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.p941-970

Cascading cognitive failures in geometric proof construction: A qualitative study of prospective mathematics teachers

Lathiful Anwar*, **Isma'ul Khusna, Cholis Sa'dijah, Santi Irawati**
Department of Mathematics, Universitas Negeri Malang, East Java, Indonesia
*Correspondence: lathiful.anwar.fmipa@um.ac.id

Received: May 5, 2026 | **Revised**: Jul 31, 2026 | **Accepted**: Aug 3, 2026 | **Published Online**: Aug 8, 2026

Abstract

Constructing a valid geometric proof requires prospective mathematics teachers to coordinate premises, deductive warrants, proof goals, and representations. Although empirical studies have documented multiple proof-related difficulties, less is known about how a breakdown at one point constrains subsequent reasoning. This qualitative descriptive-exploratory study examined four direct proof tasks and task-based semi-structured interviews with six prospective mathematics teachers from an initial sample of 20. Data were analyzed using an adapted Explorative Mathematical Argumentation (EMA) framework composed of initiation, tools, goals, and modalities. A response was classified as a failure when both written and interview data indicated that the participant possessed relevant knowledge but did not use or apply it appropriately. Failures were present in all components of the EMA framework, and were most prevalent in the application of theorems and in the use of deductive logic. Within-participant analysis indicated possible pathways, as the under-integration of a premise limited access to an appropriate warrant; the use of a theorem without support resulted in invalid intermediate assertions, and disrupted the continuity of the argument; and the use of a theorem along with a drawing resulted in a lack of checking the theorem against the conditions. These pathways were not rigid but indicated the interplay of different conceptual, strategic, representational, and communicative demands. The results of the study provide support for a stage-differentiated and functionally interrelated description of proof, and for the construction of pathways of failure as a process-oriented description of gaps across different classifications of failure.

Keywords:
Explorative mathematical argumentation, Failure analysis, Geometric proof, Proof construction, Students' difficulties

How to Cite:
Anwar, L., Khusna, I., Sa'dijah, C., & Irawati, S. (2026). Cascading cognitive failures in geometric proof construction: A qualitative study of prospective mathematics teachers. *Infinity Journal*, *15*(3), 941-970. https://doi.org/10.22460/infinity.v15i3.p941-970

1. INTRODUCTION

Mathematical proof is a key component of advanced mathematical activity because it establishes why a statement follows from accepted definitions, axioms, and previously established results. The construction of a proof requires a variety of reasoning, understanding, competence, and fluency (Kilpatrick et al., 2001). Therefore, the construction of a proof should be considered an activity involving the integration of many skills and should not be seen as the central aspect of reasoning. The ability to create a mathematically precise argument is important for future math teachers because it provides a means of understanding the content both from a conceptual and a pedagogical standpoint.

Despite the fundamental importance of constructing mathematical arguments or proofs, extensive research in teaching and learning has consistently demonstrated that students encounter significant challenges in this area. Numerous studies have examined difficulties in proof construction including theorem selection, logical and deductive argumentation, and the integration of relevant concepts (Hartono, 2025). Even in studies of inquiry learning, dynamic geometry software such as GeoGebra, and multiple representations of proof, students often struggle to construct valid mathematical proofs (Anwar & Goedhart, 2019; Anwar et al., 2024; Anwar et al., 2025; Putra et al., 2023). These findings indicate that the difficulties in proof construction cannot be solely related to the method of instruction and may be associated with the interplay of the conceptual, strategic, deductive, representational, and communicative demands.

Recent research has increasingly focused on proof-construction failure within the broader category of proof difficulty. In this study, a failure is coded when written and interview data suggest that a participant possesses relevant knowledge, but fails to activate, organize, or appropriately apply it (Hamdani et al., 2023). Justification is understood here as a deductive warrant and a legitimate transition between statements. If evidence does not show that a participant possesses relevant knowledge, then the case is interpreted as a possible limitation of knowledge rather than a failure of the proof process.

Different types of failures in proof construction have been noted in empirical research. For example, Selden and Selden (2015) show that even when students grasp executed constituent definitions, they do not produce original proofs. Likewise, students in Weber (2001) as well as Alcock and Weber (2005) do not construct appropriate deductive arguments because they fail to transform facts, and inappropriately employ definitions and theorems. Other studies show that failures come from a bad self-validation process (Kirsten & Greefrath, 2023), the reduced capacity to activate needed cognitive resources (Rabin & Quarfoot, 2022), and the absence of relevant cognition during proof making (Wulan et al., 2021). Within cognitive research, the failures are result of a disrupted coordination of the understanding, knowledge, and the reasoning (Harel & Sowder, 2007). These studies collectively identify disruptions in the activation of knowledge, instantiation of theorems, coordination of strategies, validation, and the exercise of deductive logic, which are the functional processes that are the focus of the present study.

In the context of geometry, the role of visual and spatial reasoning is task dependent. While diagrams can aid in the exploration and representation, a proof is not justified in the absence of explicit deductive warrants (Aphrodite & Rita, 2021; Sáenz-Ludlow & Athanasopoulou, 2016). Hanna (2000), Stylianides (2007), and Stylianides and Stylianides (2007) point to problems associated with considering empirical or visual evidence as adequate justification. For this reason, this research does not consider the absence of a diagram a justification for failing to reason; rather, it scrutinizes instances where diagram-based relations are either misinterpreted or accepted in the absence of a justification.

Prior studies have characterized several constraints as follows: lack of strategic knowledge, misuse of definitions and theorems, weak validation of proof steps, inability to connect premises and conclusions, and perceptual evidence reliance (Alcock & Weber, 2005; Miyazaki et al., 2017; Weber, 2001, 2004; Winer & Battista, 2022). These approaches describe relevant aspects of proof activity, but most focus on specific sources of difficulties or typify mistakes without following how an initial breakdown relates to subsequent steps in the same proof attempt. A process-level explanation is thus required to identify the difference between discrete mistakes and connected breakdowns.

The current study uses the Explorative Mathematical Argumentation (EMA) framework (Reuter, 2023) to address this gap. EMA has four functional components of argumentation: initiation, tools, goals, and modalities. These components provide a process map for locating breakdowns for each component of a proof (problem interpretation, mathematical resources, goal, and communication) in the order they appear in a proof. EMA is a qualitative organizational framework in this instance, and will not be used as a statistical model for structural dependence. The framework will be further developed based on research on proof-construction failure and proof structure (Anwar et al., 2023; Hamdani et al., 2023; Miyazaki et al., 2017) in order to provide measurable indicators for the written and interview data.

Preliminary evidence also motivated the study. A task administered on 17 April 2025 to 26 prospective mathematics teachers in the same broad educational context showed that 84.6% did not produce a valid proof of an elementary triangle-congruence statement. This result has not been used to claim a prevalence beyond that population, but did stimulate an investigation into what could be learned where breakdowns occurred in attempts to provide proof. EMA made this question analytically tractable by identifying the first observable breakdown and analyzing the reasoning that immediately followed. As with the studies mentioned above, these studies show that students appear to have an understanding of proof components but are unable to produce a full argument (Miyazaki et al., 2017; Soucy McCrone & Martin, 2004; Weber, 2002).

Accordingly, two research questions were developed for this study: (1) What types of failures for proof construction can be identified in the initiation, tools, goals, and modalities in EMA? and (2) What possible within-participant pathways might relate earlier and later failures observed in the construction of direct geometric proofs? This study provides an applicable coding system and describes the process of failure pathways. The term “pathway” refers to a process where the temporal sequence of written steps and the explanation of the participant supports the order; it does not refer to a causal relation.

2. METHOD

2.1. Research Design

A qualitative descriptive-exploratory research design was applied to capture types of geometric proof-construction failures and to understand how they interrelated in an individual proof attempt. This design enabled a framework-based analysis of failed proofs and participants' explanations, with no pretense of statistical framework for the failure categories.

This research was delimited to one Introductory Geometry class, four direct proof tasks concerning perpendicular and parallel lines, and the written and verbal data of the chosen participants. Each participant's written task/report and interview was treated as a single analytic case, and a subsequent cross-case analysis was conducted. Therefore, in the context of this research, the rationale was qualitative framework analysis.

2.2. Participants and Context

The research took place in a Mathematics Education program at a university in Malang, Indonesia, during the second semester of the 2024/2025 academic year. The first cohort of participants were 20 second-year prospective Mathematics teachers enrolled in an Introductory Geometry course.

Six participants were selected from the original group for follow-up interviews based on their ability to provide evidence-rich information. Selection criteria included: (1) responses that articulated proof-construction breakdowns pertinent to the research questions; (2) breakdowns that exhibited variations in configuration and sequences; and (3) readiness to offer explanation for their responses. The focus of sampling was on depth and variation of analysis. (Creswell, 2014).

2.3. Instruments

Two instruments were employed in this study. First, the geometric-proof task set included four direct-proof tasks related to perpendicular and parallel lines. Participants were asked to identify the givens and the conclusion, select appropriate definitions and theorems, and construct and present a complete proof. A senior expert in mathematics education evaluated the tasks for relevance to the content of the course, clarity for the givens and the conclusion, and appropriateness to elicit the proof construction process. Necessary revisions were made prior to the administration of the tasks.

Second, a semi-structured and task-based interview guide was created using the EMA components and the written responses of the participants. To this end, participants were asked to identify what they regarded as given, define the role of a definition or theorem, in the proof sequencing, describe the goal, and clarify their choice of notation or representation. The interviews focused on the written evidence, and were not intended to provide participants with guidance to complete a correct proof.

2.4. Data Collection Procedures

The methodology involves a sequential, multi-stage process designed to organize data in a manner that captures the structure of students' written proofs and their associated reasoning processes. In the first stage, participants were tasked with completing a number of tasks related to the proof of a geometric theorem. Their written responses were systematically analyzed to document initial patterns of failure, including breakdowns in reasoning, conceptual understanding, and proof structure.

To ensure a comprehensive analysis of different kinds of failure, six respondents were chosen using a purposive sampling technique. Semi-structured, task-oriented interviews were performed, during which participants verbally demonstrated their proofs, substantiated their logic, and detailed their method of solving the task. The interviews targeted the problematic steps or inconsistencies discovered in the written proofs. This method of study provided an in-depth analysis of the failure of proof construction in cognitive process of students and gave a broader perspective of their reasoning.

2.5. Data Analysis

Analyzed followed an iterative process of condensation, representation, and verification of conclusions based on the evidence (Miles et al., 2014). The first step of the analysis involved segmenting each written proof into propositions and transitions. These include premises, uses of definitions or theorems, intermediate claims, links of inference, and conclusions. Each segment in the transcripts was analyzed when a participant described a condition, chose a warrant, explained a transition, critiqued a step, or explained a choice in representation.

These segments were then analyzed using the four components of EMA and the definitions provided in Table 1. The codes were adjusted based on interpretation and familiarity of the data when a category was too broad, ambiguous, overlapped with another category, and was not supported by the total evidence. An omitted name of a theorem, non-use of a representation, and an unanswered question were not automatically interpreted as a failure.

A code was preserved when the written evidence showed a clear breakdown, and the interview either verified the breakdown or indicated that the relevant knowledge was not used. When this evidence was absent, the episode was identified as a possible knowledge limitation or an unclear case. The codes were summarized and refined under the four components of EMA, and the resulting codes (n=13) were summarized and refined under the four components of EMA (see Table 1).

Table 1. Initial coding scheme for failure analysis in geometric proof construction

Code	Initial Code	EMA Stage	Operational Definition	Indicator in Written Responses	Indicator in Interview
PIF	Premise Identification Failure	Initiation	Failure to identify, include, or correctly interpret a condition required for the proof despite evidence that the condition was available.	A required given is omitted, altered, or treated as irrelevant, and the omission affects later reasoning.	The participant misidentifies or excludes the condition. Concise omission is excluded when the premise is demonstrably used.
PrIF	Proof Initiation Failure	Initiation	Failure to coordinate available premises into an admissible first deductive step.	The proof begins from an unsupported claim, assumes the target, or produces no viable first inference.	Relevant knowledge is available but cannot be organised; a pure knowledge limitation is excluded.

Code	Initial Code	EMA Stage	Operational Definition	Indicator in Written Responses	Indicator in Interview
VF	Visualization Failure	Tools	A diagram is misread, treated as sufficient proof, or used to infer a relation without a deductive warrant.	A visually apparent relation is asserted without theorem-based support or is interpreted inaccurately.	The participant appeals to appearance as evidence. Merely not drawing or annotating a diagram is excluded.
UPAF	Universal Proposition Application Failure	Tools	A definition, axiom, or theorem is applied without satisfying its conditions, or the relevant warrant cannot be supplied despite evidence that it is known.	An irrelevant theorem is cited, its conditions are unmet, or a transition is unsupported by an available proposition.	The participant cannot explain applicability. Omission of a theorem name is excluded when the warrant is otherwise valid.
SSF	Strategy Selection Failure	Tools	Failure to select an appropriate proof strategy (e.g., forward, backward, or combined reasoning).	Random or unstructured steps; multiple attempts without direction; no clear plan.	Cannot explain reasoning strategy; justification for approach is unclear or irrelevant.
HSF	Hypothetical Syllogism Failure	Tools	Failure to construct a valid chain of logical implications linking statements.	Disconnected statements; missing logical transitions; absence of "if–then" relations.	Unable to explain how one step leads to another.
DF	Deduction Failure	Tools	A conclusion does not follow from the cited premises, or an essential transition is omitted and cannot be reconstructed.	The response contains an unsupported logical jump, invalid inference, or conclusion based on insufficient conditions.	The participant cannot supply the inferential relation; deliberate and reconstructable abbreviation is excluded.
SF	Symbolic Failure	Tools	Notation changes mathematical meaning, obscures object correspondence, or prevents verification of the proof.	Incorrect or inconsistent notation refers to the wrong object or relation.	The explanation reveals unstable object coordination; minor corrected transcription slips are excluded.
CIF	Conclusion Identification Failure	Goals	Failure to identify the proposition that must be proved.	The response pursues a target different from the stated conclusion.	The participant cannot state the proof goal; this code was not observed in the verified dataset.

Code	Initial Code	EMA Stage	Operational Definition	Indicator in Written Responses	Indicator in Interview
CUF	Condition Understanding Failure	Goals	Failure to distinguish or check the necessary and sufficient conditions for a target claim.	A conclusion is asserted from an incomplete set of conditions.	The participant cites a criterion but cannot account for all required conditions.
CFF	Conclusion Formulation Failure	Goals	Failure to formulate a final statement that corresponds precisely to the required conclusion.	The proof ends with a mathematically different or imprecise target.	The participant recognises the goal but states a different final claim.
PFF	Proof Format Failure	Modalities	The chosen representation prevents the proof structure from being communicated or verified.	The format obscures statements, warrants, or inferential direction.	Variation among valid formats is excluded; this code was not observed in the verified dataset.
VAF	Verbal Argumentation Failure	Modalities	Failure to articulate, defend, or reconstruct the warrants underlying the written proof.	Written claims lack an accessible explanation of their logical connection.	The participant repeats steps but cannot explain why they follow or why a theorem applies.

For each task, a within-participant sequence map recorded the earliest supported breakdown, subsequent proof steps, the participant's explanation, and whether later breakdowns depended on the earlier step. The six maps were then compared to identify recurring configurations and disconfirming cases. Trustworthiness was supported through triangulation of written and interview data, explicit inclusion and exclusion criteria, consideration of alternative explanations, and an audit trail linking interpretations to data segments. The counts reported below describe coded episodes in this bounded dataset and are not inferential estimates.

3. RESULTS AND DISCUSSION

3.1. Results

3.1.1. Overview of Failure Distribution Across EMA Stages

The analysis focused on the recorded interviews and written assignments of the six participants (S1-S6). In total, thirteen types of failures across the EMA components of initiation, tools, goals, and modalities were generated through the coding procedure (see Table 2). Within the bounds of this data set, the table describes the failures and does not show the frequency of failures. The greatest representations were in Universal Proposition Application Failure (UPAF = 13) and Deduction Failure (DF = 11), with Hypothetical Syllogism Failure and Strategy Selection Failure (each = 7) following.

Table 2. Distribution of failure types across participants

Code	Failure Type	S1	S2	S3	S4	S5	S6	Total
PIF	Premise Identification Failure	1	1	1	0	1	0	4
PrIF	Proof Initiation Failure	1	0	1	1	0	1	4
VF	Visualization Failure	0	1	1	1	0	1	4
UPAF	Universal Proposition Application Failure	2	2	3	2	2	2	13
SSF	Strategy Selection Failure	1	1	2	1	1	1	7
HSF	Hypothetical Syllogism Failure	1	1	2	1	1	1	7
DF	Deduction Failure	2	2	2	2	1	2	11
SF	Symbolic Failure	0	1	1	1	0	1	4
CIF	Conclusion Identification Failure	0	0	0	0	0	0	0
CUF	Condition Understanding Failure	1	1	1	1	1	1	6
CFF	Conclusion Formulation Failure	1	0	0	0	0	0	1
PFF	Proof Format Failure	0	0	0	0	0	0	0
VAF	Verbal Argumentation Failure	1	1	2	1	1	1	7

UPAF and DF likely generated the largest counts due to the occurrence of theorem application and deductive reasoning occurring multiple times throughout most proof attempts. In contrast, identification of premises and conclusions is typically done at the beginning and at the end of the proof. Therefore, the counts of the former indicate multiple points for observable breakdown, and do not indicate that these are the most difficult skills. Because of the counts and resulting distribution, it can be inferred that theorem application and deduction should be taught in relation to the interpretation of premises, checking the conditions of the theorem, and monitoring the proof goal.

3.1.2. Failures at the Initiation Stage

Two initiation-types failures were retained: Premise Identification Failure (PIF) and the Proof Initiation Failure (PrIF). PIF is related to combining the integration and interpretation of a required given. PrIF refers to the arrangement of the available premises to form the first admissible step of a deduction. A missing sentence or a failure to start was not attributed to a coding error in the absence of supporting evidence from the interview.

S1 identified $\overline{AC} \cong \overline{BC}$ as the only one premise dan confirmed in the interview that there was no other relevant premise (see Figure 1). The statement that the figure was a circle with centre O was therefore not integrated into the reasoning, although it licensed the radius relation needed later. This episode was coded as PIF because the interview showed that the centre condition was not merely omitted for brevity. The evidence supports a plausible, but not exclusive, pathway: failure to integrate the centre condition constrained the most direct warrant for $AO \cong OB$, after which S1 used an unsupported route to triangle congruence. Limited theorem retrieval and weak correspondence checking remained possible co-contributors.

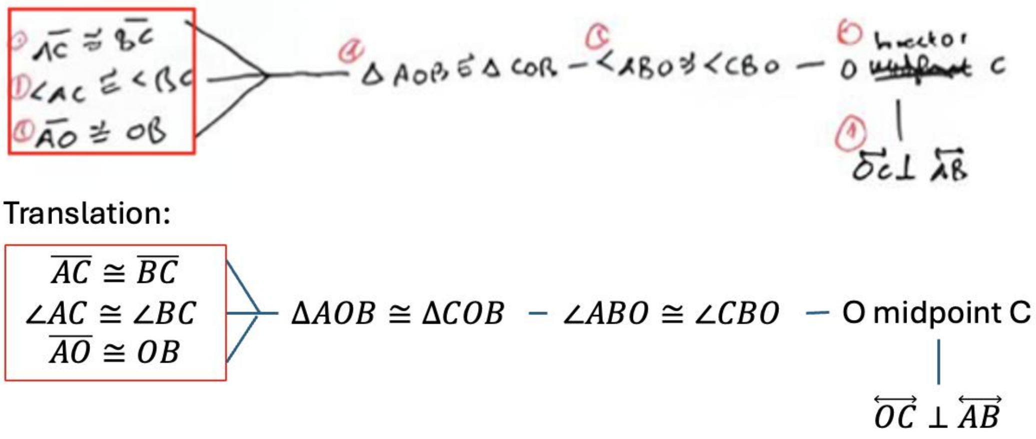

Figure 1. Premise identification failure (PIF) by S1

PrIF was applied only in cases where relevant premises or knowledge existed but were not sequenced into a valid first move. In S2's answer, the proof originated from constructed relations through O and A without having first established the properties that are allowed by the given circle (see Figure 2). Hence, the focus of the interpretation is on the coordination of the information that is available, and not on the absence of an explicitly stated premise.

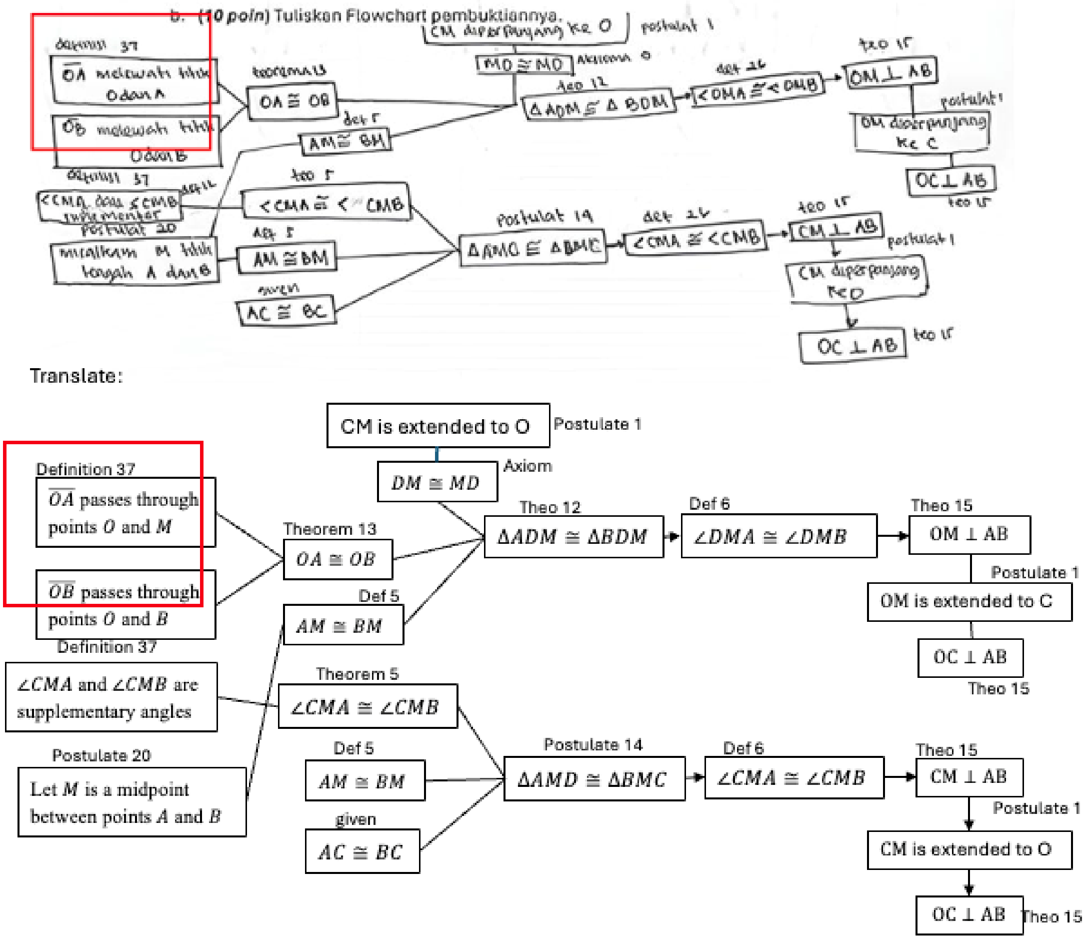

Figure 2. Proof initiation failure (PrIF) by S2

During the interview, S4 explained that “Because I had already constructed the diagram first, by extending from point O as a radius. So I no longer used the formal premise ‘a circle with center O’, but instead directly used the constructed elements.” S1 similarly stated, “I didn’t write it because I thought it was already clear from the diagram.” These statements suggest that some forms of information perceived by the senses were considered as not needing to be explicitly incorporated into the deductive argument. However, this analysis does not claim that every omission of a premise leads to a subsequent breakdown; the effect is contingent upon whether the omitted element is necessary for the justification of the subsequent statement.

Therefore, PIF and PrIF denote somewhat related yet different initiation problems: the inability to incorporate a necessary condition, and the inability to organize the available conditions into a tenable first step. Both of these problems may limit subsequent reasoning, but the effects of each are conditional rather than certain.

3.1.3. Failures at the Tools Stage

For participants S1 through S6, the tools stage had the most amount of failures and the most variation of failures, making it the most significant stage of the construction of geometric proofs that highlights that most failures occur. This stage emphasizes the understanding of both the conceptual and procedural of the construction, which includes definitions, postulates, theorems, and the techniques of proof. In this stage, a total of six failure types were identified and designated as: Visualization Failure (VF), Universal Proposition Application Failure (UPAF), Strategy Selection Failure (SSF), Hypothetical Syllogism Failure (HSF), Deduction Failure (DF), and Symbolic Failure (SF). Among these failures, Visualization Failure (VF) and Universal Proposition Application Failure (UPAF) are the most significant and are the most prevalent failures.

3.1.3.1. Visualization Failure (VF)

Visualization Failure (VF) would not be inferred simply because a participant did not include a diagram or annotation. VF was attributed only when a participant misinterpreted the given figure, considered the visual appearance as enough proof, or conjectured a spatial relationship with no logical basis. S1 indicated that he was in the habit of first drawing a diagram to help him identify the relationships and then erasing the diagram (see Figure 3). Since there were no cases in this answer for unwarranted visual inferences, absence of markings was not considered to be a case of VF.

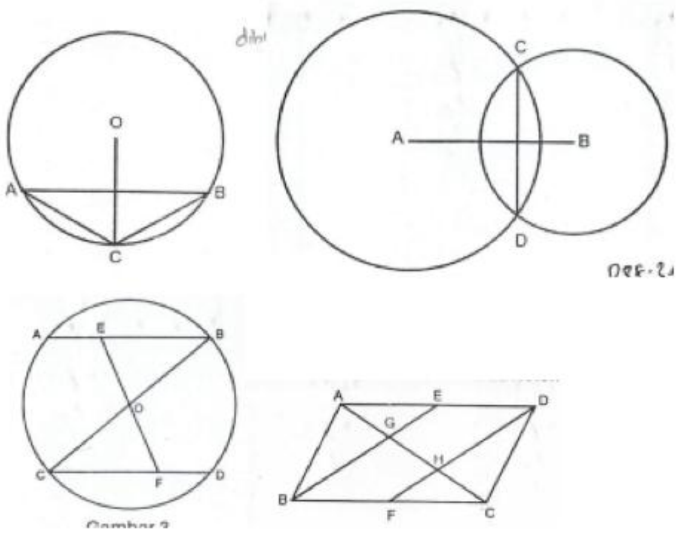

Figure 3. Visualization failure (VF) by S1

This is illustrated more clearly in S2's response (see Figure 4). S2 asserted that the midpoint relation was "already clear from the given" and so did not draw the relation as part of the proof. The relevant failure here was not the choice to not redraw the figure, but diagrammatic conviction. The relation was perceived and accepted as visually salient, and thus was used in the proof, lacking the appropriate justification.

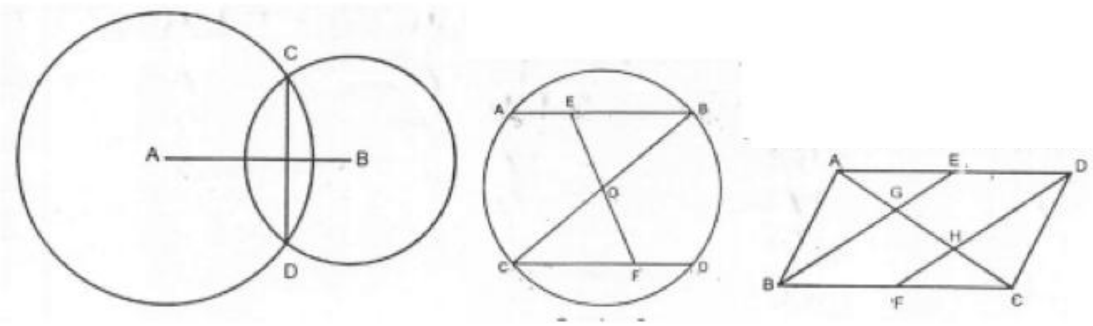

Figure 4. Visualization failure (VF) by S2

S6 provided a contrasting account: "Yes, but not always. If the diagram is simple, I only add a helper line when needed. If the existing elements are sufficient, I don't add anything." Selective use of a diagram was not considered a failure. This statement supported the exclusion rule that a participant can provide a valid deductive proof without the need for a new drawing or annotation.

In the verified VF episodes, the mechanism was not simply a case of neglected visualization. Rather, it was an epistemic shift from considering a diagram as an exploratory tool to regarding it as evidence. In this case, perceptual certainty diminished the need to seek a warrant and to check the conditions of the theorem, making it more likely that an unsupported transition would occur.

3.1.3.2. Universal Proposition Application Failure (UPAF)

Universal Proposition Application Failure was coded in two cases. The first occurs when a definition, postulate, or theorem was instantiated without satisfying its conditions. The second occurs when a participant provided no deductive warrant for a transition, even if evidence suggested that the relevant proposition was known to the participant. Simply leaving out the name of a theorem was insufficient for this code.

For instance, S1 wrote AO≅OB without giving any reason (see Figure 5). This reflects a collapsed step between the assertion and the supporting theorem. During the interview, S1 clarified, "Because I didn't know the exact theorem. There are many theorems to remember, and the exam time was limited." This captures S1's case as the inability to retrieve the relevant theorem (which was that the radii of a circle are congruent), with a therefore a gap left in the logic where the deductive step should have been, which compromised the overall rigor of the proof.

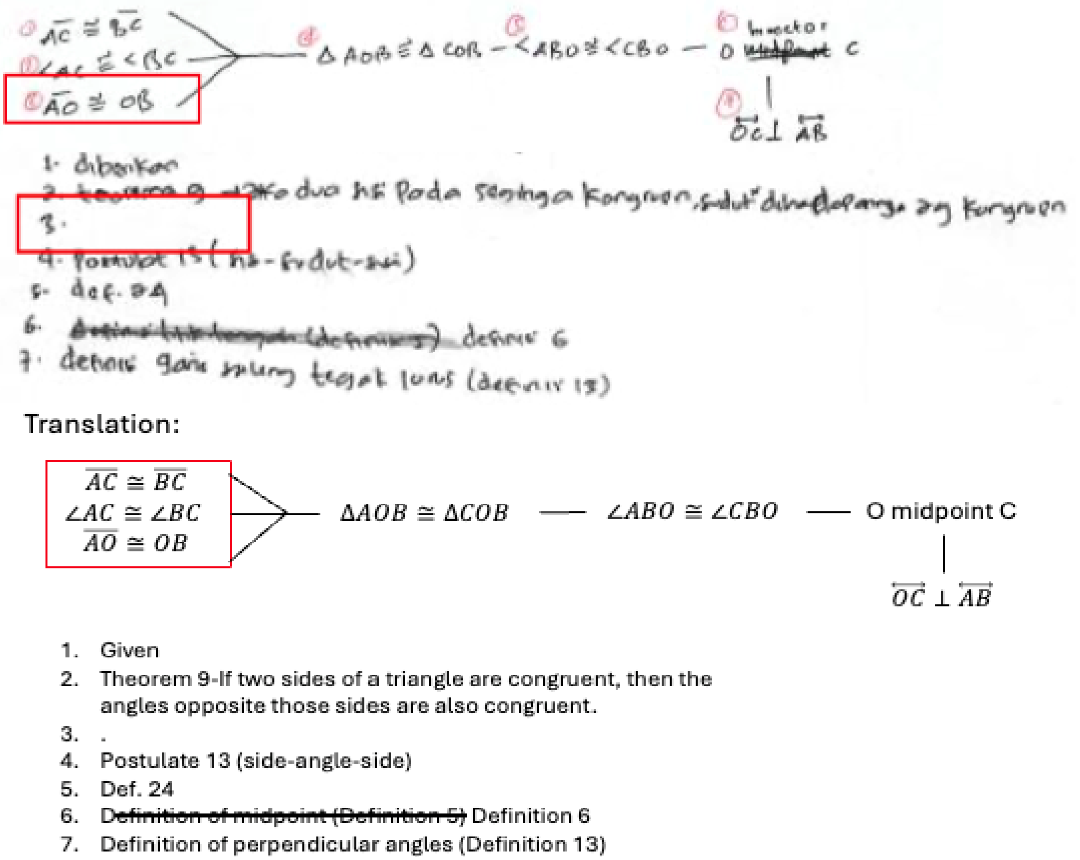

Figure 5. Universal proposition application failure (UPAF) by S1

Similar example was observed in S2, instead of using the equal-radii theorem to justify segment congruence, opted for a transitive argument (see Figure 6). Surface-level correspondence was evident in S2's equality-like relations without S2 verifying if the relations had the common element necessary for transitivity. Therefore, the case provided an instance of invalid theorem instantiation as opposed to a case whereby justification was omitted.

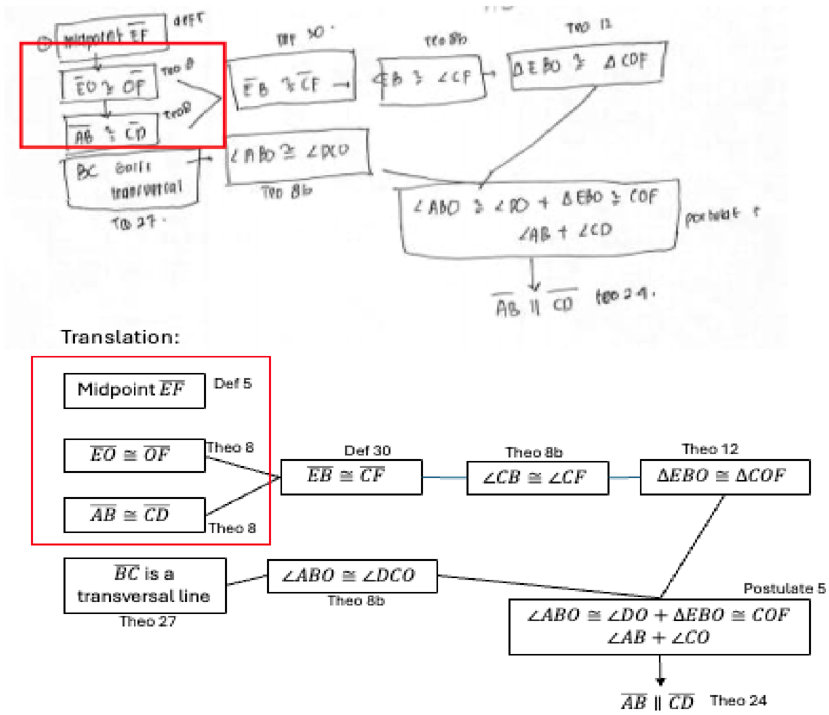

Figure 6. Universal proposition application failure (UPAF) by S2

These results illustrate indicate at least two mechanisms of UPAF: failure to retrieve a known warrant at the required moment (S1) and conceptual misapplication through surface-level matching (S2). Both can interrupt the deductive chain, but they require different instructional responses.

3.1.3.3. Hypothetical Syllogism Failure (HSF)

Hypothetical Syllogism Failure (HSF) occurs when students are unable to articulate a single coherent chain of reasoning to articulate the premises and the conclusion. HSF can be characterized by the absence of reasoning chains, loss of motivation to construct a proof, and inability to articulate the proof beyond the intermediate steps, including stopping the proof short of a conclusion.

For instance, S2 demonstrated difficulty at the *initial stage of proof construction*, particularly in identifying a starting point for the reasoning process. The absence of a written response, see Figure 7, to the proof indicates that the student did not demonstrate proof construction from a deductive starting point. This struggle is captured in S2's statement, "I still have difficulty finding the initial route of the proof for circle problems. Even though I can solve triangle problems, for circles I am still confused about how to start and continue." S2 seems to be claiming that a lack of circle proof problems has resulted in him/her not having an entry point that is clear enough to trigger the reasoning. The triangle proof problems that s/he has in comparison to circle proof problems. With the stated challenges, S2 could not make an initial deductive step, and this stymied further reasoning and the proof construction completely.

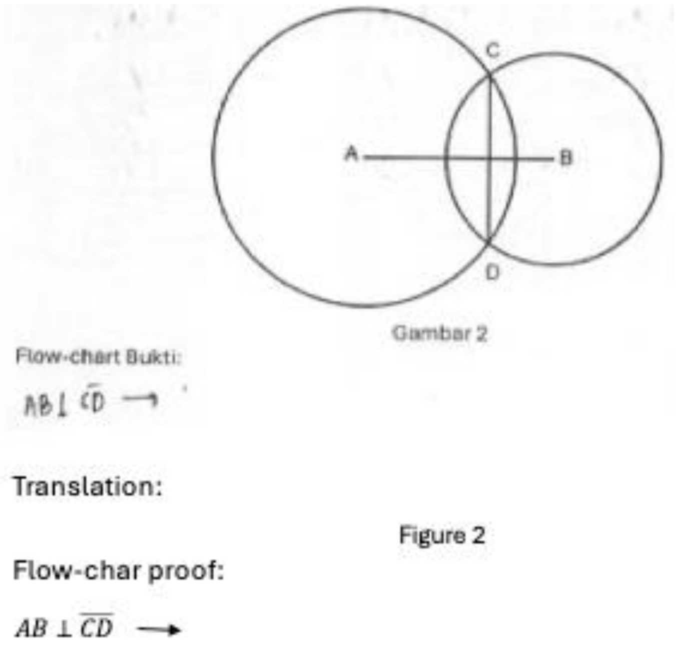

Figure 7. Hypothetical syllogism failure (HSF) by S2

The Hypothetical Syllogism Failure was also observed in S5, who was unable to relate the intermediate conclusion to the following conclusion. A snapshot of the written answer is provided in Figure 8 as noted in the documentation of the proof. In the figure, we can see that the proof ends prematurely, thus showing an incomplete deductive chain. Whereas one can see that a finding was attempted, the response was incomplete. S5 explained "This is where I got confused. I was not sure whether I could directly conclude that CD intersects AB after triangle congruence, or if there was a missing step. That is why I got stuck." This explanation indicates that the failure in connecting logically premises to intermediate conclusion to final conclusion was disrupted by the uncertainty in validating the process of drawing conclusion. Consequently, the proof process was interrupted at the stage of linking intermediate conclusions, leading to an incomplete argument.

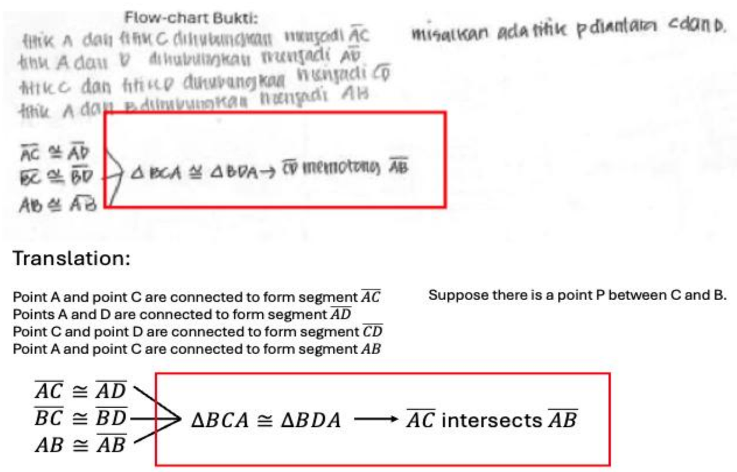

Figure 8. Hypothetical syllogism failure (HSF) by S5

This evidence confirm that the persistent instances of partial or incomplete construction, failing to show the complete reasoning from the premises to the conclusion. The recognized instances show (1) the absence of proof (S2), (2) absence of proof continuation that is a few initial steps (S3), and (3) absence of proof that has middle statement links (S5). It appears these instances show HSF in the process of proof construction in many different forms, and all these instances imply that there is no complete and coherent deductive chain.

3.1.3.4. Deduction Failure (DF)

Deduction Failure (DF) a conclusion either does not logically follow from the given premises or does not logically follow from previously established results, or when an essential transition is omitted and cannot be inferred from the interview. DF is not applied when the transition is concise and the missing step is valid and can be recovered verbally.

The DF was observed in S1's written answer (see Figure 9), who concluded that triangle AOB is congruent to triangle COB, yet the preceding statements did not establish sufficient conditions to justify this claim. The reasoning included references to selected sides and angles; however, these elements were not organized into a valid congruence configuration. This issue is reflected in S1's statement, "Because that was what came to my mind at that time," indicating that the conclusion was not derived from a systematic evaluation of the necessary conditions. Additionally, S1 used angle equality, such as ∠ABO≅∠CBO, without integrating relevant definitions (e.g., perpendicularity), suggesting that individual statements were produced without a coherent deductive foundation. As a result, the reasoning process contains discontinuities that weaken the overall logical structure.

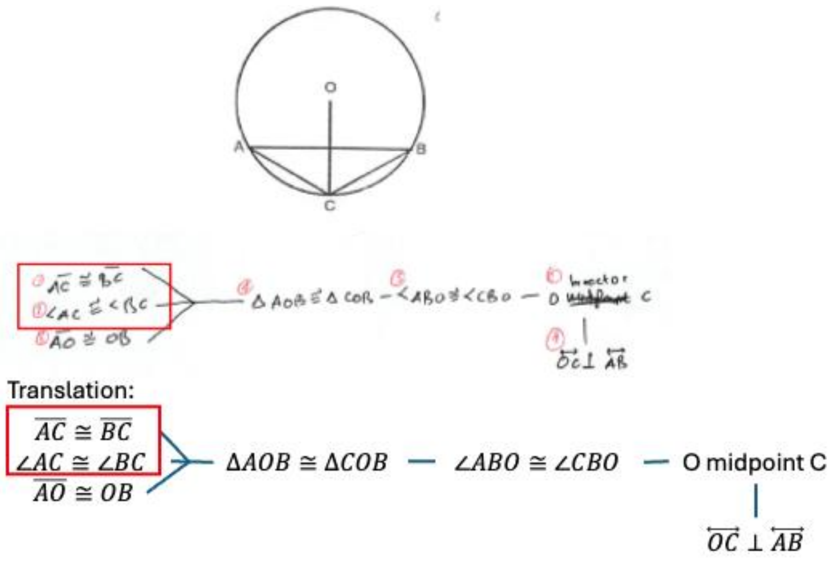

Figure 9. Deduction failure (DF) by S1

Another example of DF was identified in S3 (see Figure 10). S3 removed intermediate statements due to his perspective on redundancy. As discussed in the interview, S3 was unable to reconstruct the essential relation in which the status of AD as a radius (i.e. the necessary relation which would enable S3 to complete the transition from the premises to the conclusion). Therefore, the code reflected deductive discontinuity, not brevity.

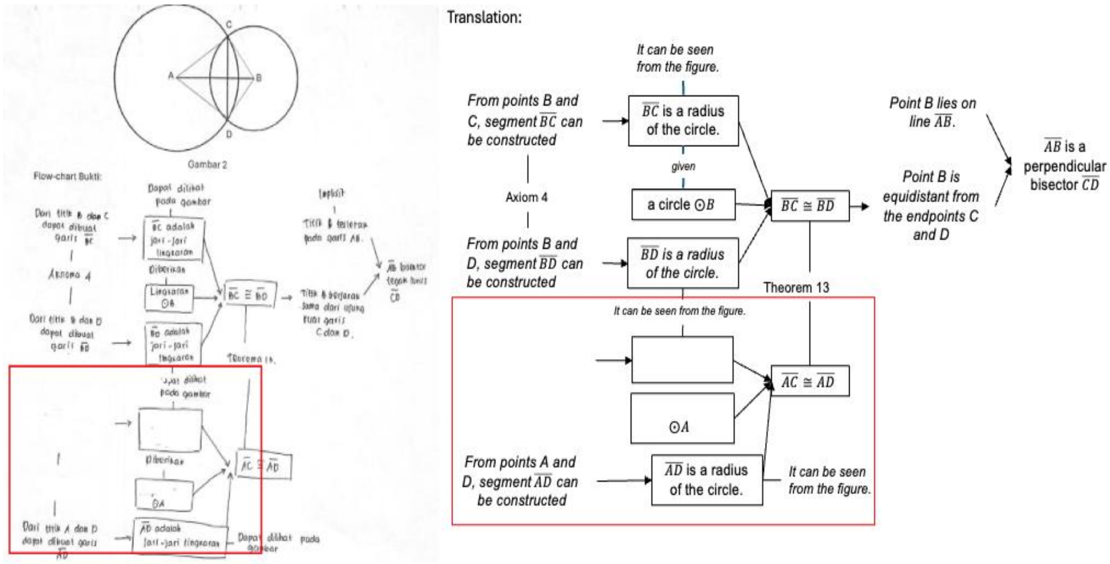

Figure 10. Deduction failure (DF) by S3

Across the verified episodes, DF showed up as unsupported inference, or as an absence of a non-reconstructable transition. Participants were able to generate statements of individual relevance, but the evidence was still invalid due to absence or incorrectness of the links allowing transition from one statement to another.

3.1.3.5.Symbolic Failure (SF)

Symbolic Failure (SF) was limited to errors in notation that caused a loss of meaning in mathematics, distorted the correspondence between geometric entities, made a theorem appear applicable to the wrong elements, or inhibited the verification of a proof. Minor transcription errors that were immediately and correctly noted in the interview were reported separately and were not considered to be cognitive failure.

For instances, in response to question 1, S2 made reference to an incomplete rubric and explained that “DO” stood for angle DFO, “EB” for angle EOB, and “CF” for angle COF (see Figure 11). In addition to making their response stylistically imprecise, these labels omitted the vertex and the pairs of figures in the congruence argument and made it difficult to determine which theorem was being used.

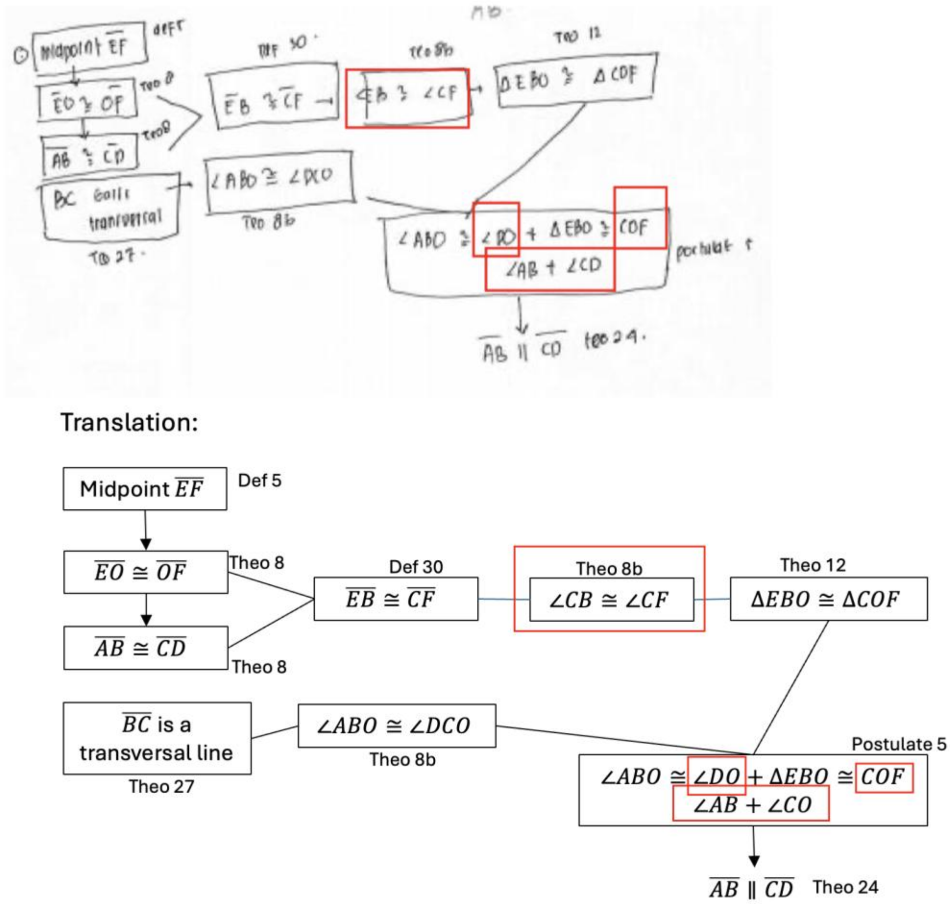

Figure 11. Symbolic failure (SF) by S2

S6 also omitted a line symbol from AD and called it an oversight (see Figure 12). Since the intended object was identified quickly and correctly, this isolated error was insufficient for SF. The code was only used when the notational ambiguity was relevant for the identification of the object or the verification of the inference.

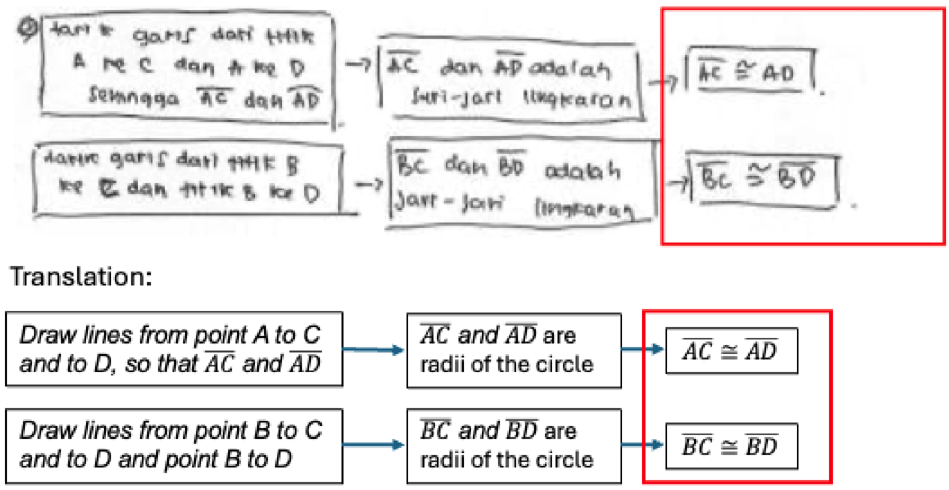

Figure 12. Symbolic failure (SF) by S6

The example of SF observed in S2, S4 and S6 show across all cases the same incomplete or unclear symbolic representations such as absent line symbols, wrong or absent angle notation, and incomplete or varying use of congruence symbols. Simplified notation was also noted in written responses, while interview data showed students could usually elaborate what each of their symbols meant. SF, in this study, is typified by the contradictions between what is written and the symbolic representation of what is geometrically meant, as captured in the students' proofs and their explanations.

Failure at the tools stage across all participants suggests that students do not only fail with isolated concepts, but also fail to integrate multiple aspects of reasoning. The combination of problems with visual reasoning, wrong application of theorems and lacking deductive structures, leads to the conclusion that students are unable to integrate conceptual understanding, procedural reasoning, and logical reasoning into a proof. The tools stage has been identified as a primary stage of concern as multiple cognitive processes are required to function simultaneously. The cycle of interrelated failures in the tools stage lays the foundation of failure to be the primary source of the problem.

3.1.4. Failures at the Goals Stage

Failures at the goal-setting stage pertain to students' ability to comprehend and formulate the objective of the proof (*to prove*). Based on the analysis of written responses and interview data, three types of failures were examined: Conclusion Identification Failure (CIF), Condition Understanding Failure (CUF), and Conclusion Formulation Failure (CFF).

3.1.4.1. Conclusion Identification Failure (CIF)

Conclusion Identification Failure (CIF) refers to an inability to determine the statement that needs to be proven. All participants identified the target statement in the focal tasks; accordingly, CIF was not observed in the verified dataset. This finding is limited to goal recognition and does not imply effective proof communication. Conclusion formulation, necessary-sufficient direction, proof format, and verbal warranting were analysed separately.

3.1.4.2. Condition Understanding Failure (CUF)

Condition Understanding Failure (CUF) denotes a student's failure to comprehend the logical connections between the requisite components/criteria and the sufficient components of geometric problems. S1 is an example of a student who has drawn an invalid conclusion of triangle congruence from inappropriate and insufficient conditions/criteria (illustrated in Figure 13). In this example, the student drew the conclusion of triangle AOB being congruent to triangle COB with no description of the requirements for congruence. The conditions cited do not fulfill the requirements for congruence (i.e., SSS, SAS, ASA).

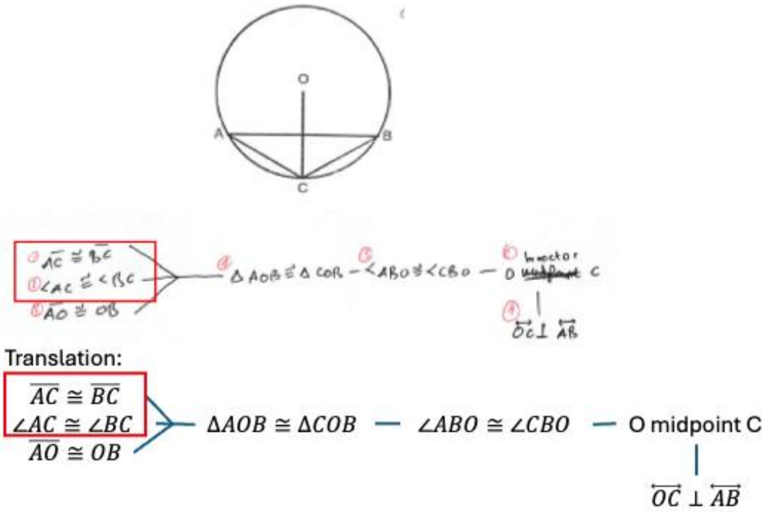

Figure 13. Condition understanding failure (CUF) by S1

This reasoning is reflected in S1's statement, "Because that was what came to my mind at that time," indicating that the conclusion was not grounded in a systematic verification of congruence criteria. Although S1 identified some relevant elements, such as pairs of sides or angles, these were not sufficient to satisfy standard congruence conditions (e.g., SSS, SAS, or ASA). As a result, the conclusion was drawn without adequate deductive support, demonstrating a failure to evaluate the sufficiency of the premises.

A similar was also identified in S2 (see Figure 14), who concluded that two triangles were congruent based solely on one pair of equal sides (AC≅BC). S2 explained that "… I think the idea was to use the Side-Side-Side (SSS) theorem. Since there are said to be congruent sides, I inferred that the triangles are congruent based on Theorem 12." According to this, it can be inferred that S2 used the theorem (SSS) but did not check whether all required conditions were satisfied. The conclusion was reached using only a few steps, so the conclusion was incomplete.

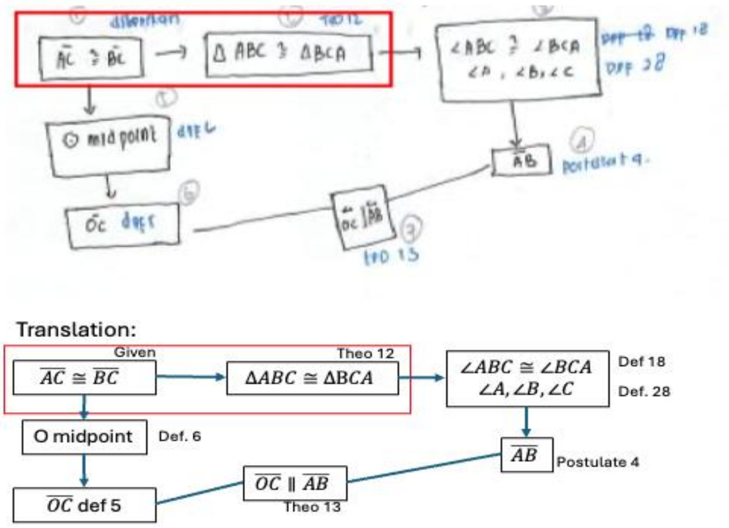

Figure 14. Condition understanding failure (CUF) by S2

Across these cases, students were able to identify relevant elements but failed to evaluate whether those elements met the logical requirements necessary to justify the conclusion. This suggests that the failure lies not only in the application of theorems but also in the validation of premise sufficiency prior to making deductive claims.

3.1.4.3. Conclusion Formulation Failure (CFF)

Conclusion Formulation Failure (CFF) refers to students' inability to accurately articulate the final conclusion of a proof in accordance with the problem requirement. This type of failure was observed to be infrequent and occurred only in Subject 1 (S1). In the written response (see Figure 15), the expected conclusion was a relationship between lines (AB≅CD).

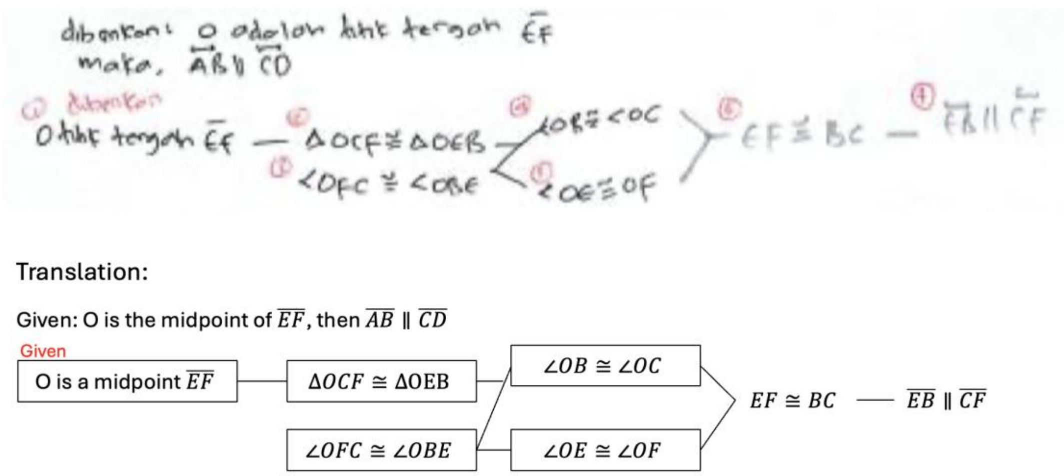

Figure 15. Conclusion formulation failure (CFF) by S1

However, S1 instead wrote the final conclusion as EB ≅ CF, which does not correspond to the required statement. This shows us that while the student attempted multiple sections of the proof, the final answer was inconsistent with the proof's goal. The mismatch is clear since the student's conclusion did not meet the requirement of the problem.

In most instances, the lack of a final concluding was not a case of Conclusion Formulation Failure (CFF), but was instead a case of students breaching the construction of the complete proof. Students generally stopped at intermediate steps. This suggests that conclusions are missing because of the breakdown of reasoning steps, even if the students realized the target. Most students, even at the goals stage, could identify the objective of the proof, but could not formulate the requisite steps for achieving the objective. Errors, as was observed through interview and written data, were more related to how steps were formulated and not because of a misunderstanding of the goal.

3.1.5. Failures at the Modalities Stage

Failures at the modalities stage relate to students' ability to communicate their proofs both in written form and through verbal explanation. Two types of failure were examined: Proof Format Failure (PFF) and Verbal Argumentation Failure (VAF).

3.1.5.1. Proof Format Failure (PFF)

Proof Format Failure, or PFF, involves students not identifying or employing an appropriate proof format, regardless of whether it is a text proof, two-column proof, or flowchart proof. In looking at students' written responses, it became clear that PFF was not present in this case. While proof formats did differ amongst students, this is not indicative of proof format failure. Of the six subjects, five utilized the flowchart proof format, while one employed the two-column proof format. They used within-each-student proof formats that were sufficiently representational of the proof's structure.

That students could demonstrate the procedural knowledge requisite for the proof formats that are accepted in the proof structure indicates that the abundant use of flowcharts likely made that format an embedded structure, or organizing system, within the students' classroom environment, which meant they could easily control proof structure at the level of format selection.

3.1.5.2. Verbal Argumentation Failure (VAF)

Students exhibit Verbal Argumentation Failure (VAF) when they are unable to articulate, defend, or reconstruct the reasoning underlying specific written proofs. While specific types of failure can be diagnosed solely from written outcomes, VAF must be identified through interview data, as verbal articulation is the key criterion for assessing VAF. Instances of VAF were reported in the current study for all interviewed students.

The VAF was observed in S1, whose written response presents a sequence of reasoning leading to a conclusion of triangle congruence, yet lacks corresponding verbal justification. In the written work, S1 constructed a chain of statements implying that triangle AOB is congruent to triangle COB; however, this reasoning was not supported during the interview. This is reflected in the statement, "From $AC \cong BC$, $\angle AC \cong \angle BC$, and $AO \cong OB$, triangle AOB is congruent to triangle COB," followed by the inability to specify the theorem used and the explanation, "Because that was what came to my mind at that time." This indicates that, although the student produced a structured written argument, the underlying rationale and justification for each step were not internally consolidated or verbally accessible.

The Verbal Argumentation Failure was also observed in S2, who was able to describe a number of steps in a procedure but was unable to give a cohesive argument with logical progression and connections. S2 described the steps with a fair level of detail and contributed with statements, "Because O is the midpoint of EF, $EO \cong OF$... then triangle EBO is congruent to triangle COF using Theorem 12... using Theorem 24, I obtained AB parallel to CD." Although the description was consistent, the logical connectors were not clear and neither were the conclusions. This is evidenced by S2's statement, "DO actually refers to angle DFO," "EB refers to angle EOB," and "CF refers to angle COF." This evidence indicates that even although the student was able to recall the steps of the solution, the inability to make logical connections among the steps and statements was clear.

There was a tendency among participants to provide extensive verbal descriptions that reflected an almost verbatim recitation of the written steps, as opposed to paraphrasing or summarizing the rationale underpinning the steps. It was common for participants to provide a sequence of steps, but many did so without linking or integrating them in a cohesive sequence. Consequently, while the written proofs incorporated accurate formal and structural elements, as well as classical and contextual elements, participants tended not to provide sufficient reasoning and justification for any of the proof elements. Participants did not face a significant problem in sequencing the steps of a written proof, but they did face a substantial problem in justifying and providing reasoning for the proof steps and elements verbally. This leads to the conclusion that the problem was not in sequencing steps in a written proof, but in verbalizing the steps and elements of the proof, respectively.

3.1.6. Failure Propagation Across EMA Stages

Three pathways across the EMA components were identified by within-participant sequence maps. The pathways illustrate temporal and functional relationships which are supported by the participants' written work and interview descriptions. The pathways do not imply deterministic causation.

The first pathway was derived from the incorporation of an insufficiently related premise into a proof. If a given premise was illustrated in the diagram and was considered as background information, the participant was unable to activate the proposition associated with the premise, and therefore, selected a different theorem, which did not completely satisfy the required conditions, and arrived at an invalid conclusion.

For the second pathway, the participants relied too heavily on the diagram. If the participants considered a certain relation to be self-evident due to the diagram, they would not check whether or not the conditions of the relevant theorem were satisfied. Consequently, the relation evident in the diagram was incorporated into the proof and considered self-evident, leading to an invalid proof.

For the third pathway, an intermediate proposition was introduced, which did not have a sufficient establishing warrant. While the subsequent propositions might have been locally plausible, the missing connection disrupted the overall continuity of the proof. The same discontinuity was evident in the participant's explanation, as the participant was unable to communicate the logical connections and the why behind the statements.

Propagation occurred only when a previous breakdown was functionally required by a subsequent step. A breakdown could be stopped when a participant provided a different valid theorem, fixed a correspondence, or filled in a missing transition. The analysis also maintained theorem-retrieval restrictions and time pressure as potential co-factors when the evidence failed to support a single explanation.

These patterns exemplify a stage-differentiated yet functionally interdependent view of proof construction. Initiation, manipulation of tools, monitoring of goals, and communication exemplify functionally distinct tasks of proof construction. However, the correctness of one proof construction step may rely on earlier steps, premises, or warrants.

3.2. Discussion

3.2.1. Failure in Proof Construction as a Structured and Sequential Phenomenon

The results build upon an existing error taxonomy by illustrating how certain breakdowns were linked to individual attempts at constructing a proof. Differentiation in stages of proof construction was due to the distinct, though functionally united, roles of initiation, tool application, goal monitoring, and communication, as one inference could rely on a prior premise or warrant.

The first proposition identifies a situation wherein the early omission of a premise would be relevant, and that is when the omitted premise is necessary to prove a later theorem. Thus, propagation would be dependent on functional necessity rather than the order in which conditions are satisfied. This would qualify the claim that the first error in a proof invariably leads to all subsequent errors. The second proposition states that when the relations in the proof are diagrammed and are perceived to be evident, the diagram may advance from an exploratory representation to an evidential substitute, overriding the verification of the conditions of the theorem, and hence result in a locally valid argument, but globally invalid.

The third proposition states that some failures in verbal and written communication are downstream effects of the integration of warrants and are not language deficits per se. Participants that failed to verbalize a transition also failed to maintain a deductive linkage in the proof. Additionally, proof schemes by Harel and Sowder (2007) help in explaining these patterns. It seems as though the students' reasoning is influenced by cognitive structures that are dynamic and context specific.

Research indicates that students activate various schemes to complete the same task, resulting in schemes that are multi-faceted and unstable (Kanellos et al., 2018). This describes the activation of fine-grained cognitive resources, not the activation of stable schemes (Hammer, 2004). Moreover, students' empirically or externally grounded schemes, especially in the earlier stages of learning, add to the instability, as students move unevenly to deductive reasoning (Akkurt & Durmuş, 2022). Thus, students' reasoning may fluctuate among empirical, intuitive, and deductive, but without achieving an integrated rationality.

Collectively, these propositions conceptualize failure pathways as an analytical account of interlinked breakdowns, as opposed to a statistical framework or a globally sequential model. The contribution specifies the evidence and boundary conditions within which a local breakdown potentially constrains subsequent reasoning.

3.2.2. Dominance of Failures at the Tools Stage: Conceptual–Procedural Gap

The largest numbers of coded episodes for UPAF and DF results from theorem instantiation and from steps of deductive inference that occur repeatedly throughout a proof. They also make evident the consequences of omissions that occur earlier in the proof. Their occurrences thus indicate breakdowns that occur repeatedly throughout the proof and are not evidence that constructed deficiency is the only predominant factor for that case and that pattern.

This interpretation aligns with empirical data showing a persistent gap between knowledge, proof knowledge, and its application in proof contexts. For example, Weber (2002) found that students, irrespective of their level, possessed the instrumental knowledge for proving but could not produce a valid proof. Similarly, Iannone and Inglis (2010) found that the main reason for the failure of students in proof tasks was a lack of valid deductive arguments, rather than a lack of adequate concepts to address the proof. Subsequent studies have also confirmed that students across different academic levels struggle to construct valid arguments in cases involving proof (Alghadari et al., 2025). Most studies on proof knowledge and its application have confirmed the existence of a logic failure gap.

The data also suggest how perceptual reasoning affects deduction. When a relationship is visually apparent, subjects sometimes develop enough confidence to cease warrant search. Thus, the problem is not visual reasoning, but rather the unmarked shift from exploration to justification.

This research aligns with most studies that document the empirical reasoning students employ when moving from informal to formal deduction (Tao & Fu, 2024), as well as their repeated failures in building valid justificatory chains, regardless of the mathematical plausibility of the individual statements (Stylianides & Stylianides, 2009).

Thus, at the tools stage, the problem is not simply conceptual; it encompasses a variety of more sophisticated challenges, such as choosing and defending a particular theorem and forming a logically acceptable conclusion. From a theoretical viewpoint, this has the potential to advance the existing literature by substantiating the argument for the necessity to reframe the understanding of learning proof not as the acquisition of separate, distinct components of knowledge, but as the incorporation of a comprehensive system of reasoning in which the elements of understanding, procedural knowledge, and reasoning are harmoniously integrated.

3.2.3. Visualization and the Risk of Perceptual Reasoning

This study does not intend to be the first to recognize the importance and danger of diagrams in geometric proofs. The more limited contribution of this study is the empirical evidence for the stages of the pathways of diagrammatic conviction that were observed: perceptual salience led to increased confidence, confidence lowered the demand for warrant, and therefore, the conditions of the theorem were less likely to be examined.

This phenomenon of duality has been recognized by a wide cross-sectional representation of the literature. For example, Barwise and Etchemendy (1996) claimed that diagrams, while they may carry cognitive benefits, also become “a breeding ground for fallacious inferences” when an individual relies on perception features and does not engage in a formal proof. In the same vein, Butcher and Aleven (2008) state that the good use of visual representation must go hand in hand with a solid conceptual framework, as a lack of focus on the critical elements of the representation may create more problems than it solves. From the perspective of mathematics, Samkoff et al. (2012) show that even experts do not use visualisation in proof uniformly, confirming that visualisation is not an uncomplicated or uniformly reliable constituent of reasoning.

The problem occurred in relation to participants treating a diagram as evidence. The participants initialized a proof with a diagram and then included a visibly plausible relation without an explicit mathematical warrant. This lack of a formal statement around the concept represented how weakly they transitioned from a conceptual representation to a deductive justification.

This mechanism also worked alongside symbolic and verbal failures. If diagram-derived entities were not properly named, participants had issues maintaining correspondence and explaining which theorem was relevant. Therefore, processes of different natures and stages of reasoning were interrelated without being too seamlessly integrated.

The studies highlighted in this paper show that diagrams should be thought of as an integrated part of proof, and should be thoroughly analyzed when considering the proof process, rather than simply being thought of as an appendage to proof. The mechanics show that the reasoning processes of mathematics aid rather than obstruct the processes of analytical visualization.

3.2.4. Logical Structure and the Breakdown of Deductive Chains

The prevalence of HSF and DF indicates that students encounter significant difficulties in constructing a coherent sequence of logical reasoning. The results suggest that while students are capable of formulating logical statements, they struggle to present these statements in a logically sequenced manner. This implies that the primary challenge lies not in a lack of understanding of individual components, but rather in the inability to structure and justify the deductive relationships involved.

This explanation is consistent with studies that illustrate students struggle to assess arguments for proof validity. Undergraduates, for example, are documented by Selden and Selden (2003) to engage arguments at the surface level, demonstrating lack of proof validation ability for the logical structure of the guide. Likewise, Miyazaki et al. (2017), when identifying deductive reasoning through hypothetical syllogism, state that students' acceptance of non-logical reasoning is an understanding gap of the reasoning chain.

The findings of this study, in addition to individual errors of inference, also describe other challenges in the reasoning structure. This is in line with Netti et al. (2024), who describe students' inability to integrate all the reasoning elements in proof construction as "schema disconnection." The findings also show that when the components of the reasoning are logically coherent, their reasoning is disjointed, lacking coherence, when the components of the reasoning are logically coherent.

The difficulty mentioned is aligned with the developmental approach to the research on deductive reasoning in the field of mathematics education. One of the issues in constructing mathematical proofs is that students need knowledge of the mathematical statements but they also need knowledge of the statements to be able to be joined through some valid structural inference (Stylianides & Stylianides, 2008). The difficulty in keeping these structural links shows that students' reasoning is more like disconnected statements, thought to be generated, but not integrated in any systematic way.

From the EMA perspective, these results pertain to continuity among functional components, as opposed to a singular, localized defect. A proof can have many correct components and still be globally invalid if the connections among premises, warrants, intermediate claims, and conclusions are not sustained.

3.2.5. Pedagogical Implications and Limitations

The mechanisms described in this study have certain implications for teaching. Premise-goal mapping can help with under-integrated givens, theorem-condition checks can help with invalid instantiation, verbal translations of visual relations can help with diagrammatic conviction, and a proof-path audit that asks, “What established statement authorizes this step?" can help with deductive gaps. The ability to orally defend and to have peers validate the proof are ways to check that concise proofs contain reconstructible warrants.

There are several limitations to this study. The study has a small, purposive sample and is limited to one course. The study is limited to a specified amount of direct geometric proofs and the nature of pathway analysis. The counts describe coded instances and not the frequency. The ordering of the incidents, in conjunction with the evidence given in the written interviews, supports the existence of the mechanisms, but not in a definitive manner. For future studies, the focus should be on the pathways in different proof domains and determine if the suggested scaffolds stop the breaks in reasoning.

4. CONCLUSION

This study identified proof-construction failures across the initiation, tools, goals, and modalities components of an adapted EMA framework. The most recurrent breakdowns involved theorem application and deductive inference, but the analysis showed that these difficulties interacted with premise interpretation, goal monitoring, representation, and communication.

The principal contribution of this research is the failure-pathway account. In the constrained dataset, three pathways were repeated. First, under-integrated premises limited access to a required warrant. Second, diagrammatic conviction reduced the pursuit of warrants and checking of theorem-conditions. Lastly, unsupported intermediate assertions disrupted deductive continuity and verbal reconstruction. These pathways further develop a taxonomy of isolated errors and demonstrate how incrementally plausible steps may result in a proof that is globally invalid.

The findings support a stage-differentiated yet functionally interdependent view of geometric proof construction. The components of EMAs perform different analyses, but later reasoning may rely on the validity of earlier premises and warrants. Therefore, instruction should integrate premise-goal mapping, checking theorems and conditions, translating diagrams to warrants, proof-path auditing, and oral defence. Teaching these components as disconnected skills will be ineffective.

Given the limits of the sample and the tasks, the results should not be interpreted as prevalence estimates nor as deterministic causal sequences. Future research should study the applicability of the pathways proposed in different mathematical domains, educational levels, and task frameworks as well as examine the possibility of interrupting the progression of locally valid proofs to globally invalid proofs by the use of specific scaffolding.

Acknowledgments

The authors gratefully acknowledge the Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang (UM), for providing financial support toward the publication of this article.

Declarations

Author Contribution	: LA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing - original draft, and Writing - review & editing; IK: Data curation, Formal analysis, Investigation, and Writing - original draft; CS: Validation, and Writing - review & editing; SI: Supervision, and Writing - review & editing.
Funding Statement	: The publication of this article was funded by the Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang (UM).
Conflict of Interest	: The authors declare no conflict of interest.
Additional Information	: Additional information is available for this paper.

REFERENCES

Akkurt, Y. Y., & Durmuş, S. (2022). Tracing proof schemes: some patterns and newperspectives. *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, *7*(1), 1–16. https://doi.org/10.23917/jramathedu.v7i1.15740

Alcock, L., & Weber, K. (2005). Proof validation in real analysis: Inferring and checking warrants. *The Journal of Mathematical Behavior*, *24*(2), 125–134. https://doi.org/10.1016/j.jmathb.2005.03.003

Alghadari, F., Arisha, B., Hidayah, N., Saparuddin, S., & Dharma, B. E. (2025). Uncovering gaps in deductive geometry thinking: Rasch-based evidence from students' work on quadratic functions. *Journal of Instructional Mathematics*, *6*(2), 117–129. https://doi.org/10.37640/jim.v6i2.2510

Anwar, L., & Goedhart, M. (2019). Understanding geometric proofs: scaffolding pre-service mathematics teacher students through dynamic geometry system (dgs) and flow-chart proof. In *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME11)*, (pp. 104–111).

Anwar, L., Goedhart, M. J., & Mali, A. (2023). Learning trajectory of geometry proof construction: Studying the emerging understanding of the structure of Euclidean proof. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(5), em2266. https://doi.org/10.29333/ejmste/13160

Anwar, L., Mali, A., & Goedhart, M. (2024). Formulating a conjecture through an identification of robust invariants with a dynamic geometry system. *International Journal of Mathematical Education in Science and Technology*, *55*(7), 1681–1703. https://doi.org/10.1080/0020739x.2022.2144517

Anwar, L., Sa'dijah, C., Listiawan, T., Utami, A. D., & Zulnaidi, H. (2025). Challenges of prospective mathematics teachers in formulating geometrical conjecture through interaction with GeoGebra. *Mathematics Teaching Research Journal*, *17*(1), 7–27.

Aphrodite, M. A., & Rita, P. (2021). Young students' ability on understanding and constructing geometric proofs. *Social Education Research*, *2*(2), 121–133. https://doi.org/10.37256/ser.222021784

Barwise, J., & Etchemendy, J. (1996). Visual information and valid reasoning. In G. Allwein & J. Barwise (Eds.), *Logical Reasoning with Diagrams* (pp. 3–26). Oxford University Press.

Butcher, K. R., & Aleven, V. (2008). Diagram interaction during intelligent tutoring in geometry: Support for knowledge retention and deep transfer. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci 2008)*, (pp. 894–899).

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage.

Hamdani, D., Purwanto, P., Sukoriyanto, S., & Anwar, L. (2023). Causes of proof construction failure in proof by contradiction. *Journal on Mathematics Education*, *14*(3), 415–448. https://doi.org/10.22342/jme.v14i3.pp415-448

Hammer, D. (2004). The variability of student reasoning, lecture 3: Manifold cognitive resources. In *Proceedings-International School of Physics Enrico Fermi*, (Vol. 156, pp. 321–340).

Hanna, G. (2000). Proof, explanation and exploration: An overview. *Educational Studies in Mathematics*, *44*(1-3), 5–23. https://doi.org/10.1023/a:1012737223465

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In J. Lester, Frank K. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 805–842). National Council of Teachers Mathematics.

Hartono, S. (2025). Evolving trends in mathematical proof research in Indonesian mathematics education: A systematic review from design to data analysis. *Multidisciplinary Reviews*, *8*(12), 2025372. https://doi.org/10.31893/multirev.2025372

Iannone, P., & Inglis, M. (2010). Undergraduate students' use of deductive arguments to solve'prove that…'tasks. In *7th Congress of the European Society for Research in Mathematics Education*.

Kanellos, I., Nardi, E., & Biza, I. (2018). Proof schemes combined: mapping secondary students' multi-faceted and evolving first encounters with mathematical proof. *Mathematical thinking and learning*, *20*(4), 277–294. https://doi.org/10.1080/10986065.2018.1509420

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. National Academies Press.

Kirsten, K., & Greefrath, G. (2023). Proof construction and in-process validation – Validation activities of undergraduates in constructing mathematical proofs. *The Journal of Mathematical Behavior*, *70*, 101064. https://doi.org/10.1016/j.jmathb.2023.101064

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis a methods sourcebook* (3rd ed.). SAGE.

Miyazaki, M., Fujita, T., & Jones, K. (2017). Students' understanding of the structure of deductive proof. *Educational Studies in Mathematics*, *94*(2), 223–239. https://doi.org/10.1007/s10649-016-9720-9

Netti, S., Abdul Rahim, S. S., & Vermana, L. (2024). Analysis of students' cognitive structures in failed mathematical proof construction. *Matematika dan Pembelajaran*, *12*(2), 127–142. https://doi.org/10.33477/mp.v12i2.8216

Putra, Z. H., Afrillia, Y. M., Dahnilsyah, D., & Tjoe, H. (2023). Prospective elementary teachers' informal mathematical proof using GeoGebra: The case of 3D shapes. *Journal on Mathematics Education*, *14*(3), 449–468. https://doi.org/10.22342/jme.v14i3.pp449-468

Rabin, J. M., & Quarfoot, D. (2022). Sources of students' difficulties with proof by contradiction. *International Journal of Research in Undergraduate Mathematics Education*, *8*(3), 521–549. https://doi.org/10.1007/s40753-021-00152-x

Reuter, F. (2023). Explorative mathematical argumentation: a theoretical framework for identifying and analysing argumentation processes in early mathematics learning. *Educational Studies in Mathematics*, *112*(3), 415–435. https://doi.org/10.1007/s10649-022-10199-5

Sáenz-Ludlow, A., & Athanasopoulou, A. (2016). The role of diagrammatic reasoning in the proving process. In *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, (pp. 542–550).

Samkoff, A., Lai, Y., & Weber, K. (2012). On the different ways that mathematicians use diagrams in proof construction. *Research in Mathematics Education*, *14*(1), 49–67. https://doi.org/10.1080/14794802.2012.657438

Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, *34*(1), 4–36. https://doi.org/10.2307/30034698

Selden, A., & Selden, J. (2015). A theoretical perspective for proof construction. In *CERME 9-Ninth Congress of the European Society for Research in Mathematics Education*, (pp. 198–204).

Soucy McCrone, S. M., & Martin, T. S. (2004). Assessing high school students' understanding of geometric proof. *Canadian Journal of Science, Mathematics and Technology Education*, *4*(2), 223–242. https://doi.org/10.1080/14926150409556607

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education*, *38*(3), 289–321.

Stylianides, A. J., & Stylianides, G. J. (2007). The mental models theory of deductive reasoning: Implications for proof instruction. In *Proceedings of CERME5, the 5th Conference of European Research in Mathematics Education*, (pp. 665–674).

Stylianides, A. J., & Stylianides, G. J. (2009). Proof constructions and evaluations. *Educational Studies in Mathematics*, *72*(2), 237–253. https://doi.org/10.1007/s10649-009-9191-3

Stylianides, G. J., & Stylianides, A. J. (2008). Proof in school mathematics: Insights from psychological research into students' ability for deductive reasoning. *Mathematical thinking and learning*, *10*(2), 103–133. https://doi.org/10.1080/10986060701854425

Tao, S., & Fu, H. (2024). An investigation into the challenges and strategies of secondary students' geometric thought processes through the lens of Van Hiele's theory. *Region - Educational Research and Reviews*, *6*(6). https://doi.org/10.32629/rerr.v6i6.2252

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, *48*(1), 101–119. https://doi.org/10.1023/a:1015535614355

Weber, K. (2002). The role of instrumental and relational understanding in proofs about group isomorphisms. In *Proceedings from the 2nd International Conference for the Teaching of Mathematics*, (pp. 1–9).

Weber, K. (2004). A framework for describing the processes that undergraduates use to construct proofs. In *Proceedings of the 28th Annual Meeting of the International Group for the Psychology of Mathematics Education*, (Vol. 4, pp. 425–432).

Winer, M. L., & Battista, M. T. (2022). Investigating students' proof reasoning: Analyzing students' oral proof explanations and their written proofs in high school geometry. *International Electronic Journal of Mathematics Education*, *17*(2), em0677. https://doi.org/10.29333/iejme/11713

Wulan, E. R., Subanji, S., & Muksar, M. (2021). Metacognitive failure in constructing proof and how to scaffold it. *Al-Jabar : Jurnal Pendidikan Matematika*, *12*(2), 295–314. https://doi.org/10.24042/ajpm.v12i2.9590