November 2020

# A Guide to Quantitative Research Proposals: Aligning Questions, Goals, Research Design, and Analysis

A Guide from

SPENCER FOUNDATION

Elizabeth Tipton, Ph.D.

# Contents


**1.0 Overview** **03**

**2.0 Articulate Clear Research Questions** **04**

**3.0 Set Clear Goals** **05**

**4.0 Shape the Study Design from the Research Question(s) and Goal(s)** **07**

**4.1** Sample: Recruitment approaches 07

**4.2** Measures: Operationalization of outcomes and variables 08

**4.3** Interventions: Explanatory studies and causality 09

**5.0 Conclusions, Samples, Estimates, and Uncertainty** **10**

**6.0 Writing a Strong Quantitative Proposal** **12**

**7.0 Appendix: Further Resources** **13**

# 1.0 Overview

In this essay, noted scholar Elizabeth Tipton elaborates on how to best articulate research design in grant proposals. This essay is a companion piece to our "A Guide to Writing Successful Field-Initiated Research Grant Proposals," which provides general information about the elements of grant writing. We also have guides for proposals focused on qualitative methods and research practice partnerships. These essays offer some insights for grant-proposal crafting as a writing genre. They are not intended to be stand-alone guides for writing grant proposals, receiving a Spencer grant, or conducting research. We recognize that there are numerous valuable approaches to conducting research and urge scholars to consider the specific concerns in their respective disciplinary and methodological fields.

The purpose of this guide is to provide a broader framework for thinking about and approaching quantitative research grant proposal writing. It is divided into four sections. The first section highlights the types of questions that can be answered with quantitative research. The second section provides further guidance in specifying the questions and goals of a study, including the population and how conclusions will be drawn. The third section focuses directly on the research design, including how the sample will be recruited, the use of appropriate measures, and questions of causality. The fourth section focuses on the modeling of outcomes and uncertainty due to sample size, providing guidelines and questions to ensure that the analyses are appropriate for the question type and goals of the study.

Importantly, although this guide presents an approach for thinking about research and a variety of examples, we do not rigidly specify *how* a researcher should convey this information in a proposal. There are a variety of formats and organizations possible, and depending upon the research question and type, proposals may need to spend more space on one part versus another.

- What are the possible conclusions or implications of the research once it is finished?
- Be sure to define the population to whom the results of your study are intended to apply, and conversely, where they may not.
- The parameter(s) of focus (i.e., to be estimated) should be clearly defined.
- It is important that you identify the *types* of conclusions you intend to draw and the *criterion* through which you will arrive at these decisions.
- Be sure to clearly explain the research design and how this design will enable you to estimate the parameters of focus.
- For a given goal and conclusion, you should explain how you arrived at the sample size that is necessary to warrant these conclusions.

## **2.0** Articulate Clear Research Questions

Well-written proposals offer clear and appropriate alignments from the guiding research question(s) through the study design, data, sample, analysis, and its implications. Quantitative research—and proposals—typically involve one or more of three types of questions: descriptive, explanatory, and predictive. **Descriptive** questions focus on understanding and describing a problem, gap, procedure, or practice. **Explanatory** questions – also called 'causal' questions – move beyond describing a problem, focusing instead on possible interventions, practices, or policies that are able to cause changes in outcomes or practices. Importantly, both descriptive and explanatory questions tend to focus on broad populations or samples and aggregate analyses across these in the past or present. **Predictive** questions, in comparison, focus on the development of algorithms and models to identify and predict future outcomes, often for individuals.

Determining if a question is descriptive, explanatory, or predictive is often more complicated than it might seem, because the same methods (e.g., regression) can often be used to answer all three questions. One helpful approach in determining the question type is to focus on the possible *claims* that can be made once the study is finished, i.e., the *implications* of the study. If the findings will be interpreted as evidence supporting the use of a policy or practice, then the question is likely explanatory.[1] If the implication is that a new algorithm or model should be used to identify students or schools in need of an intervention, then the question is predictive. Finally, if the conclusion is that a form of inequality, social structure, or problem is now better understood, the question is likely descriptive. Table 1 provides examples for each of these question types to help determine the kind of question being asked.

The type of question asked should be in clear alignment with the choice of research design, analysis plan, and potential conclusions drawn. For example, when asking a question that is inherently *explanatory* (e.g., if a policy or intervention improves outcomes for students), the research design and analysis plan should be appropriate for research (e.g., a design that appropriately addresses causality). Without clear alignment between question and design, the conclusions and goals of the proposed study will not be met using the design and analysis strategy provided.

[1]Importantly, this is true even if the method used in the study is *not* a randomized experiment. Furthermore, the fact that a change is suggested means that the intervention under study is something that can be *manipulated* and is not an *attribute* of an individual (e.g., race, gender) or group (e.g., urbanicity).

***Table 1. Three general kinds of quantitative questions***

| Question Type | Example | Possible claims |
|---|---|---|
| **Descriptive:** Who, What, When, Why, Where, How? | What types of students dropped out of high school over the past 20 years? | There is a problem/ gap/ inequality/ process/ phenomenon. |
| **Explanatory:** Does a change in X lead to an increase/ decrease in outcome Y? | Does dropping out of high school reduce future earnings? | A policy/ practice should be implemented /expanded/ removed/ changed. |
| **Predictive:** Does this model / algorithm improve our ability to predict future outcomes? | Can we identify in advance the students likely to drop out | This algorithm/ model should be implemented. |

## **3.0** Set Clear Goals

Offer a clearly defined and specific research question and goals for the study, once the question type is identified. This specification will ultimately narrow the question from a broader class of interest for study—e.g., about inequality in *all* schools—to the narrower class that can be feasibly answered in the proposed study, given the budget and real-world constraints. For example, while inequality may be a broad concern, in a particular study one might focus specifically on school funding inequality in North Carolina elementary schools in 2018-2019. This specificity does not make the research any less important—but it does help in defining the scope of the claims that can be made from the study.

In framing a study, it is important to define the **population** to whom the results are intended to apply, and conversely, where they may not. It is often helpful to turn to population-level data, such as the Common Core of Data (an annual census of schools), the American Community Survey, the U.S. Census, or national probability survey data like the Early Childhood Longitudinal Study (ECLS), to define the characteristics of this population. For example, it may be that the study focuses on rural students of color in the American South, or that the study focuses on schools in Texas. Keep in mind that the population definition is broader than simply reporting summary statistics on the *sample* in hand (unless the population of interest is the sample, which is possible). Distinguishing between a sample and population often involves asking questions like, "What is this sample a case of?" and "Where might results of this study *not* apply?"

Once a target population is defined, offer a clear definition of the **parameter(s)** of focus in the study. This requires narrowing the question and anticipating the kinds of 'output' for focus in the quantitative analyses. For an explanatory question, the goal may be to estimate the average treatment effect, subgroup treatment effects, or treatment effect differences. For a descriptive question, the goal may be to provide summary statistics (e.g., means, standard deviations, correlations), or the amount of variation in an outcome explained by different subsets of variables. For a predictive question, the goal may be to develop an algorithm that can predict well (e.g., 95% accuracy) an outcome for each student or school in the population. Again, it is important to distinguish here between the parameter, which is unknown in the population, and the *estimator,* which is based on the model and data in the study.

Often the goal of the study is not merely to report a statistic—e.g., an estimate or confidence interval—but to also draw a conclusion or make a decision based upon this estimate. For example, in an explanatory question, the study may conclude that an intervention is successful if there is (a) evidence that the *sample* effect size estimate is > 0.20 standard deviations (or another threshold), and (b) that there is evidence the effect is non-zero in the *population* (e.g., a hypothesis test based on a Type I error of 0.05). In a descriptive study, there may not be a decision per se, but instead, the goal may be to precisely estimate a measure (e.g., of poverty or achievement). In this case, a margin of error might be specified. Finally, in a predictive study, a new method or algorithm might be considered useful if it has a lower mean-squared-error than an existing measure or reaches a target accuracy metric. Importantly, these **criteria** given here are only examples. Others might be more appropriate depending on the case. What is important, however, is that criteria are stated *in advance*. Table 2 below provides examples of these for each of the types of questions.

Clear questions and goals shape research design and analytic strategies. When misaligned, research projects fall apart in their logic. For example, if a proposed explanatory study seeks to determine if an intervention can improve student learning broadly for *all students,* but then indicates that the schools in the study are all found in a single large urban school district, the results will not be able to meet the proposed goal. Persuasive research proposals are careful, therefore, to ensure that the questions asked are answerable and stated inquiry goals are feasible using the proposed research design and analysis methods.

*Table 2. Goals, Questions, and Examples*

| Topic | Questions | Descriptive | Explanatory | Predictive |
|---|---|---|---|---|
| **Population** | What population is this study about? What is it not about? (What are inclusion/ exclusion criteria?) | High school students in public schools in the U.S. that were included in NELS. | Public, regular, elementary schools serving > 50% students on free-or reduced-priced lunch in Alabama. | Elementary school students in Texas in public schools captured in the state longitudinal data system. |
| **Parameter** | What is the population quantity of interest? (There may be more than one). | What is the average and variation of X? How correlated is X with Y? How much of X is explained by Y and Z? | What is the average treatment effect of X on Y? | What is the predicted value of Y given observed characteristics? |
| **Criteria** | Is there a decision to be made? If so, what is the criterion – how will you know there is "enough" evidence or that the model is "good"? | The estimates of the mean will have a margin of error within X units, giving confidence intervals with a width of 2X. | The intervention is effective if there is sufficient evidence ($p < .05$) that the treatment effect is non-zero in the population and if the estimated effect size is medium in size (e.g., ≥ 0.20 SDs). | The algorithm is effective if it predicts Y better (e.g., has lower MSE) than other available methods. |

## **4.0** Shape the Study Design from the Research Question(s) and Goal(s)

Research design connects the specific research question to the study's claim or conclusions. The design operationalizes how the sample will be collected, how outcomes will be measured, and if relevant, how causality will be determined. Importantly, there may be multiple ways to design a study. This is a place where creativity and ingenuity show. However, some approaches are particularly well suited to specific question types. This section gives a broad overview of these problems and common approaches.

## **4.1** Sample: Recruitment approaches

Regardless of question type, nearly all studies focus on a **sample**, not the full population. As a result, how this sample is (or was) recruited has broad implications for generalizability to the target population (defined in Section 3.0). In some studies, these data may have been previously collected. For example, descriptive and predictive research often turn to large national probability samples, like the Early Childhood Longitudinal Survey (ECLS). The fact that these studies involved the random selection of units (e.g., schools) into the survey enables clear generalizations to a target population. Here the documentation to these surveys can help determine any particular concerns for generalization. For example, they may have excluded particular parts of the population (e.g., private schools), they were collected only during a specific period of time, or there may be concerns with non-response.

In other studies, the analyses focus on existing data that were *not* collected probabilistically. This type of data is particularly common for predictive and descriptive questions, where rich data on subgroups or processes are available only for those that completed a survey or who interacted with a particular product. For example, data may have been collected by a publisher on how students interacted with an online educational game or via an Mturk or Facebook survey regarding attitudes and beliefs about inequality. In these cases, the prevailing design concern is with how those that are in the data (e.g., completed the survey) might *or might not* represent the target population. How might this impact study results if those motivated to complete the survey on inequality are those with the strongest feelings? Or what if those schools that have provided data on an educational software platform tend to be highly concentrated in private schools in San Francisco? These *external validity* bias concerns should be considered.

A researcher might plan to *collect* data themself, for example, in a survey or experiment. An important consideration is how to approach recruiting a sample that represents the population of interest. A probabilistic selection plan may be ideal, but not often very feasible, particularly with uncommon subgroups. However, researchers must contend with this question even when the sample will be recruited based mainly on convenience—e.g., using Facebook or MTurk or by recruiting schools in neighboring school districts. Importantly, this may mean that budgetary and resource constraints of the study will not allow for generalizability to a target population. In this case, the target population specified may need to be narrowed, and the logics made clear in a proposal.

Just as importantly, to make claims about how a sample is related to a population, data collection is needed. In a study about college students, for example, it may be helpful to know the demographics of students in the population and to collect similar demographics for those in the sample. This will allow an *assessment* for reporting the degree of similarity between the sample and target population. It is important for a proposal of this type to offer a clear plan for data collection and analyses.

Finally, it is important to note that all studies have **limitations,** particularly based on the type of sample recruited (or data available) compared to the population more broadly. There is no perfect study, and reviewers understand this. Convincing proposals, therefore, anticipate problems, carefully examine possible approaches, and acknowledge study limitations.

## **4.2** Measures: Operationalization of outcomes and variables

Regardless of question type, all studies require careful consideration of how important outcomes and variables will be operationalized and measured within a study. For example, at a broad level, a study might focus on the effect of socio-economic status (SES) on student achievement outcomes, and particularly a sample of schools in North Carolina. But how should *socio-economic status* and *achievement* be measured? There are a variety of ways SES has been conceptualized in the literature (e.g., free- or reduced-priced lunch indicators, based on mother's income, based on mother's income and occupational prestige, etc.). Similarly, achievement can be measured in many ways—based on a proximal measure specific to the study in hand (e.g., knowledge of fractions), a researcher-developed measure, a more distal measure available in extant data (e.g., NC state second-grade math score), and so on. The relationship between SES and achievement depends upon the operationalization of each.

One concern here is with regard to *construct validity*—how well an item measures what it says it measures. This is a high-level question and requires one to think carefully about (a) what exactly the broader construct is, (b) what possible candidate operationalizations (i.e., items) there are available, and (c) how to tease apart empirically how well these candidates actually meet the study goal.

At a more specific level, there are concerns with *measurement validity*, which has to do with the level of measurement, the variable encoding, and the reliability of this measurement. The level of measurement might be, for example, at a particular time-point for a person, a general person-specific measure, at the classroom or school level. To continue this example, the relationship between SES and achievement could be measured at the *student* level or the *school* level – and again, these relationships may differ. Furthermore, SES could be encoded as a dichotomous variable (e.g., low-SES versus not), as a continuous variable, or as an ordinal variable.

Reliability has to do with the 'consistency' or 'repeatability' of selected measures. Reliability statistics (e.g., Cronbach's alpha) are often provided for existing measures, particularly test scores and other psychological constructs. However, these statistics are specific to the populations and samples studied. In this example, an achievement test (or items from it) might have been shown to be highly reliable for 8$^{th}$-grade students, but it may not be for 4$^{th}$-grade students. Similarly, the entire achievement test may have high reliability, but the sub-score focused only on fractions may not.

Just as with sample recruitment, there is no perfect study with respect to measurement. Persuasive proposals make clear how constructs will be measured and encoded, available information on the reliability of these measures, and possible limitations to these approaches. Compelling proposals anticipate problems, carefully examine possible strategies, and acknowledge study limitations.

## **4.3** Interventions: Explanatory studies and causality

In an explanatory study, the goal is to make a causal claim, by estimating the effect of an intervention.  In the literature, this is referred to as *internal validity*. Here the 'gold standard' design is a randomized experiment. In such an experiment, students, classrooms, teachers, or schools might be randomized to receive a new intervention. It is important here to carefully think through questions like: What is the comparison condition? What are the components of the intervention? How will the intervention be implemented? Is there any chance the intervention might be shared with the comparison condition?

Given the difficulties of implementing randomized experiments, researchers often turn to *quasi-experiments* instead to answer causal questions. These research designs include regression discontinuity, instrumental variables, interrupted time-series, and non-equivalent control group designs. In all of these cases, since the intervention is not randomized, it can be challenging to tease apart the effect of the intervention itself from the *selection* of the intervention. In other words, those that implement an intervention may *also* be different in a variety of ways from those that do not implement the intervention. For example, students that are retained for an additional year of Kindergarten typically differ in other ways (e.g., age, SES) from those who are not, making it difficult to tease apart the "effect" of grade retention from other experiences and characteristics.

Just as with sample recruitment and measurement, there is no perfect study with respect to causality. For a particular study, the most compelling proposals identify the most robust possible design given the budget and resource constraints. Persuasive proposals anticipate potential *threats* to causality in the study and carefully plan for approaches to mitigate these. For example, this might include planned robustness or sensitivity analyses, as well as clear statements about assumptions required and possible study limitations.

## **5.0** Conclusions, Samples, Estimates, and Uncertainty

Once the question type, goals, and study design are clear, the proposal should clearly define the models and methods that will be used to *estimate* the parameters of focus and draw conclusions based upon the specified criterion. Depending upon the question, the models and methods used might range from very simple (e.g., means, correlations) to very complex (e.g., multilevel models). Models should be chosen to estimate the target parameters (Section 3.0), as well as to appropriately account for sources of uncertainty. Table 3a presents a possible, but not exhaustive, sampling of methods and models. As highlighted before, the same methods (e.g., regression) can be used for all three question types.

For a given model, one should clearly identify the parameter **estimator.** For example, in a descriptive study, the sample correlation may be reported, while in an explanatory study, the analytic focus may be on the second-stage regression coefficient associated with an instrument in an instrumental variable analysis. Or, more commonly, a regression model may be used with many control variables to estimate the average treatment effect of an intervention in a quasi-experiment; in which case, the coefficient of focus is that associated with the intervention, not those associated with the controls.

Given this estimator, the next concern is with uncertainty and the ability of the sample in this study to answer the question and meet the goals of the study defined previously. All estimators have a measure of **uncertainty** associated with them, and these measures of uncertainty depend on the research design. For example, if schools are the unit of sampling or random assignment and outcomes are measured on students, then standard errors should take this clustering into account. Similarly, clustering and selection probabilities should be considered in the standard errors calculated using analyses of large national probability surveys (e.g., ECLS).

***Table 3a. Example models and methods***

| **Question Type** | **Example Models and Methods** |
| --- | --- |
| Descriptive | Geographic (e.g., GIS). Networks. Summary statistics. Correlations. Regression. |
| Explanatory | Average difference in group means (randomized experiment). Regression with controls. Difference-in-difference. Propensity scores. Instrumental variables. Two-stage least squares. Non-parametric regression (e.g., regression discontinuity). |
| Predictive | Regression. Machine learning. Artificial intelligence. Supervised learning. |

In most studies, to some degree the degree of uncertainty can be anticipated in advance through a **power analysis**. Measures of uncertainty can be written as functions of *design parameters* (i.e., sample characteristics that are knowable *before* a study is started). For example, the standard error of a regression coefficient is a function of the sample *size* ($n$), the variation in the covariate in the sample ($s_x^2$), the residual variation ($\sigma^2$) and the degree of multicollinearity (VIF). Or, in a group-randomized experiment, the uncertainty is a function of the intra-class correlation ($\rho$, the degree of clustering), as well as the number of student ($n$) and school ($J$) observations. Even in prediction models, the ability to classify well is a function of covariate features and sample size. In all cases, some or all of these characteristics are knowable in advance based on previous studies or extant data, and software is available to help conduct these analyses. Thus, given the decisions and conclusions that will be made from a study and the estimation strategy used, researchers should be able to determine the **sample size** that is necessary to warrant these conclusions.

Table 3b provides a handful of examples for each of these three concerns—estimator, uncertainty, and sample size. Persuasive proposals clearly convey the degree of certainty (precision) or uncertainty expected in the estimates, as a result of sample size and budgetary and other constraints, and how this degree of certainty will affect any conclusions the researcher seeks to draw. Here, the alignment between what is desired (the goal) and what is possible (the uncertainty) is important. For example, a proposed study that seeks to determine if an intervention will improve student achievement (a hypothesis test question) but has a very small sample with limited statistical power (e.g., 0.20) is unlikely to be funded. This is not to say that small studies, exploratory projects, or pilot studies cannot be considered strong and fundable proposals, but simply that the conclusions that such studies can possibly draw should be tempered by the scope of the project.

***Table 3b. Estimators, Uncertainty and Sample Size Examples***

| Topic | Questions | Descriptive | Explanatory | Predictive |
|---|---|---|---|---|
| Estimator | What is the estimation / prediction strategy (e.g., boosted regression)? | Summary statistics – means and SDs – will be calculated for each question. When there are multiple items regarding the same construct, latent variables will be estimated. | Dummy-variable regression coefficient in a hierarchical linear model that accounts for nesting of students in schools (unit of random assignment). | Bayesian predictive models using multivariate hierarchical models with X specifications will be used to predict X. Half of the data will be set aside for validation. |
| Uncertainty / Design Sensitivity | What is the measure of certainty (e.g., standard error, predictive validity, reliability)? How do design parameters affect this measure? | When estimating the mean, the margin of error is a function of sample size and the residual variation. | Standard error: a function of the intra-class correlation, sample size, and effect size. | Prediction intervals are a function of the sample size, residual variation in the population, and covariate variation in the sample. |
| Sample size | Given the goals and estimators and uncertainty, what sample size is required? | For a margin of error of X, a sample size of X is required. | The minimum detectable effect size (MDES) for power of 80% with $\alpha$ = .05 is X. | To classify students with X% error, a sample size of X is required. |

## **6.0** Writing a Strong Quantitative Proposal

In conclusion, no amount of fancy statistical modelling can make up for a poor study design. Compelling research proposals take this to heart, clearly defining the logic of the research process from beginning to end. As outlined in this guide, this involves:

- Accurately determining the type of question being asked.

- Cleary stating the specific population and parameters under study and how conclusions will be drawn from the study.

- Carefully describing the research design, including sample recruitment, measurements, and, when required, strategy for identifying causality.

- Appropriately selecting a model for estimation and indicating that the sample size of the study will be appropriate for drawing the stated conclusions and implications

Perhaps more important than any one of these steps, however, is the alignment of these steps throughout the proposal. The alignment between the question, goal, design, and analysis are of utmost importance. For additional resources on quantitative research design, please refer to the appendix, which includes articles, books, software, and training opportunities.

# 7.0
# Appendix: Further Resources

## Question

*Descriptive research questions:*

1. Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). Descriptive analysis in education: A guide for researchers. (NCEE 2017–4023). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

*Explanatory versus predictive questions:*

2. Schmueli, G. (2010) To explain or to predict? *Statistical Science, 25*(3): 289-310.

*Explanatory versus predictive questions:*

3. Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.
4. Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference.* Cambridge University Press.

## Goals

*Pre-registration:*

1. A best practice is to pre-register hypotheses, allowing confirmatory and exploratory tests to be clearly defined. In education, the Registry of Educational Effectiveness Studies (free) provides formats for pre-registering both experiments and quasi-experiments.

*General overview and guide:*

2. Duran, R. P., Eisenhart, M. A., Erickson, F. D., Grant, C. A., Green, J. L., Hedges, L. V., & Schneider, B. L. (2006). Standards for reporting on empirical social science research in AERA publications: American Educational Research Association. *Educational Researcher, 35*(6), 33-40.

## Design

*Populations:*

3. The Generalizer is a free webtool that can be used to help define a target population of schools (both K-12 and higher education) in the United States, and to develop a recruitment plan to represent this population.

*Measurement and Causality:*

4. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin,.

## Analysis

1. The Society for Research on Educational Effectiveness maintains a professional development library (with videos of prior workshops) as well as a list of software useful to researchers.

2. There are several options available for conducting analyses of statistical power including:
   - G*Power (both PC and Mac, free) – for laboratory experiments, t-tests, ANOVA, and individually randomized trials.
   - Optimal Design (PC based, free) – for cluster-randomized designs (e.g., school randomization), randomized block designs (e.g., classrooms within schools randomized), and longitudinal growth models.
   - PowerUp! (Excel-based, free) – cluster randomized designs, randomized block designs, as well as mediators and moderators of treatment effects in these designs.

3. There is a wide range of workshops available each year funded by IES and NSF intended to build methodological capacity for education researchers. Those interested may want to sign up to be on the IES and NSF listservs to receive announcements of future workshop opportunities. As of February 2020, there are five available to researchers:

- NSF funded workshops (focus on STEM education):
  - o Study Design (Summers 2021 – 2023)
  - o Introductory Meta-Analysis (Summers 2021 – 2023)
  - o Quantitative Methods and Measurement Intro (Summers 2020 – 2022)

- IES funded workshops:
  - o Advanced Meta-Analysis (Summers 2018 – 2021)
  - o Cluster Randomized Trials (Summers 2007 – 2021)

Spencer Foundation
625 North Michigan Avenue
Suite 1600, Chicago, IL 60611

**spencer.org**