Student Confidence and Certainty in Comprehension of a Proof by Induction

Michael Preheim, North Dakota State University
Josef Dorfmeister, North Dakota State University
Ethan Snow, University of Nebraska

Researchers typically utilize response correctness to interpret student proficiency in proof comprehension. However, metacognition has not been simultaneously analyzed alongside correctness to determine student competency in proof comprehension, yet it offers important information about performance behavior. The primary objective of this study is to investigate the accuracy of student confidence and certainty levels at local and holistic aspects of proof comprehension regarding a proof by induction. Students were given a three-factor proof comprehension assessment at the beginning and end of an undergraduate transition-to-proof course that collected student confidence, correctness, and certainty at each tier of an established proof comprehension framework. Results of this study highlight a critical distinction between high and low performers' metacognition throughout the host course. One outlying assessment item especially illuminates additional considerations for future application of metacognition in proof comprehension research.

Keywords: Proof comprehension, Proof by induction, Confidence and certainty, Metacognition, Multidimensional competency assessment

1 Introduction

The present study investigates the interplay between students' confidence and certainty (states of metacognition) with response correctness within proof comprehension assessments. The discussions presented in this paper operate on the following definitions of confidence and certainty. *Confidence* refers to an individual's level of belief that they *can provide* a correct response to an assessment item pertaining to a given set of clearly defined criteria, whereas *certainty* refers instead to an individual's level of belief that they *have provided* a correct response to an assessment item. Confidence and certainty are differentiated by their context and timing at which they are measured. Confidence, measured prior to viewing assessment items, is a self-assessment for anticipated performance, while certainty is a self-assessment of past performance measured after providing a response to an assessment item.

1.1 Metacognition and Multidimensional Knowledge Assessment

Metacognition includes any way of monitoring of cognition (Kitchener, 1983) and thus the term has been applied with a wide variety of interpretations. Commonly, metacognition is categorized into two broad areas – metacognitive knowledge and metacognitive skills (Veenman & Spaans, 2005). Metacognitive knowledge involves one's beliefs about the interactions of person, task, and strategy knowledge (Flavell, 1979), whereas metacognitive skills refer to the "procedural knowledge that is required for the actual regulation of and control over one's learning activities" (Veenman & Spaans, 2005). Confidence and certainty as previously defined address a subcategory of metacognitive knowledge (relating person knowledge to task knowledge) wherein an individual self-monitors their cognitive performance.

Knowledge is information which is both true and justified (Hunt, 2003). Student certainty has been well studied for providing response justification during assessment (Gardner-Medwin and Gahan, 2003; Ghadermarzi *et al.*, 2015; Hunt, 2003; Snow, 2019). The authors of the present study have investigated simultaneous evaluation of confidence, certainty, and correctness (the

“3Cs”) within an undergraduate linear algebra course and established corresponding accuracy indices for identifying outlying student metacognitive behaviors and providing new insights for assessment item analysis (Preheim *et al.*, 2022; Preheim *et al.,* 2023). The 3C assessment structure and analysis has yet to be used to investigate student competency in proof comprehension.

1.2 The Role of Confidence and Certainty in Proof Comprehension and Proof Writing

An individual’s confidence and certainty in proof comprehension may impact decisions which occur throughout proof writing tasks wherein they must make strategic choices on how to proceed (*e.g.*, deciding which method of proof to initially employ, which strategy to use to satisfy a particular case, *etc.*). For example, consider a student who does not feel confident in their comprehension of proof by contrapositive as a technique. When posed with the task of proving “if n^2 - $(n\text{-}1)^2$ is not divisible by 8, then n is even,” this student might shy away from a proof by contrapositive and be led to attempt a direct argument (which would be more tedious). Similarly, an individual’s certainty in their proof comprehension has the potential to influence their willingness to move to a next step in a proof-writing task or impact their conviction in the truth of the mathematical statement being proved.

While the role of proof in establishing mathematical conviction (*i.e.*, belief in the truth of a mathematical statement) has been discussed and investigated by researchers such as Weber or Mejía-Ramos (Weber & Mejía-Ramos, 2011; Weber *et al.*, 2014; Weber & Mejía-Ramos, 2015; Weber *et al.*, 2022), student confidence and certainty regarding proof comprehension has not been thoroughly investigated. Mejía-Ramos *et al.* (2011) developed a model by which to construct assessment items for proof comprehension assessments (PCAs) in undergraduate mathematics. This proof comprehension framework (PCF) identifies seven aspects of proof comprehension separated into two tiers: local and holistic (Mejía-Ramos *et al.*, 2012). Assessment items at the local level are intended to assess an individual’s comprehension of a small number of statements contained within a proof. At the holistic level, items assess an individual’s understanding of the proof as a whole (Mejía-Ramos *et al.*, 2012). Student metacognitive behavior at each level of the PCF has not been empirically investigated to measure student competency or assessment item validity.

The objective of this study is to investigate students’ confidence and certainty levels and their accuracies at local and holistic aspects of proof comprehension regarding a proof by induction. Distinctions between high and low performers’ metacognitive behaviors from the beginning to end of the host course are also presented and discussed.

2 Methods

2.1 Course Information and Assessment Structure

The introduction to abstract mathematics course which hosts the focused student population in the present study serves as the transition-to-proof course for undergraduate students at North Dakota State University. This course focuses on proof-writing tasks and proof comprehension amidst various topics in abstract mathematics. 3C data was collected via two assessments per section from 36 students in four sections of the course.

The PCA used in this study assesses comprehension of a proof by induction that every third Fibonacci number is even. This multiple choice PCA developed by Mejía-Ramos *et al.* (2017) which adheres to the PCF (Mejía-Ramos *et al.,* 2012) was adapted to include 3C assessment methodology. This same 40-minute in-person assessment was administered near the start of the course (the initial assessment) as well as near the end of the course (the post assessment).

Each assessment item from the PCA was mapped to one PCF subitem. Immediately preceding each assessment, student confidence levels were collected for pertinent PCF subitems. The prompt requested students to indicate their level of confidence in each PCF subitem (4=High confidence, 3=Somewhat high confidence, 2=Somewhat low confidence, and 1=Low confidence). In the first semester of collected data, students provided their confidence levels prior to viewing the proof or assessment items. During the second semester, students were given five minutes to read the proof prior to providing confidence levels.

Once the confidence self-assessments were completed, students were given an assessment packet which contained the 12 multiple-choice questions. Following each multiple-choice question, a corresponding certainty prompt (similarly structured to the confidence prompt) instructed students to record their certainty in their response before moving on to the next item.

2.2 Accuracy Indices and Metacognitive Accuracy Plots

An individual student response yields the following three values: c_1 = confidence level, c_2 = correctness (1 for correct and 0 for incorrect), and c_3 = certainty level. Two similarly defined functions were developed to measure the alignment of confidence with credit (AccConf) and the alignment of certainty with credit (AccCert). The formulas for each are shown in Figure 1.

$$\mathrm{AccConf}(c_1, c_2, c_3) = \begin{cases} 1 & \text{if } \mathrm{c}_2 = 1 \text{ and } c_1 = 3 \text{ or } 4 \\ -1 & \text{if } \mathrm{c}_2 = 1 \text{ and } c_1 = 1 \text{ or } 2 \\ -1 & \text{if } \mathrm{c}_2 = 0 \text{ and } c_1 = 3 \text{ or } 4 \\ 1 & \text{if } \mathrm{c}_2 = 0 \text{ and } c_1 = 1 \text{ or } 2 \end{cases} \qquad \mathrm{AccCert}(c_1, c_2, c_3) = \begin{cases} 1 & \text{if } \mathrm{c}_2 = 1 \text{ and } c_3 = 3 \text{ or } 4 \\ -1 & \text{if } \mathrm{c}_2 = 1 \text{ and } c_3 = 1 \text{ or } 2 \\ -1 & \text{if } \mathrm{c}_2 = 0 \text{ and } c_3 = 3 \text{ or } 4 \\ 1 & \text{if } \mathrm{c}_2 = 0 \text{ and } c_3 = 1 \text{ or } 2 \end{cases}$$

Figure 1. The metacognitive accuracy formulas AccConf and AccCert. A weighted average of these accuracy values produces the accuracy indices AIConf and AICert.

Accuracy values of 1 indicate accurate confidence/certainty while accuracy values of -1 indicate inaccurate confidence/certainty. A weighted average of accuracy values across responses was computed for each assessment to generate an accuracy index for confidence (AIConf) and an accuracy index for certainty (AICert). For this weighted average, responses reporting confidence/certainty levels 1 and 4 were weighted at three times the weight of those corresponding to levels 2 and 3. Metacognitive accuracy plots were produced to display the transitions in AIConf and AICert from the initial to post assessment among all students, high performers (students in the top 25th percentile of course standing), and low performers (students in the bottom 25th percentile of course standing).

3 Results

Among all students, AIConf increased from the initial assessment to the post assessment for all local items except item Q4 (Figure 2a). Among high performers, AIConf increased by more than 0.20 from the initial assessment to the post assessment for ten of the twelve items (excluding Q4 and one item with near maximum accuracy) (Figure 2b). Among low performers, AIConf and AICert decreased from the initial to post assessment for all holistic items (Figure 2c). Furthermore, Q4 showed substantial decrease in both AIConf and AICert among low performers (Figure 2b) and substantial increase in AICert among high performers (Figure 2c).

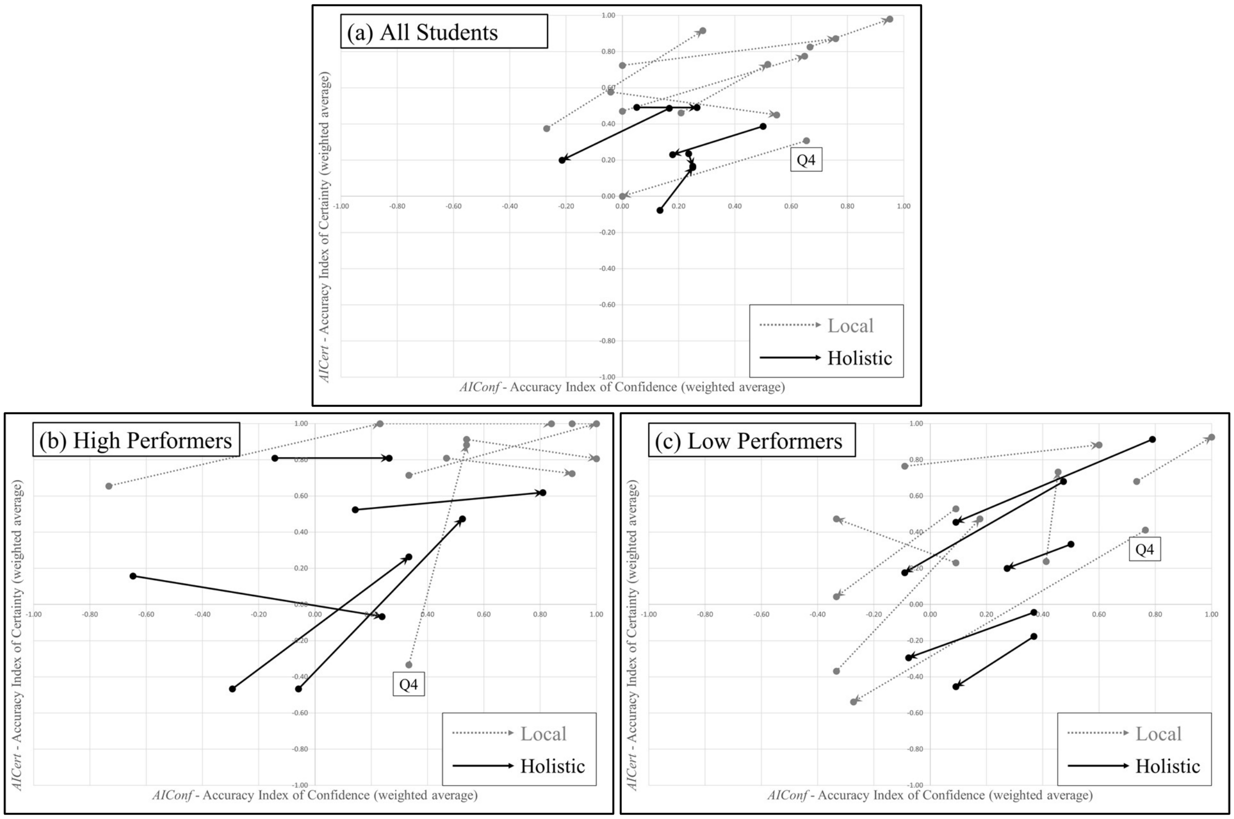

Figure 2. The transitions of AIConf and AICert for each assessment item from the initial to the post assessment are plotted as arrows on a graph of AICert vs AIConf. Items addressing local aspects of the PCF are represented with gray dotted arrows while those addressing holistic aspects are represented by black solid arrows.

Reported confidence and certainty levels demonstrated significant ($\alpha<0.05$) increase from the initial to the post assessment for all and eleven of the twelve items, respectively. Certainty levels were higher than confidence levels on average among questions pertaining to local aspects of the PCF for both the initial assessment and the post assessment. However, certainty levels among questions pertaining to holistic aspects of the PCF were lower than confidence levels during the post assessment. On average AICert was higher than AIConf among both assessments for all items, local items, and holistic items. Average confidence and certainty levels as well as correctness scores increased from the initial assessment to the post assessment among all items, local items, and holistic items. Though average AIConf and AICert among questions addressing local aspects of the PCF increased from the initial to the post assessment, both decreased among questions addressing holistic aspects of the PCF (Table 1).

Table 1. The average confidence level, certainty level, score (correctness), AIConf, and AICert on the initial and post assessments among all items, local items, and holistic items.

		Confidence	Certainty	Score	AIConf	AICert
All	Initial	2.26	2.63	0.58	0.19	0.44
	Post	3.07	3.14	0.67	0.37	0.50
Local	Initial	2.56	2.92	0.70	0.17	0.53
	Post	3.26	3.40	0.75	0.53	0.67
Holistic	Initial	1.95	2.21	0.40	0.22	0.30
	Post	2.88	2.77	0.54	0.15	0.25

4 Discussion

Transitions from the initial to the post assessment highlight a critical distinction between questions addressing local aspects of the PCF and those addressing holistic aspects. Increased accuracy of confidence regarding local aspects of the PCF indicates that students' increase of confidence levels at the local tier (Table 1) was accurate to their performance. Though questions at the holistic tier also saw an improvement of confidence levels and score (Table 1), the varying shifts in students' accuracy of confidence (Figure 2a) suggest that the increase of confidence was not necessarily indicative of an increase in performance. This may suggest that students feel more comfortable reading mathematical proof by the end of the course, but it does not guarantee they are more competent in comprehending proof at the holistic level.

A further distinction between local and holistic aspects is made when comparing the shifts of AIConf and AICert among high versus low performers. The results shown in Figure 2 (b and c) demonstrate that while high performers exhibited increased accuracy of confidence among holistic aspects of proof comprehension, low performers experienced a decrease in metacognitive accuracy among holistic aspects of the PCF. These preliminary findings suggest that training student metacognition with respect to proof comprehension may be beneficial for improving student performance in abstract mathematics. Further investigation of this idea is merited.

3C data identified outlying metacognitive behavior on assessment item Q4 which would not have been detectable by correctness-only assessment methods. Qualitative evaluation of the question prompt and multiple-choice options indicates that while the prompt seems to address a local aspect of the PCF, the correct option requires holistic understanding of the proof. As these PCA items originated from open-response questions (Mejía-Ramos *et al.*, 2017), this discrepancy is not particular to the multiple-choice structure. Moreover, the most selected incorrect option includes a correct description of the proof's procedural structure followed by an incorrect conceptual justification for the final claim of the proof. This outlying behavior therefore supports an observation by Piatek-Jiminez (2004) that procedural knowledge and conceptual knowledge of proof by induction are not necessarily reliant upon each other.

5 Conclusions, Future Directions, and Questions for the Audience

The 3C assessment distinguished metacognitive behavior at the holistic level between high- and low- performing students. Development and use of methods for enhancing students' metacognitive accuracy in proof comprehension could improve student performance in abstract mathematics. Future directions of this research are focused on employing the 3C assessment methodology to improve students' metacognitive accuracy among holistic aspects of proof.

Adapting proof comprehension assessments to utilize 3C assessment methodology allows for the detection of interesting outlying metacognitive behavior. Investigating such outliers can aid in revising assessment items, generating new assessment items, and guiding directions for continued research. For example, the investigation of outlying behavior observed in this study made the distinction between conceptual and procedural knowledge observable. To more comprehensively analyze proof comprehension within the PCF, future studies might consider methods to further distinguish conceptual and procedural knowledge and study the 3C methodology in other related and unrelated domains.

The authors conclude this proposal with some intended questions for the attending audience:

1. How might 3C data be used to guide assessment item revision or PCA construction?
2. How might metacognitive behavior differ if PCAs utilized open-response questions?
3. Are there other methods of proof which you expect would generate critically different metacognitive behavior than that demonstrated in the present study?

References

Flavell, J.H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *The American Psychologist, 34* (10), 906-911.

Gardner-Medwin, A.R. & Gahan, M. (2003). Formative and Summative Confidence-Based Assessment. In *Proceedings of the 7th International Computer-Aided Assessment Conference in Loughborouh, UK, July 2003* (pp. 147-155)

Ghadermarzi, M., Yazdani, S., Pooladi, A., & Hosseini, F. (2015). A Comparative Study between the Conventional MCQ Scores and MCQ with the CBA Scores at the Standardized Clinical Knowledge Exam for Clinical Medical Students. *Journal of Medical Education (Tehran), 14* (1).

Hunt, D.P. (2003). The concept of knowledge and how to measure it. *The Journal of Intellectual Capital, 4* (1), 100-113.

Kitchener. (1983). Cognition, Metacognition, and Epistemic Cognition: A Three-Level Model of Cognitive Processing. *Human Development, 26* (4), 222-232.

Mejía-Ramos, J.P., Fuller, E., Weber, K., Rhoads, K., & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics, 79* (1), 3-18.

Mejía-Ramos, J.P., Lew, K., de la Torre, J., & Weber, K. (2017). *Developing and validating proof comprehension tests in undergraduate mathematics. Research in Mathematics Education, 19* (2), 130-146.

Piatek-Jimenez., K.L. (2004). Undergraduate mathematics students' understanding of mathematical statements and proofs. *ProQuest Dissertations Publishing*.

Preheim, M., Dorfmeister, J., & Snow, E. (2022). Efficacy of a Three Factor Assessment Method in Teaching Undergraduate Linear Algebra. In *Proceedings of the 24th Annual Conference on Research in Undergraduate Mathematics Education in Boston, Massachusetts* (pp. 1269–1270).

Preheim, M., Dorfmeister, J., & Snow, E. (2023). Assessing Confidence and Certainty of Students in an Undergraduate Linear Algebra Course. *Journal for Stem Education Research.* https://doi.org/10.1007/s41979-022-00082-6

Snow, E. (2019). Effects of a Confidence-Based, Individualized Remediation Strategy on Student Learning and Final Grades in a Multi-Campus Human Anatomy Curriculum, Dissertation in the UND Department of Biomedical Sciences – Doctor of Philosophy.

Veenman, M.V.J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences, 15* (2), 159–176.

Weber, K., Inglis, M., & Mejía-Ramos, J. P. (2014). How Mathematicians Obtain Conviction: Implications for Mathematics Instruction and Research on Epistemic Cognition. *Educational Psychologist, 49* (1), 36-58.

Weber, K., & Mejía-Ramos, J.P. (2011). Why and how mathematicians read proofs: an exploratory study. *Educational Studies in Mathematics, 76* (3), 329-344.

Weber, K., Mejía-Ramos, J.P. (2015). On Relative and Absolute Conviction in Mathematics. *For the Learning of Mathematics, 35* (2), 15-21.

Weber, K., Mejía-Ramos, J.P., & Volpe, T. (2022). The Relationship Between Proof and Certainty in Mathematical Practice. Journal for Research in Mathematics Education, 53 (1), 65–84.