PROOF SUMMARY AND PROOF COMPREHENSION

Katharina Kirsten

University of Münster, Germany

Proof summaries have recently been discussed from two perspectives: as a strategy to promote proof comprehension and as an instrument to measure it. To gain deeper insight into the construct of proof summaries, this study investigates their relation to prior knowledge and performance on a multiple-choice proof comprehension test. Sixty advanced undergraduates watched a proof video and completed either a summary followed by a test or the reverse. Analyses showed no meaningful correlation between summary scores and test performance, suggesting the two assessments capture different constructs. Regression analyses indicate that summarizing may be beneficial for students with higher prior knowledge, which has implications for teaching practice.

INTRODUCTION

It is widely known that for many undergraduate students difficulties arise not only in the construction of proofs, but also in reading and understanding the proofs presented in a textbook or lecture (e.g., Hodds et al., 2014; Inglis & Alcock, 2012). In this context, the writing of summaries is discussed as a means of promoting deeper understanding and bringing out higher-level ideas more clearly (e.g., Demeke, 2025). However, research on the effectiveness of the summarizing strategy in connection with proof is limited and ambiguous. Moreover, in recent work, proof summaries appear not as a proof-reading strategy but as an instrument for assessing proof comprehension (Davies & Jones, 2022). This raises questions about the nature of proof summaries and their relationship to proof comprehension. The aim of this study is therefore to provide new theoretical insights into the concept of proof summaries that can inform methodological decisions, as well as practical implications regarding their potential for teaching. The study is embedded in a larger project on proof videos. In proof videos, a proof is presented step-by-step by writing down the relevant mathematical steps while the instructor thinks aloud. Doing so, the video gives insight into an expert's behavior which can help students to better understand mathematical concepts and to learn new methods for proof construction. However, as multiple-choice tests demonstrated a poor understanding of the proofs shown in the video (Wirth et al., 2024), this learning environment is well suited for analyzing the potential and impact of proof summaries.

THEORETICAL FRAMEWORK

Undergraduates' proof comprehension

According to Mejia-Ramos et al. (2012), a proper proof comprehension involves both local and holistic understanding of the proof presented. Students demonstrate local

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 843-850. PME.

comprehension when they understand the meaning of terms and statements, the justification of claims, the logical status of the proof's statements and the proof framework. Holistic comprehension of the proof is indicated when students are able to summarize high-level ideas, identify the modular structure, transfer the general ideas or methods to another context, and understand how the proof relates to examples. Previous studies showed that undergraduate students' proof comprehension is generally moderate or low. This applies to written proof (Neuhaus & Rach, 2019) as well as proofs presented in videos (Wirth et al., 2024). Possible reasons include the importance of sufficient prior content knowledge and the use of reading strategies (Neuhaus & Rach, 2019). For example, Alcock and Weber (2005) observed that some students focus more on local elements, such as specific implications, instead of analysing the proof's high-level ideas. Consistent with this, Inglis and Alcock (2012) found no reading behaviour that would allow students to identify implicit links between statements or attend to the global structure. Based on these findings, students' proof comprehension is likely to depend on how they engage in the proof presented. Accordingly, Samkoff and Weber (2015) reported that teaching proof reading strategies can support students' proof comprehension.

Summarizing as a strategy to improve proof comprehension

The ICAP framework posits that the mode of engagement can influence learning outcomes: While *passive* engagement may result in limited understanding, characterized by verbatim repetition of knowledge, *active* engagement enables students to apply knowledge on a superficial level. Conversely, *constructive* or *interactive* engagement fosters deeper, transferable understanding (Chi & Wylie, 2014). One strategy for dealing *constructively* with the learning material is to summarize its content. The aim of this strategy is to create a summary that reflects the essence of a text on the one hand and significantly reduces the information on the other (King, 1992). In the context of mathematical proofs, which are already very short and concise, summarizing is less about reduction and more about focusing on the most important aspects and higher-level ideas. According to Davies and Jones, "a proof summary invokes an appropriate level of rigour or detail to generate a mapping to a formal proof." (Davies & Jones, 2022, p. 472). As undergraduates' often have difficulties focusing on the global structure of a proof (Inglis & Alcock, 2012), proof summaries in particular could support a holistic understanding. At least, certain aspects such as summarizing higher-level ideas or recognizing the modular structure from the assessment model of Mejia-Ramos et al. (2012) are directly addressed by the summary task. In general, summarizing the learning material in one's own words supports a deeper understanding and establishes connections to one's prior knowledge (King, 1992). However, while there is evidence that summarizing can enhance learning in general, depending on the quality of the summary and students' motivation (Donoghue & Hattie, 2021; King, 1992), research specific to mathematics or proof is scarce. Yang and Lin (2012) showed that a combination of reading-oriented tasks, including summarizing, can promote Year 9 students' comprehension of geometry proofs more

effectively than tasks involving writing and applying proofs. Demeke (2025) reported that summarizing proofs is an effective proof-reading strategy for experts, but not necessarily for undergraduates. In this study, many undergraduates failed to grasp the key ideas of the proof, merely repeating the sentences contained within it. Davies and Jones (2022) also analyzed undergraduates' proof summaries, viewing them less as a proof-reading strategy and more as a means of measuring it.

Summaries as a measure for proof comprehension

Based on the assessment model of Mejia-Ramos et al. (2012), various studies have developed multiple-choice tests, which assess the different facets of proof comprehension (e.g., Hodds et al., 2014; Neuhaus & Rach, 2019; Wirth et al., 2024). As the development of a reliable and valid multiple-choice test is time-consuming and demanding, Davies and Jones (2022) instead used summaries (restricted to 40 words) within a comparative judgement approach to measure proof comprehension. Comparing both approaches, they found only low correlations between the results of a multiple-choice test and the summary-rating ($r = .28, p < .001$) and concluded: "the low correlations could be part of a meaningful pattern in which the assessments simply assess different things" (Davies et al., 2020, p. 195). While producing a proof summary is a production task, taking a multiple-choice test is a recognition task. This, can lead to biases in terms of familiarity with the task or other underlying covariates (Davies et al., 2020). Furthermore, traditional proof comprehension tests contain items that address different aspects of proof comprehension at the local and global level (Mejia-Ramos et al., 2012), whereas proof summaries can focus on individual aspects and are generally considered as a means of capturing higher-level ideas (Demeke, 2025; King, 1992). Thus, summaries may be more likely to capture holistic comprehension.

To address criterion validity, Davies et al. (2020) also reported a moderate correlation between the results of the multiple-choice test and the final course scores ($r = .55, p < .001$), and a low correlation between the summary scores and the final course scores ($r = .23, p < .001$). Aside from the fact that only one specific proof was considered in this study, these correlations raise questions about the relationship between proof comprehension and prior content knowledge, and how this is reflected in the respective measurement approaches. The present study addresses these questions and discusses the potential of proof summaries as a proof-reading strategy or a measuring instrument, taking prior knowledge into account.

THE CURRENT STUDY

At first glance, the two perspectives on summarizing seem contradictory: If summaries are considered as an effective proof-reading strategy, using this strategy could improve proof comprehension, and lead to distorted measurement results. Thus, the aim of this study is to gain a more detailed insight into the nature of proof summaries and their relationship with prior content knowledge and proof comprehension, as assessed by

multiple-choice tests. The first research question addresses the comparability of the two assessment approaches.

RQ1: *To what extent are summary scores, multiple-choice test scores and prior knowledge related? Does the order of processing affect the relationship between summary scores and test scores?*

Based on the findings of Davies et al. (2020), a low correlation between summary scores and test scores is expected. As both the quality of a summary and the performance on multiple-choice tests have been shown to depend on prior content knowledge (King, 1992; Neuhaus & Rach, 2019), positive correlations are expected. Different correlations resulting from the order of processing are exploratively analyzed. The second research question addresses the possible directional effects of writing a summary. Because previous findings are ambiguous (Demeke, 2025; Yang & Lin, 2012), this research question is being investigated in an exploratory manner.

RQ2: *To what extent does using the summarizing strategy improve proof comprehension, as measured by a multiple-choice test? Does prior knowledge moderate this effect?*

MATERIAL AND METHODS

Procedure

To achieve meaningful results and avoid floor effects, the study was aimed at advanced Bachelor students (5th semester) who were attending a revision course on the lectures "Analysis" and "Linear Algebra" in preparation for an oral examination. Just like the original lectures that students attended in their first year of study, the revision course was proof based. The study was conducted in the middle of the semester and a total of 60 students took part on a voluntary basis. All students were first presented with a proof video on the following statement: "Let $f: [0,1] \to [0,1]$ be a continuous function. Show that f has a fixed point, that is, there exists $x \in [0,1]$ with $f(x) = x$." The video presents the development of the proof of this statement step by step in about seven minutes, ending with a valid proof (Wirth et al., 2024). The students had up to 25 minutes to watch the video, stopping and rewinding in between. After watching the video, one group (n=34) was asked to write a short summary of the proof (with no word limit) before taking a multiple-choice test to assess their proof comprehension. A second group (n=26) completed the multiple-choice test first, then wrote a summary. The groups were assigned randomly. The multiple-choice test consisted of six items, each with four response options. Three of the items targeted the local level of proof comprehension, and three items addressed the holistic level. Both the summary and the multiple-choice test had to be completed without accessing the video. The students were informed in advance about the entire procedure. Their examination grade in "Analysis" was used as an indicator of prior knowledge (1.0-4.0), while the length of time spent working with the video (in minutes) was used as an indicator of engagement.

Data coding

A coding scheme was developed based on Davies and Jones' (2022) analysis of the aspects that influenced mathematicians' evaluations of proof summaries. This scheme assessed the proof summaries against five key elements: (1) correct paraphrase of the proving task (existence of a fixed point or an intersection of $f(x)$ and the identity), (2) turning the intersection problem into a zero-point problem, (3) consideration of the auxiliary function $g(x) = f(x) - x$ and justification of the continuity of the same, (4) consideration of the special cases $g(0) = 0$ and $g(1) = 1$, (5) applying the intermediate value theorem. Each key element was given a score of 0, 1 or 2, where 0 meant it was not mentioned, 1 meant it was mentioned superficially or as verbatim repetition, and 2 meant it was explained satisfactorily. For the multiple-choice test, each of the six items was coded using a partial credit system (2 – four correct answers, 1 – three correct answers, 0 – two or less correct answers).

Preliminary analyses

Because the multiple-choice test covered both local and holistic comprehension, the model fit was tested by comparing a one-dimensional model with a two-dimensional model. To this end, confirmatory factor analyses were conducted using the R package lavaan (Rosseel, 2012) and a robust ML estimator to account for the partial credit scores (Li, 2016). While the one-dimensional model showed satisfactory fit indices, $(\mathrm{X}^2(9) = 10.816, p = .283, CFI = .939, RMSEA = .056, SRMR = .067)$, the two-dimensional model did not fit the data well $(\mathrm{X}^2(8) = 12.117, p = .146, CFI = .876, RMSEA = .084, SRMR = .067)$. For this reason, this study, like other studies (Davies et al., 2020; Mejía Ramos et al., 2025), uses an overall rating scale to measure proof comprehension. The reliability ($\alpha = .61$) is sufficient and reflects the broad scope covered by the items included.

As both the quality of proof summaries and the test performance can be influenced by prior content knowledge, a randomisation check was performed. There was no significant difference in the final grades achieved by the two groups of students on the "Analysis" course $(M_{SMC} = 3.41, SD_{SMC} = 0.58, M_{MCS} = 3.36, SD_{MCS} = 0.74, T(56) = 0.24, p = .809)$. However, the group that wrote a summary first spent statistically significantly more time watching the proof video $(M_{SMC} = 10.47, SD_{SMC} = 3.04, M_{MCS} = 8.69, SD_{MCS} = 1.57, T(51.73) = 2.94, p = .005)$.

RESULTS

The descriptive results showed moderate average summary scores (maximum: 10) for both groups $(M_{SMC} = 4.84, SD_{SMC} = 2.48, M_{MCS} = 4.65, SD_{MCS} = 2.26, T(49) = 0.27, p = .785)$. The average scores in the multiple-choice test (maximum: 12) were slightly above the theoretical mean $(M_{SMC} = 7.06, SD_{SMC} = 2.98, M_{MCS} = 7.15, SD_{MCS} = 2.52, T(58) = -0.13, p = .897)$.

For RQ1, correlation analyses were conducted (see Table 1). The summary scores and the multiple-choice test scores showed no meaningful correlation, indicating that multiple-choice tests and proof summaries measure different constructs. This assumption is supported by considering students' prior knowledge. While the multiple-choice test scores correlated statistically significantly with the students' examination grade, this was not the case for the summary scores. Considering the order of procession, there was a very weak, non-significant positive correlation for the summary-MC group ($r_{SMC} = .08, p = .679$), and a weak non-significant negative correlation for the MC-summary group ($r_{MCS} = -.22, p = .345$). Although these results could be due to motivational issues and are not statistically significant, the order of processing could be a relevant factor when comparing different approaches to measuring proof comprehension.

Table 1: Correlation coefficients (Pearsons' r)

	Summary	MC test	Prior knowledge	Duration
Summary	-	-.02	.13	.13
MC test		-	-.29*	.09
Prior knowledge			-	-.12

Note: $n = 49$, $*p < .05$

For RQ2, three regression models were calculated, with the multiple-choice test score as the dependent variable (see Table 2). Model 1 included the students' group and prior knowledge as predictors. In this model, the prior knowledge was a statistically significant negative predictor ($\beta = -.29, p = .027$), indicating that students with a lower (=better) grade achieved a higher score in the proof comprehension test. Model 2 additionally contained the interaction between group and prior knowledge. The interaction effect was marginally statistically significant ($\beta = .24, p = .059$), indicating that students' prior knowledge moderated the effects of writing a summary on test scores. To control for the time on task, Model 3 included the time students spent watching the video. Including this variable had no effect on the results.

Table 2: Standardized regression coefficients

	Model 1	Model 2	Model 3
Group	-0.02	-0.01	0.01^{+}
Prior knowledge	-0.29*	-0.35*	-0.34*
Interaction (G x PK)		0.24^{+}	0.24^{+}
Duration			0.06
R^2 / R^2 adjusted	0.085 / 0.052	0.144 / 0.097	0.148 / 0.083

Note: $n = 58$, $*p < .05$, $^{+}p < .1$

DISCUSSION

This study investigated the relationship between proof summaries and multiple-choice comprehension tests, taking prior knowledge into account. The results replicate findings from Davies et al. (2020), indicating that proof summaries and multiple-choice tests are measuring different constructs. In particular, proof summaries may be less (or differently) related to prior knowledge. This provides new perspectives for measuring proof comprehension, as a high correlation could also indicate low discriminatory power between proof comprehension and prior knowledge. However, future studies should analyze the differences between the two approaches in more detail to accurately describe the constructs measured. Although the key elements of the coding scheme for proof summaries were explicitly addressed in the multiple-choice test, the multiple-choice test may cover a broader range of facets of proof comprehension. In addition, motivational issues or familiarity with the task may be considered, especially as the summary scores were quite low. The same applies to the reported effects of the order of the tasks. An unexpected negative correlation was observed when the multiple-choice test was performed first. The observed effects could also indicate that both preparing a summary and taking a comprehension test stimulate cognitive processes (King, 1992; Yang & Lin, 2012). This study partly confirms the effectiveness of the summarizing strategy, showing that students with a higher level of prior knowledge who wrote a summary first demonstrated a better proof comprehension. Similar to self-explanations, which have already been shown to support proof comprehension (Hodds et al., 2014), summarizing might therefore be worth investigating in more detail. However, as Demeke's (2025) findings suggested, students still need training in using this strategy. Yet, training on proof comprehension strategies should combine strategy training and content repetition to realise its full potential.

As only one proof was analysed in this study using a small sample, the results are not necessarily transferable to other proofs or target groups. In addition, the proof was presented in a video that also provided insight into the process of constructing the proof. Although the results of the proof comprehension test are comparable to those of other studies that use written proof, the medium and presentation may have influenced the outcome. Despite these limitations, the study provides new insight into the potential and limitations of proof summaries as a measurement instrument, as well as the conditions under which the summarizing strategy can effectively promote proof comprehension. Further analyses and intervention studies will have to follow.

References

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Davies, B., Alcock, L., & Jones, I. (2020). Comparative judgement, proof summaries and proof comprehension. *Educational Studies in Mathematics*, *105*(2), 181–197. https://doi.org/10.1007/s10649-020-09984-x

Davies, B., & Jones, I. (2022). Assessing proof reading comprehension using summaries. *Int. Journal of Research in Undergraduate Mathematics Education*, *8*(3), 469–489.

Demeke, E. (2025). Investigating strategies undergraduates and mathematicians use for enriched proof-reading experience. In K. Weber & M. Savić (Eds.), *Research in Mathematics Education. New Directions for Mathematics Education Research on Proving* (pp. 217–237). Springer. https://doi.org/10.1007/978-3-031-85004-2_10

Donoghue, G. M., & Hattie, J. A. C. (2021). A meta-analysis of ten learning techniques. *Frontiers in Education*, *6*, Article 581216. https://doi.org/10.3389/feduc.2021.581216

Hodds, M., Alcock, L., & Inglis, M. (2014). Self-explanation training improves proof comprehension. *Journal for Research in Mathematics Education*, *45*(1), 62–101.

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education*, *43*(4), 358–390. https://doi.org/10.5951/jresematheduc.43.4.0358

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal*, *29*(2), 303–323. https://doi.org/10.3102/00028312029002303

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data. Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Res. Methods*, *48*(3), 936–949.

Mejía Ramos, J. P., Weber, K., Lew, K., La Torre, J. de, & Ma, W. (2025). Student performance on proof comprehension tests in transition-to-proof courses. In K. Weber & M. Savić (Eds.), *Research in Mathematics Education. New Directions for Mathematics Education Research on Proving* (pp. 151–176). Springer.

Mejia-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. https://doi.org/10.1007/s10649-011-9349-7

Neuhaus, S., & Rach, S. (2019). Proof comprehension of undergraduate students and the relation to individual characteristics. In U. T. Jankvist, M. van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Educatio.* Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.

Rosseel, Y. (2012). lavaan : An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Samkoff, A., & Weber, K. (2015). Lessons learned from an instructional intervention on proof comprehension. *The Journal of Mathematical Behavior*, *39*, 28–50. https://doi.org/10.1016/j.jmathb.2015.05.002

Wirth, L., Kirsten, K., Serpé, C., & Greefrath, G. (2024). Comparing different types of instructional videos in a flipped proof-based classroom. *Frontiers in Education*, *9*, Article 1438355.

Yang, K.-L., & Lin, F.-L. (2012). Effects of reading-oriented tasks on students' reading comprehension of geometry proof. *Mathematics Ed. Research Journal*, *24*(2), 215–238.