Models for proof comprehension in secondary and tertiary education: Uniting the perspectives

VERENA SPRATTE, ROSTOCK

***Zusammenfassung**: In der Hochschuldidaktik ist das Beweisverständnis zuletzt zum Gegenstand eines eigenen, aufstrebenden Forschungszweiges geworden. Obwohl die dort genutzten Modelle weitgehend aufeinander aufbauen, zeigt dieser Artikel Inkonsistenzen in ihrer Betrachtung verschiedener Textebenen auf. Anschließend führt er die bisherigen Beiträge zu einem einheitlichen Modell des Beweisverständnisses zusammen. Dazu greift er auf aktuelle Modelle der kognitionspsychologischen Leseforschung zurück. Am Beispiel unterschiedlicher Lehr- und Lernziele zu Beweisen in Schule und Hochschule erweist sich die Anschlussfähigkeit und Adaptivität des einheitlichen Modells.*

***Abstract:** In tertiary education didactics, proof comprehension has recently established itself as a growing research field. Although the models used there largely build upon each other, this article highlights inconsistencies in the consideration of different text levels. It merges the previous contributions into a unified model of proof comprehension, integrating current models from cognitive-psychological reading research. Using the example of different teaching and learning goals concerning proofs in school and university, the connectivity and the adaptivity of the unified model are demonstrated.*

1. Introduction

Research on mathematics education at the tertiary level has been an emerging field of study for the last two decades. Its growth came along with an increasing interest in the teaching and learning of proof (Mejía-Ramos et al., 2012) as

> *Proofs are the heart of mathematics.* [...] What is proved today is true – today, tomorrow and in a thousand years. This distinguishes mathematics from all other sciences. (Grieser, 2018, p. 14, emphasis in original)

Deductive reasoning is an essential characteristic of the text genre of proof and the specific way to gain knowledge within mathematics (Davis et al., 2012; Selden & Selden, 2013). Thus, ever since proofs were systematically introduced to mathematics in the ancient Greek culture, learning mathematics includes the learning of proof (Reid & Knipping, 2010). How to introduce novices to the deductive \`heart of mathematics´ has been a central question to researchers and lecturers in mathematics education since then (Reid & Knipping, 2010). One possibility to foster proof competencies in educational settings is to analyze existing proofs given to the learners as models to represent the acceptable deductive structure, the use of precise language, and specific proving techniques (Conradie & Firth, 2000; Fischer & Malle, 1985; Weber, 2012).

Despite the long tradition of presenting proofs in mathematical courses – be it oral or written, at the secondary or tertiary level – research on students' perception of those proofs is still in its infancy. In 2009, Mejía-Ramos and Inglis pointed out and in 2015, Sommerhoff et al. confirmed that the presentation and the reception of proofs are still underrepresented areas of research compared to proof production. Their findings initiated several studies focussing on the reading of proof (e.g., Mejía-Ramos et al., 2012; Mejía-Ramos & Weber, 2014; Neuhaus-Eckhardt, 2022; Panse et al., 2018; Spratte, 2022).

In their literature review, Mejía-Ramos and Inglis (2009) considered proof validation, proof evaluation, and proof comprehension as three different reading goals. By now, their distinction is well established in mathematics education research on the reading of proof (Selden & Selden, 2015). As the proofs presented in learning contexts serve as models for the genre of proof, assuming their correctness and educational value is a reasonable learners' habit (G. Harel & Sowder, 1998). Consequently, research in proof comprehension gained significance, resulting e.g. in a significant number of recent dissertation projects related to proof comprehension (for example Davies, 2020; Hodds, 2014; Neuhaus-Eckhardt, 2022).

Yet the scope of adequate proof comprehension is a topic of ongoing discussion (Neuhaus-Eckhardt, 2022; Zazkis & Zazkis, 2015). A question closely related is about the role proofs are supposed to play in schools, in teacher education, or in mathematical service lectures for students of other (mainly scientific or economic) disciplines[1]. This topic has often been discussed, but remains current (Neumann et al., 2017). What kind of comprehension shall students in different educational stages and settings gain when reading a certain proof?

For students of university mathematics, Neuhaus-Eckhardt and Rach (2023) emphasized the importance of understanding the general ideas and methods of a proof to construct similar proofs later on. Meanwhile, Bauer and Skill (2020) promoted the explanatory power of proof in university mathematics for non-mathematicians regarding the relation between calculatory methods and the underlying models. Reading similar proofs, students from both groups seem to be well-advised to focus on different aspects of the text, resulting probably in varying reading behaviors (Mejía-Ramos & Weber, 2014).

So far, researchers in mathematics education mostly defined proof comprehension operationally using lists of indicational tasks or questions created from the normative point of view of mathematics professors and educators (e.g., Mejía-Ramos et al., 2012; Yang & Lin, 2008). These approaches to the comprehension-oriented reading of proofs followed a psychometric research tradition: They essentially treated the reading process as a black box, abstracting from its details and focusing on the resulting proof comprehension (Schnotz & Dutke, 2004). Doing so, they tried, for example, to identify individual characteristics that facilitate proof comprehension (Neuhaus-Eckhardt, 2022), to describe the type of understanding gained by specific groups of readers (Lin & Yang, 2007), or to evaluate support measures (Hodds, 2014).

Besides the psychometric reading research, there is a second main tradition in research on general reading known as the cognitive-psychological approach. It analyzes how readers extract information from the text and build up a coherent mental text model. While psychometric reading research identifies individual prerequisites for successful reading, cognitive-psychological reading research aims at explaining how these prerequisites enter the reading process and result in varying degrees and kinds of reading success. Its focus lays on structures and processes that in combination build the comprehension of a given text. Thus, a well-founded cognitive-psychological model of reading is necessary to ensure the validity of psychometrical tests for reading comprehension (Schnotz & Dutke, 2004). Nevertheless, the current discourse on proof comprehension rarely takes the more recent cognitive-psychological findings of reading research into account, even though numerous points of contact exist (Neuhaus-Eckhardt, 2022).

The main interest throughout this paper will be to revise the existing operational indicators and models of proof comprehension from a cognitive-psychological perspective (Chapter 2). I will focus on three conceptualizations that rely on each other: The model of reading comprehension for geometric proof by Yang and Lin (2008), the assessment model of proof comprehension presented by Mejía-Ramos et al. (2012), and the recent dissertation project of Neuhaus-Eckhardt (2022). A detailed comparison in Sections 3.1 and 3.2 results in a comprehensive model for the reading of proofs that applies to different institutional levels and distinct reading goals in Section 3.3. It reveals that in the current research on tertiary education, proof comprehension is narrowed to the epistemic function of proof (Rav, 1999). In secondary school contexts, a wider notion of reading proofs is inherent in the models. These (at first glance probably unintuitive) findings are discussed in Chapter 4 in the light of different educational purposes in schools and universities. Chapter 4 further elaborates on the additional value the unified model provides for both teaching and research concerning the reading of proof. It shows how the model might enlighten discussions about reading processes that need to accomplish research on the resulting comprehension. Section 4.2 discusses how the model relates to the roles of proof in secondary education, university mathematics, and university mathematics for non-mathematicians, while Section 4.3 discusses its limitations and related open questions.

2. Proof comprehension as a specific reading comprehension: The state-of-the-art

Despite the significance of proofs for mathematics as a scientific discipline, the notion of \`proof´ is not as clear to mathematicians and researchers in mathematics education as one might hope for (Czocher & Weber, 2020; Reid & Knipping, 2010). The popular definition of Stylianides (2007) refers to proof as "a connected sequence of assertions for or against a mathematical claim" (p. 291) that relies on knowledge, argumentative modes, and forms of expression acceptable in the (classroom) community. According to this definition, the rules a proof has to follow are essentially shaped by the mathematical (sub-)community the proof addresses.

Indeed, the socio-mathematical norms concerning proofs varied and still vary over time and between different sub-fields of mathematics (Reid & Knipping, 2010). Weber and Czocher (2019) showed that even among professional mathematicians there are substantially different judgments concerning a given specific text that claims to be a proof. Especially, the amount of necessary or dispensable detail gives rise to inconsistent judgments on a given prooftext (Inglis et al., 2013). Czocher and Weber (2020) go as far as to define proof as a cluster category, for which no fixed set of criteria constitutes membership.

Despite all dissent, there is a kind of consensus among mathematicians about *prototypical proofs* (Czocher & Weber, 2020): Proofs verify a mathematical statement using deductive logic strict enough such that all gaps might be closed in principal. They convince mathematicians and show them why a theorem is true. Finally, proofs need to be accepted by the mathematical community they address. Selden and Selden (2013), who restrict their analysis to the written representation of such prototypical proofs, consider these and similar proofs' characteristics as those of a text genre. This paper follows their approach and defines a proof as a written version of a prototypical proof. This is in line with what Reid and Knipping (2010) call *proof texts*.

Throughout, I will consider reading a proof first of all as reading any text. Surprisingly, this is not the way many researchers in mathematics education approach this topic. Going back to the influential literature review by Mejía-Ramos and Inglis (2009) already discussed in the introduction, there is a now well-established distinction between three kinds of reading a proof:

1) If the reader aims at judging the proof in terms of its correctness, his or her reading is considered as *proof validation* (Selden & Selden, 2015).
2) Using any criterion other than correctness to assess the given text (e.g., beauty, innovation, or explanatory power) is called *proof evaluation* (Selden & Selden, 2015).
3) If the rationale in reading is not a statement about the proof, but to gain some insight from reading the proof, we talk about *proof comprehension* (Mejía-Ramos et al., 2012).

From the beginning of research on students' reading of proof until the 2010s, proof validation was the most important criterion to measure students' reading success in both secondary and tertiary mathematics education (Davies, 2020; Weber, 2012). The findings were alarming: Students from both secondary and tertiary levels considered the symbolic representation of mathematics as a central indication for the validity of proofs. They accepted flawed deductive arguments or examples as proofs, sometimes performing at chance rate in their validity judgments (G. Harel & Sowder, 1998; Panse et al., 2018; Reid & Knipping, 2010, pp. 59-72).

These poor results may at least partly be caused by the non-assessment of students' proof comprehension and proof validation in typical mathematics courses: The expected assessment strongly guides students' learning process (Conradie & Firth, 2000). If proofs play a minor role in the course and remain unassessed at all, or if only proof construction is required in final examinations (as it has traditionally been in proof-related mathematics courses at all stages), it will not be the students' focus to develop substantial proof comprehension (Bauer & Skill, 2020; Davies, 2020; Mejía-Ramos et al., 2012).

It has been the urge to improve assessment in proof-oriented mathematics classes that moved Conradie and Firth (2000) to present a first proof comprehension test to replace reproductive tasks (‘State and prove Theorem X’) in written exams. Their suggestions were refined first by Yang and Lin (2008) for the secondary and later by Mejía-Ramos et al. (2012) and Neuhaus-Eckhardt (2022) for the tertiary level, all building upon the previous contributions. Within this Chapter, I will trace the development from Yang and Lin (2008) via Mejía-Ramos et al. (2012) to Neuhaus-Eckhardt (2022) with special emphasis on the cognitive-psychological potential of each conceptualization. While Yang's and Lin's pioneering work offered some kind of competence model for proof comprehension[2], the most popular contribution by Mejía-Ramos et al. (2012) restricted its model to a systematic list of operational indicators. Both Yang and Lin (2008) and Mejía-Ramos et al. (2012) still focussed on the assessment of proof comprehension in teaching situations, whereas Neuhaus-Eckhardt (2022) very recently presented the first model for proof comprehension as a latent construct itself.

2.1 Yang's and Lin's Model for Reading Comprehension of Geometric Proof (RCGP)

In this section, I will present the model for Reading Comprehension of Geometric Proof (RCGP) by Yang and Lin (2008) in line with its close relations to modern reading research. Similar to most current approaches in general text comprehension, Yang and Lin (2008) considered proof comprehension as the product of a cyclic process, in which the reader includes new information from the text into his pre-information and thereby constantly (re)forms a mental representation of the proof (Lenhard et al., 2017; OECD, 2019). Doing so, Yang and Lin (2008) implicitly built their model on a cognitive-psychological basis, even though they did not explicitly relate their considerations to modern models or results from reading research. Instead, they referred to Duval (1998) who distinguished a micro, local, and global level in which information is organized when working on geometry problems. The underlying literature on cognitive-psychological reading research might be unfamiliar to readers with more mathematical backgrounds. However, it is essential in order to capture the particular strengths of the RCGP model. Thus, I will present it a little more in detail.

One of the most popular models for reading comprehension is the *Construction-Integration-Model* introduced by van Dijk and Kintsch (1983). It describes comprehending a text as the formation of a mental text representation and distinguishes between the *text surface*, the *text base*, and the *situation model* (van Dijk & Kintsch, 1983). This influential distinction might be illustrated using the example of the daily weather forecast in a local newspaper: After reading it once in the morning, one will most likely be able to reproduce the main information when asked for it. One might also remember the structure of the text, for example the order in which the information was presented. Both are part of the *text base*, a mental representation of the propositions' semantic content (Kintsch, 1998). What one will most likely not remember is the *text surface*, i.e. the exact wording the information was presented in (van Dijk & Kintsch, 1983). As soon as the information from different sections is related to each other ("Considering the wind and the lot of rain in the west, tomorrow might be rainy...") or to information not given in the text ("... and my umbrella got lost, so I better buy one today."), integration of the text base into a coherent and adequate mental representation (the *situation model*) takes place (Kintsch, 1998).

Depending on the prior information and the purpose of the reader, the situation model will be of different scope and type. Superfluous information might be dropped, and implicitly related information be included. While the text base relates information within sentences or very short passages to reproduce the main content, the situation model connects information from the whole text to form a globally coherent representation (Kintsch, 1998). Thus, the integration of information takes place on different text levels, which are usually assumed to be structured hierarchically in terms of the amount of information under consideration (Lenhard et al., 2017):

On a *micro level*, single words and symbols are analyzed by the reader. Decoding symbols and single words to grasp their semantic content is a prerequisite to constructing coherence within small text segments, i.e. to build a text base (OECD, 2019). Micro-level processes thus lay the basis of all reading processes. Note that this linguistic notion of micro-level resembles Duval's (1998) micro-level as both contain only one piece of information and do not establish any context yet.

Referring to the deductive structure of proof, Yang and Lin (2008) now claim that proof comprehension requires two such micro-level processes: On the one hand, the content needs to be decoded *semantically*. In line with van Dijk and Kintsch (1983), they call the corresponding semantic micro-level the *surface level* of reading comprehension. On the other hand, the same information may take different *logical* roles in a proof, for example, when a statement is first proved and later built on. Note that the argumentation steps are not yet reconstructed on the micro level. Rather, signal words such as "assumption", "implies" or conjunctions give preceding or following information a certain logical status. To identify this status, it is not necessary to decode the information itself. Thus, there exists a logic micro-level called the *level of recognizing elements* besides the semantic surface level discussed in reading research (Lenhard et al., 2017).

In line with (but without explicit reference to) the famous Toulmin-Scheme for the analysis of arguments, Yang and Lin (2008) consider the logic micro-level as the foundation to construct *local* coherence within one argumentation step in the proof (Toulmin, 2003). According to Toulmin, information considered as an evident fact operates as data that underpins the claim, whereas a warrant clarifies how the claim is supported by the data. Toulmin further considers backings that support the warrant, modal qualifiers that clarify the degree of conviction for the claim, and rebuttals to state exceptional cases.

Analyzing proofs that operate with strictly deductive conclusions, researchers in mathematics education often reduce the Toulmin-Scheme to data, warrant, and conclusion only (Inglis et al., 2007). Accordingly, Yang and Lin (2008) consider chaining of “premises, properties and conclusions in this proof" (p. 62), if necessary with reference to essential figures, as the characteristic operation on the local text level of their RCGP model.[3] Therefore, this level is called the *level of chaining elements*.

On the three levels described so far, coherence was constructed within sentences to grasp their meaning including the presented single argumentative steps. This resembles Kintsch's notion of the text base. In contrast, to construct *global* coherence the reader integrates the different steps into a situational model of the text, here the (geometric) proof. According to Yang and Lin (2008), this requires attention to both the semantic and the logical value of information. Only on this global level, the reader might for example grasp the importance of different assumptions for the theorem and distinguish them from similar assumptions of related theorems.

Altogether, Yang and Lin (2008) distinguish four levels, namely the

1) surface level (semantic micro-level),
2) recognizing elements (logic micro-level),
3) chaining elements (logic local level), and
4) encapsulation of the proof (semantic and logic global level),

that form the basis of their RCGP model. Analogously to text levels in general reading comprehension, the levels are not considered in a strict hierarchy: As argued above, the semantic and the logic micro-level are considered to be independent of each other. Global coherence can (at least partly) be constructed even if certain local steps are not fully understood. Further, contextual information from the global level may be used to reconstruct the meaning of an unknown term (Lenhard et al., 2017) on the semantic micro level. Even though proofs usually rely at least partly on dense symbolic language and offer little redundancy[4] (Österholm, 2008; Selden & Selden, 2013), the reader might refer to the context to approach unknown terms or notation, for example when investigating the role of a symbol within a formula.

One might have noticed that, so far, the RCGP model is not very specific to mathematical proof. Even though Yang and Lin include a logic micro-level, this should not be considered a unique feature of proof. As well as the Toulmin-Scheme, this text level could be applied to most arguments from many argumentative fields (Toulmin, 2003). Instead, the four levels offer the opportunity to consider proof comprehension in a still quite general framework based on reading and argumentation research.

From an extensive literature review on (sometimes geometry-specific) proofs and from interviews with mathematicians and mathematics educators, Yang and Lin (2008) collected 19 so-called *contents* of what it means to understand a proof to specify the levels concerning proof comprehension. Doing so, they provided an important step towards operationalizing proof comprehension.[5] Yang and Lin (2008) structured the 19 contents into six *facets*:

1) *Basic knowledge* includes knowledge of symbols, terms, figures, and the proof's methodology.
2) The *Logical status* refers to recognizing the logical value of statements within the proof or the proven theorem and checking proof steps for their deductive correctness.
3) *Integration / Summarization* takes place when the proof is structured into lemmata and the main idea of the proof is identified. It also includes making implicit warrants explicit, probably adding information that is not given within the text.
4) *Generality* is constituted by checking the validity of the proof or the proven statement, which may (but does not necessarily) lead to a judgment of the theorem's epistemic value as generally true or false.
5) *Application / Extension* expands the proven statement by adding other well-established statements that allow a generalization or specification. Another kind of application is to use the proof method in a different context, to compare different proof methods, or to apply the statement or its proof, e.g. to construct examples.
6) *Appreciation / Evaluation* refers to the perception of beauty or the mathematical and cultural value of the proof or the proven statement.

In line with their theoretical perception of text levels, Yang and Lin (2008) placed the six facets between the four levels as shown in Figure 1.[3] To measure secondary school students' comprehension of a geometric[6] proof, they operationalized the facets (note: not the levels) by means of the concrete proof text. Students' answers to the resulting comprehension questions specifying a certain facet indicate their text comprehension of the corresponding next-lower level: Students who give correct answers to questions on Logical Status are supposed to have gained an understanding of the level of recognizing elements etc. An exception is the facet of Appreciation / Evaluation, which was not operationalized due to its subjective character that opposes objective assessment.

As Yang and Lin (2008) pointed out, the facets resemble Bloom's famous taxonomy of cognitive educational goals. This increased their confidence in the integrity of the RCGP as well as it strengthens the RCGP's hierarchic tendency. According to Yang and Lin (2008), a multidimensional scaling analysis indeed confirmed to group the facets as shown in Figure 1. Yet, the methodology chosen by the authors does not allow conclusions about the nature of the separating layers. Therefore, their correspondence to the text levels still needs to be considered as a theoretical assumption.

By introducing the RCGP model, Yang and Lin (2008) innovated research in the reading of proof in several ways: They were the first to consider different text levels of coherence and gave the first extensive list of operational indicators for the successful reading of proof. Further, their work made secondary school students' geometric[6] proof comprehension measurable and thereby initiated research about the role of individual characteristics and about the development of proof-reading skills (e.g., Lin & Yang, 2007).

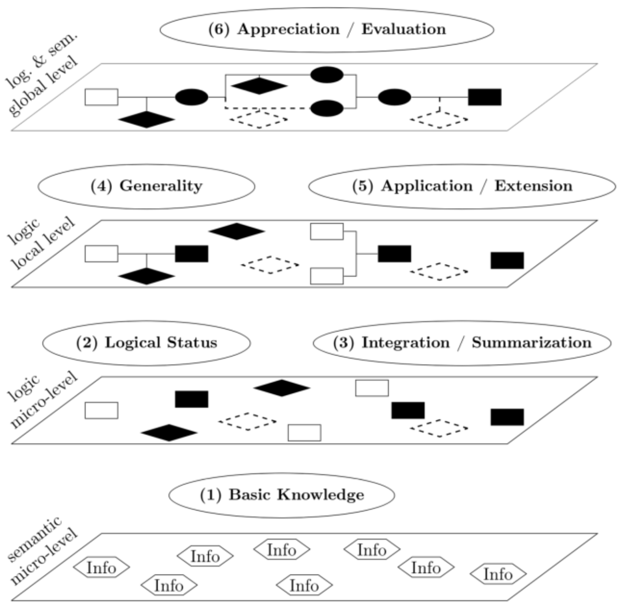

Fig. 1: Reading Comprehension of Geometric Proof by Yang and Lin (2008). Visualization of the Toulmin-Scheme within text levels adapted from Knipping and Reid (2019).[3]

Unfortunately, the RCGP model is mostly referred to in terms of its levels only, denying its original complexity and the huge effort to establish a relation between the independently developed facets and levels (Hodds, 2014; Mejía-Ramos et al., 2012; Zazkis & Zazkis, 2015).

2.2 The Assessment Model for Proof Comprehension in Undergraduate Mathematics (AMPC)

Even though the RCGP model was developed for the context of secondary school geometry classes, Mejía-Ramos et al. (2012) claimed its potential to be expanded to a general Assessment Model for Proof Comprehension (AMPC) in undergraduate mathematics. Recall that the levels presented in the RCGP model fit the current cognitive-psychological reading research well (as shown in Section 2.1) and that this branch of research claims to describe reading processes for readers of different expertise, age, and in various kinds of reading situations (OECD, 2019). Thus, it seems reasonable that the RCGP levels apply to the reading of proofs in tertiary education as well.

According to Mejía-Ramos et al. (2012), the rising complexity and length of the proofs used in university mathematics courses[1] increase the relevance of proof comprehension on a global text level (Selden, 2012). Thus, the authors developed an operationalization of the encapsulation level, which remained unassessed in the RCGP. Leaving the facets unmentioned, Mejía-Ramos et al. (2012) referred to the first three levels of RCGP as different dimensions of *local proof comprehension*:

1) The *Meaning of terms and statements* corresponds to the RCGP surface level. It is operationalized by stating definitions or examples for given terms in the proof and trivial reformulations and implications of statements within the proof.
2) The dimension of the *Logical status of statements and proof framework* requires the identification of the top-level logical structure, like proof by induction or contradiction, and the recognition of the logical status of statements within this framework. It refers to the RCGP level of recognizing elements.
3) The *Justification of claims* resembles the RCGP level of chaining elements and includes identifying all claims supported by some given data or vice-versa all data supporting a given claim, and making implicit warrants explicit.

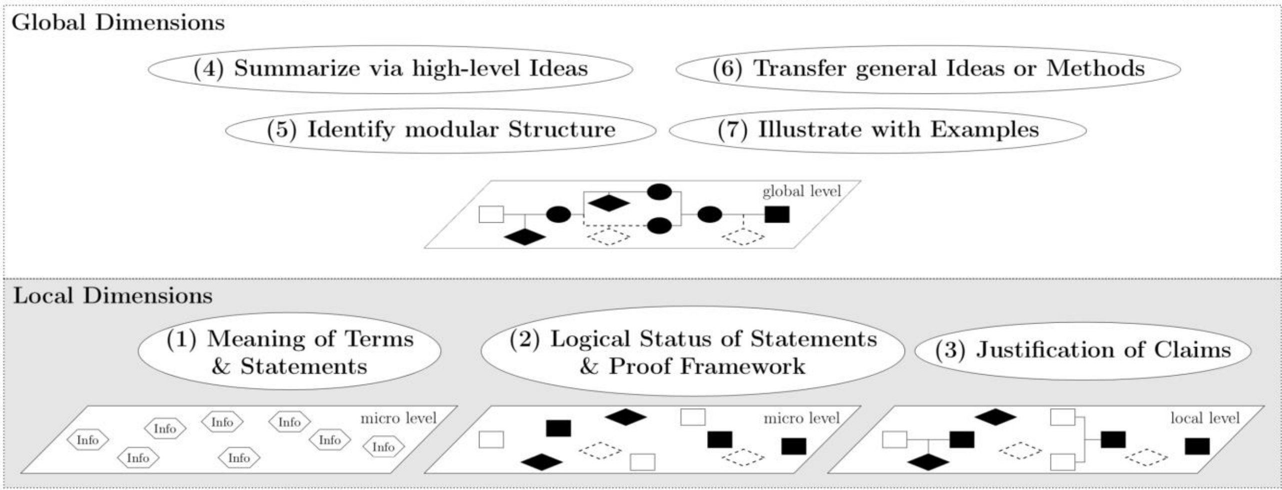

Fig. 2: Assessment Model of Proof Comprehension by Mejía-Ramos et al. (2012) as based on the model for RCGP.

Note that the AMPC presumes the correspondence between the RCGP's levels and facets on the surface level, as the first dimension of AMPC closely relates to the first facet of RCGP. While the second AMPC dimension resembles both the second RCGP facet and level, the third AMPC dimension shows a better fit to the third RCGP level than to any of the facets. In particular, Mejía-Ramos et al. (2012) did not consider aspects from the RCGP's facet of Integration and Summarization as indicators for local proof comprehension. Further, they used the term *local* in a broader way than usual for reading research by subsuming the micro-level(s) into the local dimensions.

As an expansion of the encapsulation level, Mejía-Ramos et al. (2012) presented four different *global dimensions of proof comprehension*, based on another literature review and interviews with university mathematicians about their reasons for reading and presenting proofs and their notion of understanding proof.

4) *Summarizing* the whole proof or a sub-proof *via high-level ideas* highlights – in contrast to the logical proof framework – the content-related core ideas that might well be specific to the mathematical domain the proof is placed in.

5) To partition the proof into modules and to identify their purpose and the logical relations between each other belongs to the dimension *Identifying the modular structure*.

6) Identifying and transferring the proof method to another proving task is summarized in the dimension of *Transferring the general ideas or methods to another context*. It resembles the RCGP facet of Application / Extension.

7) *Illustrating* a sequence of inferences (and not only a term as in the first dimension) *with examples* or a diagram constitutes the last dimension.

The methodology used by Mejía-Ramos et al. (2012) to derive the new dimensions is very similar to the RCGP's development of facets. It seems legitimate to interpret the global AMPC dimensions in analogy to the RCGP's sixth facet. This leads to a visualization of the AMPC in Figure 2 that adheres closely to Figure 1.[3] It reveals a good fit between the cognitive-psychologically based text levels and the AMPC dimensions, even though these relations are not discussed by Mejía-Ramos et al. (2012).

As Yang and Lin (2008) did for their RCGP model, Mejía-Ramos et al. (2012) presented both the local and the global comprehension as distinct, but not hierarchical. In the same way, the authors did not claim that the seven dimensions cover all aspects of proof comprehension or that they were empirically separable. Their aim was not to build a competence model but to allow a suitable assessment of proof comprehension. Their carefully designed proof comprehension tests should enhance teaching quality for undergraduate mathematics and help to evaluate the impact of teaching experiments. Therefore, and in the light of the ambiguous labeling of Integration or Summarization as local in the RCGP and global in the AMPC, it is not surprising that all attempts to empirically separate local and global comprehension known to the author of this paper so far have failed (e.g., Neuhaus-Eckhardt, 2022 and an unpublished one by the author herself).

In the decade after being published, the AMPC received a lot of attention among mathematics education researchers. Compared to the RCGP model with levels and facets, the AMPC has a slim, thus attractive design. Its strength lies in its broad applicability to various mathematical domains[2] of undergraduate mathematics and to various contexts: While Mejía-Ramos et al. (2017) gave detailed instructions to create proof comprehension tests on the highest standard for research and high-stakes examinations, Bickerton and Sangwin (2021) offered a pragmatic manual to construct automatically assessed proof comprehension tasks in line with the AMPC, for example for weekly exercise classes.

2.3 Components of proof comprehension as a situative text model

As seen in Section 2.2, the AMPC deviated from the implicit cognitive-psychological roots of the RCGP and focused on the operationalization of global reading comprehension (the RCGP's encapsulation level) for mathematical proofs. In her dissertation, Neuhaus-Eckhardt (2022) recombined the AMPC's focus on proofs in university mathematics courses[1] with cognitive-psychological models from general reading research.

Referring back to Kintsch’s notion of situative mental text models, Neuhaus-Eckhardt (2022) defined proof comprehension as a coherent mental model of a proof built after reading the text. She distinguished between a theoretical mental model offered by the text and an individual mental model developed by a specific reader in a certain reading situation. To specify the components such a mental model may[7] consist of, Neuhaus-Eckhardt first referred back to the operational indicators given in the RCGP and the AMCP, enriching them with contributions by Zazkis & Zazkis (2015), Pracht (1979), and others. In a second step, she characterized what a mental representation of the proof needs to contain to successfully respond to these indicators.

In the first step, Neuhaus-Eckhardt (2022) collected and sorted the different indicators for proof comprehension according to three levels of coherence. Doing so, she adopted the AMPC's wider notion of local comprehension as presented in Section 2.2 which incorporates the logic and semantic micro-levels. This results from their shared focus on university students in mathematical programs and the role and characteristics of proofs in these courses (as has been discussed in Section 2.2 and will be explored in more depth in Section 4.2.2).

In line with Bickerton and Sangwin (2021), Neuhaus-Eckhardt reasonably noted that the dimensions labeled as global by Mejía-Ramos et al. (2012) actually operate on different text levels: The sixth dimension, namely to transfer the general ideas or methods to another context, leaves the scope of the proof itself. It requires capturing the scope and possible limitations of proving methods and ideas, which can be seen as a form of proof evaluation according to usefulness or innovative potential. Transferring proof methods further necessarily includes some kind of proof production, even if in a reduced form as suggested by Mejía-Ramos et al. (2012).

Following Neuhaus-Eckhardt (2022), I will call this new text level the *level beyond the text*. Note that the remaining three global dimensions Mejía-Ramos et al. (2012) propose do indeed operate on the global level as defined in Section 2.1 and do not go beyond the text. Neuhaus-Eckhardt's collection of operational indicators for proof comprehension thus distinguishes between local, global, and beyond-the-text elements as shown in Figure 3.[3]

Neuhaus-Eckhardt (2022) motivated the distinction of the level beyond the text from the global level with the common practice among mathematicians to read proofs for good proving ideas and to present proofs to teach their university mathematics students how to prove (Mejía-Ramos & Weber, 2014; Weber, 2012). Neuhaus-Eckhardt herself pointed out that including aspects at least very close to proof construction in a model of proof comprehension is a disputable question (Neuhaus-Eckhardt, 2022). Whether it is appropriate or not depends on how close the constructed proof is to the given original, as this determines the relevance of text-external information.

General reading research discusses the importance of information and applications external to the text as well. Lenhard et al. (2017) presented a reading comprehension test including questions that can only be answered by deducing an instance that is itself not directly described in the text. The PISA reading framework lists the assessment of the quality and credibility as well as reflection on content and form as advanced processes of reading comprehension (OECD, 2019). In line with this, proof methods including their scopes and limitations have a huge impact on mathematicians' judgment about a proof's quality (Weber & Mejía-Ramos, 2011).

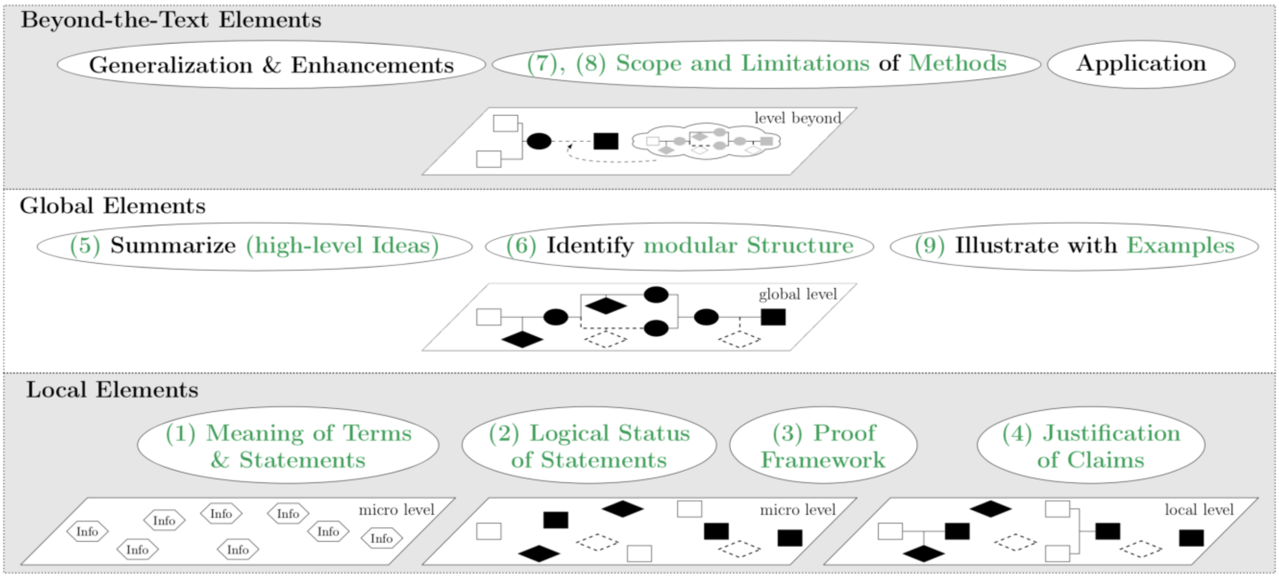

Fig. 3: Operational indicators for proof comprehension according to Neuhaus-Eckhardt (2022) as based on the AMPC. Written in green are the possible components of proof comprehension as a mental model.

Yet in the OECD's definition of reading literacy, "doing something with what we read" (2019, p. 28), is not a part of understanding the text, but placed on the same rank. This includes using the text to create something new such as a new proof. Text-generating actions are one of those aspects that distinguish reading literacy from mere text comprehension.[8]

Neuhaus-Eckhardt (2022) reacted to these nuances in her second step, where she synthesized the operational indicators of proof comprehension to nine possible[7] components of proof comprehension as a mental model. To that end, she excluded explicit text production such as summaries or generalizations as well as strategic knowledge but incorporated the proof method including its scope and limitations. She further detached the components from the text levels, arguing that the distinction between the local and the global level becomes meaningless as the situation model does not preserve the text order. Nevertheless, as the following consideration of the nine components will show, there is still variety in the amount of information to be considered.

Four of the nine components arose from the AMPC's local dimensions. As already seen in Figure 3, Neuhaus-Eckhardt split the proof framework from the logical status of statements, resulting in the four components

1) *Meaning of terms and statements* that refers to definitions and visualizations of terms in the proof and the semantic meaning and trivial consequences of statements, including the proven statement itself,
2) *Logical status of statements* within the proof, especially to distinguish assumptions from claims,
3) *Proof framework* that includes the proof method such as (in)direct proof or proof by induction, the relation between single statements and the claim to be proven and the purpose of a single statement for the proof, and
4) *Justification of claims* where single steps of the proof are reconstructed and proof gaps are explained, including identifying all statements supported by a specific statement and justifying the completeness of case distinctions.

Following Kintsch (1998), these local components are not parts of a mental text model. They rather belong to the text base. Neuhaus-Eckhardt (2022) nevertheless included them in her theoretical model of proof comprehension, arguing that the mental model emerges from the text base and enriches it with global coherence without losing those components necessary to understand the proof.

As parts of a theoretical mental model of the proof in a strict sense, i.e. related to the global text level, proof comprehension according to Neuhaus-Eckhardt (2022) contains (if suitable)[7]

5) the *main idea(s) of the proof* that allows to summarize the given text,

6) the *proof's modular structure* including the connection between different modules and their purpose for the complete proof, but not the order in which they were originally presented, and
7) the *proof method* used within the proof that is intended by the lecturers or readers to be transferred to similar proofs. Note that this transferable proof method could differ from the main idea and the proof framework, e.g. if it contains a specific formulation of a definition that is especially useful in certain proof settings.[9]

The proof method already points at the level beyond the text, where it is motivated from. Still, it does not leave the scope of the proof itself, whereas knowledge of the

8) *scope and limits of the proof method* requires experience with multiple proofs, so clearly goes beyond the global text level.

The last component presented in Neuhaus-Eckhardt (2022) does not clearly relate to one of the text levels already discussed. The author referred to the illustration of several proof steps with examples and diagrams required in the seventh AMPC component and Zazkis' and Zazkis' (2015) suggestion to include a similar, but local visual illustration of notions. Therefore, she claimed that

9) *useful examples and visualizations* for notions or for techniques used within the proof should be incorporated into sound proof comprehension. Similar to the main ideas that form the basis of sustainable proof summaries – which may serve as indicators of proof comprehension, but are not themselves part of the mental model of the proof (Davies, 2020) – the knowledge of useful illustrations is a prerequisite for some operational indicators.

At first sight, the differences between the operational indicators for and the components of proof comprehension seem rather insignificant. But Neuhaus-Eckhardt's (2022) detailed analysis sharpened the awareness for the huge borderland between proof production and proof comprehension in everyday teaching and learning of proof in university mathematics. Her notion of the level beyond the text might help to structure this grey area, even though its practical value still needs to be proved empirically. Considering the rising interest in new ideas for the teaching of proof, such as faded worked examples or spelling out the proof for a specific example (Bickerton & Sangwin, 2021; Kempen, 2018), the interplay of proof comprehension and proof production might gain even more attention in the future (Neuhaus-Eckhardt & Rach, 2023).

3. Comparing, reflecting, and uniting the models of proof comprehension

So far, I have shown the successive development of models for proof comprehension from Yang and Lin (2008) via Mejía-Ramos et al. (2012) to Neuhaus-Eckhardt (2022). Yang's and Lin's original consideration closely related to different scopes of information considered during the reading process, thus offering lots of links to cognitive-psychological reading research. Focussing on assessment in university mathematics, Mejía-Ramos et al. (2012) stressed the importance of global proof comprehension while merging micro and local text levels. Neuhaus-Eckhardt (2022) carried this development further by introducing a level beyond the text that draws attention to the borderland between proof comprehension and proof production, which is of great importance, especially among university mathematicians.

Altogether, the models offer five different ranges of information to be considered while constructing a coherent situation model of the text: a logical and a semantical micro level, a local and a global level of coherence, and a level-beyond-the-text to establish connections to related or similar proofs. In light of all five text levels, I will now revise the different models presented in Chapter 2 and show inconsistencies in the connections between text levels and operational indicators for proof comprehension. This investigation will lead to a unification of the models in Section 3.3.

3.1 A closer look at the text levels in the models of proof comprehension

I will start the comparison with the two micro-levels introduced by Yang and Lin (2008). Both the AMPC and Neuhaus-Eckhardt's model adopted the surface level from the RCGP model. On this semantic micro-level, readers grasp the meaning of terms and statements within the proof. Among all three models, there is also consensus about the logic micro-level that is expressed in recognizing the logical status of a statement within the proof.

The RCGP model measured understanding on this logic micro-level via the second and third RCGP facet. Recall that the second facet included checking the correctness of a single argumentative step. In contrast, the other models see such a check as an indicator for the local level (not for a micro-level). As already discussed in Section 2.2, the AMPC further refused Yang's and Lin's allocation of the third, fourth, and fifth facet, which motivates a more detailed analysis.

While Mejía-Ramos et al. (2012) considered summarizing the proof according to its main idea a global activity, Yang and Lin (2008) placed it as part of Integration / Summarization between the second and third level, i.e. even below the level of chaining elements (local coherence). This difference could be caused by the characteristics of proofs used in secondary schools, which are usually shorter than those on the tertiary level and have a less complex, more often than not chained argumentative structure (Selden, 2012). To extract the main idea of such a short, linear proof might sometimes be the same as understanding the one central argumentative step, thus being a matter of local coherence. Nevertheless, in line with the PISA reading framework, I argue that identifying which of the (few) chained steps is central is a matter of global coherence: OECD (2019) points out that identifying central ideas or summarizing longer or more complex passages (such as proofs used in university mathematics) requires integration of the text base into a situation model, thus is a matter of global coherence. Thus, just as Toulmin (2003) suggested, the line between local and global coherence might best be drawn when more than one argumentative step comes into play – implying that a local and a global coherence level can be distinguished even for most proofs used on the secondary level (Brunner, 2014).

The position of summarizing below the local level is only one example of vagueness within the RCGP model regarding the fit of the facets to the levels they measure. These ambiguities can at least partly be explained by the fact that the RCGP model lacks a level of global coherence: Even though Yang and Lin (2008) themselves considered the encapsulation level as one of global coherence, it inherently goes beyond the text.

> The characteristic of the top level is to interiorize this proposition and its proof as a whole, where one can apply this proposition and its proof, and distinguish different premises from similar propositions. (Yang & Lin, 2008, p. 63)

Here, Yang and Lin (2008) indicated that an encapsulated proof comprehension allows comparisons between similar statements with slightly different premises and applications of the proposition and its proof, e.g. to further proof constructions. A global level that stays inside the scope of the proof to be read is missing in the RCGP model. This may again be caused by characteristics of proofs used on the secondary level.

Revising the RCGP facets in the light of all five text levels, some aspects of single argumentative steps from both the second and the third facet require proof comprehension on a level of local coherence. The AMPC and Neuhaus-Eckhard combined these aspects in their dimension `justification of claims´, which seems more appropriate for the assessment of how readers establish local coherence than the RCGP facets of Generality and Application / Extension. Throughout, in case of ambiguity, I will follow the AMPC's structure in questions regarding the line between local and global coherence levels.[10]

Just as the RCGP model lacked attention to questions concerning global coherence, both the AMPC and its expansion by Neuhaus-Eckhardt diminished the importance of the micro-levels by subsuming them as local aspects (see Section 2.2). Consequently, Yang's and Lin's distinction between semantic and logic information was not considered either. In Section 4.2.2, I will point out why this is not appropriate, even for proofs in university mathematics education.

3.2 The relation of proof comprehension to evaluation and validation

Additionally, the models differ in the way they relate proof validation and proof evaluation to proof comprehension. Yang and Lin (2008) considered the evaluative act of "appreciating the beauty of mathematical structure" (p. 67) in their seventh facet of proof comprehension. This reaffirmed Selden's and Selden's (2015) claim that proof evaluation must be based on substantial proof comprehension. Following this point of view, it should be possible to use a reader's evaluated judgment of a given proof to gain information about the quality of his or her proof comprehension. Similarly, the PISA reading framework refers to evaluative tasks as

> drawing upon one's knowledge, opinions or attitudes beyond the text in order to relate the information provided within the text to one's own conceptual and experiential frames of reference. (OECD, 2019, p. 35)

Analogously, Yang and Lin (2008) used proof validation as an operational indicator for proof comprehension in the facet of Generality. Both Mejía-Ramos et al. (2012) and Neuhaus-Eckhardt (2022) did not keep up with this approach, but followed the suggestion of Mejía-Ramos and Inglis (2009) to treat different reading goals separately. They based their decision on the assumption that different reading goals will result in different reading behaviors, as suggested by Weber and Mejía-Ramos (2011). At least three arguments challenge this point of view: On the one hand, current cognitive-psychological approaches to general reading describe all kinds of reading processes, including different readers' intentions, using the same reading framework (OECD, 2019; Schnotz & Dutke, 2004). Describing the reading of proof from the perspective of the different text levels should therefore not be restricted to proof comprehension.

On the other hand, Panse et al. (2018) compared the eye movements of novices and experts when reading proofs known to be correct and when reading proofs of unknown correctness. They found no significant differences regarding the assumed truth or fallibility of the proof. Therefore, they raised the question of whether proof comprehension and validation are necessarily intermingled as

> perhaps the majority of validation effort is actually directed at comprehension, because comprehension must be attained before a validity judgment can be made. Or perhaps a sincere comprehension attempt provides validation 'for free', because good comprehension would flag up invalid inferences. (Panse et al., 2018, p. 370)

This challenges the practical use of the theoretical distinction between validation and comprehension as reading goals. Further, far more than the common three reading intentions are needed to appropriately describe the proof-reading behavior of undergraduate students in university mathematics (Spratte, 2022, 2023): Several students read proofs for example to gain a deeper conceptual understanding of notions within the proof. The distinction between the different reading goals according to Weber and Mejia-Ramos (2011) therefore runs the risk of promoting blind spots in research regarding students' reading behavior.

Finally, Neuhaus-Eckhardt's (2022) definition of proof comprehension itself established links to proof validation and proof evaluation. According to her, to understand a written proof means to construct a coherent mental model for it, and proof comprehension is the resulting mental model. Especially, her mental model incorporated micro- and local-level components such as the meaning of terms and statements. This implies that there is hardly any reading of proof without building some proof comprehension: Any information taken from the text is processed on the base of existing pre-information and forms (part of) a mental representation of the text. Thus, the question of interest is not the existence of proof comprehension, but its scope and its adequacy to the text and to the sociocultural reading situation including the reader's intention (Neuhaus-Eckhardt, 2022; see also Section 4.2). Proof comprehension hence can be described for every process of reading proof, independent of the reader's intentions.

3.3 The Unified Model for Proof Comprehension (UMPC)

Altogether, the distinction of the global text level from the level beyond the text allows uniting both the RCGP model and the AMPC with the refinements presented in Neuhaus-Eckhardt (2022) to a new, generalized model of proof comprehension which I will call the Unified Model for Proof Comprehension (UMPC). As suggested in the RCGP model and by Toulmin (2003), the unified model distinguishes the semantic content of the text from its logical structure and thus includes two micro-levels placed next to each other. As it is common in the literature for general reading comprehension, four text levels form a second distinctive feature (Lenhard et al., 2017): Micro, local, and global text levels are distinguished from each other and from the level beyond the text. From the local level on, both logic and semantic information need to be taken into account simultaneously (though not evenly distributed) to construct coherence. Thus, the UMPC consists of five levels for the perception of new information from the text.

Constructing local coherence in single argumentation steps is viewed as a mainly (though not purely) logical process and is therefore placed closer to the dimension of logic than to the semantic dimension (Toulmin, 2003; Yang & Lin, 2008). In contrast, global coherence requires more semantic understanding. This leads to a pyramid shape of the UMPC, as seen in Figure 4.[3]

As the unified model is fundamentally based on components of the three models presented in Chapter 2, their operational indicators are sorted into the UMPC in Figure 4. In brackets are those indicators that widely rely upon the readers' subsequent proof productions, i.e. those not to be interpreted as components of a theoretical mental model according to Neuhaus-Eckhardt (2022). The dimensions suggested as local in the AMPC and adopted by Neuhaus-Eckhardt are distributed among both micro levels and the local level. This is in line with the first three levels of the RCGP model, though the arrangement of the micro levels next to each other prevents the hierarchical impression inherent to the RCGP model, which I discussed in Section 2.1. Following Neuhaus-Eckhardt (2022), the AMPC's global dimension of Transferring is now referred to on the level beyond the text.

This more detailed view including all text levels might explain why the previous attempts to empirically confirm the distinction between a local and a global proof comprehension as defined in Mejía-Ramos et al. (2012) failed. Even if one hopes to identify the text levels as latent variables in proof comprehension tests[11], one must question the unity of the local factors grouped in the AMPC critically.

In line with the RCGP model, the unified model in Figure 4 includes the evaluation and the validation of proof, challenging the well-established distinction of the different reading activities established in Selden and Selden (2015). Reasons for a more holistic view were already discussed in Section 3.1. Note that the two activities are included in the UMPC similarly, but not analogously to facets in the RCGP model. In the RCGP model, both were used as operationalizations for comprehension on the next-lower text level. In the UMPC, this is adopted only for proof evaluation and the level beyond the text. For proof validation, all four levels inherent to the text itself need to be taken into account.

4. Contributions by and limitations of the UMPC

Throughout this chapter, I will discuss the additional value the UMPC offers for both teaching and research related to the reading of proof in different educational settings. Due to the limited scope of this paper, I will not provide empirical data to support the UMPC's practical value.[12] I will show the usefulness of the UMPC first by considering proof reading processes from a researcher's perspective. Second, I will argue for the need to consider all presented levels in both secondary and tertiary education.

4.1 Adequacy to proof reading processes

As argued in Chapter 1, psychometrical measurements of proof comprehension need to be based on sound cognitive-psychological models of reading proof to ensure the assessment's validity. All three models of proof comprehension in Chapter 2 offered links to established cognitive-psychological reading research. Nevertheless, their common main focus was to provide efficient and valid assessment instruments for proof comprehension as the product of a successful reading process.

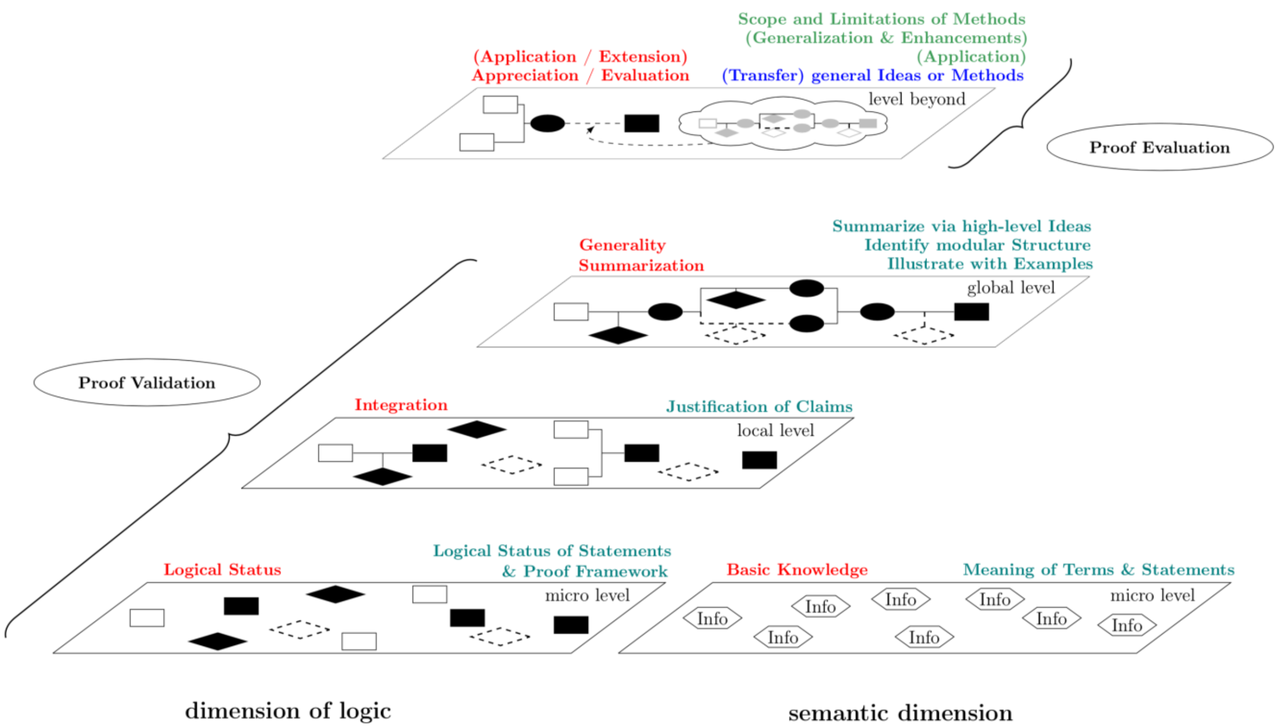

Fig. 4: Unified model for proof comprehension (UMPC) including proof validation and proof evaluation. Colors for operational indicators: Red: RCGP, Blue: AMPC, Green: Neuhaus-Eckhardt (2022), Teal: both AMPC and Neuhaus-Eckhardt (2022). In brackets are those indicators that widely rely upon the readers' subsequent proof productions.

So far, there are only a few attempts to describe the proof-reading process itself. Exceptions are some current approaches to reading strategies going back to Weber (2015) and Eye-Tracking studies such as Panse et al. (2018).[13] One reason for the sparse research situation could be the higher effort that is necessary to observe the reading process instead of its product: Studies involving eye-tracking or videographic analysis of reading processes are more time-consuming and more expensive than a questionnaire or an interview. Thus, they usually restrict to small numbers of participants, while the psychometrical assessment of proof comprehension easily allows studies of larger scale.

Another reason not to investigate the process of reading proofs might be the lack of an appropriate theoretical model to be applied in the analysis of the data. The existing models collect and arrange facets/components of proof comprehension named by experts, which seems plausible to ensure the validity of a psychometrical approach to proof comprehension. But to investigate the reading process itself the genesis of these components in a reader's mind needs to be considered. The UMPC follows a common way to address this problem in general reading research by raising attention to the text levels. Even though the levels were inherent in the existing contributions by Yang and Lin (2008), Mejía-Ramos et al. (2012), and Neuhaus-Eckhardt (2022), the UMPC restructures and states them explicitly, incorporating current cognitive-psychological reading research and Toulmin's argumentation theory. Thus, the UMPC seems more appropriate to describe proof reading processes than any of the preexisting models from Chapter 2.

Yet the UMPC is not the only process-oriented model for reading proofs: Ahmadpour et al. (2019) offer a proof-specific model of reading processes by "describing the transitions between different states of understanding when reading a proof" (p. 1). Their model includes seven states, organized in three different paths a reader will likely (but not necessarily) follow through a proof text: On the path of structure, the reader first uses examples to follow the text, from which he subsequently generalizes, abstracts and formalizes his understanding. On the path of procedure, formalization takes place without abstraction, i.e. the proof is read as a formalized general procedure instead of an abstract structure (Sfard, 1991). Finally, readers on the path of form focus on the manipulation of symbols only.

Ahmadpour et al. (2019) do not relate their model to research about general reading processes, but base it on theories from mathematics education. They consider, for example, the distinction on what establishes conviction for prospective mathematicians by G. Harel and Sowder (1998) and Sfard's (1991) thoughts on reification of mathematical concepts. Thus, their model seems to relate better to understanding the concept of proof in general than to proof comprehension of a single text. Especially for readers who are already used to operating with abstract mathematical concepts, the model presented by Ahmadpour et al. (2019) might not be appropriate to describe the reading of a current proof text. For less experienced readers, the two models could complement each other: It seems likely that students who read proofs as procedural proofs focus on the local coherence, heading from one argumentative step to the other and illustrating each step with exemplifying instances. To further investigate the types of proof perception and their influence on the different levels of proof comprehension, empirical studies are missing so far.

4.2 Adequacy to different educational stages and degrees of formalism

As seen in Chapter 2, each of the preexisting conceptions of proof comprehension lacked attention to at least one of the five text levels. In this section, I will argue that both the global level and the level beyond the text enrich the discussions about reading proof in secondary school. To that end, a brief discussion about the role of proof in this educational setting is inevitable. Analogously, I will show the importance of the micro-levels for proof reading in university mathematics education. Some brief thoughts on proofs in university service teaching of mathematics[1] wind up this section.

4.2.1 Reading proofs on the secondary level

There is a long tradition of calls from mathematics educators to include mathematical reasoning and proving in mathematical learning processes at all stages (G. Harel & Sowder, 1998; Sommerhoff et al., 2015). Indeed, many national curricula require secondary students to follow another person's mathematical argument, sometimes explicitly including written versions (Bundesministerium für Unterricht, Kunst und Kultur, 2009; Council of Chief State School Officers [CCSSO], 2022; Ständige Konferenz der Kultusminister der Bundesrepublik Deutschland [KMK], 2012, 2022). The nature of these arguments is left quite open. In the German context, they reach from visual argumentation in lower secondary to formal proof at the end of upper secondary school (KMK, 2012, 2022), where proof is approached through a continuous increase in formality and deductive rigor over time (Brunner, 2014). Consequently, as common in many international contexts as well, proving as the highest form of mathematical argumentation appears as a marginal and extraordinarily difficult activity throughout students' school careers (Bausch et al., 2014; Fesser & Rach, 2022; G. Harel & Sowder, 1998). Instead, visual elements and generic examples usually play an important role (Kempen, 2018).

Compared to the use of the established models, there is a first advantage of the UMPC to be noted here: Its roots in cognitive-psychological reading research allow to easily incorporate how readers deal with information from various visualizations and the text simultaneously. The PISA reading framework considers continuous texts, non-continuous texts (lists, diagrams, etc.), and mixtures of both (OECD, 2019). Adopting this approach to the reading of proof could lead to more sophisticated operational indicators by taking into account that the difficulty of proof comprehension tasks rises with a growing number of sources to be considered. Thus, the UMPC responds to the rising interest of mathematics education researchers in visual (elements of) proof (R. Harel & Marco, 2023).

A second strength of the UPMC, especially compared to the RCGP model, is the global text level it contains. As already discussed, most advanced arguments and proofs on the secondary level have a rather simple, chained argumentative structure (Selden, 2012). Nevertheless, this does not diminish the importance of the global text level in secondary contexts, even if not all operational indicators such as summarization or the modular structure are appropriate. Completeness of the (simple) argumentative chain as well as the generality of the mathematical statement need to be addressed on the secondary level (KMK, 2022; Yang & Lin, 2008) and this requires the global text level as I have discussed in Section 3.1.

A third argument for the UMPC's usefulness in secondary school contexts is the relevance of the level beyond the text for mathematics that fosters general education. This claim requires first of all to investigate the contribution of proof to general education. According to Heinrich Winter (1995), mathematics education becomes meaningful when it enables students to

> 1) notice and to understand phenomena of the world that (should) concern everyone in a specific way, be it from nature, society or culture,
>
> 2) get to know and to comprehend mathematical objects and facts, represented in language, symbols, pictures, and formulas, as mental creations and a deductively structured system of its own kind, and
>
> 3) gain heuristics and problem-solving strategies that may be applied in- and outside of mathematics when dealing with mathematical exercises. (Winter, 1995, p. 35, translated, V. S.)

All of these three *basic experiences* have a distinct connection to proofs: Learning to prove is considered to contribute substantially to the learning of heuristics, especially to logic and deductive reasoning (Grieser, 2018; Reid & Knipping, 2010, pp. 79-80; Vohns, 2016). This does not only include the construction of convincing arguments (as part of Winter's third experience) but also comprehending and questioning existing arguments (as part of Winter's third and first experience). Following Toulmin's (2003) argumentation theory, an argumentative step itself establishes conviction to a certain degree expressed in a modal qualifier. Thus, the local text level needs to be addressed carefully when teaching students how to argue convincingly. This is already suggested in the RCGP model and in many national curricula that focus first on single arguments and later on the chaining of steps (CCSSO, 2022; KMK, 2022; Yang & Lin, 2008).

Toulmin (2003) further draws our attention to the data, warrants, and backings suitable within an argumentational field. Focussing on the nature of warrants and data acceptable within mathematics, i.e. from a local text level, students might experience the intrinsic norms of mathematical reasoning as part of Winter's second basic experience. Indeed, any insight into mathematics as a deductively ordered system of its own kind with time-independent truth may not take place without exposure to proof (Rolfes et al., 2022; Weber, 2012).

Mathematical statements may serve as warrants only after they have been proven themselves (or been accepted as axioms). Thus, proofs "organize logically unrelated individual statements which are already known to be true, into `a coherent unified whole´" (De Villiers, 1990, p. 21). In doing so, proofs help to structure axioms, definitions, and theorems on both a local and a global level of the mathematical world. This systematization includes experiencing the fruitfulness and scope of certain mathematical statements – even if it takes place only locally without reference to certain axiomatic systems

(Kempen, 2018). It is echoed in the RCGP level of encapsulation, where Yang and Lin (2008) refer to the distinction of different premises from similar propositions. Thus, getting to know mathematics as a deductively structured system of its own kind may start from the local perspective, but will inevitably lead to proof comprehension on the level beyond the text.

Finally, experiencing a need for proof fosters critical thinking, which is a precondition of responsible citizenship. It is inherent, for example, in our modern judicial proceedings, where only complete chains of evidence allow discarding the presumption of innocence (Vohns, 2016; Winter, 1995). Therefore, engaging with mathematical proof might contribute to general education by understanding and challenging the argumentative culture(s) surrounding each of us. Standing above the obstacles of modal, i.e. (un)certain conclusions, mathematical proofs might not be representative of everyday argumentation. Yet for many people and many argumentative fields, such as medicine, law, or ethics, mathematical proofs serve as an ideal model for what to consider as proven and how to cope with uncertainties (Toulmin, 2003, p. 118). Again, this touches on the evaluative facet of the level beyond the text.

Another reason to focus on proof evaluation and therefore consider the level beyond the text in secondary education is suggested by Reid and Knipping (2010). They argue that on a level of general education, students shall be

> prepared not to practice in the field [of mathematics, V.S.], but rather to appreciate the products of it. In that case a focus on proof reading rather than writing, and a greater emphasis on roles such as aesthetics might be called for. (p. 222)

As mathematicians consider mathematical beauty in terms of a proof's innovative power, its creativity, and the use of surprising components, this kind of perception requires, for example, comparisons of different proofs and a sense of the scope of proving methods, thus calling upon the level beyond the text (Inglis & Aberdein, 2014).

4.2.2 Reading proofs in university mathematics

The presentation of the AMPC and its extension by Neuhaus-Eckhardt (2022) in Figures 2 and 3 already revealed that both models consider the micro-levels as local dimensions of proof comprehension. Their importance for proof comprehension among university mathematics is undoubted and has been demonstrated empirically several times: Bauer and Skill (2020) asked third-semester students in a more-dimensional analysis course in the middle of a proof presentation how it might continue. A third of all given ratings fell on one of four possible answers that assumed the statement to be proven. Similarly, Zazkis and Zazkis (2015) report that preservice teachers in their last term still face difficulties in distinguishing the claim to be proven from supporting data, confirming the need for more attention to the logic micro-level in a university context.

The novelty of the UMPC in contrast to the existing models of proof comprehension for the tertiary level is to consider the micro levels as individual levels (again). Using the same text levels describing proof comprehension in secondary and tertiary education offers advantages especially for research in the transition from secondary to tertiary level, which is a field of special interest for many mathematics educators (Bausch et al., 2014; Gueudet, 2008; Kempen, 2018).

Another argument shows why the semantic micro-level is at least as important as the logical one in the context of university mathematics. In the still predominant way of teaching university mathematics in a definition-theorem-proof format, proofs become the main carriers of mathematical knowledge and gain an epistemic function by introducing the reader to proving methods, mathematical concepts, and strategies (Davis et al., 2012; Kempen, 2018; Rav, 1999). This epistemic role of proof is prominent in the AMPC's extension by Neuhaus-Eckhardt (2022), as shown in Section 2.3.

But the mathematical knowledge carried within proofs is not restricted to their epistemic function: Proofs on the tertiary level also deepen the understanding of the thematic field and serve in creating new and systematizing existing knowledge (Davis et al., 2012, p. 167; Kempen, 2018). As Spratte (2022, 2023) shows, many students of university mathematics claim to read proofs for the sake of understanding mathematical concepts in the thematic field better, e.g. to understand definitions. Implicitly referring to the explanatory function of proof, the students make use not of the proof methods to transfer them to another context, but of one of the proof's semantic ingredients. Similar to an encapsulated comprehension of the proven statement, the deepened knowledge about the definition reaches beyond the text, for example when evaluating the usefulness of certain equivalent definitions or identifying critical examples for the notion under consideration. Because understanding a proof (as understanding any text) is a circular process, this knowledge is probably applied when semantically decoding the same notion within the text later again. Thus, in a cyclic comprehension process, deepened content knowledge and pre-knowledge seem to act indivisibly and therefore connect to both the semantic micro-level and the level beyond the text.

Further, Reid and Knipping (2010, p. 76) and Weber and Mejía-Ramos (2011) give examples of proofs actually showing an even stronger statement than originally given. The reader may discover the stronger theorem when reading the proof and reflecting on the importance of certain assumptions, working mainly on a local or global text level to expand the theorem. Yang and Lin (2008) also call upon the extension of a theorem by relaxing its premises or considering specific cases to deduce more specific statements. Reviewing the components of proof comprehension in Figures 2 and 3 under this perspective, I conclude that proof comprehension tests on the tertiary level so far have narrowed the role proofs play for mathematics to the functions of validation, explanation, and the epistemic function, resulting in an overemphasised, yet narrowed global level and level beyond the text. Discovering mathematics when reading a proof and evaluating the text are two major aspects, but not the only ones underestimated in the ongoing discussion.

4.2.3 Reading proof in university service courses

Little is known so far about lecturers' rationals to (or not to) present proofs in their service mathematics lectures[1], for example in engineering, physics, or economy. Besides the explanatory function and an insight into mathematics as a deductivly ordered system, Bauer and Skill (2020) mention the link between mathematical models and their application that is revealed in proofs as one possible reason. Analogously to the discussion on proofs in general education contexts in Section 4.2.1, these reasons turn the attention to the level beyond the text. The explanatory function of proof further calls upon central properties and ideas of the proof, thus operating on a global text level (Steiner, 1978). The level of local coherence and the micro levels seem to play a minor role in service mathematics. Nevertheless, this hypothesis lacks an empirical foundation, yet.

4.3 Limitations and open questions

Despite the numerous benefits of the UMPC discussed above, there are certain limitations inherent in the model and several open questions remain. As already mentioned frequently, the UMPC’s appropriateness to empirical data still has to be shown. Another limitation is the restriction to written proofs without explicitly including multiple sources such as diagrams accompanying the proof. I have discussed in Section 4.2 that the UMPC might be adapted to include visualizations in line with the PISA reading framework, but have not presented a specific operationalization or detailed description yet.

A third constraint is implied by the work of Ahmadpour et al. (2019) as presented in Section 4.1: The UMPC assumes that the reader receives the proof in terms of its deductive argumentative structure. This kind of reading corresponds to only one of the three paths suggested by Ahmadpour et al. (2019). How well the UPMC fits, for example, a reading relying on form and symbolic elements of a proof is an open question.

The probably greatest limitation lies in a hidden assumption: As common among researchers in mathematics education, the UMPC refers to proof comprehension as a single construct, i.e. stable over different mathematical fields.[2] But does reading a proof from algebraic number theory follow the same (or at least sufficiently similar) processes as one from stochastics? The generality of the UMPC's foundations seems to apply to any piece of mathematics, but further investigation and empirical evidence are needed.

5. Conclusions

Throughout this paper, I have summarized different concepts of proof comprehension into a unified model. The model is based on Toulmin's argumentation theory, text levels from cognitive-psychological reading research, and preliminary works on proof comprehension by Yang and Lin (2008), Mejía-Ramos et al. (2012) and Neuhaus-Eckhardt (2022). It defines proof comprehension as a coherent mental model for a given written proof, which is the result of a cyclic reading process. Thus, I closely related proof validation and proof evaluation to the UMPC as the results of proof reading processes with different intentions that also lead to the building of a mental model of the proof under consideration.

Different functions and roles of proof in secondary and tertiary educational contexts were discussed in light of the UMPC and revealed its additional value to teaching and research. Especially, discussions on proof comprehension at the secondary level need to include both the global text level and the level beyond the text, which has not been discussed in the light of general education so far. Similarly, the AMPC as the so-far predominant model of proof comprehension on the tertiary level respects the level beyond the text only in a limited way. It restricts to the proof method as part of the epistemological proof function and underestimates the consequences of proof for the knowledge of the thematic field. It also lacks evaluative aspects of proof reading, which play an important role in reading processes of mathematical experts (Mejía-Ramos & Weber, 2014). The unified model, therefore, allows a more integral way to look at the reading of proof than the models found in the existing literature.

Comments

[1] The importance of proof as the actual bearers of mathematical knowledge (Weber & Mejía-Ramos, 2011) leads to a prominent role of proof in mathematics education at the tertiary level – at least for students in mathematics programs or teacher programs for (upper) secondary education, who are supposed to learn how to prove in their first term(s) (Kempen, 2018). Throughout, I will talk about these students as *university mathematics students*. Several non-mathematics university programs also include some mathematics education, such as mathematics for engineering, where the focus is less on proof and more on theorems, application, and algorithms (Bauer & Skill, 2020). Students in these courses are considered as *service mathematics students*.

[2] The assumption that proof comprehension is a unidimensional construct is not trivial. The appropriateness of one model of proof comprehension among various mathematical subdisciplines is often implicitly assumed. Davies (2020) discusses it in more detail: Mejía-Ramos and Weber (2016) developed tests based on the AMPC for two numbertheoretical proofs and one on the incountability of an interval. They report high correlations between any two of these tests and deduce that "proof comprehension can be a meaningful single-dimensional construct" (p. 5). To assume unidimensionality also fits the failed attempts to separate local from global dimensions of proof comprehension as discussed in Section 2.2. The assumption further is forstered by the data provided by Davies (2020).

[3] Throughout this paper, the reduced version of the Toulmin scheme is used to visualize different levels of proof comprehensions. In Figures 1 to 4, a hexagon represents a semantic piece of information of unknown logical status. White rectangles represent data, while black ones symbolize claims. Black diamonds are explicitely given warrents, dashed white diamonds are used for implicit warrants. Ovals indicate former claims that further serve as data for a subsequent argumentative step. The symbols follow Knipping and Reid (2019).

[4] As Österholm (2008) points out, the lack of redundancy is not specific to proof, but characteristic for any technical prose.

[5] Note that an operationalization requires a ready theoretical construct and aims to make it measurable. As all three contributions discussed in Sections 2.1, 2.2, and 2.3 provide severe contributions to the concept of proof comprehension themselves, they should not be considered as mere operationalizations.

[6] At first glimpse, the RCGP model might not seem very specific to geometry. Nevertheless, I already pointed out how it deeply relies on Duval's (1998) levels of information processing when solving geometric problems. Also, some of the RCGP's contents implicitely assume that the proof can be visualized. This holds for almost any proof from geometry on the secondary level, but might be non-trivial for proofs from other mathematical domains.

[7] Neuhaus-Eckhardt (2022) emphasized that not all components may be suitable for every proof, e.g. if a proof is rather short and contains no modules. For the theoretical mental model, the proof-text determines the adequacy of the components, while in the individual mental model, also individual learner's resources, his or her reading goals and the sociocultural context influence the components' adequacy.

[8] "The PISA 2018 reading framework considers writing to be an important correlate of reading literacy. However, test design and administration constraints prohibit the inclusion of an assessment of writing skills, where writing is in part defined as the quality and organization of the production. However, a significant proportion of test items require readers to articulate their thinking into written answers. Thus, the assessment of reading skills also draws on readers' ability to communicate their understanding in writing, although such aspects as spelling, quality of writing and organization are not measured in PISA." (OECD, 2019, p. 49)

[9] Neuhaus-Eckhardt (2022) did not specify the relation between the main idea(s), the proof framework, and the proof method. A clear distinction between those components was not her focus; she rather included the component as the transferable proof method could differ from the main idea and the proof framework.

[10] Indeed, the RCGP's facet of Generality relies on the complete proof but stays within the proof itself. It therefore corresponds to the global text level. The facet of Application / Extension, originally placed next to Generality, goes beyond the proof by considering new proof construction processes.

[11] To identify the text levels as latent variables in a factor analysis for a proof comprehension test still seems rather unlikely to the author of this paper. Even though these levels are regularly used to construct valid tests for (general) reading comprehension, the latent variables underneath these tests are for example the individual's prior knowledge or use of reading strategies. See, for example, Cromley et al. (2010).

[12] Expectations regarding an empirical confirmation should not be set too high anyway as "all models are approximations. Essentially, all models are wrong, but some are useful." (Box & Draper, 1987, p. 424)

[13] A detailed overview of eye tracking in mathematics education research including a critical discussion on the methodology is given by Strohmaier et al. (2020).

References

Ahmadpour, F., Reid, D., & Fadaee, M. R. (2019). Students' ways of understanding a proof. *Mathematical Thinking and Learning, 21*(2), 85–104.

Bauer, T., & Skill, T. (2020). Mid-Proof Peer Instruction: Zum diagnostischen Potential von Votingfragen inmitten von Beweisen. [Mid-Proof Peer Instruction: On the diagnostic potential of voting questions in the middle of proofs.] *Hanse-Kolloquium zur Hochschuldidaktik 2020*.

Bausch, I., Biehler, R., Bruder, R., Fischer, P. R., Hochmuth, R., Koepf, W., Schreiber, S., & Wassong, T. (2014). *Mathematische Vor- und Brückenkurse. Konzepte, Probleme und Perspektiven.* [Mathematical preparational and bridging courses. Concepts, problems, and perspectives.] Springer.

Bickerton, R. T., & Sangwin, C. (2021). Practical online assessment of mathematical proof. *International Journal of Mathematical Education, 53*(1), 1–24.

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces.* Wiley.

Brunner, E. (2014). *Mathematisches Argumentieren, Begründen und Beweisen. Grundlagen, Befunde und Konzepte.* [Mathematical Argumentation, Reasoning and Proof. Foundations, Findings, and Concepts] Springer.

Bundesministerium für Unterricht, Kunst und Kultur. (2009). *Bildungsstandards für Mathematik 8. Schulstufe.* https://www.iqs.gv.at/themen/nationale-kompetenzerhebung/grundlagen-der-nationalen-kompetenzerhebung/grundlagen-der-bildungsstandards

Conradie, J., & Firth, J. (2000). Comprehension Tests in Mathematics. *Educational Studies in Mathematics, 42*, 225–235.

Council of Chief State School Officers. (2022). *Common Core State Standards for Mathematics*. https://learning. ccsso.org/wp-content/uploads/2022/11/ADA-Compliant-Math-Standards.pdf

Cromley, J., Snyder-Hogan, L. E., & Dubas, U. (2010). Reading comprehension of scientific text: A domain-specific test of the Direct and Inferential Mediation Model of Reading Comprehension. *Journal of Educational Psychology, 102*(3), 687–700.

Czocher, J. A., & Weber, K. (2020). Proof as a Cluster Category. *Journal for Research in Mathematics Education, 51*(1), 50–74.

Davies, B. (2020). *Comparative Judgment and Proof.* PhD thesis. Loughborough University.

Davis, P. J., Hersh, R., & Marchisotto, E. A. (2012). *The Mathematical Experience. Study Edition*. Springer.

De Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras, 24*, 17–24.

Duval, R. (1998). Geometry from a cognitive point of view. In C. Mammana & V. Villani (Eds.), *Perspectives on the Teaching of Geometry for the 21st Century* (pp. 37-62). Kluwer Academic Publishers.

Fesser, P., & Rach, S. (2022). Wissenschaftspropädeutik als eine Zieldimension von Mathematikunterricht in der gymnasialen Oberstufe. Normative und analytische Perspektiven. [Scientific propaedeutics as a target key dimension of mathematics in upper secondary education.] In T. Rolfes, T., S. Rach, S. Ufer, & A. Heinze (Eds.), *Das Fach Mathematik in der gymnasialen Oberstufe* (pp. 47–74). Waxmann.

Fischer, R., & Malle, G. (1985). *Mensch und Mathematik. Eine Einführung in didaktisches Denken und Handeln.* [Human and mathematics. An introduction into didactical thinking and practice.] BI Wissenschaftsverlag.

Grieser, D. (2018). *Exploring Mathematics. Problem-Solving and Proof.* Springer.

Gueudet, G. (2008). Investigating the secondary-tertiary transition. *Educational Studies of Mathematics, 67*, 237–254.

Harel, G., & Sowder, L. (1998). Students' Proof Schemes: Results from Exploratory Studies. *CBMS Issues in Mathematics Education, 7*, 234–283.

Harel, R., & Marco, N. (2023). Visual Proofs from High School Students and Teachers' Perspectives. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel., & M. Tabach (Eds.), *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education 1* (p. 254). University of Haifa.

Hodds, M. (2014). *Improving Proof Comprehension in Undergraduate Mathematics.* Loughborough University Press.

Inglis, M., & Aberdein, A. (2014). Beauty Is Not Simplicity: An Analysis of Mathematicians' Proof Appraisals. *Philosophia Mathematica, 23*(1), 87–109.

Inglis, M., Mejía-Ramos, J. P., & Simpson, A. (2007). Modelling mathematical argumentation: the importance of qualification. *Educational Studies in Mathematics, 66*, 3–21.

Inglis, M., Mejía-Ramos, J. P., Weber, K., & Alcock, L. (2013). On Mathematicians' Different Standards When Evaluating Elementary Proofs. *Topics in cognitive science, 5*(2), 270–282.

Kempen, L. (2018). *Begründen und Beweisen im Übergang von der Schule zur Hochschule.* [Reasoning and proving in the transition from school to college.] Paderborn University Library.

Kintsch, W. (1998). *Comprehension. A Paradigm for Cognition.* Cambridge University Press.

Knipping, C., & Reid, D. A. (2019). Argumentation analysis for Early Career Researchers. In G. Kaiser & N. Presmeg (Eds.), *Compendium for Early Career Researchers in Mathematics Education. ICME-13 Monographs* (pp. 3–31). Springer.

Lenhard, W., Lenhard, A., & Schneider, W. (2017). *Ein Leseverständnistest für Erst- bis Siebtklässler II (ELFE II).* [A reading comprehension test for grades 1 to 7: ELFE II] Hogrefe.

Lin, F.-L., & Yang, K.-L. (2007). The Reading Comprehension of Geometric Proofs: The Contribution of Knowledge and Reasoning. *International Journal of Science and Mathematics Education, 5*, 729–754.

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., & Inglis, M. (2009). Argumentative and proving activities in mathematics education. In F.-L. Lin, F.-J. Hsieh, G. Hanna, M. de Villiers (Eds.), *Proceedings of the ICME Study 19 conference: Proof and Proving in Mathematics education, Volume 2* (pp. 88–93). The Department of mathematics, National Taiwan Normal University.

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.

Mejía-Ramos, J. P., & Weber, K. (2014). Why and how mathematicians read proofs: Further evidence from a survey study. *Educational Studies in Mathematics, 85*, 161–173.

Mejía-Ramos, J. P., & Weber, K. (2016). Student performance on proof comprehension tests in transition-to-proof courses. Paper presented at *19th annual Special Interest Group of the Mathematical Association of American on Research in Undergraduate Mathematics Education.* Pittsburgh, PA.

Neuhaus-Eckhardt, S. (2022). *Beweisverständnis von Studierenden. Zusammenhänge zu individuellen Merkmalen und der Nutzung von Beweislesestrategien.* [Proof comprehension of undergraduate students. Relations to individual characteristics and to the use of proof-reading strategies.] Waxmann.

Neuhaus-Eckhardt, S., & Rach, S. (2023). Constructing a proof after comprehending a similar proof – relation and examples. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel., & M. Tabach (Eds.), *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education 3* (pp. 403-410). University of Haifa.

Neumann, I., Heinze, A., & Pigge, C. (2017). *Welche mathematischen Lernvoraussetzungen erwarten Hochschullehrende für ein MINT-Studium? Eine Delphi-Studie.* [What mathematical learning requirements do college teachers expect for STEM studies? A Delphi study.] Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik.

OECD. (2019). PISA 2018 Reading Framework. In: *PISA 2018 Assessment and Analytical Framework* (pp. 21–71), OECD Publishing.

Österholm, M. (2008). Do students need to learn how to use their mathematics textbooks? The case of reading comprehension. *Nordisk matematikkdidaktikk, 13*(3), 53–73.

Panse, A., Alcock, L., & Inglis, M. (2018). Reading Proofs for Validation and Comprehension: an Expert-Novice Eye-Movement Study. *International Journal for Research in Undergraduate Mathematical Education, 4*, 357–375.

Pracht, E. (1979). Beweisverständnis und dessen Überprüfbarkeit. In Dörfler, W. (Ed.), *Beweisen im Mathematikunterricht. Vorträge des 2. Internationalen Symposiums für "Didaktik der Mathematik" von 26.9. bis 29.9.1978 in Klagenfurt* (pp. 349–356). Teubner.

Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica, 7*(3), 5–41.

Reid, D. A., & Knipping, C. (2010). *Proof in Mathematics Education. Research, Learning and Teaching.* Sense Publishers.

Rolfes, T., Rach, S., Ufer, S., & Heinze, A. (Eds.). (2022). *Das Fach Mathematik in der gymnasialen Oberstufe.* [Mathematics in German upper secondary schools.] Waxmann.

Schnotz, W., & Dutke, S. (2004). Kognitionspsychologische Gundlagen der Lesekompetenz: Mehrebenenverarbeitung anhand multipler Informationsquellen. [Cognitive-psychological foundations of reading literacy: multilevel processing using multiple sources of information.] In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Eds.), *Struktur, Entwicklung und Förderung von Lesekompetenz: vertiefende Analysen im Rahmen von PISA 2000* (pp. 61–99). VS Verlag für Sozialwissenschaften

Selden, A. (2012). Transitions and Proof and Proving at Tertiary Level. In G. Hanna & M. de Villiers (Eds.), *Proof and Proving in Mathematics Education. The 19th ICMI Study*. Springer.

Selden, A., & Selden, J. (2013). The Genre of Proof. In: *Mathematics and Mathematics Education: Searching for Common Ground* (pp. 248–251). Advances in Mathematics Education. Springer

Selden, A., & Selden, J. (2015). A comparison of proof comprehension, proof construction, proof validation and proof evaluation. *Didactics of Mathematics in Higher Education as a Scientific Discipline Conference Proceedings*, 339–345.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics, 22*, 1–36.

Sommerhoff, D., Ufer, S., & Kollar, I. (2015). Research in mathematical argumentation: A descriptive review of PME proceedings. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of 39th Psychology of Mathematics Education conference, Vol. 4* (pp. 193–200). Hobart, Australia: PME.

Spratte, V. (2022). (Why) Do students read the proofs we show them? In J. Hodgen, E. Geraniou, G. Bolondi & F. Ferretti (Eds.), *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)*. Free University of Bozen-Bolzano and ERME.

Spratte, V. (2023). (Warum) lesen Studierende Beweise? [(Why) do students read proofs?] In J. Härterich, M. Kallweit, K. Rolka, & T. Skill (Eds.), *Hanse-Kolloquium zur Hochschuldidaktik der Mathematik 2021. Beiträge zum gleichnamigen Symposium am 12. November 2021 an der Universität Duisburg-Essen* (pp. 256-273). WTM-Verlag.

Ständige Konferenz der Kultusminister der Bundesrepublik Deutschland. (2012). *Bildungsstandards im Fach Mathematik für die allgemeine Hochschulreife.* [Educational standards in mathematics for the general qualification for university entrance.] https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf.

Ständige Konferenz der Kultusminister der Bundesrepublik Deutschland. (2022). *Bildungsstandards für das Fach Mathematik. Erster Schulabschluss (ESA) und Mittlerer Schulabschluss (MSA).* [Educational standards for mathematics. Lower secondary education.] https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-ESA-MSA-Mathe.pdf.

Steiner, M. (1978). Mathematical Explanation. *Philosophical Studies, 43*(2), 135–151.

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics, 104*, 147–200.

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

Toulmin, S. E. (2003). *The uses of argument.* Cambridge University Press.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Academic Press

Vohns, A. (2016). Mathematical Literacy as a Civil Right and/or a Civic Duty? Two General Education Approaches. *Proceedings of the 13th International Congress on Mathematical Education (TSG-23: Mathematical Literacy)*, Hamburg.

Weber, K. (2012). Mathematicians' perspectives on their pedagogical practice with respect to proof. *International Journal of Mathematical Education in Science and Technology, 43*(4), 463–482.

Weber, K. (2015). Effective Proof Reading Strategies for Comprehending Mathematical Proofs. *International Journal of*

Research in Undergraduate Mathematics Education, 1, 289–314.

Weber, K., & Czocher, J. (2019). On mathematicians' disagreements on what constitutes a proof. *Research in Mathematics Education, 21*(3), 251–270.

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

Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. [Mathematics instruction and general education.] *Mitteilungen der GDM 61*, 37–46.

Yang, K.-L., & Lin, F.-L. (2008). A model of reading comprehension of geometry proof. *Educational Studies in Mathematics, 67*(1), 59–76.

Zazkis, D., & Zazkis, R. (2015). Prospective Teachers' Conceptions of Proof Comprehension: Revisiting a Proof of the Pythagorean Theorem. *International Journal of Science and Mathematics Education, 14*(4).

Address of the author

Verena Spratte
Universität Rostock
Institut für Mathematik
Ulmenstraße 69, Haus 3
18057 Rostock
verena.spratte@uni-rostock.de