Infinity Journal of Mathematics Education
Volume 15, No. 3, 2026
p-ISSN 2089-6867
e–ISSN 2460-9285
https://doi.org/10.22460/infinity.v15i3.p831-862

Fostering methodological reasoning in pre-service mathematics teachers: A hypothetical learning trajectory for quasi-experimental design

Jeinne Mumu[1*], Lisnani[2], Benidiktus Tanujaya[1]
[1]Department of Mathematics Education, Universitas Papua, West Papua, Indonesia
[2]Department of Mathematics Education, Universitas Negeri Surabaya, East Java, Indonesia
[*]Correspondence: j.mumu@unipa.ac.id

Received: May 2, 2026 | **Revised**: Jul 22, 2026 | **Accepted**: Jul 24, 2026 | **Published Online**: Aug 4, 2026

Abstract

Quasi-experimental designs (QEDs) are widely used in educational research, however pre-service mathematics teachers often struggle to justify their methodological choices when ideal experimental conditions cannot be met. This study designed and refined a pedagogical approach to support mathematics education students' conceptual understanding of QEDs. Employing educational design research, the study involved 41 third-year undergraduates enrolled in an experimental design in education course. The intervention was conducted across two iterative cycles: a teaching experiment (n = 9) and a classroom experiment (n = 32). A Hypothetical Learning Trajectory (HLT) was developed to structure learning activities around the critical analysis of authentic research excerpts. Analysis of students' written design proposals and clinical interviews revealed a conceptual progression from viewing interventions as true experiments to recognizing QEDs as contextually grounded compromises. Specifically, students demonstrated transformed reasoning by shifting from rigidly demanding random assignment to strategically utilizing intact groups to manage confounding variables. Retrospective analysis further highlighted the need for strengthened scaffolding regarding the notion of control. Consequently, the refined HLT yielded a proposed Local Instructional Theory (LIT) comprising four empirically grounded design principles: (1) contextualizing the need for comparison, (2) negotiating intact-group constraints, (3) evaluating threats to validity, and (4) reflecting on methodological compromises. In practice, this study provides mathematics teacher education programs with a structured framework to foster pre-service teachers' critical methodological reasoning, directly enhancing their capacity to conduct context-sensitive educational research.

Keywords:
Design research, Hypothetical learning trajectory, Mathematics education, Quasi-experimental design, Research methodology education

How to Cite:
Mumu, J., Lisnani, L., & Tanujaya, B. (2026). Fostering methodological reasoning in pre-service mathematics teachers: A hypothetical learning trajectory for quasi-experimental design. *Infinity Journal*, *15*(3), 831-862. https://doi.org/10.22460/infinity.v15i3.p831-862

1. INTRODUCTION

Preparing pre-service teachers to become competent educational researchers has become an increasingly important component of teacher education programs worldwide. As well as developing their pedagogical skills, future teachers must be able to critically evaluate research evidence, design rigorous classroom studies, and make informed methodological decisions that support evidence-based educational practice. In this context, the "Experimental Design in Education" course in Mathematics Education program plays a pivotal role in equipping pre-service mathematics teachers with the knowledge and reasoning skills required to design, evaluate, and interpret educational research. As well as providing procedural knowledge, the course develops methodological reasoning, enabling students to justify research design decisions, evaluate threats to validity, and critically interpret causal claims rather than merely applying methodological terminology (Creswell & Creswell, 2018; Shadish et al., 2002).

Despite these objectives, research methods education consistently reports that students find it hard to apply abstract methodological principles to real-life research situations, especially when conducting their undergraduate thesis (Boote & Beile, 2005; Earley, 2014). These challenges are particularly evident when students learn about experimental and quasi-experimental designs, where methodological decisions must strike a balance between theoretical rigor and the practical realities of educational settings (Shadish et al., 2002). For pre-service mathematics teachers, this challenge may be even greater, given that their academic training focuses primarily on deductive reasoning, formal proof and deterministic thinking. By contrast, educational research requires probabilistic reasoning, contextual judgement and the ability to evaluate methodological trade-offs in uncertain conditions (Feldon, 2007; Garfield & Ben-Zvi, 2008).

Traditionally, experimental research is founded on three core methodological principles: randomization, replication and local control (blocking). Collectively, these principles reduce bias, control error variance and strengthen causal inference (Fisher, 1935; Smucker et al., 2023; Tanujaya, 2013). In true experimental designs, random assignment allows researchers to isolate the effects of a treatment while minimizing selection bias. However, genuine random assignment is often impractical or unethical in educational settings because students are usually organized into fixed classrooms that cannot be rearranged at random (Pyott, 2021; Rahayu et al., 2022). Consequently, educational researchers frequently employ quasi-experimental designs (QEDs).

Crucially, the distinction between true experimental designs and QEDs encompasses more than just the absence of random assignment. Instead, it represents a fundamentally different approach to establishing causal inference (Cook & Campbell, 1979; Reichardt, 2019). Unlike true experiments, QEDs require researchers to identify, measure and address alternative explanations through design features and statistical controls that mitigate threats to internal validity (Rahayu et al., 2022). Furthermore, QEDs encompass diverse methodological approaches, including non-equivalent control group designs, interrupted time series, regression discontinuity designs and matched comparison group designs. Each approach requires different assumptions, analytical procedures and validity considerations (Gopalan et al., 2020; Shadish et al., 2002). Consequently, understanding QEDs requires substantially more than memorizing research design labels. Students must be able to reason about how methodological decisions influence causal inference in different contexts, making quasi-experimental design one of the most conceptually challenging topics in research methodology education.

Developing this form of methodological reasoning draws upon several complementary strands of educational research. Studies in statistics education have consistently documented students' misconceptions regarding variability, statistical inference, uncertainty, and evidence-based reasoning (Garfield & Ben-Zvi, 2008; Tanujaya & Mumu, 2020). Research on methodological reasoning similarly demonstrates that novice researchers frequently struggle to evaluate research validity, identify potential sources of bias, and justify methodological trade-offs in authentic research situations (Feldon, 2007). More recently, research in causal inference education has shown that learners often rely on intuitive rather than principled reasoning when interpreting causal relationships from observational and non-randomized evidence (Grosz, 2023; Martínez-Abad & León, 2023). Meanwhile, design research has demonstrated the value of learning trajectories for supporting students' conceptual development across mathematics, statistics, and science education by systematically anticipating students' thinking, misconceptions, and conceptual change (Clements & Sarama, 2004; Cobb et al., 2003; Siemon, 2021; Simon, 1995).

Despite these substantial contributions, a significant gap remains at the intersection of these four research domains. While existing learning trajectory research has focused primarily on conceptual development in mathematics, statistics, and scientific enquiry, considerably less attention has been devoted to understanding how students progressively develop the reasoning skills needed to evaluate quasi-experimental research. In particular, little is known about how students' progress from making intuitive, superficial classifications of research designs — such as equating any study involving a treatment with a true experiment — to principled reasoning about causal inference, internal validity and methodological justification. Without a theoretically grounded learning trajectory to describe this progression, instruction often relies on students memorizing research design terminology, which leaves them insufficiently prepared to justify their own methodological decisions in educational research.

To address this gap, the present study uses a design research methodology to investigate and iteratively refine a learning trajectory for teaching QEDs. The study focuses on developing a Hypothetical Learning Trajectory (HLT) to enhance students' methodological reasoning in three key areas: (1) identifying and mitigating threats to internal validity (e.g., confounding variables); (2) evaluating the suitability of different QED variants for particular research contexts; and (3) constructing justified causal claims from non-randomized data. Accordingly, the study is guided by the following objectives: (1) To map the empirical learning trajectory, identifying students' cognitive processes, anticipated learning obstacles, and necessary instructional; (2) To design, implement, and refine a sequence of learning activities based on the critical analysis of authentic thesis excerpts and journal articles employing experimental approaches; and (3) To examine how students construct their understanding of QEDs, progressing from an initial conflation with true experimental designs toward a nuanced evaluation of causal claims and methodological trade-offs.

Ultimately, this study contributes to the field of statistics and research methodology education by providing an empirically grounded learning trajectory for teaching quasi-experimental designs. The resulting refined HLT constitutes a local instructional theory (Cobb et al., 2003) that offers a structured pedagogical framework for supporting students' development of critical and principled methodological reasoning. These findings have the potential to inform instructional practices in contexts where complex methodological concepts must be learned within real-world constraints.

2. METHOD

This study employed a design research (DR) approach, which is also referred to as design-based research (Cobb et al., 2003) or developmental research (Freudenthal, 1991; Gravemeijer, 1994). DR is particularly appropriate for this study, which seeks to develop contextually grounded theories of learning — commonly termed local instructional theories (LITs) — through iterative cycles of design, enactment, analysis and refinement conducted in authentic educational settings (Bakker, 2004; Prediger et al., 2015). DR is characterized by its interventionist, iterative, process-oriented, theory-generating and pragmatic nature (Barab & Squire, 2004; Plomp, 2013). Rather than merely evaluating the effectiveness of an intervention, this approach aims to co-develop instructional designs alongside theoretical insights into how learning processes unfold within those designs (Van den Akker et al., 2006).

In line with these principles, the present study had two main objectives: (1) designing and iteratively refining a learning trajectory for understanding quasi-experimental designs (QEDs); and (2) developing an empirically grounded LIT that explains students' conceptual development in this domain. The study followed the prototypical DR cycle consisting of three interconnected phases: (1) preliminary design, (2) design experiment and (3) retrospective analysis (Gravemeijer & Cobb, 2006; Van den Akker et al., 2006). Consistent with established design research traditions, the two implementation cycles are explicitly termed the teaching experiment (a small-group, micro-level exploration) and the classroom experiment (a whole-class, macro-level validation), as illustrated in Figure 1.

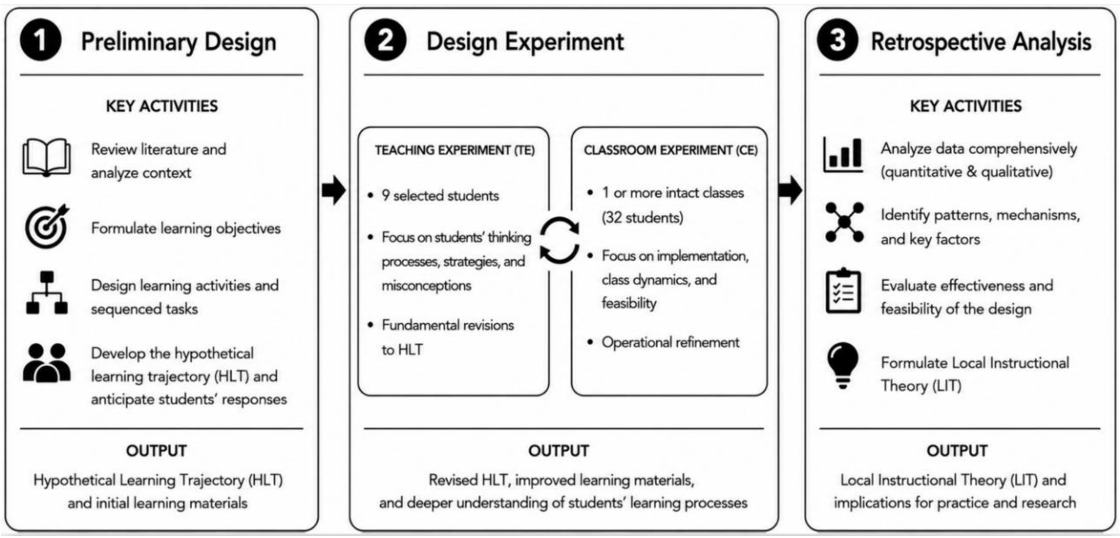

Figure 1. The phases of design research

2.1. Research context, participants and positionality

The study was conducted at the University of Papua, specifically within the Mathematics Education department. The educational context was the compulsory "Experimental Design in Education" course, which focuses on research methodology and statistical design for pre-service mathematics teachers.

Participants comprised two distinct groups across the iterative cycles. For the first cycle (the teaching experiment), nine undergraduate students were purposively selected. They were selected based on their academic performance in a prior foundational statistics course to ensure a mix of students with high, medium and low prior knowledge (categorized by scores above 80, 60-79, and below 60, respectively), capturing a wide range of cognitive responses. For the second cycle (the classroom experiment), an intact class of 32 undergraduate students enrolled in the regular course participated in the intervention.

In terms of researcher positionality, the author acted as the facilitating lecturer in both cycles. Recognizing the inherent power dynamics and potential observer bias associated with this dual role, the researcher kept a reflexivity journal throughout the study to set aside personal assumptions and preconceptions about student learning. To mitigate the influence of the lecturer's authority on student responses and ensure objective data collection, an independent co-researcher (a faculty member from the same department not involved in course instruction) was appointed as the primary observer.

2.2. Phases of Design Research

2.2.1. Phase 1: Preliminary Design and HLT Development

The preliminary design phase involved an extensive literature review on experimental design pedagogy. Based on this review, a Hypothetical Learning Trajectory (HLT) was formulated consisting of learning goals, conjectures about students' cognitive processes and proposed instructional activities.

To ground these activities in reality, authentic manuscripts were purposely selected to serve as core instructional materials. This corpus included fifteen excerpts from previous undergraduate theses and ten published journal articles. The manuscripts were selected based on the following criteria: (a) they contained common methodological flaws in applying QEDs (to serve as catalysts for cognitive conflict and critical analysis), or (b) they exemplified rigorous, context-sensitive applications of QEDs (to serve as models of best practice). These materials were integrated into a series of six structured learning activity sheets (LAS), which guided students through progressively more challenging tasks involving the identification, evaluation and design of QEDs.

2.2.2. Phase 2: Design Experiment (Teaching and Classroom Experiments)

The design experiment phase involved the empirical testing and refinement of the proposed HLT through two successive implementations. In the teaching experiment, the initial HLT was implemented with a small group of nine students. This phase enabled in-depth observation of the students' reasoning processes and the identification of ambiguities in the task design.

Based on the findings from the teaching experiment, three key revisions were made to the HLT before the classroom experiment began. Firstly, an explicit scaffolding module was added to Learning Activity Sheet 2 (LAS 2) to distinguish between 'control as procedural equivalence' (achieved via random assignment) and 'control as a comparative reference' (utilizing intact groups). Secondly, targeted prompting questions were integrated into the lesson plans to prevent students from equating the mere presence of two groups with a valid control condition. Thirdly, the conceptual abstraction activity in Stage 4 was extended to include sentence frames, helping students to articulate the positive pragmatic features of QEDs rather than relying solely on deficit-based definitions. Following these revisions, the refined HLT was implemented in a classroom experiment involving 32 students.

The classroom experiment was conducted over six weekly sessions. Each session lasted 100 minutes (two 50-minute periods with a break). During this phase, the researcher acted as the facilitating lecturer while an independent observer monitored classroom dynamics and documented learning processes in a naturalistic setting.

2.2.3. Phase 3: Retrospective Analysis

Following the completion of the design experiments, a retrospective analysis was conducted to evaluate the relationship between the designed HLT and the actual learning trajectory that emerged during implementation (Gravemeijer & Cobb, 2006). This phase served two primary functions: (1) to analyze the alignment and deviations between the hypothesized and observed learning processes, and (2) to formulate an empirically grounded Local Instructional Theory (LIT).

Rather than merely confirming whether the instructional sequence worked, the retrospective analysis examined how students' evolving reasoning corresponded with, diverged from, and ultimately reshaped the initial design conjectures. This analysis focused on identifying the specific instructional mechanisms that supported students' progression from surface-level classification of research designs toward principled methodological reasoning. The insights gained from this phase informed the iterative refinement of the HLT and culminated in the abstraction of four core instructional principles that constitute the study's LIT.

2.3. Data Analysis

To ensure a rigorous examination of the learning processes, data analysis was conducted using a multi-layered approach that integrated qualitative coding, constant comparison, and triangulation.

Analytical Units and Coding Procedure

The primary analytical units were defined as learning episodes (coherent sequences of interaction focused on a specific methodological idea) and individual written justifications. The coding procedure was conducted in two rounds using the constant comparative method (Corbin & Strauss, 2015). First, open coding was applied to transcripts and artifacts to identify emerging learning trajectories, misconceptions, and shifts in reasoning. Second, axial coding was used to map these empirical findings onto the initial HLT conjectures, specifically categorizing data based on alignment (where student reasoning matched predictions) and deviation (where unexpected reasoning emerged).

Comparing HLT and Actual Learning Trajectory

A central component of the analysis involved systematically comparing the HLT with the actual learning trajectory observed in the classroom. This comparison focused on three aspects: (1) the starting point of students' reasoning (pre-conceptions), (2) the learning process (how students interacted with the tasks and each other), and (3) the end point (the quality of their methodological reasoning). Discrepancies between the HLT and the actual trajectory were specifically analyzed to understand the underlying causes, such as unanticipated student difficulties or the effectiveness of specific scaffolding strategies.

Reliability and Inter-Rater Agreement

To ensure the reliability of the qualitative coding, two researchers independently coded a subset of the data (approximately 20% of the transcripts and artifacts). Following independent coding, consensus meetings were held to discuss discrepancies and refine the coding scheme. Inter-rater reliability was calculated using Cohen's kappa, yielding a coefficient of $\kappa = 0.82$, indicating a high level of agreement between coders (Cole, 2024; ten Hove et al., 2024). The remaining data were then coded using the agreed-upon scheme.

Triangulation and Integration of Data Sources

Methodological triangulation in this study went beyond the mere collection of multiple data sources; it involved the systematic integration and cross-verification of these sources to construct a robust account of learning. Specifically:

a. *Video Transcripts*: Analyzed to trace the chronological, interactional progression of students' reasoning (the 'how' and 'why' of their conceptual shifts).
b. *Student Artifacts*: Worksheets and concept maps were cross-referenced with video data to verify the consolidation of individual understanding. When discrepancies arose between what students articulated during discussions (video) and what they documented individually (artifacts), these instances were specifically analyzed to uncover hidden misconceptions or performative understanding.
c. *Field Notes*: Notes from the independent observer were integrated to contextualize the video data, capturing non-verbal cues and off-camera interactions that fixed cameras might have missed.

This multi-layered integration ensured that the reconstructed learning trajectory was a comprehensive synthesis of observed, documented, and contextualized learning processes, providing a solid empirical basis for the formulation of the Local Instructional Theory.

2.4. Data Collection and Data Corpus

To capture the complexity of students' conceptual development during the design experiment, multiple complementary sources of evidence were collected throughout the two research cycles. In line with the principles of design research, the focus of the data collection was not only on learning outcomes, but also on the processes through which students' understanding evolved during the implementation of the HLT. This enabled iterative refinement of the instructional design and supported triangulation across different forms of evidence.

Instrument Development and Validation

The instructional materials, including pre- and post-intervention diagnostic tasks, LAS, interview protocols, and classroom observation guides, were developed based on the theoretical principles underpinning the HLT and DR (Bakker, 2018; Van den Akker et al., 2006). The instruments were specifically designed to elicit students' methodological reasoning concerning QEDs rather than simple factual recall. Content validity was established through expert review by three specialists in mathematics education research methodology and educational research design. Their feedback led to revisions in task wording, the sequencing of learning activities, and the inclusion of additional prompts encouraging students to justify their methodological decisions. Prior to the classroom experiment, all instruments were piloted during the teaching experiment to identify ambiguities, estimate task completion time, and evaluate whether the activities elicited the anticipated reasoning processes.

Data Sources

a. *Diagnostic Tasks (Pre- and Post-Intervention)*: Students completed identical diagnostic tasks before and after the intervention. The tasks required participants to analyze authentic educational research scenarios, determine whether each study represented a true experiment or a QED, and justify their decisions using methodological reasoning related to group assignment, control, and causal inference. A total of 32 paired responses were collected during the classroom experiment.

b. *Video Recordings*: All learning sessions in both cycles were video-recorded. Approximately 2 hours of video were collected during the teaching experiment and 10 hours during the classroom experiment. These recordings captured classroom interactions, whole-class discussions, collaborative group work, and instructor scaffolding. Selected episodes illustrating significant conceptual change were transcribed verbatim for qualitative analysis.

c. *Student Learning Artifacts*: Students' written work was systematically collected throughout the intervention, including completed LAS, annotations of authentic research articles, group discussion summaries, and revised research design analyses. These artifacts documented the progression of students' thinking across the six instructional stages.

d. *Semi-Structured Interviews*: To gain deeper insight into students' reasoning, semi-structured clinical interviews were conducted with selected participants following major instructional activities. These interviews explored students' interpretations of QEDs, their perceived differences from true experiments, and the reasoning underlying their methodological decisions.

e. *Field Notes*: Throughout both cycles, the researchers maintained detailed field notes documenting classroom interactions, emerging misconceptions, unexpected learning trajectories, and modifications made during implementation.

2.5. Trustworthiness and Research Ethics

To ensure the trustworthiness of the qualitative findings, four criteria were applied: (1) Credibility was ensured through prolonged engagement, peer debriefing with the independent observer, and member checking with a subset of students; (2) Transferability was supported by providing a thick description of the educational context and participant characteristics; (3) Dependability was maintained through a comprehensive audit trail of raw data, coding memos, and analysis logs; and (4) Confirmability was achieved through the researcher's reflexivity journal to bracket biases.

Regarding research ethics, this study was conducted in strict accordance with ethical standards. Informed consent was obtained in writing from all participants prior to the study. For the teaching experiment, additional specific consent was obtained for intensive audio-visual recording. Participants were explicitly informed that their participation was voluntary and that they could withdraw from the audio-visual recording at any time without any penalty to their academic grades. To ensure confidentiality, all participants were assigned pseudonyms (e.g., Student 1, Group A). Finally, the student work utilized in this study (specifically thesis excerpts and scientific journal articles) was obtained from public internet repositories and is therefore in the public domain; thus, formal written permission was not required. These materials were used strictly for research purposes, and all identifying information regarding the original authors was omitted to ensure complete anonymity.

3. RESULTS AND DISCUSSION

3.1. Results

This study developed and implemented a structured pedagogical sequence framed as a HLT to help students understand the fundamental principles of experimental design, such as replication, blocking and randomization, and their critical role in distinguishing true experimental from quasi-experimental designs. The HLT was constructed from a synthesis of the literature, which identified a key prerequisite for learning: a robust understanding of quasi-experimental methodology requires a solid grasp of the fundamental principles of true experiments and an awareness of the practical limitations that result in methodological compromise.

3.1.1. Pre- and Post-Intervention Assessment Outcomes

To evaluate the shift in students' methodological reasoning, pre- and post-intervention tasks were administered, requiring students to classify research designs and justify their reasoning. The results demonstrated a significant transition from surface-feature classification to principled methodological reasoning.

In the pre-intervention task, the majority of students (87.5%, n = 28) relied on surface features to classify designs. They predominantly equated the mere presence of a "*treatment*" and a "*control group*" label with a true experimental design, regardless of how the groups were formed. Only 12.5% (n = 4) correctly identified the absence of random assignment in quasi-experimental scenarios. Representative pre-test justifications reflected this surface-level reasoning: "*This is a true experiment because there is an experimental class that gets the new method and a control class that uses the old method.*"

In the post-intervention task, a substantial conceptual shift was observed. The proportion of students utilizing principled methodological reasoning—specifically evaluating the mechanism of group formation (random assignment vs. intact groups) and the mitigation of confounding variables—increased to 78.125% (n = 25). Only 21.875% (n = 7) continued to rely on surface-feature classifications. Post-test justifications reflected a deeper understanding: "*Although the author calls it an experiment and uses two intact classes, it is actually a quasi-experimental design (non-equivalent control group) because the students were not randomly assigned, meaning initial equivalence cannot be guaranteed.*" This transition matrix indicates that the instructional sequence successfully disrupted initial misconceptions and fostered principled reasoning.

3.1.2. Preliminary Design and HLT Development

At this stage, the researcher designed HLT based on the literature review. There is a critical point that students must pass in studying the concept of quasi-experimental design. Students need to understand the concept of true-experimental design. There are three basic principles in experimental research: randomization, replication, and control. Randomization Ensures that subjects or experimental units are assigned to different groups randomly, eliminating selection bias. Replication increases the precision of estimates by repeating the experiment on multiple subjects. Control isolates the effect of treatments by using control groups or baseline conditions to minimize confounding factors (Fisher, 1935; Rahayu et al., 2022; Tanujaya, 2013). Therefore, the following HLTs have been developed.

a. Learning Objective

The HLT developed in this study was intended to support students' understanding of experimental and quasi-experimental research designs through a sequence of six interconnected learning domains. The trajectory was designed to guide students from understanding the fundamental principles of true experimental design toward developing a more informed understanding of quasi-experimental methodology. Each domain represents a specific learning objective, highlights key methodological ideas, and describes the expected learning outcome.

The first domain, *Conceptual Understanding of Experimental Design*, focuses on the three fundamental principles of true experimental design: randomization, replication, and control. Students are expected to distinguish these principles and understand their methodological roles in educational research. Randomization is introduced as a strategy for reducing selection bias through random assignment, replication as a means of improving the consistency and generalizability of findings, and control as a way of isolating the effects of an intervention. Through examples from educational research, students are expected to develop a clearer understanding of the methodological characteristics of true experimental designs.

The second domain, *Critical Analysis of Research Literature*, encourages students to examine authentic research reports, including undergraduate theses and peer-reviewed journal articles. At this stage, students analyze research designs, methodological procedures, and the reporting of research methods to determine whether the conclusions presented are supported by the methodological evidence. The emphasis is placed on evaluating the consistency between research questions, design implementation, and reported findings rather than relying solely on the terminology used by the authors.

The third domain, *Methodological Differentiation*, focuses on distinguishing quasi-experimental designs from true experimental designs based on their methodological characteristics. Particular attention is given to the absence of random assignment as a defining feature of quasi-experimental research and its implications for selection bias. Students are also encouraged to recognize that research designs should be classified according to their methodological features rather than the labels assigned by researchers.

The fourth domain, *Conceptual Formulation of Quasi-Experimental Design*, aims to help students construct a more complete understanding of quasi-experimental methodology. Rather than defining quasi-experimental design only in terms of what it lacks, students identify its defining characteristics, including the use of existing groups, intervention, comparison, and its practical application in educational settings where random assignment is often difficult to implement. This stage is intended to support a more balanced understanding of both the strengths and limitations of quasi-experimental research.

The fifth domain, *Evaluation of Causal Inference and Internal Validity*, extends students' understanding to the implications of methodological decisions. Students examine how the absence of random assignment may influence internal validity and causal inference by considering common threats such as selection bias and confounding variables. They are expected to evaluate how these methodological limitations affect the interpretation of research findings and the strength of causal conclusions.

The final domain, *Reflection and Validation*, provides opportunities for students to integrate the concepts developed throughout the trajectory. Students reflect on methodological decisions, apply relevant principles to authentic research situations, and justify their evaluations of research designs and findings. This stage is intended to encourage students to connect theoretical understanding with practical research contexts and to make methodological judgments that are supported by evidence.

Taken together, these six learning domains describe a progressive learning trajectory that moves from understanding the basic principles of true experimental design toward the ability to analyze, differentiate, evaluate, and justify quasi-experimental designs in educational research. As illustrated in Figure 2, the trajectory reflects a gradual development of methodological understanding through successive stages of conceptual learning, critical analysis, conceptual refinement, methodological evaluation, and reflective application. Rather than assuming that conceptual understanding develops in a linear manner, the trajectory provides a structured framework for supporting students' increasingly sophisticated reasoning about research methodology.

DOMAIN	LEARNING OBJECTIVE	CORE FOCUS	EXPECTED OUTCOME
1 Conceptual Understanding of Experimental Design	Distinguish randomization, replication, and control; relate to educational research.	• Randomization reduces selection bias • Replication improves precision and generalizability • Control isolates intervention effect	Students clarify concepts of true experimental design.
2 Critical Analysis of Research Literature	Critically analyze research manuscripts and evaluate alignment of principles and practice.	• Examine design, procedures, and reporting • Evaluate alignment between reported methods and research claims	Students evaluate methodological rigor within authentic research reports.
3 Methodological Differentiation	Differentiate experimental, quasi-experimental, and non-experimental designs.	• No random assignment → selection bias • Misclassify design type (true experiment vs. quasi-experiment)	Students classify designs by characteristics, not terminology.
4 Conceptual Formulation of Quasi-Experimental Design	Define quasi-experimental design and contrast with true experimental design.	• Key features: non-random assignment, existing groups, comparison, intervention • Value in educational settings	Students build a sound understanding of quasi-experimental design.
5 Evaluation of Causal Inference and Internal Validity	Assess threats to internal validity and strengthen causal claims.	• Threats: selection bias, confounding • Strategies to improve causal inference and internal validity	Students evaluate validity threats and improve causal inference.
6 Reflection & Validation	Justify design choices and validate interpretations using principles.	• Apply methodological principles • Justify design decisions and interpretation of results	Students make justified judgments in designing and interpreting research.

Figure 2. The component of learning objective

b. Learning Activities

The learning sequence was organized into a series of six interconnected activities designed to guide students gradually towards a deeper understanding of quasi-experimental design in educational research. This instructional trajectory is summarized in Figure 3.

The first activity, *Problem Orientation*, focused on activating prior knowledge. Students were given selected extracts from the methodology sections of undergraduate theses and scholarly journal articles which claimed to use a quasi-experimental research design. Using these excerpts as a starting point, students explored the problem while the lecturer posed guiding questions to stimulate discussion. For example, they asked whether all of the studies could be considered experimental and what methodological evidence could justify such claims. At this stage, the lecturer's role was primarily as a facilitator of problem exploration and initial discussion. The learning product of this stage was a list of preliminary assumptions and a defined research focus.

The second activity, *Guided Analysis*, involved a more systematic examination of research designs. Working in small groups, students analyzed one undergraduate thesis and one journal article each using guiding questions. They examined several methodological aspects, including participant selection procedures, the presence or absence of randomization, the use of control or comparison groups, and the implementation of pre-test–post-test structures. To support this process, students completed a structured research design analysis table. Throughout this stage, the lecturer acted as a facilitator and coach, scaffolding analytical thinking to help students focus on the methodological features most important for distinguishing between designs.

The third activity, *Conceptual Confrontation*, emphasized comparison. Each group presented the results of their analysis, prompting a broader class discussion about similarities, differences, and issues in the designs. This dialogue highlighted studies that met the criteria for true experimental designs compared to those that lacked randomization yet still examined instructional effects. At this stage, the lecturer's role was primarily that of a moderator of comparative and conceptual discussion. The outcome of this stage was a comparative map illustrating the different types of research designs and the key issues identified.

The fourth activity, *Directed Cognitive Conflict & Initial Abstraction*, encouraged students to think more critically about the contradictions found. Building on the previous comparisons, students were guided to identify contradictions between their initial assumptions and the actual research designs analyzed. They questioned underlying assumptions and derived initial abstractions about the nature of these designs. The lecturer served as a facilitator of this cognitive conflict and abstraction process. The learning product at this stage consisted of initial abstractions and key ideas derived from the cognitive conflict.

The fifth activity, *Concept Construction*, moved students toward synthesis. Using the abstractions from the previous stage, students constructed their own understanding and definitions of quasi-experimental design. The lecturer acted as a guide and supporter of concept construction and integration, encouraging students to synthesize their observations into coherent methodological statements. The outcome was student-generated definitions and conceptual models.

The sixth and final activity, *Reflection & Validation*, involved students reflecting on and evaluating their constructed understanding. They validated their concepts with evidence and criteria, comparing their definitions with those found in standard research methodology textbooks. A reflective discussion addressed which aspects of the students' formulations aligned with established definitions and why quasi-experimental designs are frequently used in educational research. At this stage, the lecturer facilitated reflection and validation, providing feedback and clarification. The final outcome was a validated understanding and refined definitions with supporting evidence. Finally, a schematic summary of this instructional sequence is presented in Figure 3.

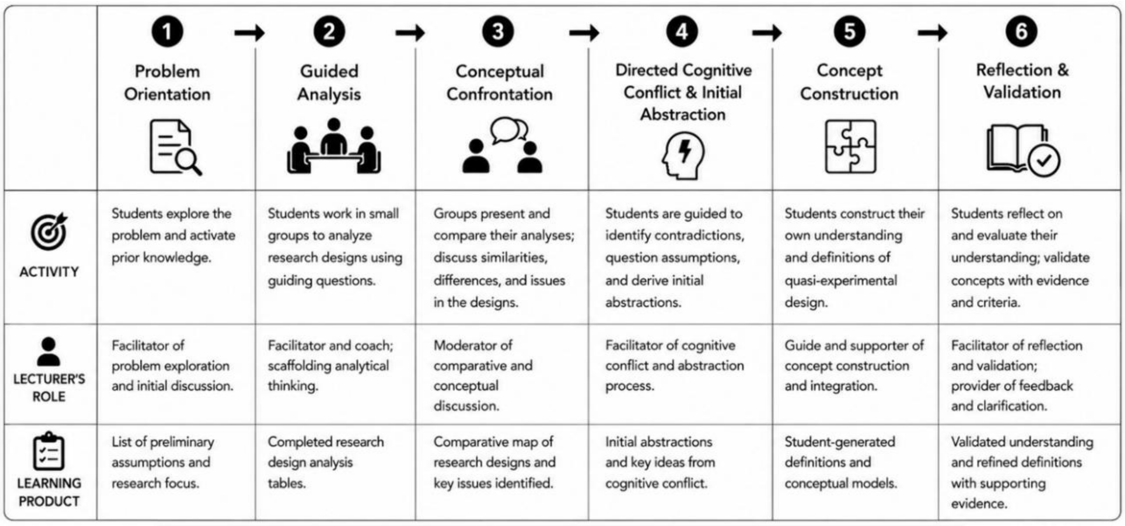

Figure 3. The six-activity of learning trajectory

c. *Hypothetical Conjecture*

The conjectures formulated in this study describe the anticipated development of students' understanding as they engage with each stage of the six-stage learning trajectory. These conjectures served as theoretically informed hypotheses about how students' methodological reasoning might evolve, the misconceptions they were likely to encounter, and the instructional supports required to facilitate conceptual development.

Stage 1: Problem Orientation. Initially, many students were conjectured to equate any study involving a treatment or intervention with a true experimental design. At this stage, they were expected to exhibit "*Label Bias*"—assuming that any study labeled "*experimental*" or "*quasi-experimental*" by the author is methodologically sound—and "*Treatment Equals Experiment,*" conflating the mere presence of an intervention with a true-experimental design. They were likely to pay limited attention to methodological aspects such as randomization and the role of control groups, reflecting a rather simplified, surface-level view of experimental research.

Stage 2: Guided Analysis. As students proceeded to systematically analyze excerpts from undergraduate theses and peer-reviewed journal articles using structured analysis tables, a gradual shift in their understanding was anticipated. However, they were expected to encounter several methodological confusions: mistaking the random selection of intact classes (sampling) for the random assignment of individuals to conditions; overlooking the critical issue of initial group equivalence by assuming that the mere presence of a "*control group*" was sufficient; and misunderstanding replication by equating a large sample size or the use of two intact classes with true methodological replication. Through guided analysis, students were expected to begin focusing on the mechanism of group formation rather than just the structural presence of groups.

Stage 3: Conceptual Confrontation. Further conceptual development was expected to emerge during the comparative discussion stage, where students examined different research designs side-by-side. At this stage, students were conjectured to experience "*Synthesis Error*"—concluding that a design is a "*true experiment*" despite clear evidence of non-random assignment, often due to over-reliance on the researcher's label—and "*Binary Thinking,*" struggling to categorize studies that have a treatment and a control group but lack randomization (e.g., asking, "*If it's not a true experiment, is it just bad science?*"). Through comparative discussion, students were expected to move beyond the simple distinction between experimental and non-experimental studies, starting to construct broader conceptual categories that accommodate intermediate forms of research design.

Stage 4: Directed Cognitive Conflict & Initial Abstraction. During the subsequent abstraction stage, students were expected to experience a "*Conceptual Gap*"—confusion about how to classify a methodologically rigorous non-randomized study—and initially resort to a "*Negative Definition*", defining quasi-experiment solely by what it lacks (e.g., "*an experiment without randomization*"). Through narrative scaffolding that framed quasi-experimental design as a pragmatic and valid methodological compromise, students' reasoning was expected to shift from relying primarily on definitional statements toward a more contextualized and argumentative form of methodological understanding. They were anticipated to begin identifying the positive, intentional design features (e.g., pre-testing, matched comparison groups) that strengthen causal inference under practical constraints.

Stage 5: Concept Construction. As students attempted to formulate their own definitions of quasi-experimental design, they were conjectured to encounter "*Overgeneralization*"—creating definitions that were either too narrow (only describing one type, like non-equivalent groups) or too vague—and "*Strengths Omission,*" focusing only on limitations while failing to articulate the design's utility and rationale. Through guided prompts and sentence frames, students were expected to refine their conceptual understanding, defining quasi-experimental design not by what it isn't, but by its purpose, its core methodological characteristics, and the typical contexts for its use.

Stage 6: Reflection & Validation. Finally, after engaging with established theoretical explanations presented in research methodology literature, students were conjectured to potentially exhibit "*Textbook Authority*"—accepting the textbook definition passively without integrating it with their constructed understanding—and a "*Justification Gap,*" struggling to articulate why quasi-experimental designs are so prevalent in educational research beyond "*it's easier.*" Through comparative reflection and critical application, students' conceptualizations were expected to become more stable and coherent. By this point, students should be able to explain how specific design choices affect the validity of a study and the strength of any causal inferences that can be made, demonstrating principled methodological reasoning. The students' conjectures across the HLT are presented in Figure 4.

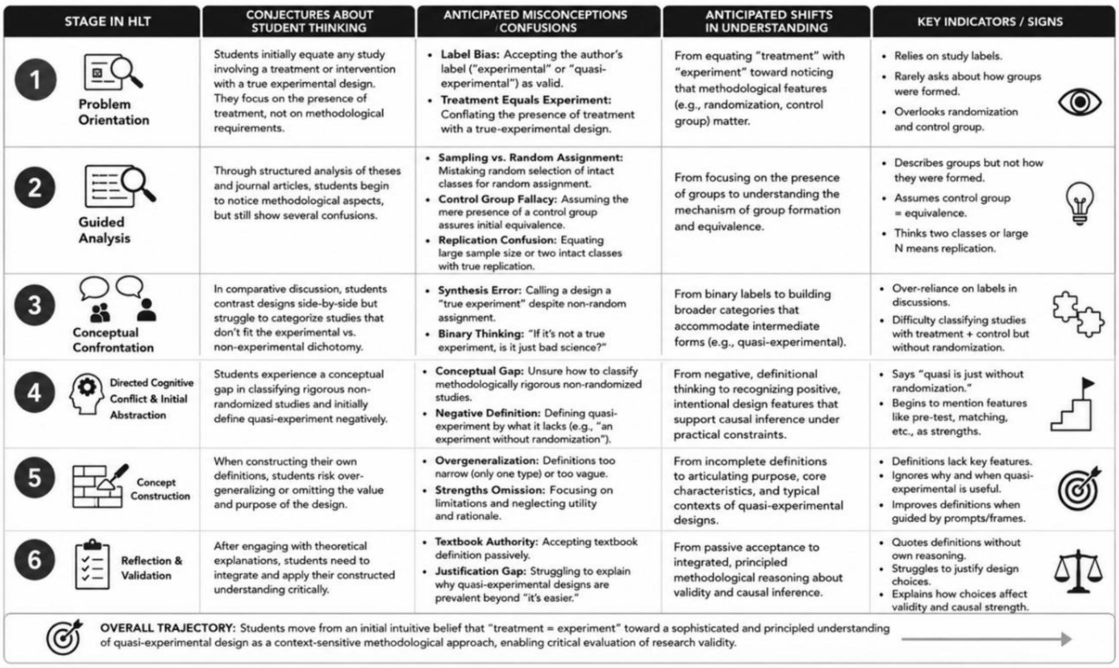

Figure 4. Students thinking conjectures across the HLT

3.1.3. Design Experiment: Implementation and Findings

a. Teaching Experiment

The teaching experiment constituted the first stage of empirical testing of the HLT. In accordance with the principles of design research, this phase aimed not only to examine the feasibility of the instructional design but also to investigate the extent to which the conjectured learning processes corresponded with students' actual patterns of reasoning (Bakker, 2018; Gravemeijer & Cobb, 2006). The learning activities embedded in the HLT were therefore implemented and observed with particular attention to how students interpreted key methodological concepts related to experimental and quasi-experimental designs.

Overall, the instructional sequence unfolded largely in line with the expectations formulated during the preliminary design phase. Many students demonstrated patterns of reasoning that closely reflected the initial conjectures. In particular, when examining excerpts from research manuscripts, several students experienced a form of conceptual tension upon recognizing that numerous studies labeled as "*experimental*" did not employ random assignment. This realization often became a starting point for deeper methodological discussion, leading students to reconsider their initial assumptions about what constitutes an experimental study. As the learning activities progressed, many students began to articulate a more refined operational understanding of quasi-experimental design, describing it in terms of its methodological characteristics rather than merely as the absence of randomization.

At the same time, the teaching experiment revealed an important nuance that had not been fully anticipated in the preliminary conjectures. A subset of students encountered difficulty in interpreting the roles of treatment and control within different experimental paradigms. In several instances, students appeared to equate the presence of two groups with the existence of a valid control condition, without examining the methodological basis on which those groups had been formed. This interpretation suggested that some students were relying on a simplified structural understanding of experimental design rather than engaging with the underlying logic of group equivalence.

Further analysis indicated that this difficulty was closely related to the shifting meaning of the concept of "*control*" across methodological traditions. In the context of a true experimental design, the control condition is typically constructed through random assignment, which seeks to distribute pre-existing differences among participants and thereby ensure initial comparability between groups. Through this procedure, the control condition functions as a systematically generated baseline against which the effects of the intervention can be evaluated.

In quasi-experimental research, however, the notion of control is often less procedural and more comparative in nature. Comparison groups frequently exist as intact units within educational settings, for instance, parallel classes or cohorts of students, and are not assumed to be fully equivalent at baseline. Consequently, researchers rely on alternative strategies, such as pretesting, statistical adjustment, or matching procedures, to estimate and address potential differences between groups. For students encountering these methodological distinctions for the first time, the conceptual transition from "*control as procedural equivalence*" to "*control as comparative reference*" proved challenging.

The emergence of this misconception provides valuable insight into how students interpret experimental terminology in educational research. It also suggests that greater attention should be given to scaffolding students' analysis of group formation procedures during the manuscript analysis stage. Future iterations of the HLT may therefore benefit from incorporating more explicit prompts that direct students' attention to issues of initial group equivalence, such as asking whether comparison groups were created through random assignment or were pre-existing entities within the educational context. By foregrounding these methodological questions, the instructional design may better support students in developing a more sophisticated understanding of internal validity in quasi-experimental studies.

b. Classroom Experiment (Pilot Implementation)

The classroom experiment phase extended the testing of the refined HLT to a broader, authentic classroom context. In design research, this stage examines the practicality and robustness of the instructional design under naturalistic conditions (Mumu et al., 2021; Prahmana, 2017). The complete set of learning activities, conjectures, and assessment instruments was implemented with a cohort of 32 students enrolled in the "*Experimental Design in Education*" course over six weekly sessions.

Pre- and Post-Intervention Evidence of Conceptual Shift

The viability of the learning trajectory was strongly supported by empirical shifts in students' written justifications. In the pre-intervention task, the majority of students relied on surface-feature classification. For example, Student A initially classified a study as a "*true experiment*" with the justification: "*It is a true experiment because there is an experimental class that receives the new method and a control class that uses the old method.*" However, in the post-intervention task, the same student demonstrated principled methodological reasoning, writing: "*Although the author uses two classes, this is a quasi-experimental design (nonequivalent control group). The intact classes were not randomly assigned, meaning initial equivalence cannot be guaranteed, and the author must use a pre-test to control for selection bias.*" This transition from surface-level labeling to evaluating the mechanism of group formation was observed in 84.38% (n = 27) of the cohort.

Classroom Interactions and Cognitive Conflict

Observations from the classroom experiment provided rich insights into students' conceptual development, particularly during the "*Conceptual Confrontation*" stage (Stage 3). Video transcripts revealed that students initially struggled with the concept of "*control.*" The following excerpt from a classroom discussion illustrates how the lecturer's scaffolding redirected students from surface-level assumptions to principled reasoning:

Lecturer : *Look at this thesis excerpt. The author claims this is a true experiment because there is a "control group." Do you agree based on the methodology described?*
Student 1 : *Yes, because there are two groups being compared, so one is the control.*
Lecturer : *But how were those two groups formed? Were the individual students randomly assigned to them, or were they already in those classes?*
Student 2 : *They were already in those classes. So, they might have different abilities or motivations from the start.*
Lecturer : *Exactly. So, does the "control" here function the same way as a procedurally equivalent baseline in a randomized true experiment?*
Student 1 : *No, it is just a comparison group. We cannot be sure they are equal at the baseline without random assignment.*

This dialogue demonstrates the productive cognitive conflict generated by the HLT, moving students away from binary thinking toward a nuanced understanding of internal validity.

Analysis of Student Artifacts

Further evidence of conceptual development was captured in students' written artifacts, specifically the structured research design analysis worksheets completed during Stage 2 and refined in Stage 4. Initially, students' analyses were deficit-oriented. For instance, an early worksheet entry simply stated: "*Weakness: No randomization.*" However, following the "*Directed Cognitive Conflict & Initial Abstraction*" scaffolding, students' later artifacts reflected a pragmatic reframing of quasi-experimental designs. A revised artifact from the same student noted: "*Strength: Although random assignment was not feasible in this school setting, the researcher used a pre-test to establish baseline equivalence and applied ANCOVA to statistically control for initial differences. This makes it a robust quasi-experimental design.*"

Refinement of the HLT Based on Classroom Findings

While the overall structure of the HLT proved effective, the classroom experiments highlighted areas for further refinement. A recurring challenge emerged when students were asked to design hypothetical quasi-experimental studies. Some initially proposed forms of randomization that were unrealistic in actual educational settings (e.g., randomly reassigning students across different schools). Only after guided reflection did they adjust their proposals to accommodate contextual constraints, such as using intact classes with matched pre-test scores. This observation confirmed that the logic of "*methodological compromise*" requires repeated, scaffolded application across diverse scenarios before it is fully internalized.

Taken together, the triangulated findings from pre/post-tests, classroom dialogue, and written artifacts provide robust empirical support for the six-stage learning trajectory. They confirm that the HLT successfully guides students from surface-feature classification toward principled, context-sensitive methodological reasoning.

Table 1 details the six-stage learning trajectory. It outlines each Activity Stage, the Hypothesized Student Misconceptions and Error (Conjectures), the Specific Instructor Interventions (Scaffolding) designed to address them, while the Primary Learning Goal of the stage.

Table 1. The learning trajectory on comprehending quasi-experiment design

Activity Stage	Conjecture: Students Misconception & Errors	Designed Scaffolding: Teacher Intervention
Problem Orientation	• **Label Bias:** Assuming any study labelled "experimental" or "quasi-experimental" by the author is methodologically sound. • **Treatment Equals Experiment:** Conflating the mere presence of a treatment or intervention with a true-experimental design.	• Present contradictory examples (a strong randomized controlled trial (RCT) vs. a weak pre-post design). • Pose a reflective question: "Does the author's label, or the presence of a treatment, automatically make a study a true experiment? What

Activity Stage	Conjecture: Students Misconception & Errors	Designed Scaffolding: Teacher Intervention
		methodological evidence should we look for?”
Guided Analysis	• **Randomization Confusion:** Mistaking the random *selection* of intact classes (sampling) for the random *assignment* of individuals to conditions. • **Group Equivalence:** Overlooking the critical issue of initial group equivalence, assuming a “control group” is sufficient. • **Replication Misunderstanding:** Equating a large sample size or the use of two classes with true methodological replication.	• Provide a structured analysis table focusing on: Selection, Allocation, Comparison, Measurement. • Ask targeted questions: “Who or what was randomized? Individuals or pre-existing groups?” “How similar were the groups *before* the treatment began?” • Clarify that two intact classes are not replicates but independent, potentially non-equivalent units.
Conceptual Confrontation	• **Synthesis Error:** Concluding a design is a “true experiment” despite clear evidence of non-random assignment, often due to over-reliance on the researcher’s label. • **Binary Thinking:** Struggling to categorize studies that have a treatment and a control group but lack randomization (“If it’s not a true experiment, is it just bad science?”).	• Facilitate a comparative discussion where groups present side-by-side analyses of a true experiment and a non-randomized study. • Pose a metacognitive question: “If the core principle of randomization is missing, what are the direct consequences for the study’s internal validity and causal conclusions?” • Introduce key threats (selection bias, maturation) that become paramount.
Directed Cognitive Conflict & Initial Abstraction	• **Conceptual Gap:** Experiencing confusion about how to classify a methodologically rigorous non-randomized study. • **Negative Definition:** Initially defining quasi-experiment solely by what it lacks (e.g., "an experiment without randomization").	• Introduce the term “Quasi-Experiment” as a formal methodological category. • Provide narrative scaffolding: Frame it as the pragmatic and valid methodological compromise for establishing plausible causal inference when random assignment is not feasible.

Activity Stage	Conjecture: Students Misconception & Errors	Designed Scaffolding: Teacher Intervention
		• Guide students to identify the positive, intentional design features they observed (e.g., use of a carefully matched comparison group, pre-testing, statistical controls).
Concept Construction	• **Overgeneralization:** Creating definitions that are either too narrow (only describing one type, like non-equivalent groups) or too vague. • **Strengths Omission:** Focusing only on limitations, failing to articulate the design's utility and rationale.	• Prompt: "Define quasi-experimental design not by what it *isn't*, but by its *purpose*, its *core methodological characteristics*, and the *typical contexts* for its use." • Offer a sentence frame: "*A quasi-experimental design is a research approach used to [PURPOSE] by [KEY METHODOLOGICAL ACTIONS], often employed when [PRACTICAL CONSTRAINT] because it allows for [STRENGTH].*"
Reflection & Validation	• **Textbook Authority:** Accepting the textbook definition passively without integrating it with their constructed understanding. • **Justification Gap:** Struggling to articulate *why* this design is so prevalent in educational research beyond "it's easier."	• Have students compare their definitions with authoritative sources (e.g., Campbell & Stanley). Discuss: "What aspects did you capture well? What did the textbook add?" • Pose a final application question: "Using your refined definition, re-evaluate one initial manuscript. Justify: Is it a *strong* or *weak* example of quasi-experimental design, and why?"

The proposed learning trajectory was structured around a sequence of progressively evolving goals intended to support students' conceptual development. At the initial stage, the instructional activities focused on activating students' prior knowledge and eliciting their initial assumptions about experimental research. These early discussions were intended to reveal students' existing conceptions, which in many cases reflected a simplified understanding of experimental design.

Building on this foundation, the second stage was designed to strengthen students' analytical skills through the examination of authentic research manuscripts. By engaging with methodological sections of theses and journal articles, students were encouraged to identify concrete evidence of how research designs are actually implemented in practice. This activity allowed students to move beyond abstract definitions and begin grounding their understanding in empirical examples.

The third stage sought to create a productive form of cognitive tension. Through guided comparison between the ideal characteristics of true experimental designs and the methodological procedures commonly found in educational research, students were encouraged to recognize discrepancies between theoretical expectations and practical research conditions. This contrast highlighted the methodological compromises that often arise when experimental procedures are implemented in real educational settings.

In the fourth stage, the instructional focus shifted toward resolving this tension through conceptual reconstruction. Students were guided to reconsider their earlier assumptions and to recognize quasi-experimental design not merely as an imperfect form of experimentation but as a purposeful methodological alternative developed to address contextual constraints. This stage encouraged students to expand their conceptual categories and to interpret research designs in terms of their methodological logic.

The fifth stage aimed to integrate the insights generated during earlier discussions. Through collective analysis and reflection, students synthesized their observations into more coherent explanations of quasi-experimental design. Importantly, the conceptual understanding that emerged during this stage was largely articulated by the students themselves, reflecting the cumulative outcomes of the preceding analytical activities.

Finally, the sixth stage was intended to stabilize and refine the newly constructed concept. Students compared their emerging interpretations with established theoretical descriptions found in research methodology literature and applied the concept in evaluating research examples. Through this process of theoretical comparison and critical application, students' understanding of quasi-experimental design became more coherent and internally consistent.

c. Retrospective Analysis

In accordance with design research methodology (Bakker, 2018; Gravemeijer & Cobb, 2006), the retrospective analysis systematically compared the HLT with the actual learning trajectory observed across both the teaching experiment and the classroom experiment. Rather than merely evaluating whether the instructional sequence functioned as intended, this phase examined how students' evolving reasoning corresponded with, diverged from, and ultimately reshaped the initial design conjectures. These empirical insights informed iterative refinements of the HLT and culminated in the formulation of a LIT for fostering methodological reasoning about quasi-experimental design.

Alignment and Deviation from Hypothetical Conjectures

Overall, the observed learning trajectory closely aligned with the original HLT. The initial conjectures regarding Label Bias and Treatment Equals Experiment were consistently confirmed during the pre-intervention tasks and early classroom discussions. Students predominantly classified research designs using superficial indicators, such as the author's terminology or the mere presence of an intervention, rather than examining the underlying methodological logic. The structured analysis worksheets and lecturer questioning effectively disrupted these surface-level heuristics by directing students' attention toward group formation, comparison strategies, and causal inference.

Nevertheless, the retrospective analysis also revealed an important deviation from the original conjectures. The HLT anticipated that students' principal difficulty would involve understanding the absence of random assignment. Instead, classroom evidence indicated a deeper conceptual misconception. Many students interpreted the mere existence of a control group as sufficient evidence of methodological rigor, overlooking the fact that intact groups may differ systematically before the intervention. Consequently, they failed to recognize the need for compensatory procedures—such as pre-test equivalence, matching, statistical adjustment (e.g., ANCOVA), or other design-based controls—to strengthen causal claims in quasi-experimental research.

This finding suggests that the original HLT underestimated the conceptual transition required for students to reconceptualize "*control*." Initially, students viewed control as a procedural guarantee automatically provided by randomization. Through the intervention, however, effective learning required a more sophisticated understanding of control as a methodological strategy for reducing alternative explanations when random assignment is infeasible.

Empirical Refinement of the HLT

These findings prompted iterative refinement of the instructional trajectory during the classroom experiment. The most substantial revisions occurred in Stage 2 (Guided Analysis) and Stage 4 (Directed Cognitive Conflict & Initial Abstraction).

Rather than asking students simply to identify whether a study included a control group, the revised HLT required them to explain how equivalence between groups had been established or approximated. Lecturer prompts therefore shifted from descriptive questions (e.g., "*Was there a control group?*") toward explanatory and evaluative questions, such as "*How did the researchers reduce selection bias without random assignment?*" and "*What evidence supports the comparability of the groups?*"

Similarly, the concept-construction activities in Stage 5 (Concept Construction) were redesigned to emphasize the reasoning underlying quasi-experimental designs. Sentence frames no longer asked students merely to define quasi-experiments, but instead required them to explain the compensatory mechanisms that justified methodological decisions. As a result, students increasingly described quasi-experimental designs not as inferior versions of true experiments, but as context-sensitive research designs that balance practical constraints with strategies for strengthening causal inference.

These refinements transformed the HLT from a trajectory primarily focused on identifying the absence of randomization into one explicitly supporting students' understanding of how methodological compromises can still produce credible evidence when accompanied by appropriate design and analytical controls.

Formulation of the Local Instructional Theory (LIT)

The retrospective analysis identified four interconnected instructional principles that consistently supported students' progression from surface-feature classification toward principled methodological reasoning. Together, these principles constitute the emergent LIT for teaching quasi-experimental design:

1) *Authentic Epistemic Confrontation*: Conceptual change is initiated when students critically examine authentic research manuscripts containing realistic methodological constraints, rather than simplified textbook examples. Exposure to authentic cases encourages students to evaluate methodological decisions instead of memorizing definitions.
2) *Comparative Cognitive Conflict*: Systematic comparison between true experimental and quasi-experimental studies creates productive cognitive conflict that challenges intuitive assumptions about randomization, control, and causal inference. This comparison encourages students to identify the underlying logic of research design rather than relying on surface characteristics.
3) *Pragmatic Reframing*: Students develop a more sophisticated understanding of quasi-experimental design when instructional scaffolding emphasizes its methodological strengths, compensatory strategies, and contextual appropriateness, instead of portraying it merely as an imperfect substitute for randomized experiments.
4) *Metacognitive Justification*: Deep methodological understanding emerges when students are required to explicitly justify their design decisions using methodological evidence rather than declarative definitions. Explaining why particular controls, matching procedures, or statistical adjustments are necessary promotes transferable methodological reasoning.

The retrospective analysis demonstrates that the primary value of the HLT lies not in confirming predetermined instructional conjectures, but in exposing unexpected aspects of students' reasoning that require instructional redesign. The iterative refinement process therefore produced more than an improved sequence of learning activities; it generated a context-sensitive LIT explaining how authentic comparison, structured cognitive conflict, pragmatic reframing, and metacognitive justification collectively support the development of principled methodological reasoning about quasi-experimental design. This theoretical contribution extends beyond the present instructional setting and offers transferable design principles for research methodology education in higher education.

3.2. Discussion

3.2.1. From Surface-Feature Classification to Methodological Reasoning

The findings indicate that students' understanding of experimental and quasi-experimental research evolved substantially. They progressed from recognizing superficial features to evaluating methodological characteristics on principle. Before the intervention, many students categorized research designs based on explicit labels provided by authors or the mere presence of an intervention, without examining the underlying logic of group formation, control mechanisms, and causal inference. These reasoning patterns align with established research on conceptual development, which demonstrates that novices typically rely on salient surface features, whereas deeper understanding requires attention to the structural principles that govern a domain (Vosniadou, 2013). Similar patterns have been documented in research methods education, where novice researchers frequently focus on procedural labels rather than the methodological principles that underpin research design decisions (Earley, 2014).

The learning trajectory implemented in this study was explicitly designed to disrupt students' reliance on surface-level classification by engaging them with authentic research reports that required systematic analysis of methodological decisions rather than research terminology alone. Instead of asking whether a study was labelled *experimental* or *quasi-experimental*, students were encouraged to examine how participant groups were formed, whether random assignment was feasible, how comparison groups were established, and how these design features affected the validity of causal claims. Through repeated engagement with authentic research cases, students gradually shifted from identifying design labels to evaluating the methodological logic underlying each study. This progression was evident in the comparison between the pre- and post-intervention tasks, where students increasingly justified their classifications using evidence related to group formation, control of confounding variables, and threats to internal validity rather than superficial characteristics.

These findings align with methodological literature emphasizing that rigorous evaluation of experimental research depends on understanding the logic of design rather than simply recognizing design labels (Creswell & Creswell, 2018). From a design research perspective, the findings further suggest that authentic case analysis functioned as a productive learning environment by creating opportunities for cognitive conflict between students' initial intuitions and the methodological evidence embedded in published research reports (Bakker, 2018). As students repeatedly justified and defended their decisions, they progressively constructed more principled methodological reasoning that integrated theoretical knowledge with contextual judgment. Such reasoning is particularly important in educational research, where practical constraints frequently necessitate quasi-experimental designs and where the quality of causal inference depends on researchers' ability to evaluate methodological trade-offs rather than pursue idealized experimental conditions (Shadish et al., 2002).

Ultimately, this trajectory reflects a broader imperative in contemporary research education. Recent research suggests that developing methodological reasoning requires learners to move beyond procedural recognition toward the epistemic evaluation of research evidence. Rather than memorizing methodological terminology, students need opportunities to justify causal claims by examining design assumptions, alternative explanations, and threats to validity. Such epistemic engagement represents a critical shift from procedural competence to evidence-based methodological thinking (Grosz, 2023; Horton, 2023; Martínez-Abad & León, 2023), indicating that future curricula must prioritize the critical evaluation of causal assumptions over the rote application of design labels.

3.2.2. Conceptual Development of Quasi-Experimental Design

Another important finding concerns students' evolving understanding of quasi-experimental design. At the beginning of the learning trajectory, many students conceptualized quasi-experimental research primarily in terms of what it lacked—particularly the absence of random assignment. Consequently, they frequently regarded quasi-experiments as inherently inferior to true experiments. This perception is unsurprising, as research methods textbooks commonly introduce quasi-experimental designs by contrasting them with the methodological ideal of randomized experiments (Creswell & Creswell, 2018; Shadish et al., 2002).

The instructional trajectory was deliberately designed to challenge this deficit-oriented perspective. During the Conceptual Confrontation and Directed Cognitive Conflict phases, students examined authentic educational studies in which random assignment was either impractical or ethically inappropriate. Rather than evaluating research quality solely through the presence or absence of randomization, students were encouraged to analyze the rationale underlying design decisions, including the use of intact classrooms, comparison groups, and statistical controls to strengthen causal inference. These activities prompted students to reconsider their initial assumptions and progressively evaluate quasi-experimental designs in relation to their methodological appropriateness for particular educational contexts.

The observed conceptual progression is consistent with conceptual change theory, which argues that durable learning occurs when existing conceptions become inadequate for explaining new evidence and learners actively reconstruct more coherent conceptual frameworks (Posner et al., 1982; Vosniadou, 2013). In this study, authentic research cases served as conceptual counterexamples that exposed the limitations of defining quasi-experimental research solely by the absence of random assignment. As students repeatedly justified their design classifications using methodological evidence, they gradually reframed quasi-experimental designs from being "incomplete experiments" to being contextually appropriate approaches for investigating causal relationships under realistic educational conditions.

This interpretation aligns with contemporary methodological scholarship, which emphasizes that the quality of causal inference depends less on adherence to an idealized experimental model than on the extent to which a design addresses threats to internal validity within practical constraints (Creswell & Creswell, 2018; Shadish et al., 2002). Recent discussions in educational and social science methodology similarly argue that quasi-experimental designs provide robust opportunities for causal investigation when random assignment is infeasible, provided that researchers explicitly justify design choices and employ appropriate strategies to minimize bias (Cook et al., 2008; Johnson & Christensen, 2020). Consequently, the present findings suggest that methodological reasoning develops not by encouraging students to seek "perfect" research designs, but by helping them evaluate the appropriateness of alternative designs in relation to specific research contexts and methodological trade-offs.

Ultimately, the present findings resonate with recent discussions in educational methodology, which argue that contemporary research training should emphasize design justification rather than design hierarchy. Increasingly, educational researchers are encouraged to evaluate how methodological decisions address practical, ethical, and contextual constraints instead of assuming that randomized experiments always represent the optimal solution. Consequently, understanding quasi-experimental design involves learning to balance internal validity with ecological validity and implementation feasibility (Gopalan et al., 2020; Grosz, 2023).

3.2.3. Implications for the Local Instructional Theory

The iterative refinement of the HLT across teaching and classroom experiments yielded a LIT, which explains how methodological reasoning can be fostered through systematic analysis of authentic research cases. In line with the design research tradition, these principles should be viewed as empirically grounded design principles that account for the learning processes observed in this specific instructional context (Bakker, 2018; Gravemeijer & Cobb, 2006; Tanujaya et al., 2018), rather than as universal prescriptions. Together, they describe the instructional mechanisms through which students progressed from recognizing research designs at a surface level to principled methodological reasoning.

The first design principle, authentic epistemic confrontation, emphasizes the importance of engaging students with authentic research reports rather than simplified textbook examples. These authentic cases required students to evaluate the methodological evidence embedded in real studies, including participant selection, group formation, intervention procedures, and threats to validity, instead of relying on design labels provided by the authors. The findings indicate that repeated exposure to authentic methodological problems encouraged students to adopt a more evidence-based approach to evaluating research designs, shifting their focus from procedural terminology to methodological reasoning. This principle aligns with contemporary perspectives on case-based learning, which show that authentic problems encourage deeper engagement with disciplinary practices and help students to develop expert-like reasoning skills (Hmelo-Silver et al., 2007; Jonassen, 2011; Mumu & Tanujaya, 2019).

The second principle, comparative cognitive conflict, emphasizes the importance of systematically comparing true experimental and quasi-experimental studies. Direct comparison enables students to recognize that methodological differences extend beyond the presence or absence of random assignment, encompassing broader considerations such as internal validity, feasibility, and causal inference. In line with conceptual change theory (Posner et al., 1982; Vosniadou, 2013), presenting conflicting methodological explanations appears to stimulate conceptual restructuring by revealing the limitations of students' initial ideas and encouraging them to construct more coherent explanatory frameworks. Recent research in science and mathematics education also supports this principle, showing that analysing contrasting cases in a comparative way makes underlying principles more obvious and facilitates conceptual change (Alfieri et al., 2013; Schwartz & Bransford, 1998).

The third principle, pragmatic reframing, illustrates how students gradually shifted their perception of quasi-experimental design from a deficit-based viewpoint to a perspective centred on methodological suitability. Rather than viewing quasi-experimental research as an inferior substitute for true experimentation, students increasingly recognised its legitimacy as a strategy for investigating causal relationships within the constraints of real-world education. This finding suggests that learners develop a methodological understanding when they evaluate research designs in relation to their underlying purposes, contextual feasibility and the trade-offs involved in balancing methodological rigour with practical realities (Creswell & Creswell, 2018; Shadish et al., 2002). This principle aligns with recent scholarship emphasising that research design education should prepare students to navigate real-world constraints rather than pursue idealised experimental conditions (Cook et al., 2008; Johnson & Christensen, 2020).

The fourth principle, metacognitive justification, relates to students' ability to articulate and defend methodological decisions using theoretical arguments. By the end of the instructional sequence, not only had students identified research designs more accurately, they had also justified why particular designs were appropriate, discussed threats to internal validity and explained how methodological decisions influenced the credibility of causal claims. These findings suggest that methodological reasoning extends beyond procedural knowledge to include reflective justification, evaluation of evidence and explicit consideration of alternative methodological choices. This interpretation is consistent with contemporary perspectives on scientific reasoning, which identify explanation, argumentation and evidence-based justification as key indicators of conceptual understanding and disciplinary proficiency (Kuhn, 2010; Osborne, 2014).

Taken together, these four principles form a local instructional theory that describes how methodological reasoning can be developed through a series of carefully designed activities involving authentic case analysis, conceptual confrontation, contextual reframing and reflective justification. Although the theory was developed for a specific research methods course for pre-service mathematics teachers, it offers transferable design principles that could inform future instructional interventions aimed at supporting methodological reasoning in educational research. More broadly, this study contributes to design research by demonstrating how the iterative refinement of a HLT can produce instructional principles that can be applied in different contexts while remaining sensitive to contextual variation (Bakker, 2018; Van den Akker et al., 2006).

4. CONCLUSION

This study involved the development and iterative refinement of a six-stage Hypothetical Learning Trajectory (HLT) to improve the understanding of quasi-experimental design among pre-service mathematics teachers. The findings demonstrate that the instructional trajectory successfully facilitated a conceptual progression from classifying surface features towards principled methodological reasoning. Rather than relying on research labels or the mere presence of an intervention, students increasingly evaluated methodological characteristics such as group formation, random assignment, internal validity and the logic of causal inference. A particularly significant outcome was the shift in students' understanding of quasi-experimental design — from viewing it as an imperfect form of true experimentation, to recognising it as a contextually appropriate methodological approach for investigating causal relationships within the constraints of real-world education.

The principal contribution of this study lies in the development of a Local Instructional Theory (LIT), which comprises four empirically grounded design principles: Authentic Epistemic Confrontation, Comparative Cognitive Conflict, Pragmatic Reframing and Metacognitive Justification. Together, these principles explain how authentic research cases, structured comparisons, contextual reasoning and evidence-based justifications can progressively foster methodological reasoning. As well as providing a practical framework for teaching research methodology, the proposed LIT also contributes to design research by offering an explanatory account of how instructional design can support the conceptual development of experimental and quasi-experimental research designs. Specifically for mathematics teacher education program, the LIT provides a structured approach to cultivating the critical methodological reasoning that pre-service teachers require to become evidence-based practitioners and reflective researchers.

Despite these contributions, this study has several limitations. Firstly, the study was conducted with a single cohort of pre-service mathematics teachers in one institutional context, which may restrict how widely the findings can be applied. Secondly, the study examined students' conceptual understanding during instruction, rather than their long-term ability to independently design, implement and critically evaluate quasi-experimental research. Thirdly, while several strategies were employed to enhance trustworthiness, the researcher's dual role as instructor and investigator may have affected classroom interactions and data interpretation.

Future research should examine whether the proposed Local Instructional Theory can be applied across different disciplinary, institutional and cultural contexts. Longitudinal investigations are also required to ascertain whether the methodological reasoning developed through the HLT is maintained and applied in real research projects, particularly undergraduate thesis work. Subsequent design research should also refine instructional support for complex methodological concepts, especially baseline equivalence, control of confounding variables and causal inference in non-randomized settings, to further strengthen students' methodological reasoning. Ultimately, these efforts could contribute to a broader vision of research methods education, preparing future teachers to critically consume research evidence and produce contextually grounded, methodologically sound studies that advance educational practice.

Acknowledgments

The authors would like to express our gratitude to the Mathematics Education Study Programme at the Faculty of Teacher Training and Education at Universitas Papua for their financial support and institutional assistance, which made this research possible. They would also like to thank the undergraduate students enrolled on the *Experimental Design in Education* course at the Mathematics Education Study Programme for their active participation, thoughtful engagement and valuable contributions throughout the implementation of the learning trajectory and the data collection process. Their willingness to critically analyse research designs and reflect on their methodological reasoning was essential to the successful completion of this study.

Declarations

Author Contribution : JM: Conceptualization, Investigation, Methodology, and Writing - original draft; L: Formal analysis, Validation, and Writing - review & editing; BT: Investigation, Supervision, Visualization, and Writing - review & editing.

Funding Statement : This research was funded by the Program Studi Pendidikan Matematika at Universitas Papua through the Institutional Development Fund, under Grant Number 070-F/UN42.10.1/PP/2025.

Conflict of Interest : The authors declare no conflict of interest.

Additional Information : Additional information is available for this paper.

REFERENCES

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Dissertation. Utrecht University. https://dspace.library.uu.nl/handle/1874/893

Bakker, A. (2018). *Design research in education: A practical guide for early career researchers*. Routledge. https://doi.org/10.4324/9780203701010

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, *13*(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1

Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, *34*(6), 3–15. https://doi.org/10.3102/0013189x034006003

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical thinking and learning*, *6*(2), 81–89. https://doi.org/10.1207/s15327833mtl0602_1

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13. https://doi.org/10.3102/0013189x032001009

Cole, R. (2024). Inter-rater reliability methods in qualitative case study research. *Sociological Methods & Research*, *53*(4), 1944–1975. https://doi.org/10.1177/00491241231156971

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin Boston.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*(4), 724–750. https://doi.org/10.1002/pam.20375

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Sage.

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

Earley, M. A. (2014). A synthesis of the literature on research methods education. *Teaching in Higher Education*, *19*(3), 242–253. https://doi.org/10.1080/13562517.2013.860105

Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist*, *42*(3), 123–137. https://doi.org/10.1080/00461520701416173

Fisher, R. A. (1935). *The design of experiments*. Oliver & Boyd.

Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Kluwer Academic Publishers.

Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

Gopalan, M., Rosinger, K., & Ahn, J. B. (2020). Use of quasi-experimental research designs in education research: Growth, promise, and challenges. *Review of Research in Education*, *44*(1), 218–243. https://doi.org/10.3102/0091732x20903302

Gravemeijer, K., & Cobb, P. (2006). Design research from a learning design perspective. In J. Van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 29–63). Routledge. https://doi.org/10.4324/9780203088364-12

Gravemeijer, K. P. E. (1994). *Developing realistic mathematics education*. CDBeta Press.

Grosz, M. P. (2023). Should researchers make causal inferences and recommendations for practice on the basis of nonexperimental studies? *Educational Psychology Review*, *35*(2), 57. https://doi.org/10.1007/s10648-023-09777-7

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Horton, N. J. (2023). Teaching causal inference: moving beyond 'correlation does not imply causation'. *Journal of Statistics and Data Science Education*, *31*(1), 1–3. https://doi.org/10.1080/26939169.2023.2178778

Johnson, R. B., & Christensen, L. (2020). *Educational research: Quantitative, qualitative, and mixed approaches* (7th ed.). SAGE Publications.

Jonassen, D. H. (2011). *Learning to solve problems: A handbook for designing problem-solving learning environments*. Routledge.

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*(5), 810–824. https://doi.org/10.1002/sce.20395

Martínez-Abad, F., & León, J. (2023). Inferencia causal en investigación educativa: Análisis de la causalidad en estudios observacionales de carácter transversal [Causal inference in educational research: Causal analysis in cross-sectional observational studies]. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, *29*(2), 3. https://doi.org/10.30827/relieve.v29i2.26843

Mumu, J., Prahmana, R. C. I., Sabariah, V., Tanujaya, B., Bawole, R., Warami, H., & Monim, H. O. L. (2021). Students' ability to solve mathematical problems in the context of environmental issues. *Mathematics Teaching Research Journal*, *13*(4), 99–111.

Mumu, J., & Tanujaya, B. (2019). Analysis of mathematical connection in abstract algebra. *Journal of Physics: Conference Series*, *1321*(2), 022105. https://doi.org/10.1088/1742-6596/1321/2/022105

Osborne, J. (2014). Scientific practices and inquiry in the science classroom. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (Vol. 2, pp. 579–599). Routledge.

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research* (pp. 10–51). Netherlands Institute for Curriculum Development (SLO).

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

Prahmana, R. C. I. (2017). *Designing mathematics learning trajectory: An introduction*. Lambert Academic Publishing.

Prediger, S., Gravemeijer, K., & Confrey, J. (2015). Design research with a focus on learning processes: an overview on achievements and challenges. *Zdm*, *47*(6), 877–891. https://doi.org/10.1007/s11858-015-0722-3

Pyott, L. (2021). Tennis anyone? Teaching experimental design by designing and executing a tennis ball experiment. *Journal of Statistics and Data Science Education*, *29*(1), 22–26. https://doi.org/10.1080/10691898.2020.1854638

Rahayu, W., Tanujaya, B., & Iriyadi, D. (2022). *Desain eksperimen untuk bidang pendidikan [Experimental design for education]*. Jakad Media Publishing.

Reichardt, C. S. (2019). *Quasi-experimentation: A guide to design and analysis*. Guilford Publications.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475–522. https://doi.org/10.1207/s1532690xci1604_4

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

Siemon, D. (2021). Learning progressions/trajectories in mathematics: Supporting reform at scale. *Australian Journal of Education*, *65*(3), 227–247. https://doi.org/10.1177/00049441211045745

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*(2), 114–145. https://doi.org/10.5951/jresematheduc.26.2.0114

Smucker, B. J., Stevens, N. T., Asscher, J., & Goos, P. (2023). Profiles in the teaching of experimental design and analysis. *Journal of Statistics and Data Science Education*, *31*(3), 211–224. https://doi.org/10.1080/26939169.2023.2205907

Tanujaya, B. (2013). *Penelitian percobaan [Experimental research]*. Remaja Rosdakarya.

Tanujaya, B., & Mumu, J. (2020). Students' misconception of HOTS problems in teaching and learning of mathematics. *Journal of Physics: Conference Series*, *1657*(1), 012081. https://doi.org/10.1088/1742-6596/1657/1/012081

Tanujaya, B., Prahmana, R. C. I., & Mumu, J. (2018). Designing learning activities on conditional probability. *Journal of Physics: Conference Series*, *1088*(1), 012087. https://doi.org/10.1088/1742-6596/1088/1/012087

ten Hove, D., Jorgensen, T. D., & van der Ark, L. A. (2024). Updated guidelines on selecting an intraclass correlation coefficient for interrater reliability, with applications to incomplete observational designs. *Psychological Methods*, *29*(5), 967–979. https://doi.org/10.1037/met0000516

Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational design research*. Routledge. https://doi.org/10.4324/9780203088364

Vosniadou, S. (2013). *International handbook of research on conceptual change* (2nd ed.). Routledge.