harm to student participants. These could include power differentials, vulnerability, and participant burden (Elgie, p. 16). Owen (2006) also notes issues of conflict of interest between the roles of teacher and researcher, and the complexity of obtaining meaningful informed consent.

These critical issues will be raised explicitly by official ethics review processes, so that if you do go through that process, you will be asked to address them. However, considering them at the start of your project may help you determine whether an official review is necessary or appropriate, and will help you prepare for such a review if it is necessary.

Owen (2006) also outlines several specific areas of concern to be addressed: confidentiality of participant data, data collection, voluntary participation and informed consent.

**i) Confidentiality**
An essential consideration when working with student data is student privacy, whether the data is gathered in a course or through other types of educational programs. But it is more than just keeping a student's name confidential; it also includes protecting data from unauthorized access, loss or theft. There are numerous possible strategies to assist in maintaining confidentiality, for example: anonymizing data so that no one, including you, the researcher, can identify individual students; keeping data secured in a locked cabinet or password-protected computer and avoid transmitting or transporting files on a laptop or USB key that could be stolen; and creating a plan to destroy data when it is no longer needed.

**ii) Data collection**
Your method of data collection may directly affect the need for ethics review. One consideration is if your data collection process is invasive, e.g. videotaping students is more invasive than conducting a perception survey. Another consideration is if you anticipate using some deception in your research where too much information might skew the results. In this case you might provide less than full disclosure to participants, thus ethics review will be required. Also, there might be situations where a normally minimally invasive process, might be more invasive or outright inappropriate if there is some vulnerability on the part of the participants (Guidelines, p. 4).

**iii) Voluntary participation**
Many educational programs, like writing tutorials, are ones in which students choose to participate, and are not tied to grades or requirements. You might have problems such as selection and recruitment of volunteers, tracking students' progress over time, or students' refusal to participate or dropout of the study (Elgie, p. 27). Elgie provides a useful tipsheet (p. 28) outlining various strategies that can be used, such as sampling many more participants than you would otherwise expect to, offering small incentives, or appealing to their pride in their program/university.

A different situation is where you might want to experiment in your course or program with an innovation that will support student learning. You might wonder if it's ethical to deliberately restrict the use of this innovation to one group of students and not another. In practice, one way teachers/researchers get around this is to make optional the innovation whose impact they want to study. Of course, this self-selection raises the possibility that motivation becomes a confounding variable of the study, which then needs to be addressed in the data interpretation.

The University of Toronto ethics review office addresses this issue, stating that if a teacher/ researcher wants to involve some students and not others, *….a third party should be involved in recruitment and selection to provide some distance between teacher/researcher and student/participant. The teacher/researcher should not be aware of who has agreed to participate while the teacher-student relationship still exists.* In the context of a course, it is recommended that identifiable data be analyzed only after grades have been submitted to the school so that *"a real or appearance of potential for evaluative effect on student/participants no longer exists"* (Teacher-Researcher Conflicts of Interest, 2003).

**iv) Informed consent**

If you are gathering data outside normal educational requirements, you need to inform participants about the research project, and obtain their consent. Your informed consent letter would include information such as a statement about the research project itself, the identity of the researcher, the expected duration and nature of participation, a description of research procedures, the responsibilities of the participant, and a description of all reasonably foreseeable risks and potential benefits, both to the participants and in general. You may also need to include information about if and how participants can with draw from the study. Drafting the letter may be challenging, so you might want to obtain examples of consent forms from the REB office, colleagues or other educators/researchers on your campus. TSCPS 2 provides specific information about informed consent letters.

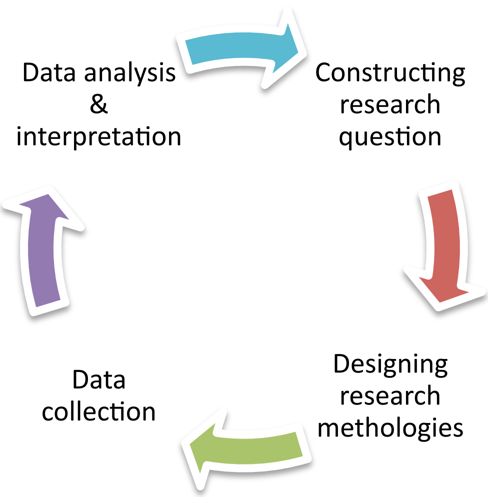
Ethics is a key consideration of any educational research project, regardless of whether ethics review is ultimately needed. By reviewing these issues right from the start, you can be assured your research is ethical and appropriate, and save yourself time and complications later on.

## Section 5: Looking Forward

In the previous sections, we've provided a brief overview of some of the key questions, statistical tests and approaches that you might use as you start conducting educational research, as well as a brief overview of ethical considerations. Our approach of working backwards is designed specifically to provide suggestions for how faculty and program coordinators can maximize the use of their existing data for formative and reporting purposes. However, once you have gone through this process, you may discover that you are not satisfied: perhaps you have a different research question you want to examine, or another teaching innovation that you want to investigate. Or perhaps you would simply like to improve your methodologies and measures and re-run your study to eliminate some of the uncertainties or limitations you have encountered. We have a few suggestions.

### a. Improving Research Design

Research is a continuous and iterative process where our results inform our future questions.




To achieve more robust conclusions and results, we can only stress the importance of planning. Hopefully, by working through this booklet, you are now more aware of the gaps and limitations in your existing process, and have a clearer sense of the different ways that you may improve in the future.

To counter some of the problems of intervening and confounding variables, discussed above, there are several possible tools that can help.

- Pre- and post- tests: The advantage to implementing pre- and post- tests is that data is being gathered from the same students both before and after the intervention. This reduces the problem of intervening variables such as motivation and baseline skill levels because improvement can be measured, rather than just performance on a single test.

  When designing pre- post- tests, aim to test for more than recall or knowledge-based items; ask questions that require students to demonstrate skills and values. This will help you assess student improvement on a deeper level.

- Perception surveys: To help overcome problems related to self-selection biases and to make the most of evaluation forms and other perception instruments, the following tips can help:

  - Include negatively phrased questions, to ensure that students are reading the questions, and not simply plugging in any answer to get to the end.
  - Use language carefully and ensure that questions are unambiguous and that the terminology is familiar to respondents.
  - Add the option "other" because it is important to know if person deliberately did not answer or accidentally missed the question.
  - Given that respondents will often have a variety of perspectives, it is useful to have clarifying questions that will help pinpoint the reasoning behind the responses.

**b. Improving Statistical Analysis**
This workbook is intended as only an introduction to the basics of quantitative analysis, aimed primarily towards faculty and staff that have little or no background either in quantitative analysis

or education research. As such, there are many possibilities for developing more sophisticated approaches that are beyond the scope of what we are able to cover here. For example, your research questions may lead you to wish to analyze and control for multiple variables (multivariate analysis). Or perhaps your data is not normally distributed, and you are seeking statistical tests that can provide more accurate results (e.g. non-parametric statistics).

To address these further questions, there are a number of resources available, some of them likely right on your own campus. Statistics departments often offer consulting services, either formally or informally. Departments of Education and Psychology also often have friendly neighbourhood stats specialists that can help you troubleshoot or find appropriate tests. Another valuable resource would be your local faculty development or teaching and learning centre.

The internet is also a wonderful resource, and if you have questions about how to run specific calculations or use specialized software programs such as Excel, OpenCalc, or SPSS, there is very likely a video or webpage explaining the process.

## Section 6: Conclusion

This workbook is meant to be a practical introduction to educational research practices for educators who are interested in conducting research within their courses or programs. As indicated in the introduction, it has been written primarily by practitioners without formal training in either educational research or quantitative analysis. We have been developing our knowledge through trial and error and we would love to hear from you. Has this workbook been useful for you? What problems have you encountered and how have you resolved them? What questions remain? Please do not hesitate to contact us:

Allyson Skene askene@utsc.utoronto.ca
Sarah King sking@utsc.utoronto.ca

## GLOSSARY

*Quantitative research is a type of educational research in which the researcher decides what to study; asks specific, narrow questions; collects quantifiable data from participants; analyzes these numbers using statistics; and conducts the inquiry in an unbiased, objective manner. (Cresswell, p. 46).*

**Confounding variable** - a variable which is not a factor being considered in an observational study or experiment, but which may be at least partially responsible for the observed outcomes. Experimental design methods use randomization to minimize the effect of confounding variables, but that is not possible in observational studies (Dictionary of Mathematics).

**Correlation** - a number that summarizes the direction and degree of relationship between two or more dimensions or variables…….correlation coefficients can range from -1.00 to +1.00. The value of -1.00 represents a perfect negative correlation while a value of +1.00 represents a perfect positive correlation. A value of 0.00 represents a lack of correlation (Statsoft).

**Descriptive statistics** – present information that helps a researcher describe responses to each question in a database as well as determine overall trends and the distribution of the data (Cresswell, p. 638).

**Effect Size** – is a means for identifying the strength of the conclusions about group differences or about the relationship among variables….. (Cresswell, p. 639).

**Experiments** – …(test) an idea (or practice or procedure) to determine whether it influences an outcome of dependent variable…..(used) when you want to establish probable cause and effect between your independent and dependent variables (Cresswell, p. 299).

**Inferential statistics** – enable a researcher to draw conclusions, inferences or generalizations from a sample to a population of participants (Cresswell, p. 640).

**Intervening variable** – is an attribute or characteristic that “stands between” the independent and dependent variables and exercises an influence on the dependent variable apart from the independent variable. Intervening variables transmit (or mediate) the effects of the independent variable on the dependent variable (Cresswell, p. 641).

**Measurement scales** – ….variables are classified as:

a. **Nominal** variables allow for only qualitative classification. That is, they can be measured only in terms of whether the individual items belong to some distinctively different categories, but we cannot quantify or even rank order those categories…..Typical examples of nominal variables are *gender*, *race*, *color*, *city*, etc.

b. **Ordinal** variables allow us to rank order the items we measure in terms of which has less and which has more of the quality represented by the variable, but still they do not allow us to say "how much more." A typical example of an ordinal variable is the socioeconomic status of families….. Also, this very distinction between nominal, ordinal, and interval scales itself represents a good example of an ordinal variable.

c. **Interval** variables allow us not only to rank order the items that are measured, but also to quantify and compare the sizes of differences between them. For example, temperature, as measured in degrees Fahrenheit or Celsius, constitutes an interval scale.

d. **Ratio** variables are very similar to interval variables; in addition to all the properties of interval variables, they feature an identifiable absolute zero point, thus, they allow for statements such as x is two times more than y. Typical examples of ratio scales are measures of time or space (Royce, p. 594).

**Population** – the total membership of a defined class of people, objects or events.

**Regression analysis** - The general purpose of multiple regression is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable…..In the social and natural sciences multiple regression procedures …. allows the researcher to ask (and hopefully answer) the general question "what is the best predictor of ...". For example, educational researchers might want to learn what are the best predictors of success in high-school (Statsoft).

**Reliability** – means that the individual scores from an instrument should be nearly the same or stable on repeated administration of the instrument, and that they should be free from sources of measurement error and consistent (Cresswell, p. 646).

**Sample** - is a subgroup of the target population that the researcher plans to study for the purpose of making generalizations about the target population (Cresswell, p. 646).

**Statistical Significance** – (represented as p-value) The statistical significance of a result is the probability that the observed relationship (e.g. between variables) or a difference (e.g., between means) in a sample occurred by pure chance ("luck of the draw"), and that in the population from which the sample was drawn, no such relationship or differences exist. Using less technical terms, we could say that the statistical significance of a result tells us something about the degree to which the result is "true" (in the sense of being "representative of the population")……..The higher the p-value, the less we can believe that the observed relation between variables in the sample is a reliable indicator of the relation between the respective variables in the population. (Statsoft).

**Skewed distribution** – If the distribution of a variable is not symmetrical about the median or the mean it is said to be skewed. The distribution has **positive skewness** if, in some sense, the tail of high values is longer than the tail of low values, and **negative skewness** if the reverse is true. (Upton)

**Tests** – statistical tests ……frequently used in educational research ….. for hypothesis testing, e.g. t-test, Chi-square, ANOVA. (Cresswell, p. 199)

**Validity** – means that researchers can draw meaningful and justifiable inferences from scores about a sample or population. (Cresswell, p. 649)

**Variables** – are characteristics of units that vary, taking on different values, categories or attributes for different observations….The dependent variable is the one the researcher is interested in explaining and predicting…….The explanatory variables that do the influencing and explaining are called independent. (Singleton, p. 84)

## References

The Concise Oxford Dictionary of Mathematics. Christopher Clapham and James Nicholson. Oxford University Press, 2009. Oxford Reference Online. Oxford University Press. University of Toronto Libraries. 10 September 2012. http://www.oxfordreference.com.myaccess.library.utoronto.ca/views/ENTRY.html?subview=Main&entry=t82.e581

Cresswell, J. W. (2008). Educational research: planning, conducting and evaluating quantitative and qualitative research. 3rd edition. Pearson Education Inc., New Jersey.

Elgie, S., Childs, R., Fenton, N., Levy, B.A., Lopes, V., Szala-Meneok, K., WIggers, R. D. (2012). Researching Teaching and Student Outcomes in Postsecondary Education: A Guide. Higher Education Quality Council of Ontario.

Owen, M. (2006). Conflict and convergence: The ethics review of action research. Journal of Academic Ethics 4 (1-4).

Singleton, R. A. Jr., Straits, B.C. (2010). Approaches to Social Research. 5th edition. Oxford University Press.

Statsoft.com. http://www.statsoft.com/textbook/basic-statistics/#Correlationsa. Accessed October 21, 2012.

*Tri-Council Policy Statement – Ethical Conduct for Research Involving Humans* (TCPS 2). (2010). Government of Canada. www.pre.ethics.gc.ca.

University of Toronto. (2009). Committee on Human Subjects in Research. Guidelines for Deception and Debriefing in Research. www.research.utoronto.ca/wp-content/uploads/2009/03/Deception_and_Debriefing_Guidelines.pdf

University of Toronto. (2003). Teacher-Researcher Conflicts of Interest. www.research.utoronto.ca/wp-content/uploads/2010/01/teacher-researcher-role-based-conflict-c2003.pdf

## Appendix A

The following are some beginning answers to the scenario exercises on p. 22. They are not intended to be comprehensive, as solutions would depend on what data you might already have, and what other sources are easily available in your specific context. But hopefully these ideas will provide a starting point to spark your imagination.

***Scenario 1****: You've developed a new approach to teaching a co-curricular writing workshop that involves more active engagement from the students. Students signed a participation sheet at the beginning of the workshop and filled out a survey evaluation form at the end. What statistical tests might you employ to analyze the information you have currently? What additional data might you obtain, and how might this change your analysis?*

**Given data**: attendance, responses to perception surveys for one workshop

With so little data, the possibilities for analysis are limited. Frequencies counts of responses to the perception surveys would be primary. You could also use chi-square t-test to determine whether the distribution of answers in each question is significant (or random).

If your perception survey also asks for demographic information (year, major, first language, etc.), you can also examine the frequency counts of these variables for each of the perception responses. Chi-square would again be the significance test of choice.

**Possible additional data that would allow for more in depth analysis**: the most likely in this scenario would be responses to the perception surveys from previous iterations of the workshop.

Assuming the perception surveys from past workshops asked the same (or at least some of the same) questions, you would be able to use these to create a control group and test group situation to see what if any difference there is in student responses to the new workshop material.

a) Chi-square would be most appropriate for testing significance in this instance because perception surveys give ordinal data.
b) Effect size would allow you to calculate the magnitude of the difference between the two groups. For a chi-square, phi and Cramer's V are the appropriate calculations.

Another possibility in a co-curricular workshop is that you can partner with the instructor of the course and use course grades to create a control group (students who didn't attend) and test group (students who did attend). If you have access to this data you could go forward with the following:

a) histogram of grades from those who participated and those who did not
b) comparison of mean grades from participants and non-participants, with standard deviation, and a t-test to confirm significance
c) effect size calculation to capture magnitude of effect on test group (Cohen's d would be most appropriate for t-test)

***Scenario 2**: You have introduced a new assignment into your class, requiring students to participate in online discussions of weekly readings to help support their learning. Although students did not receive marks for these assignments directly, you have tracked participation through the learning management system. You know how many times each student has posted, and how long each post is. All grades for the course are in. How would you go about analyzing this data statistically?*

**Given data**: grades, participation stats, number of posts from each participant, length of post from each participant

There are a number of options in this scenario both for descriptive and inferential statistics with just the given data:

(a) histogram of grades from those who participated and those who did not
(b) comparison of mean grades from participants and non-participants, with standard deviation, and a t-test to confirm significance
(c) effect size calculation to capture magnitude of effect on test group (Cohen's d would be most appropriate for t-test)
(d) correlation between grades and number of posts, with t-test confirming significance of the correlation
(e) correlation between grades and length of posts, with t-test confirming significance of the correlation