Worksheet: Educational Research Design - Quantitative Focus

Created by Lauren Margulieux, PhD

If you are interested in me giving a talk to your research group about research methods and design in computing or engineering education, please contact me at lmargulieux@gsu.edu.

Introduction

This worksheet will help you plan each of the following aspects of an educational research project with a quantitative focus.

- Crafting research questions that are interesting and answerable
- Selecting research methods that are appropriate and thorough
- Identifying or designing measurements that provide reliable and valid data
- Conducting appropriate analysis of data based on the type of data and the research questions

The worksheet is meant to be a companion to my research design blog series, where the concepts are explained.

https://laurenmarg.com/2022/06/26/research-design-series-introduction/

For in-depth information about specific designs, see the Research Methods Knowledge Base, https://conjointly.com/kb/table-of-contents/

The worksheet will help you pick appropriate statistical analysis, but it does not provide step-by-step instructions for conducting them. For step-by-step guides to analyzing data in SPSS, I like Laerd Statistics, statistics.laerd.com

For advice on heavily qualitative research design, I recommend finding a different resource.

Crafting a Research Question

https://laurenmarg.com/2022/07/05/research-design-research-questions/

A research question should make clear the knowledge that your study contributes. Thus, a good educational research question has several elements

- identifies the group that you are studying, such as students in an online course
- the construct (dependent variable) that you intend to measure, such as number of forum posts
- the independent variable or learner characteristic that might predict differences among people (if applicable), such as whether students are in a fully online or hybrid program

If you are measuring multiple dependent variables, determine whether you are measuring one construct or multiple constructs. For example, you might use multiple measures to examine one construct, such as self-efficacy. In contrast, you might use multiple measures that examine multiple constructs, such as self-efficacy, performance, and learning experience. Create a new research question (or sub-question) for each construct.

Create a new research question (or sub-question) for each independent variable or learner characteristic that is being treated as a predictor variable.

Last, determine whether you research question is descriptive, relational, or causal.

Type of Question	Definition	Example	Explanation
Descriptive	Asks about characteristics of groups/variables	What is the average number of forum posts in an online course for students in fully online and hybrid programs?	The answer to this question describes a forum behavior
Relational	Asks about relationship among variables	How do men differ from women in the number of forum posts in an online course?	The answer to this question relates the gender of students to a forum behavior
Causal	Asks about cause-and-effect relationship among variables	How does number of forum posts by the instructor affect the average number of posts by students in online courses?	The answer to this question establishes cause and effect between instructor and student behavior

Write your research question here:

My question is (check one): ☐ descriptive ☐ relational ☐ causal

Selecting Research Methods

Pre-Tests and Post-Tests

https://laurenmarg.com/2022/07/18/research-design-pre-and-post-tests/

Based on your research question, do you need a pre-test (i.e., are you measuring a change from before an intervention to after an intervention, such as knowledge or skill learned)?

☐ Yes, I need a pre-post design ☐ No, I need a post-only design

Based on your research question, do you need multiple post-tests (i.e., are you measuring a change at multiple points after an intervention, such as for retention)?

☐ Yes, I need more than one post-test ☐ No, I need only one post-test

	Definition	Example	Explanation
Pre-post design	Takes measurements before and during/after intervention to capture change	Give the same test at the beginning (pre-test) and end (post-test) of a course	By comparing the pre- and post-tests, the learning gains can be determined
Post-test	Takes measurements during/after intervention to capture outcomes	Give a survey (post-test) to measure students' opinions at the end of a course	Survey responses show students' opinions at the end of the course
Multiple post-tests	Takes measurements at multiple points during/after intervention	Give the same test at the beginning (pre-test), middle (post-test), and end (post-test) of a course	Multiple post-tests allow researchers to track progress throughout the course

Non-Experimental vs. Experimental Design

https://laurenmarg.com/2022/07/28/research-design-non-experimental-and-experimental-designs/

Research designs exists on a spectrum between high ecological validity (i.e., similar to conditions in normal life) and high experimental validity (i.e., the environment is controlled to isolate the effects of predictor variables). This spectrum and major research paradigms are pictured below.

	Observational Studies	Design Experiments	Quasi-Experiments	Randomized control trials	
High Ecological Validity	No control over the environment	Intervention is designed within the context of an environment	Intervention is tested among pre-made or self-selected groups	Intervention is tested among randomly assigned groups	**High Experimental Validity**

The type of research design that you need depends on the type of research question that you have. Descriptive and relational questions can be answered with non-experimental designs, and causal questions must be answered by experimental designs. Note: these design categories are independent from pre-test and post-test designs, so you can have a pre-post non-experimental design or a pre-post experimental design.

Design	Definition	Example	Research Question
Non-experimental design	Researchers do not manipulate anything about the learning experience	Observing interactions on a discussion board	Appropriate for descriptive and relational questions
Design experiment	Researchers design manipulations within the context of a learning experience	Creating a CS unit for a math class at a local school	Be careful about drawing causal conclusions for a generalized audience
Quasi-experimental design	Researchers manipulate some variables and not others	Exploring the interaction of teaching style and gender on test scores	Be careful about drawing causal conclusions about non-manipulated variables
Experimental design	Researchers manipulate a variable to determine whether it affects the outcome	Teaching different sections of a course with different styles and measuring test scores	Appropriate for causal questions

What research paradigm is appropriate for your research question and your desired balance of ecological validity and experimental validity? _____

How can you improve your weaker validity? For example, what can you measure to better understand the learning environment if you are doing a randomized control trial? What different scenarios can you observe to improve the generalizability of an observational study?

Identifying or Designing Variables and Measurements

Operationalization

https://laurenmarg.com/2022/08/24/research-design-levels-of-measurement/

Operationalization refers to how you define your terms within the context of a research study. For example, if you are measuring self-efficacy, what exact measurements are you using? An important part of operationalization is deciding which related features are you including or excluding. For example, if you are using the New Generalized Self-Efficacy scale, which measures general self-efficacy, then you might get a different result than if you used the Computer Programming Self-Efficacy Scale, which is domain specific. Neither is right or wrong, but you should consider which is more appropriate and how the decision affects the conclusions that you can draw.

What are you operational definitions for your independent (if applicable) and dependent variables? What features are included and excluded in your operational definition?

Independent Variables

https://laurenmarg.com/2022/08/03/research-design-dependent-and-independent-variables/

For each independent variable, determine how many levels you need. For discrete variables, such as online or hybrid, this is easy. For continuous variables, such as score on a self-efficacy test, what are important distinctions based on how you operationalize self-efficacy? Perhaps it makes sense to distinguish between low, medium, and high self-efficacy, or perhaps it makes sense to keep it as a continuous variable.

**It rarely makes sense to use a split-mean (i.e., all participants above the mean are classified as high, and all participants below the mean are classified as low). This technique groups people who are slightly above or below the mean, which is the majority of people, with those who are multiple standard deviations above or below the mean, even though they have much more in common with those who are just on the other side of the mean. It causes significant validity issues.

List each of your independent variables (usually there are 1-3), list the levels of each, whether they are between-subjects, within-subjects, or mixed (see blog post), and indicate whether they are fixed or random (see blog post). You don't need to include participant characteristics (e.g., gender, ethnicity, religion) unless your research question explicitly includes them.

Independent variable	Levels or Continuous	Between/Within/Mixed	Fixed/Random

Dependent Variables

Based on your research questions and your research paradigm, determine whether you will use quantitative, qualitative, or mixed dependent variables.

https://laurenmarg.com/2022/08/17/research-design-qualitative-quantitative-and-mixed-data/

For quantitative dependent variables, determine the level of measurement for each.

https://laurenmarg.com/2022/08/24/research-design-levels-of-measurement/

Level of Measurement	Definition	Example	Explanation
Nominal	Numbers are placeholders for categories	0 = male 1 = female	Data don't provide information about relationship between values
Ordinal	Numbers provide rank-order but not exact difference between categories	1 = low 2 = medium 3 = high	Data provide information about rank but not exact differences
Interval	Numbers provide information about difference between categories	1 = Strongly Disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly Agree	The difference between 1 and 2 are equal to the difference between 2 and 3
Ratio	Numbers provide information about difference between categories and zero means an absence	0 = 0 forum posts 1 = 1forum posts 2 = 2 forum posts 3 = 3 forum posts 4 = 4 forum posts	The difference between 1 and 2 are equal to the difference between 2 and 3, and 0 means no posts were made

To determine when in your research protocol you will measure dependent variables, you should place the dependent variable as close to the thing that it measures as possible. For example, if you are asking about the learning experience while using instructions, ask directly after the participant uses the instructions. If you wait until after the participant does something else, like solving problems, then your results might be confounded. If you collect learner characteristics, collect them at the end to avoid stereotype threat issues and to avoid participant making assumptions about the study.

Dependent variables should include both process and product data. Process data measures information about the learning process. Product data measures information about the outcomes.

List each of your dependent variables, grouped by construct, when it will be measured, how long it will take to complete on average, and whether it provides process or product data.

Dependent variable	Construct	When Measured	Time to Complete	Process/Product

How long will it take to complete the entire research protocol? This includes giving instructions for the research, the intervention or activity, time to complete any measurements, and debriefing. If the session will take longer than 60-90 minutes, participants might need breaks or lose attention. Consider whether each element of the research design is important enough for this trade-off.

What are the limitations of your research protocol? What relevant constructs are you not measuring? You don't have to include everything in one study, but consider what you are excluding and whether there are options to provide information about that construct.

Demographic Data

https://laurenmarg.com/2022/08/30/research-design-survey-design-demographics-validity-and-reliability/ The table gives examples of common demographic questions and options.

Questions	Level of Measurement	Response Options
What is your gender?	Nominal	Man, Woman, Other
What is your age?	Ratio	_____
What is your employment status?	Nominal	Employed for wages Self-employed Unemployed Homemaker Student Retired Unable to work
What is the highest level of education that you have completed?	Ordinal	Some high school High school diploma/GED Some college Associate's degree Bachelor's degree Some graduate school Master's degree Professional degree (M.D., J.D., Pharm.D., etc.)

		Doctoral degree (Ph.D., T.D., etc.)
Are you a domestic or international student?	Nominal	American (domestic), International
How do you describe yourself? (pick one or check all that apply)	Nominal	American Indian Pacific Islander East Asian South Asian Middle Eastern Black Hispanic or Latino Non-Hispanic White or Caucasian
What is your major?	Nominal	_____
Please report your high school GPA if you remember it	Interval	____ out of _____ (e.g., 4.0)
Which institute are you currently attending?	Nominal	_____ or list of choices
Which best describes how many years you've been in college?	Ratio	First-year Second-year Third-year Fourth-year Fifth-year Other
Please report your college GPA if you know it	Interval	_____ out of 4.0
Have you taken any (domain) courses in high school and/or college? (might split into two questions if warranted)	Nominal	Yes, No
If so, how many?	Ratio	_____
What were their names?	Nominal	_____
Do you consider English to be your primary language?	Nominal	Yes, No
If not, what is your primary language?	Nominal	_____
How comfortable do you feel using a computer?	Interval	1 – Not comfortable at all 2 3 4 – Neutral 5 6 7 – Very comfortable
How difficult do you think learning to (course outcome) will be?	Interval	1 – Very difficult 2 3 4 – Neutral 5 6 7 – Very easy

Improve your Design before Collecting Data

Now that you have defined your research design, make your hypotheses.

For each level of your independent variable(s), what results do you expect on the dependent variable(s)? Use a table to help you plan it out. In the table below, the hypothesis is that the control groups performs average-ly on the process DV and the product DV. For the first level of the IV, we expected the learning process to improve (e.g., take less time) with similar outcomes as the control group. For the second level of the IV, we expect the learning process to be worse but with better outcomes (e.g., retention).

	DV Process	DV Product
Control	=	=
IV level 1	↑	=
IV level 2	↓	↑

Making a table like this should be difficult if you have an interesting research question (e.g., that the result isn't obvious). When deciding what to place in each box, take note of any tradeoffs/decision making factors. For example, on IV level 1, maybe you expect the product DV to be equivalent to the control ONLY IF they take less time to complete the instruction. If they take the same amount of time, then you might expect the product DV to improve.

You want to keep note of these deciding factors because they will help you anticipate alternative results and possible explanations for those results. When you have a possible alternative results and likely explanations, make sure you are measuring something about that likely explanation (e.g., time to complete instruction). Then if you get to the end of the study with alternative results, you already have the data you need to examine likely explanations. This forethought makes null results much more compelling.

Another strategy to create a better research design is to pre-write your limitations section, either for the results you expect or for null results. Then analyze that limitations section for anything that you can improve in the research design before collecting data.

Conducting Appropriate Analyses of Data

Part of the data cleaning process includes determining whether your data are normally distributed. If so, use parametric tests. If not, use non-parametric tests, which are more conservative.
https://laurenmarg.com/2022/10/18/research-design-preparing-data-for-quantitative-analysis/

Refresher on the goals of statistical analysis –
https://laurenmarg.com/2022/10/25/research-design-what-statistical-significance-means/

Descriptive Statistics

If you have a descriptive research question, use descriptive statistics to analyze the central tendency, variability, and any other features of interest. Keep in mind the levels of measurement for your dependent variables to ensure that your conclusions make sense.

https://laurenmarg.com/2022/11/07/research-design-descriptive-statistics/

Inferential Statistics

If you have a **relational research question**, use correlation to quantify the relationship between two variables.

https://laurenmarg.com/2022/11/14/research-design-inferential-statistics-for-relational-questions/

Correlation

- If both variables are continuous (i.e., interval or ratio), use Pearson's *r* correlation (this is most common and is typically the default option for analysis software).
- If one variable is truly dichotomous and the other is continuous, use point biserial correlation.
- If one variable is artificially dichotomous and the other is continuous, use biserial correlation.
- If you have discrete data with more than two values (polytomized instead of dichotomized), then you'll have to use some of the more obscure tests, such as tetrachoric, polychoric, polyserial, Spearman's rho, or Kendall's Tau-B.

The resulting correlation coefficient will be between -1 and 1. Negative numbers indicate a negative/inverse relationship. Positive numbers indicate a positive relationship. Numbers closer to -1 or 1 are stronger relationships, and numbers closer to 0 are weaker relationships.

Value of correlation coefficient	Strength of relationship
r >.5 or >-.5	Strong
r = .5 to .3 or -.5 to -.3	Moderate
r = .3 to .1 or -.3 to -.1	Weak
r <.1 or <-.1	No relationship

If you have a **causal research question**, pick the appropriate test based on the rest of your research design. All of these tests assume that your dependent variable is at the interval or ratio level of measurement.

https://laurenmarg.com/2022/11/21/research-design-inferential-statistics-for-causal-questions/

T-Test

- Compare two discrete groups – t-test
    - For comparing groups that are between-subjects (participants in groups are mutually exclusive), use an independent-samples t-test.
    - For comparing groups that are within-subjects (participants are the same in both groups), use a paired-samples t-test.
    - For comparing a group to the mean from literature or another study, use a one-sample t-test (this is not common).

The effect size of t-tests are calculated using Cohen's *d.*

Value of *d*	Size of effect
d > .8	Large effect
d = .8 to .5	Medium effect
d = .5 to .2	Small effect
d < .2	No effect

ANOVA

- Compare more than two discrete groups – ANOVA
    - For comparing groups that are between-subjects, use a standard ANOVA.
    - For comparing groups that are between-subjects AND have a covariate (such as a pre-test), use ANCOVA.
    - For comparing groups that are between-subjects AND when measuring a construct with multiple dependent variables, use MANOVA.
    - For comparing groups that are within-subjects or mixed, use repeated-measures ANOVA.

Use appropriate post-hoc analyses to examine differences among individual groups

- To compare multiple levels of one independent variable, use Tukey's HSD test.
- To examine differences among multiple independent variables, use simple main effects.

When running multiple post-hoc analyses, use Bonferroni's correction to adjust the error rate.

The effect size of F-tests (ANOVAs) are calculated using Cohen's *f*.

Value of *f*	Size of effect
f > .4	Large effect
f = .4 to .25	Medium effect
f = .25 to .1	Small effect
f < .1	No effect

Regression

- Compare groups with at least one continuous independent variable – linear regression

Regression is necessary to analyze the predictive value of independent variables that are continuous, but it is also very useful for analyzing multiple fixed and random independent variables at once. It allows

you to determine how much unique variance each variable explains in the results. There are many options for regression depending on exact designs, so finding a more detailed guide is recommended.

Effect Sizes

Effect sizes not only describe the differences between groups, but they examine how much of the variance in a dependent variable is attributable to an independent variable.

For more information about interpreting and calculating effect sizes - https://laurenmarg.com/2022/11/28/research-design-interpreting-and-calculating-effect-sizes/

Additional Analyses

When rating qualitative/open-ended data, especially to quantify it, using multiple raters increases reliability. To examine reliability among raters, use interrater reliability.

https://laurenmarg.com/2022/12/05/additional-analyses-interrater-reliability-and-demographics/

- For binary data – use Cohen's Kappa
- For rated or ranked data – use inter-class correlation coefficients
    - Intra-class correlation coefficient of consistency says that both raters put items in the same order
    - Intra-class correlation coefficient of absolute agreement says that both raters gave items the same scores

To improve the validity of your conclusions, use demographic characteristics to describe the sample that you studied. You may also explore the relationships between demographic characteristics and outcomes to determine if there are any differences based on demographic groups.

Describe the results of your study, using statistical analyses as evidence for your claims. Consider the validity of alternative explanations for your results and discuss those that are compelling. Also describe your sample and justify how it is similar or different than the general population to make arguments about how generalizable your results are.