Planning Defensible Quantitative Studies in Senior High School

Genesis Catalino

Abstract

Senior high school quantitative-research instruction often presents research designs, sampling methods, instruments, and statistical procedures as separate topics. Students may consequently select a test from a list after data collection rather than build a coherent chain from the research question to a warranted conclusion. This conceptual-pedagogical article analyzes a Practical Research 2 instructional text and develops the Articulate-Link-Inspect-Gauge-Narrate (ALIGN) framework for quantitative research planning. ALIGN asks student researchers to Articulate the question, target quantity or relationship, and intended claim; Link the claim to a defensible design, population, variables, comparison, and time structure; Inspect measurement validity, sampling, data provenance, and ethical protections; Gauge the analysis through estimands, descriptive displays, assumptions, effect sizes, uncertainty intervals, sensitivity checks, and appropriately interpreted statistical evidence; and Narrate conclusions whose language, scope, and transparency match the evidence. The framework corrects three recurrent misconceptions: that a large sample automatically produces representativeness, that reliability alone establishes validity, and that statistical significance proves practical importance or causation. Two instructional tools accompany the framework: a stage-by-stage alignment audit and a question-to-claim matrix for description, estimation, group comparison, paired change, association, prediction, and causal inquiry. The article argues that statistical technique should be chosen only after the target claim, design, measurement scale, dependence structure, and data quality are specified. For school-based studies, methodological rigor also requires proportionate ethics: voluntary participation, understandable consent or assent procedures, privacy protection, minimized burden, and clear limits on access and reuse. ALIGN is offered as an instructional design framework, not as an empirical demonstration of improved research performance. It can support proposal development, protocol review, data-analysis planning, results writing, oral defense, and revision. Its central contribution is a visible claim-evidence chain that allows teachers, students, and reviewers to locate exactly where a quantitative conclusion is strengthened, weakened, or exceeded by the design.

Keywords: quantitative research; research alignment; statistical reasoning; senior high school; research methods education; evidence-based claims

Introduction

A quantitative study is not made rigorous by the presence of numbers, a large respondent count, a reliability coefficient, or a familiar statistical test. Rigor depends on whether the question, design, measures, sample, analysis, and conclusion support one another. A perfectly computed result can answer the wrong question. A representative sample cannot repair an instrument that fails to measure the intended construct. A statistically significant association cannot by itself justify a causal claim.

These distinctions are difficult for beginning researchers because methods courses necessarily divide a project into teachable components. Students learn variables, designs, sampling, validity and reliability, data collection, descriptive statistics, inferential procedures, tables, and discussion. The source Practical Research 2 text follows this useful progression and provides proposal prompts, literature-review guidance, examples of research questions, sampling descriptions, instrument criteria, statistical procedures, and reporting activities (Gudelos, 2023). Its breadth makes it a strong instructional foundation. The risk is fragmentation: each component may be completed correctly in isolation while the overall inference remains misaligned.

A common symptom is the test-first question: 'What statistical treatment should I use?' before the target claim and data structure are clear. Another is formula-based sample selection without a stated purpose for precision, power, or representation. A third is treating validity and reliability as labels obtained through one expert review or one coefficient. The Standards for Educational and Psychological Testing emphasize that validity concerns the evidence and theory supporting interpretations of scores for their proposed uses, while reliability concerns consistency and measurement error (American Educational Research Association et al., 2014). Neither is a decorative property of an instrument.

This article proposes that quantitative-research instruction should be organized around a claim-evidence chain. The chain begins with what a researcher wants to know and ends with what the evidence permits the researcher to say. Statistical analysis is an important link, but it cannot compensate for weak design, ambiguous measurement, biased selection, missing data, or overextended language. Making the chain visible allows beginning researchers to revise a study before collecting data and to interpret results with appropriate restraint afterward.

Purpose of the Article

This article has three objectives: (a) to distinguish a coherent quantitative inference from a collection of completed research components; (b) to develop the ALIGN framework for planning and reviewing senior high school studies; and (c) to provide practical criteria for matching questions, designs, analyses, and claims. The guiding question is: How can beginning researchers construct a transparent path from a research problem to a defensible conclusion without treating statistical procedures as automatic decision rules?

Conceptual-Pedagogical Approach

Instructional Source and Analytical Unit

The article uses conceptual synthesis and instructional design analysis. The Practical Research 2 text supplies the curricular sequence: nature of quantitative inquiry, topic and problem development, literature review, research design, sampling, instruments, validity and reliability, data collection, statistical analysis, presentation, interpretation, and reporting. Research-design scholarship, testing standards, statistical guidance, ethics principles, reproducibility recommendations, and journal reporting standards supply criteria for strengthening alignment. No participant sample, classroom intervention, dataset, statistical estimate, or learning-effect claim is reported.

The unit of analysis is the claim-evidence chain: a traceable account of how a stated question becomes a target quantity or relationship, design, operational measure, sample, dataset, analysis, and bounded conclusion. The chain can be documented in a one-page protocol table before data collection and updated during analysis. Each link must be inspectable enough that another reader can identify assumptions, alternatives, and limits.

Alignment Before Technique

Questions Imply Different Targets

A research topic is not yet a research question, and a research question is incomplete until its target is specified. 'Social media and academic performance' names a domain. 'How many hours do Grade 12 learners report using social media on school days?' targets a distribution. 'Is use associated with grades?' targets an association. 'Does reducing use improve grades?' targets a causal effect and requires stronger design conditions. These questions may use some of the same variables, but they do not support the same analysis or claim.

The target should identify the population, variables or outcomes, comparison or relationship, time frame, and intended quantity. In contemporary statistical language, this quantity is often called the estimand: what the study aims to estimate. Beginning researchers need not use advanced notation, but they should be able to complete the sentence, 'The study aims to estimate or describe...' This prevents a vague objective from being retrofitted to whatever result appears interesting.

Design Determines the Kind of Inference

Descriptive, correlational, comparative, predictive, quasi-experimental, and experimental designs do different inferential work (Creswell & Creswell, 2018). A cross-sectional survey can describe a sampled population at a particular time and estimate associations under stated assumptions. It usually cannot determine temporal order or eliminate confounding. A randomized experiment is designed to support causal contrasts by creating comparable groups in expectation, but its conclusions still depend on implementation, attrition, measurement, and analysis. Shadish, Cook, and Campbell (2002) organize validity threats around statistical conclusion, internal, construct, and external validity; alignment requires considering all four rather than focusing only on the final test.

Language should follow design. 'Related to,' 'associated with,' and 'predicts within this model' are not synonyms for 'causes.' 'Improved after' does not establish that an intervention produced the improvement when no credible comparison exists. The design should be chosen because it can address the intended claim under feasible and ethical conditions, not because its name sounds more advanced.

The ALIGN Framework

ALIGN has five recursive stages: Articulate the question and target claim; Link the claim to design and data structure; Inspect measurement, sampling, provenance, and ethics; Gauge analysis and uncertainty; and Narrate a bounded, transparent conclusion. A weakness discovered in a later stage should trigger revision upstream. For example, a measure that captures only frequency may require narrowing a claim about 'engagement,' while clustered classroom data may require revising an analysis that assumed independent observations.

Articulate the Question and Target Claim

The researcher states the problem as an answerable question and identifies the intended descriptive, associational, predictive, or causal claim. The protocol names the population, units of observation, variables, comparison or relationship, time period, and target quantity. It also states what finding would count as practically meaningful. This last step separates a research purpose from a search for any nonzero difference.

Hypotheses can discipline predictions, but they should not replace a well-specified question. Nor should a null hypothesis become the only scientific statement. Estimation asks how large an effect or difference may be and how uncertain that estimate is. The American Statistical Association cautions that a p-value does not measure the probability that a hypothesis is true, the size or importance of an effect, or the quality of the study (Wasserstein & Lazar, 2016).

Link the Claim to Design and Data Structure

The researcher maps the target claim to a design capable of producing relevant evidence. The map includes exposure or predictor, outcome, comparison, timing, assignment or selection mechanism, repeated observations, clustering, and plausible confounders. It distinguishes independent groups from matched or repeated measurements because dependence changes both the interpretation and the analysis.

A design diagram can expose omissions before data collection. For a pretest-posttest intervention, it should show who receives the intervention, whether a comparison group exists, when measures occur, and what events could affect groups differently. For a survey, it should show the sampling frame, contact process, eligibility, response, and exclusions. The objective is not to draw a complex figure but to make the origin of every analyzed value visible.

Inspect Measurement, Sampling, Provenance, and Ethics

Measurement begins with a construct definition and the proposed interpretation and use of scores. Items, scoring rules, response options, administration conditions, language, and population all affect the evidence needed. Expert review can contribute content evidence, but it does not alone establish construct validity. Internal consistency addresses one aspect of score reliability and cannot prove that the instrument measures the intended construct. Flake and Fried (2020) show how unreported measurement choices create flexibility that weakens inference.

Sampling begins with a defined population and a usable sampling frame. Sample size and representativeness are separate questions. A large convenience sample may estimate the wrong population precisely, while a smaller probability sample may provide better population coverage. Sample-size justification should match the study's aim - precision, power, resource constraints, or detection of a meaningful effect - rather than rely on a universal formula (Lakens, 2022). The justification should state assumptions and acknowledge nonresponse, exclusions, and clustering.

Data provenance records who collected what, from whom, when, where, with which version of an instrument, under what conditions, and through which cleaning decisions. Ethical review is not an appendix to this process. The Belmont principles of respect for persons, beneficence, and justice direct attention to voluntary participation, understandable information, risk and benefit, and fair selection (National Commission, 1979). School-based studies also require age-appropriate assent or consent processes, privacy safeguards, and clear boundaries on teacher authority and student participation.

Gauge Analysis, Assumptions, and Uncertainty

Analysis should begin with data quality and description: missingness, outliers, distributions, coding, group sizes, measurement ranges, and graphical displays. The researcher then chooses an analysis family based on the target quantity, variable types, design, dependence, and assumptions. The choice between parametric and nonparametric procedures is not determined by sample size alone. Distributional shape, measurement level, robustness, estimand, and design all matter.

Results should foreground estimates, effect sizes, and uncertainty intervals, with tests used as one source of evidence. A threshold such as $p < .05$ should not divide findings into true and false or important and unimportant. Wasserstein, Schirm, and Lazar (2019) recommend moving beyond bright-line statistical-significance declarations toward consideration of context, design, evidence quality, and practical consequences. Sensitivity analyses can show whether conclusions change under reasonable coding, exclusion, or model choices.

Narrate a Bounded and Transparent Conclusion

The conclusion answers the original question using language supported by the design and analysis. It reports the estimate and uncertainty, identifies the analytic sample, describes relevant assumptions, and separates observed results from interpretation. Generalization is bounded by the sampled population and recruitment process. Causal language is bounded by design and implementation. Practical importance is argued from magnitude, context, costs, benefits, and stakeholder consequences rather than p-values alone.

Transparent reporting makes the research auditable. The APA Journal Article Reporting Standards identify information needed across quantitative manuscripts, including design, sampling, measures, procedures, analysis, and results (Appelbaum et al., 2018). The National Academies distinguish reproducibility of computations from replication through new evidence and emphasize sufficiently detailed methods, data, and analytic records (National Academies of Sciences, Engineering, and Medicine, 2019). For student projects, transparency can include a dated protocol, instrument version, codebook, cleaned-data log, analysis syntax or spreadsheet formulas, and explanation of deviations.

Table 1

ALIGN stages, required products, and diagnostic checks

Stage	Research product	Diagnostic question	Characteristic misalignment
Articulate	Question; population; variables; time; target quantity; meaningful magnitude	What exactly is to be described, estimated, compared, predicted, or caused?	Broad topic or test named before the claim
Link	Design map; timing; assignment or selection; dependence; comparison; confounders	Can this design generate evidence for the intended claim?	Causal language from descriptive or cross-sectional evidence
Inspect	Construct definitions; instrument evidence; sampling frame; provenance; ethics plan	Who or what is represented, measured, omitted, and protected?	Large sample assumed representative; reliability treated as validity
Gauge	Data-quality audit; estimand; display; assumptions; effect size; interval; sensitivity	Why does this analysis fit the target and data structure?	Procedure chosen by a test menu or p-value threshold

Stage	Research product	Diagnostic question	Characteristic misalignment
Narrate	Bounded result; uncertainty; scope; practical meaning; transparent methods and records	Does every conclusion remain within the design and evidence?	Overgeneralized, causal, or importance claims unsupported by data

A Question-to-Claim Matrix

Table 2 illustrates how question type constrains design, analysis, and language. It deliberately names analysis families rather than prescribing one test. The final choice depends on the exact outcome, scale, distribution, dependence, sample, assumptions, and target quantity. For example, comparing two independent groups may involve a difference in means, medians, proportions, or distributions; these are related but different estimands.

Table 2

Question-to-claim alignment for common quantitative inquiries

Question purpose	Design and data requirement	Analysis family	Defensible claim boundary
Describe a distribution	Defined population or dataset; one or more measured variables	Counts, proportions, center, spread, quantiles, and plots	Describes observed data; population claim depends on sampling
Estimate a population value	Sampling frame and defensible selection; measurement evidence	Point estimate and interval; design-based or model-based estimation	Precision and coverage depend on sampling and assumptions
Compare independent groups	Distinct groups; common outcome; confounding considered	Difference or ratio of means, proportions, medians, or model parameters	Association or difference; causation needs credible assignment/control
Assess paired change	Same or matched units measured twice; timing and attrition recorded	Paired differences, change scores, repeated-measure model	Change is observed; attribution requires a comparison and design
Study association/prediction	Variables measured on same units; temporal and model structure stated	Correlation, contingency analysis, or regression with validation	Association or predictive performance, not causation by default
Estimate a causal effect	Explicit intervention/exposure; counterfactual comparison; confounding strategy	Design-appropriate effect estimate with uncertainty and sensitivity	Causal scope limited by design, implementation, and population

Teaching ALIGN in Practical Research 2

Use an Alignment Protocol Before Data Collection

Students should complete the five-stage alignment audit before administering an instrument. Teachers can require one sentence for the target claim, one design diagram, a construct-to-item map, a sampling and ethics flow, and an analysis shell with planned tables or figures. An empty results table is useful because it forces students to name variables, groups, units, and summaries before seeing the data. This reduces the temptation to choose questions and analyses after inspecting results.

Protocol review should be consequential. If reviewers cannot determine the target population, the project should not proceed to sample-size calculation. If an outcome is undefined, the instrument should not proceed to validation. If group dependence is unclear, the statistical plan should not be approved. These checkpoints convert alignment from advice into an observable part of research quality.

Teach Statistical Decisions Through Contrastive Cases

Test selection is better learned through contrasted research situations than through a one-to-one lookup chart. Students can compare two studies with identical variable names but different designs: independent classes versus the same learners before and after instruction; probability sampling versus volunteer response; association measured once versus change following a randomized intervention. They then explain why the data structure and claim, not the title of the variables, determine the analysis.

A useful exercise presents an apparently impressive result and asks students to locate the weakest link. A tiny p-value may accompany a trivial effect, a biased sample, an unreliable measure, selective exclusion, or an overclaimed causal interpretation. Conversely, a wide interval in a small feasibility study may be informative when the purpose is to estimate procedures or variance rather than declare effectiveness. Such cases teach judgment rather than rule recitation.

Assess Revision and Transparency

The student portfolio should preserve versions of the question, instrument, sampling plan, codebook, analysis plan, and conclusion. A change log records what changed, why, and whether the change occurred before or after results were known. This practice reflects the transparency principles promoted by open-science frameworks, which encourage clear documentation of research materials and analytic decisions (Nosek et al., 2015).

An analytic rubric can score five dimensions corresponding to ALIGN. Emerging work presents components without connection. Developing work links the question and design but leaves measurement or sampling assumptions implicit. Proficient work states a target claim, provides relevant design and measurement evidence, justifies analysis, reports magnitude and uncertainty, and bounds conclusions. Extending work anticipates alternative explanations, documents deviations, performs sensitivity checks, and explains the practical and ethical consequences of uncertainty.

Discussion

Why a Test Menu Is Not a Research Method

Statistical procedures are often taught as pairs: two groups and a t-test, three groups and ANOVA, association and correlation. These pairings can be useful reminders after design is known, but they are unsafe as a primary decision system. 'Two groups' does not reveal whether observations are independent, outcomes are continuous or binary, the target is a mean or distribution, assumptions are tenable, or the comparison supports causation. ALIGN places these questions before technique.

The framework also reframes limitations. A limitation is not a ritual paragraph that excuses a study after completion. Potential threats are design inputs. Selection bias affects recruitment; measurement uncertainty affects instrument development; confounding affects design and analysis; missing data affects collection and modeling; limited precision affects sample-size planning and interpretation. Recognizing a threat early creates options to reduce it. Recognizing it late should narrow the claim.

Ethics and Statistical Quality Are Connected

Poor alignment can be an ethical problem because participants bear burden without producing useful evidence. An unnecessarily large sample exposes more people than needed, while an inadequately justified sample may be unable to address the question. Collecting identifying information without analytic purpose increases privacy risk. Asking sensitive questions through an untested instrument can create harm and measurement error simultaneously. Ethical design therefore includes statistical purpose, data minimization, security, fair recruitment, and honest uncertainty.

Research Agenda

ALIGN can be examined through design-based research, protocol audits, rubric validation, think-aloud studies, and comparisons of conventional versus alignment-centered instruction. Researchers could investigate whether students specify clearer estimands, distinguish reliability from validity, justify sample size by purpose, choose analyses that reflect dependence, and reduce causal overclaiming. Studies should preserve student protocol versions so improvement can be traced to particular feedback and revisions.

Further work may compare ALIGN across academic strands, group and individual projects, spreadsheet and statistical-software environments, and proposal versus completed-report assessment. It may also test whether alignment audits improve oral defenses by helping students explain why a design and analysis fit the claim. The framework remains a conceptual instructional proposition until such effects are evaluated with transparent evidence.

Conclusion

A defensible quantitative study is a coherent argument, not a sequence of forms and formulas. The question identifies the target; the design determines what evidence can be generated; measurement and sampling determine what the observations represent; analysis estimates patterns and uncertainty; and reporting states only what the chain supports. ALIGN makes this structure explicit through five recursive stages: Articulate, Link, Inspect, Gauge, and Narrate.

For senior high school research, the framework preserves the broad curricular value of the source Practical Research 2 text while strengthening the connections among its components. It does not require advanced mathematics for every project. It requires precision about the claim, transparency about choices, respect for participants, and proportional interpretation of evidence. When students can show where each conclusion comes from and where it must stop, quantitative research becomes more than statistical procedure: it becomes disciplined reasoning under uncertainty.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. American Psychologist, 73(1), 3-25. https://doi.org/10.1037/amp0000191

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. Advances in Methods and Practices in Psychological Science, 3(4), 456-465. https://doi.org/10.1177/2515245920952393

Gudelos, J. T. (2023). Practical research 2: Senior high school quantitative research [Instructional text].

Lakens, D. (2022). Sample size justification. Collabra: Psychology, 8(1), Article 33267. https://doi.org/10.1525/collabra.33267

National Academies of Sciences, Engineering, and Medicine. (2019). Reproducibility and replicability in science. National Academies Press. https://doi.org/10.17226/25303

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). The Belmont report. U.S. Department of Health and Human Services. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. Science, 348(6242), 1422-1425. https://doi.org/10.1126/science.aab2374

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

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. The American Statistician, 70(2), 129-133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond p < 0.05. The American Statistician, 73(S1), 1-19. https://doi.org/10.1080/00031305.2019.1583913