

## measuring the quality of philosophical dialogue: a high-inference rating instrument for research and teacher education

deborah bernhard<sup>1</sup>

university of teacher education lucerne, switzerland  
orcid id: 0000-0002-8990-8060

dominik helbling<sup>2</sup>

university of teacher education lucerne, switzerland  
orcid id: 0000-0002-6898-8532

### abstract

Various studies have shown that philosophizing with children at school can have a positive effect on cognitive, language and social skills. However, previous studies have not considered how the quality of the dialogue influences these outcomes. Addressing this gap, our article introduces a high-inference rating instrument to assess the quality of philosophical dialogue. This instrument features four quality dimensions: Philosophical Richness, Co-construction, Focus, and Restrained Facilitation. It was applied to evaluate 63 class dialogues from a Swiss study involving secondary-school students. The article presents the instrument using excerpts from the study, along with initial validation information, showing its reliability in measuring philosophical dialogue quality. The significance of this research lies in its focus on the quality of philosophical dialogue, an aspect often overlooked in educational settings. By providing a tool to assess this quality, we open up new pathways for analyzing the effectiveness of philosophical dialogues in schools. This is especially relevant as educational systems increasingly recognize the importance of developing students' critical thinking and discussion skills. The findings highlight the potential of high-quality philosophical dialogues to enrich students' educational experiences and offer insights for educators and researchers into enhancing these interactions. This study contributes to a deeper understanding of how philosophical dialogue can be used as a powerful educational tool, not just for engaging students but for fostering meaningful and impactful learning.

**keywords:** high-inference rating instrument; philosophy with children; quality of philosophical dialogue.

### medir la calidad del diálogo filosófico: un instrumento de calificación de alta inferencia para la investigación y la formación docente

### resumen

Varios estudios han demostrado que filosofar con niños y niñas en la escuela puede tener un efecto positivo en las habilidades cognitivas, lingüísticas y sociales. Sin embargo, estudios anteriores no han considerado cómo la calidad del diálogo influye en estos resultados. Abordando esta brecha, nuestro artículo introduce un instrumento de evaluación de alta inferencia para valorar la calidad del diálogo filosófico. Este instrumento presenta cuatro dimensiones de calidad: Riqueza Filosófica, Co-construcción, Enfoque y Facilitación Restringida. Se aplicó para evaluar 63 diálogos de clase de un estudio suizo con estudiantes de secundaria. El artículo presenta el instrumento utilizando extractos de ese

---

<sup>1</sup> E-mail: [deborah.bernhard@phlu.ch](mailto:deborah.bernhard@phlu.ch)

<sup>2</sup> E-mail: [dominik.helbling@phlu.ch](mailto:dominik.helbling@phlu.ch)

análisis, junto con información inicial de validación, mostrando su fiabilidad en la medición de la calidad del diálogo filosófico. La importancia de esta investigación radica en su enfoque en la calidad del diálogo filosófico, un aspecto a menudo pasado por alto en los entornos educativos. Al proporcionar una herramienta para evaluar esta calidad, abrimos nuevos caminos para analizar la eficacia de los diálogos filosóficos en las escuelas. Esto es especialmente relevante ya que los sistemas educativos reconocen cada vez más la importancia de desarrollar habilidades de pensamiento crítico y discusión en los estudiantes. Los hallazgos destacan el potencial de los diálogos filosóficos de alta calidad para enriquecer las experiencias educativas de los estudiantes y ofrecer perspectivas para educadores e investigadores sobre cómo mejorar estas interacciones. Este estudio contribuye a una comprensión más profunda de cómo el diálogo filosófico puede utilizarse como una herramienta educativa poderosa, no solo para involucrar a los estudiantes, sino para fomentar un aprendizaje significativo e impactante.

**palabras clave:** instrumento de valoración de alta inferencia; filosofía con niños; calidad del diálogo filosófico.

### **medindo a qualidade do diálogo filosófico: um instrumento de avaliação de alta inferência para a pesquisa e a formação de professores**

#### **resumo:**

Diversos estudos têm demonstrado que filosofar com crianças nas escolas pode ter um efeito positivo nas habilidades cognitivas, linguísticas e sociais. No entanto, estudos anteriores não consideraram como a qualidade do diálogo influencia esses resultados. Abordando essa lacuna, nosso artigo introduz um instrumento de avaliação de alta inferência para averiguar a qualidade do diálogo filosófico. Esse instrumento apresenta quatro dimensões de qualidade: Riqueza Filosófica, Co-construção, Enfoque e Facilitação Restringida. Ele foi aplicado na avaliação de 63 diálogos de classe de um estudo suíço realizado com estudantes do ensino médio. O artigo apresenta o instrumento utilizando extratos desse estudo, junto com informações de validação inicial, mostrando sua confiabilidade na medição da qualidade do diálogo filosófico. A importância dessa pesquisa reside em seu foco na qualidade do diálogo filosófico, um aspecto frequentemente negligenciado nos ambientes educacionais. Ao oferecer uma ferramenta para avaliar essa qualidade, abrimos novos caminhos para a análise da efetividade dos diálogos filosóficos nas escolas. Isso é especialmente relevante à medida que os sistemas educacionais reconhecem cada vez mais a importância de desenvolver o pensamento crítico e as habilidades de discussão dos alunos. Os resultados destacam o potencial que os diálogos filosóficos de alta qualidade têm de enriquecer as experiências educacionais dos estudantes e oferecer novas ideias e percepções aos professores e pesquisadores sobre como melhorar essas interações. Este estudo contribui para uma compreensão mais profunda de como o diálogo filosófico pode ser utilizado como uma ferramenta educativa poderosa, não apenas para engajar os estudantes, mas também para promover um aprendizado significativo e impactante.

**palavras-chave:** instrumento de avaliação de alta inferência; filosofia com crianças; qualidade do diálogo filosófico.



measuring the quality of philosophical dialogue: a high-inference rating  
instrument for research and teacher education

Philosophical dialogue with children and adolescents has found its way into schools all over the world. The diverse teaching methods and approaches that have become established due to different authors and educational contexts are often subsumed under the term *Philosophy with Children* (PwC). The roots of this form of classroom discourse lie in the work of Matthew Lipman, who started the *Philosophy for Children* (P4C) program in the 1970s (Lipman, 1988, 2009). Since then, the effects of philosophical praxis in the classroom have been examined in numerous empirical studies. These studies show that the regular implementation of PwC has a positive effect on the cognitive development of students (García-Moriyón et al., 2005; Yan et al., 2018). Various studies further suggest that this method contributes to significant improvements in reasoning and argumentation skills (Collins, 2007; Säre et al., 2016; Walker et al., 2013), language ability (Alt, 2018; Schleifer & Courtemanche, 1996), social skills (Daniel, 2012, 2021) and critical thinking skills (Colom et al., 2014; Siddiqui et al., 2019, 2022; Worley & Worley, 2019). Previous research, however, has not included consideration of whether and to what extent the quality of collaborative philosophical inquiry had an impact on participants' progress in the aforementioned areas. To facilitate such consideration, this article presents a high-inference rating instrument that measures the quality of philosophical dialogue. The instrument was developed as part of an intervention study and tested in 21 secondary-school classes in Switzerland. The paper first addresses the characteristics of philosophical dialogue and then describes the development and indicators of the high-inference rating scale using excerpts from classroom discussion. It then presents the results of the validation of the instrument carried out as part of the study. Finally, it addresses the limitations of the instrument, possible further developments and practical applications in research and teacher training.

### *characteristics of philosophical dialogue*

This article bases its understanding of philosophical dialogue primarily on Matthew Lipman's ideas (Lipman, 1988, 2003) but also on the publications of important German-speaking authors such as Ekkehard Martens (2003), Barbara Brüning (2003) and Kerstin Michalik (2013), thus connecting to the tradition of the Socratic Method. Children and young people explore philosophical questions collaboratively and expand their insights in a co-constructive process using specific philosophical methods. Philosophical dialogue thus differs from everyday conversations and discussions because certain philosophical methods are used to promote philosophical reflection. Reviewing the philosophical practice of Socrates and Aristotle, Ekkehard Martens defined five core philosophical methods that support successful philosophising: Phenomenology, Hermeneutics, Analysis, Dialectics and Speculation. Children participating in philosophical inquiry perceive and describe facts and phenomena in a differentiated way (phenomenological method). They try to understand the statements of their peers and to grasp their perspectives (hermeneutic method). Concepts and terms are scrutinised, and arguments are examined and questioned (analytical method). This happens in a collaborative, co-constructive exchange (dialectical method). Furthermore, participants in philosophical inquiry formulate hypotheses, consider possible consequences and discuss creative ideas (speculative method) (Marsal, 2009, 2014; Martens, 2003; Meerwaldt et al., 2018).

The teacher's task is to promote critical and independent thinking in students using specific questions and statements. In the Socratic tradition, this can be called the maieutic method since the teacher helps 'give birth' to ideas and thoughts. The aim of such questions is to deepen the students' philosophical reflection (Fisher, 2007; Haynes, 2008; Michalik, 2016). With increasing philosophical practice, children become able to use such questions and prompts for themselves (Kennedy, 2013). Table 1 presents different types of questions that can support philosophical reflection (Helbling, 2018).

Describe facts and clarify terms	<ul style="list-style-type: none"> <li>• Can you explain that? What do you mean by that?</li> <li>• What do you understand by this term? In what situations do you use it?</li> <li>• Can you give an example?</li> </ul>
Identify differences and similarities	<ul style="list-style-type: none"> <li>• What are the features of x and y?</li> <li>• What is the difference between x and y?</li> </ul>
Justify opinions and reflect on evaluations	<ul style="list-style-type: none"> <li>• Can you give a reason? How would you justify that?</li> <li>• How do you know that?</li> <li>• Can you give an example?</li> </ul>
Creating hypotheses and considering consequences (thought experiments)	<ul style="list-style-type: none"> <li>• What would happen if...?</li> <li>• What could be done so that...?</li> <li>• Does anyone have another idea?</li> </ul>

Table 1: *Types of questions in philosophical dialogue*

In addition to the depth of philosophical reflection, the type of discourse is crucial for the quality of philosophical dialogue. In a community of inquiry (Sharp, 2009), students should collaboratively and co-constructively share thoughts and ideas so that the content of the dialogue evolves. Participants refer to each other and repeat statements, rephrase them, add to them, question them, contradict them, or ask questions if they are not understood (Daniel, 2008; Fisher, 2007; Kennedy, 2013; Michalik, 2016). Progress becomes visible since views are not simply expressed additively (Daniel et al., 2005; Jackson, 2013). The teacher can encourage co-construction through moderation questions, as exemplified in Table 2 (Fisher, 2007).

Moderation questions and statements to promote co-construction	<ul style="list-style-type: none"> <li>• Can you agree with x? Why/why not?</li> <li>• Did everyone understand what x said? Can you repeat what you meant?</li> <li>• Would anyone like to add something to this question/statement? Has anyone thought of something similar? Does anyone disagree?</li> <li>• What do others think about this?</li> <li>• So you agree that...</li> <li>• Some of you think that ... and others think it is rather ...</li> <li>• What have we found out together?</li> </ul>
----------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Table 2: *Questions and statements to promote co-construction*

Another feature of quality is that a previously defined question becomes focused and clarified. It is easy, instead, for students to digress and get lost in detail.

To counteract this, the teacher can regularly summarise the status of the dialogue with regard to the question and explicitly revisit the underlying philosophical question (Brüning, 2003; Jackson, 2013). In order to be able to focus on the question, the participants must also comply with the rules of discourse so that no disturbances or interruptions occur (Haynes, 2008; Jackson, 2013).

Therefore, facilitating a philosophical dialogue requires great skill and sensitivity. This becomes clear in Haynes' description of the teacher's role:

The teacher's role is to prompt children into giving reasons for their ideas, making distinctions and connections, constructing arguments and developing hypotheses and analogies. A balance has to be struck between a sense of adventure and exploration and a sense of direction and progress. The teacher helps the group to clarify and build on ideas whilst allowing the discussion to follow its own course. (Haynes, 2008, pp. 30–31)

Despite being expected to promote philosophical reflection and co-construction, teachers must also moderate with restraint (Krüger & Schick, 2012; Sharp, 2018). As a result, they should not express any opinions or attitudes of their own and should try to limit their share of the speaking: "In a community of enquiry teachers are facilitators, holding back on their own views, tactfully supporting children's thought and talk and encouraging the dialogue to flow between participants" (Haynes, 2008, p. 35). In addition, teachers should not steer the dialogue in a direction that they have set but should follow the students' train of thought (Jackson, 2013). The teacher's tasks in a philosophical dialogue can be summarised as follows (Brüning, 2003; de Boer, 2015; Kennedy, 2004):

- Promoting philosophical reflection and co-construction with questions and prompts
- Summarising different contributions and positions; sharpening opinions
- Keeping the focus on the topic/question
- Monitoring the dialogue for inclusion of all participants and compliance with rules
- Summarising results (at the end)
- Providing prompts to reflect on the dialogue

Philosophical dialogue can therefore be regarded as successful if a deeper philosophical reflection takes place through the application of philosophical methods and if the students gain new insights in a co-constructive and collaborative way while maintaining the focus. Teachers support these processes and act as facilitators, holding back their own views and limiting their share of the speaking.

### ***the development of a high-inference rating instrument***

As already mentioned, until now there has been no instrument for measuring the overall quality of philosophical dialogue. Studies dealing with quality have exclusively addressed individual factors. For example, Cassidy and Christie (2013) used categories to examine what types of contributions participants made to philosophical inquiry. This made it possible to determine how often children used (philosophical) talk moves such as formulating examples or hypotheses, defining or differentiating words or phrases or introducing new ideas. A Canadian research team led by Marie-France Daniel described the development of complexity and critical thinking in dialogues (Daniel, 2008, 2021). Helzel and Michalik (2015) addressed, among other things, the multiple perspectives and references of students. Santi (1993) analyzed and described philosophical dialogues based on the criteria of *spontaneous philosophical content, argumentative structure, epistemic categories, and the roles of teachers and peers*. However, assessing the impact of philosophical dialogue requires a more comprehensive measurement tool. Therefore, we developed a high-inference rating instrument based on the rating system of Hugener et al. (2006) as well as Kunter (2005) to capture the overall quality of a philosophical dialogue. The first step consisted of a literature analysis, from which central quality characteristics emerged via a deductive procedure. These were then condensed into dimensions and operationalised with indicators. The preliminary instrument was adapted and supplemented in dialogue and based on an example class discussion, in accordance with an inductive procedure.

For this study, we used a high-inference procedure, which requires a high degree of inference and interpretation. Such procedures are used to assess the quality of instructional sequences – or, in our case, conversational sequences (Clausen et al., 2003; Seidel, 2005). In contrast, low-inference methods yield statements about easily observable behaviour.

Due to the need for intensive interpretation, the conversational features to be assessed must be presented precisely. Therefore, the content of each dimension is described in the rating instrument. This basic idea is then further elaborated using detailed indicators that are as close to the behaviour level as possible. The

description and the indicators reflect a typical ideal philosophical inquiry. The rater (or, in our case, the two raters) must then rate the extent to which the dialogue corresponds to this specification on a four-point scale, assessing how often a certain behaviour occurs and how pronounced it is. Usually, the rater also examines the extent to which different students are involved in the class discussion. In our case, this third feature could not be measured or rated since only audio recordings were involved. The aim is for the judgments to reflect the overall impression in every dimension (Hugener et al., 2006).

Each category was supplemented by various remarks to clarify any ambiguities that arose. A complete philosophical dialogue forms the unit of analysis. The use of a highly inferential method requires intensive study of the instrument and many joint discussions for assessment (Rakoczy & Pauli, 2006).

Based on the literature, we defined four dimensions of the quality of philosophical dialogue: *Philosophical Richness*, *Co-construction*, *Focus* and *Restrained Facilitation*. Table 3 provides an overview of the dimensions and indicators. The complete instrument with the corresponding literature can be found in the Appendix (Supplementary File 1: High-Inference Rating Instrument).

Dimension	Indicators
<b>Philosophical Richness</b> The criterion captures the extent to which different questions and contributions are illuminated in a differentiated manner using philosophical methods.	<ul style="list-style-type: none"><li>• Terms, phenomena and situations are described and clarified.</li><li>• Similarities, differences and connections are formulated.</li><li>• Justifications or examples supporting opinions are expressed and evaluations reflected upon.</li><li>• Hypotheses are formulated, consequences assessed and speculations expressed.</li><li>• Diverse and substantial thoughts are expressed.</li></ul>
<b>Co-construction</b> The criterion captures the extent to which the teacher's and students' statements and questions relate to each other and build on each other.	<ul style="list-style-type: none"><li>• Statements refer to previous statements by repeating, rephrasing, supplementing or differentiating them.</li><li>• Statements relate to previous statements by questioning or contrasting with them.</li><li>• Unclear statements are taken up and clarified by (follow-up) questions.</li><li>• Progress becomes evident in successive statements. Statements are not primarily made in an additive manner.</li></ul>



<p><b>Focus</b></p> <p>The criterion captures the extent to which the statements and questions from the teacher and students focus on answering the philosophical question.</p>	<ul style="list-style-type: none"> <li>• Questions and statements contribute to answering the underlying question. The conversation does not digress in content.</li> <li>• Summaries, transitions and questions that focus on the underlying issue create a common thread in the conversation.</li> <li>• The underlying philosophical question is explicitly taken up again during or at the end of the dialogue.</li> <li>• Basic dialogue rules are followed so that the dialogue is not disturbed or interrupted.</li> <li>• The teacher recognises statements with potential for the topic and deepens them with specific questions.</li> </ul>
<p><b>Restrained facilitation</b></p> <p>The criterion captures the extent to which the teacher succeeds in promoting co-constructive dialogue and philosophical richness without assuming too dominant a role in the dialogue.</p>	<ul style="list-style-type: none"> <li>• The teacher refrains from expressing personal opinions or evaluations and does not provide any answers.</li> <li>• The teacher follows the students' trains of thought and deepens them by adapting questions flexibly, rather than leading the dialogue towards a specific goal.</li> <li>• The teacher creates space so that the students can talk to each other and limits her or his share of the speaking.</li> <li>• The teacher creates space so that the students can develop their own trains of thought without being too quickly restricted by the focus of the topic.</li> </ul>

Table 3: Dimensions and indicators of the high-inference rating instrument

In the next section, we will take a closer look at the dimensions using examples from philosophical dialogues.

### *presentation of the instrument using excerpts from philosophical dialogue*

In order to make the instrument more comprehensible, two academic excerpts from philosophical dialogues taken from the Swiss study mentioned above are presented in this section and linked to the dimensions.

The intervention study aimed to examine the influence of philosophical dialogue about the nature of science on the acceptance and understanding of evolution. In the teaching unit, philosophical dialogue was thus linked to the content of the subject of biology. The teaching approach of using philosophical reasoning in school subjects goes back to Kerstin Michalik (2009, 2013) and holds great potential for the students' (interdisciplinary) learning. The focus of these dialogues was not on the development of questions by the students but rather on the conversation itself and its progression. Accordingly, the instrument is exclusively centred on the discussion phase.

The teaching unit on evolution was carried out in 21 secondary-school classes, with students mainly aged between 13 and 15. The classes were divided on three occasions, during which half took part in a philosophical dialogue while the other half delved further into the subject for about 25 minutes. In the philosophical dialogues, the students dealt with the following three questions: Does the theory of evolution make belief in God and religion superfluous? Can we rely on science? What distinguishes humans and animals? The first two dialogues thus focused on epistemological questions and in particular on the relationship between religion and science. These topics are part of the concept of the nature of science (Hofheinz, 2010; Lederman et al., 2002; McComas & Olson, 1998) and have proven to be important factors influencing the acceptance and understanding of evolution. The third dialogue focused on anthropological and ethical questions, and its discussion required recourse to the first two dialogues. The students in the classes had generally known each other for one or two years, but this was their first time participating in a philosophical dialogue in this group or at all. Research findings indicate that experience in philosophical praxis and the duration of the group's existence have a significant impact on the quality of the dialogues (Daniel, 2008).

All class discussions were assessed using the instrument described. This allowed the pursuit of the question whether the quality of the philosophical dialogue influences the acceptance or understanding of evolution. Further information on and results of the study can be found in Bernhard (2022).

In the following simplified excerpt, students consider the question of whether one can believe in God and hold the theory of evolution to be true at the same time. Due to the audio recording, the students speaking cannot be identified. The statements are thus simply numbered and assigned either to the teacher (identified as 'T' for 'teacher') or to the students.

- S1: So, I think yes, that is possible. Believing in God does not only mean that he created the world and everything. It could also just help if you have a problem or something. You can still count on God.
- S2: So, I think you can. So, you can do everything, but you must try to separate it. You can't just, so you must separate it. So, these are two different issues, because I think that belief is believing it is so. And the other is simply a fact, so it's about facts. Facts that can be proven, so to speak. With God, it's just believing.



- S3: I don't know if that's true, either. Those giraffes used to look like this.<sup>3</sup> There's no proof of it, it's just a guess.
- S4: You can prove that.
- S5: If you find dead animals, you can use the genes to look up what kind of animal it was; then you have proof that it had a short neck.
- S6: But we don't know exactly how it came about.
- S7: You can do experiments. It's definitely different now. It's always changing. You can see this in all animals.
- S8: They also only learned about the dinosaurs from the bones they found.
- S9: Well, there are enough facts, I think, or evidence.
- S10: Yes, I think it is possible that one can be a believer and a scientist. Because if you find out something in science, for example, and believe in God and read it somewhere in the Bible and it is confirmed again, then you are actually even safer.
- T1: So, they can confirm each other.
- S11: Yes, exactly.
- T2: That would be great. Are there also subject areas where this is not possible?

### *philosophical richness*

Here, students name two important reasons that one can be a scientist and religious at the same time. First, belief in God involves more than just the question of whether God created the world; it can, for example, fulfil the need for help or protection (S1). It is also emphasised that science and faith are two things that must be clearly distinguished. The important distinction between knowledge and belief is mentioned (S2). A brief discussion develops from this point, clarifying in part how science can find evidence. During the discussion, the students seek to support their views with concrete examples and thus provide justifications (S5/S7/S8). However, there is no conclusive clarification or statement of the extent to which faith and science are two different domains or of what distinguishes knowledge from faith. Here, the teacher fails to follow up specifically or summarise. She does not elaborate, merely asking about the last point mentioned (T1).

---

<sup>3</sup> In the science lesson, the class dealt with the development of the neck of giraffes, among other things.

### *co-construction*

In the excerpt, it becomes clear that the students' statements relate to each other and are not uttered additively. The second statement is even questioned by a student so that a longer exchange develops (S3). Various participants complement each other in explaining why the idea of the development of the giraffe is based on evidence (S5/S7/S8/S9). The teacher also refers in her statements to statements made by the students. She paraphrases and sharpens the students' statements (T1) and adapts her question to them (T2).

### *focus*

The statements by the students and the teacher all focus on the underlying question. There are no distractions, and the rules of conversation are observed. No summaries or transitions are visible in this section. A summary would have been helpful, especially regarding the important distinction between knowledge/facts and belief. The teacher does not recognise the importance of this issue for the underlying question; consequently, no deepening takes place.

### *restrained facilitation*

The teacher gives the students plenty of space to talk to each other and keeps her share of the speaking to a minimum. Her questions are adaptive and refer to statements made by the students (T1/T2). However, deepening the central point of the students' discussion falls short. Furthermore, the teacher – presumably unconsciously – expresses an evaluation by positively assessing a possible mutual confirmation of religion and science (T2).

In another excerpt, the class discusses similarities and differences between humans and animals. The philosophical dialogue deals with the relationship between humans and animals and the consequences that follow from this relationship.

- T1: Let's go to the basic question, what do animals and humans have in common? Or maybe also what are the differences – both.
- S2: So, animals and humans have feelings; both eat, both drink.
- T2: And what are the differences?



- S3: No idea.
- S4: We are smarter.
- T3: How do you know?
- S5: We build houses, and we build roads, and we have cars, and they still live in the wild, and we build things, and they are, yes.
- T4: We develop things, for example, tools and such. Don't animals do that?
- S6: No.
- S7: Monkeys do. I've seen a video of a monkey using a rock to crack a coconut. That's also a tool that it built.
- S8: Or the ants also build their houses.
- S9: Yes, that's right.
- T5: Hmm, could an animal now also build a complex vehicle like a car?
- S10: Maybe.
- T6: So, they both need tools, and animals build houses too, but we might agree that humans can build or make more complex things, develop tools to use them.
- T7: Do you see any other differences between humans and animals? Something we do and can do and animals don't do?
- S11: So, I think we can also plan for the future, and animals don't even know that they are dying.
- T8: Plan the future. Give me an example. For example, what are your plans now for the future, maybe the next few weeks or something?
- S12: Yes, actually nothing. But I don't know if the animal knows if it's going to die anyway or something.
- T9: So, you think animals don't even realise that life eventually comes to an end?
- S13: Yeah, I don't know what they are thinking.
- S14: I have the feeling that they don't study around at all. Well, they don't ask themselves that, they just live it and take it as it is. They don't even think about how it could be otherwise. I have a feeling they can't even think that far.
- T10: Hmm, think so far. You said something important. Thinking about the future, planning vacations, thinking about what it would be like to live in America. Do you think animals can do that too?
- S15: I do not think so.
- T11: So, is that actually something that distinguishes us from animals?
- S16: Yes.

### *philosophical richness*

In response to the teacher's question (T1), a student formulates various similarities between humans and animals (S1). However, the topic of similarities is abandoned rather abruptly and not further differentiated. Various, substantial factors are proposed on the question of differences: intelligence (S4), building things (S5), using tools (T4/S7), planning one's own future (S11), being aware of one's own death or life (S11/S14). In this dimension, it is irrelevant whether the teacher or the students introduce or elaborate on these points. Although these key areas are mentioned, they are not further described or clarified. It remains unclear how the use of tools differs between animals and humans. It remains just as open to what extent human and animal intelligence, thinking, consciousness and planning can be differentiated. The areas are therefore only treated superficially. At various points, views are supported with examples (S5/S7/S8). Elsewhere, students find it difficult to further justify suggestions (S12/S13).

### *co-construction*

The teacher often behaves in a co-constructive manner. She paraphrases statements made by the students (T4/T9/T10) and asks specific questions (T4/T8/T11). Sometimes, however, it seems as if the teacher is putting statements into the students' mouths that go beyond anything they have expressed (T4/T6). Two students contradict a statement by a colleague who attributes the use of tools exclusively to humans (S7/S8). Otherwise, the students do not behave in a co-constructive manner without the support of the teacher. The conversation is like a teacher-student ping-pong. Some progress is visible, but the topics change too quickly and are therefore not ultimately clarified. In terms of content, there is hardly any progress in relation to the question of using tools or planning for the future.

### *focus*

Most of the statements contribute to answering the underlying question. However, the question about one's own plans for the future in the next few weeks serves no purpose and tends to lead away from the topic (T8). The teacher tries to differentiate or summarise statements (T6, T10). It would make sense to relate or

contrast statements more often and to let the students react to them. The philosophical potential of individual comments is only partially recognised. The characteristics of planning for one's own future and being aware of one's own existence should have been deepened, given their importance for the difference between humans and animals. Discussion rules are followed, and there are no distractions.

### *restrained facilitation*

As already mentioned, the teacher puts certain statements in the students' mouths that go beyond anything they actually expressed (T6). It sometimes seems as if she had been waiting for the opportunity to introduce certain keywords, such as 'tools' (T4). Thus, she pursues a certain goal (T5) instead of following the students' trains of thought. At various points, however, the teacher asks specific questions and adapts them to the statements made by the students (T3/T8/T9). The teacher is rather dominant in terms of content and takes up a large part of the discussion. It would be more effective to ask more moderate questions and to invite the opinions of other students.

### *evidence of the high-inference rating instrument's validity*

According to the American Educational Research Association et al. (2014), validity describes "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11). There are various sources of evidence: evidence based on test content, evidence based on response processes, evidence based on internal structure, evidence based on relations to other variables and evidence for the validity and consequences of testing (American Educational Research Association et al., 2014). Since the development of the instrument and its validation were not the focus of the Swiss research project, only individual sources of evidence are addressed in this chapter.

The results of the rating of the 63 philosophical dialogues in the study in Figure 1 show that the instrument is able to record different levels of quality.

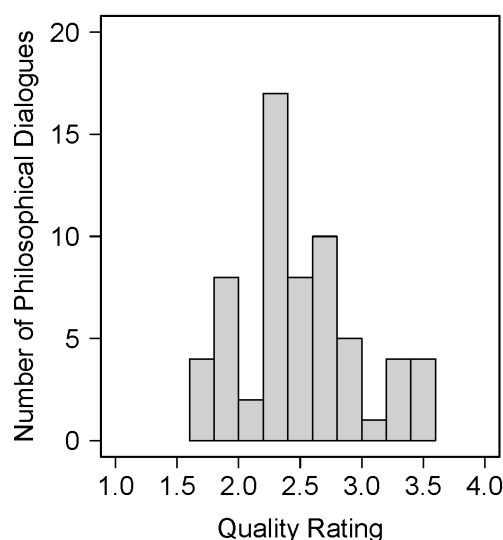


Figure 1: Histogram of the quality of philosophical dialogues,  $N = 63$ .

The crucial question now is whether these differences can be validly measured. In the following sections, we therefore briefly discuss evidence based on test content, evidence based on internal structure and evidence based on relations to other variables. In addition, we address interrater reliability in the assessment of the dialogues as an aspect of objectivity.

#### *evidence based on test content*

To ensure the fit of content and construct, the instrument was developed based on existing literature on philosophical dialogue and improved in discussions with experts. The references to literature on the dimensions and indicators are listed in the Appendix (Supplementary File 1: High-Inference Rating Instrument).

#### *evidence based on internal structure*

The dimensionality of the instrument was tested with the help of an exploratory factor analysis (maximum likelihood) with oblique rotation (oblimin) (Bühner, 2011; Preacher & MacCallum, 2003). A factor with an eigenvalue  $> 1$  could be detected, which explains 43% of the variance. Cronbach's alpha, i.e. internal consistency, was 0.72. This value is acceptable. The values in the factor analysis indicate how strongly a particular dimension correlates with the underlying construct, the quality of philosophical dialogue. A high value shows a high



correlation and indicates that the variable is strongly connected to the dimension. This means that the variable makes a significant contribution to explaining this construct. Table 4 shows that each dimension makes a relevant contribution to the construct of quality, though to varying degrees.

Dimension	Factor
	1
Philosophical Richness	.86
Co-construction	.74
Focus	.46
Restrained Facilitation	.45
Eigenvalue	1.71
Explained variance	0.43

Table 4: Results of the maximum likelihood factor analysis with oblimin rotation to capture the dimensionality of the quality of philosophical dialogue, N = 63

### *evidence based on relations to other variables*

If theoretically based connections with other constructs can also be determined empirically, this can provide a source of evidence for validity. In our research project, based on numerous theoretical and empirical works, we assumed that the quality of philosophical dialogue has a positive influence on acceptance of evolution. Figure 2 presents a latent change score model (Kievit et al., 2018; McArdle & Grimm, 2010) of the impact of the quality of philosophical dialogue on the change in acceptance of evolution. A higher quality of dialogue leads to a greater increase in acceptance of evolution ( $\beta = .214$ ,  $p = .009$ ). The model explains 19.9% of the variance in the change in acceptance of evolution. The theoretically assumed connection was thus empirically confirmed, providing evidence of the instrument's validity. The acceptance of evolution was measured using a questionnaire before and after the intervention (Bernhard, 2022)

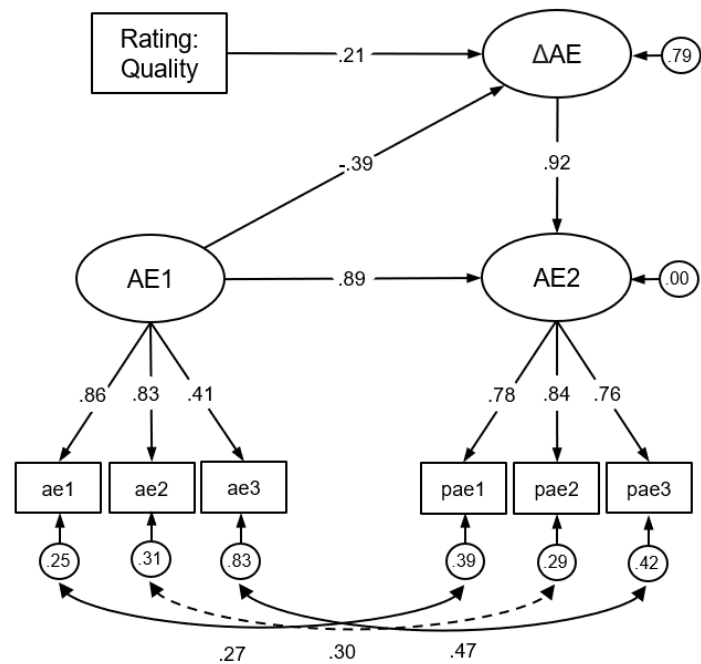


Figure 2: Latent change score model for the philosophy group<sup>4</sup> ( $N = 180$ ) with the dependent variable acceptance of evolution and the independent variable quality of philosophical dialogue. Solid lines indicate significant relationships, dashed lines non-significant relationships. AE1 = pre-test acceptance, ae1/ae2/ae3 = pre-test acceptance items, AE2 = post-test acceptance, pae1/pae2/pae3 = post-test acceptance items,  $\Delta AE$  = change in acceptance, Rating: Quality = quality of the philosophical dialogue.

### *interrater reliability*

Two evaluators, the author and a master student, rated the philosophical dialogues from the study. The master student familiarised herself with the introductory materials for the teachers of the first dialogue as well as the rating instrument and discussed these with the author. Subsequently, various discussions on the first philosophical question were jointly evaluated, allowing the resulting judgments to be discussed and clarified. The intraclass correlation (ICC) was calculated to determine interrater reliability. Since 27 of the 63 philosophical dialogues were assessed independently by both raters, values were evaluated using a two-factor analysis of variance. The rating aimed to achieve the greatest possible agreement, so an unadjusted ICC was calculated (Koo & Li, 2016; Wirtz & Caspar, 2002). The results may be seen in Table 5.

<sup>4</sup> This refers to the group that participated in three philosophical dialogues.

Overall quality	0.903 (Confidence interval: 0.800 < ICC < 0.954)
Philosophical richness	0.705 (Confidence interval: 0.455 < ICC < 0.853)
Co-construction	0.849 (Confidence interval: 0.697 < ICC < 0.928)
Focus	0.662 (Confidence interval: 0.379 < ICC < 0.831)
Restrained facilitation	0.851 (Confidence interval: 0.701 < ICC < 0.929)

Table 5: Intraclass correlation (ICC) for overall quality and each dimension,  $N = 27$

There is no precise definition of when the ICC is considered sufficient. According to Koo and Li (2016), values above 0.5 can be interpreted as moderate reliability, values above 0.75 as good reliability and values above 0.90 as excellent reliability. It can be stated that overall agreement on assessments for this instrument is very high. However, reliability is lower for the focus dimension.

### *conclusion*

In this section, we critically discuss the potential of the developed high-inference rating instrument. We consider the limitations of our study, possible further developments and the opportunities for application in research and teacher training.

The instrument presented offers an innovative approach to holistically assessing philosophical dialogues and reflects many key features of quality. The analyses of the 63 class discussions show that the instrument enables reliable measurement. In addition, the use of the instrument is practical and comparatively simple, even if training is necessary.

Because the instrument was developed during the study, no cross-validation could take place. Consequently, it would be important to have dialogues evaluated both by different groups and by additional raters in order to obtain reliable results. Another limitation of both the instrument and the study is the omission of the distribution of student statements. Since the study analysed audio recordings, it was not possible to track which students, or how many, participated. However, with video recordings and an additional indicator in the dimension of co-construction, this factor could be taken into account in the future.

Furthermore, a sharpening and (further) differentiation of the individual dimensions, especially philosophical richness, is necessary. As can be seen in the exploratory factor analysis, this dimension is a kind of 'super factor' that seems to contain a lot of information from the other dimensions. Differentiation could be supported, for example, by formulating the levels of the individual indicators.

Another challenge is the distinction between statements made by the teacher and those made by the students. With the present instrument, this distinction is only partially taken into account in the dimensions of co-construction and restrained facilitation. Nevertheless, it is only partially possible to make statements about the contribution of the teacher to the quality of the dialogue. To examine this area more closely, low-inference methods could be helpful. Lotz et al. (2013) advocate a combination of the two procedures, which would allow for mutual validation and provide a more comprehensive picture of a dialogue. Thus, a possible further development would be the addition of a low-inference instrument in which, for example, the types of talk moves (introducing new content perspectives, formulating hypotheses, etc.) by teachers and students are recorded. This type of classroom discourse analysis is already widespread, as can be seen in the contributions of Cassidy and Christie (2013) or Moser et al. (2022).

The potential of a high-inference rating instrument such as the one introduced here is evident both in teacher training and in research. When pre-service teachers engage in the method of PwC, they need criteria for what constitutes a dialogue of good philosophical quality. The instrument can be used to analyse and reflect on conversations and can serve as a basis for discussion. In this way, the pre-service teachers' ideas about philosophical dialogue can be sharpened. This is especially important because learning the competencies required to apply PwC has proven challenging (Akkocaoğlu Çayır, 2019; Mathis & Conrad, 2015).

An effective means of measuring the quality of philosophical dialogue is a key contribution to the research. As with all teaching methods, the quality of implementation is decisive in determining success. The instrument offered here can be used to examine the effects of PwC in a more differentiated way. In addition, it



opens up an opportunity to check the effectiveness of teacher training in the field of philosophising.

## references

- Akkocaoğlu Çayır, N. (2019). Philosophy for Children in Teacher Education: Effects, Difficulties, and Recommendations. *International Electronic Journal of Elementary Education*, 11(2), 173–180. <https://doi.org/10.26822/iejee.2019248591>
- Alt, K. (2018). Sprachbildung im Gedankenaustausch – Welchen Beitrag kann das Philosophieren mit Kindern zur Sprachbildung von Kindern im Vorschulalter leisten? In H. de Boer & K. Michalik (Eds.), *Philosophieren mit Kindern - Forschungszugänge und -perspektiven* (pp. 81–90). Verlag Barbara Budrich.
- American Educational Research Association; American Psychological Association; National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Bernhard, D. (2022). *Philosophieren im naturwissenschaftlichen Unterricht: Eine Interventionsstudie in Biologie zum Thema Evolution* [Dissertation]. Pädagogische Hochschule Heidelberg, Heidelberg. <https://opus.ph-heidelberg.de/frontdoor/index/index/docId/437>
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3., aktualisierte und erw. Aufl.). PS Psychologie. Pearson Studium. <http://lib.myilibrary.com/detail.asp?id=404890>
- Brüning, B. (Ed.). (2016). *Ethik/Philosophie Didaktik*. Cornelsen Verlag GmbH. <http://ebookcentral.proquest.com/lib/gbv/detail.action?docID=4776663>
- Brüning, B. (2004). Philosophieren in der Grundschule: Methoden und internationale Bilanz. In H.-J. Müller & S. Pfeiffer (Eds.), *Denken als didaktische Zielkompetenz: Philosophieren mit Kindern in der Grundschule; Tagung "Philosophieren mit Kindern in der Grundschule" 23./24. April in Oldenburg* (pp. 32–41). Schneider-Verl. Hohengehren.
- Brüning, B. (2003). *Philosophieren in der Sekundarstufe: Methoden und Medien*. Beltz.
- Cam, P., & Beck, U. (1996). *Zusammen nachdenken: Philosophische Fragestellungen für Kinder und Jugendliche; eine praktische Einführung*. Verl. an der Ruhr.
- Camhy, D. G. (2013). Philosophieren mit Kindern aus kosmopolitischer Perspektive.: Komplexes Denken und das Konzept der Community of Inquiry. *Pädagogische Rundschau*, 67(6), 731–740.
- Cassidy, C., & Christie, D. (2013). Philosophy with children: Talking, thinking and learning together. *Early Child Development and Care*, 183(8), 1072–1083. <https://doi.org/10.1080/03004430.2013.773509>
- Clausen, Marten, Reusser, Kurt, & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen. Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31(2), 122–141.
- Collins, C. (2007). *The Role of Dialogue-Based Ethical Inquiry in Educating for a Just Democracy: An Intervention Study*. <https://ep.liu.se/ecp/021/vol1/005/ecp2107005.pdf>
- Colom, R., García Moriyón, F., Magro, C. M., & Morilla, E. (2014). The Long-term Impact of Philosophy for Children: A Longitudinal Study (Preliminary Results). *Analytic Teaching and Philosophical Praxis*, 35(1).
- Daniel, M.-F. (2008). Learning to Philosophize: Positive Impacts and Conditions for Implementation. *Thinking: The Journal of Philosophy for Children*, 18(4), 36–48. <https://doi.org/10.5840/thinking200818415>

- Daniel, M.-F. (2012). Pupils' Age and Philosophical Praxis: Two Factors Influence the Development of Critical Thinking in Children. *Childhood and Philosophy*, 8(15), 105–130.
- Daniel, M.-F. (2021). In the Footsteps of Matthew Lipman: Dialogue among Peers and Dialogical Thinking. *Analytic Teaching and Philosophical Praxis*, 41(1), 62–79.
- Daniel, M.-F., Lafortune, L., Pallascio, R., Splitter, L., Slade, C., & La Garza, T. de (2005). Modeling the Development Process of Dialogical Critical Thinking in Pupils Aged 10 to 12 Years. *Communication Education*, 54(4), 334–354. <https://doi.org/10.1080/03634520500442194>
- de Boer, H. (2015). Partizipation und Bildung im Gespräch - Studierende philosophieren mit Kindern. In H.-J. Fischer, H. Giest, & K. Michalik (Eds.), *Probleme und Perspektiven des Sachunterrichts: Vol. 25. Bildung im und durch Sachunterricht* (pp. 159–166). Klinkhardt Julius.
- Fisher, R. (2007). Dialogic teaching: Developing thinking and metacognition through philosophical discussion. *Early Child Development and Care*, 177(6-7), 615–631. <https://doi.org/10.1080/03004430701378985>
- García-Moriyón, F., Rebollo, I., & Colom, R. (2005). Evaluating Philosophy for Children. *Thinking: The Journal of Philosophy for Children*, 17(4), 14–22. <https://doi.org/10.5840/thinking20051743>
- Haynes, J. (2008). *Children as philosophers: Learning through enquiry and dialogue in the primary classroom*. Routledge/Falmer.
- Helbling, D. (2018). Der Fraglichkeit der Welt mit nachdenklichem Lernen begegnen: Philosophieren in Natur, Mensch, Gesellschaft. In D. Helbling & P. Trevisan (Eds.), *Nachdenken und vernetzen in Natur, Mensch, Gesellschaft (E-Book): Studienbuch für den kompetenzorientierten Unterricht im 1. und 2. Zyklus* (1st ed., pp. 59–88). hep verlag.
- Helzel, G., & Michalik, K. (2015). Kindliche Entwicklungsprozesse beim Philosophieren mit Kindern: Eine empirische Untersuchung zu Mehr-Perspektivität und Ungewissheitstoleranz. In H.-J. Fischer, H. Giest, & K. Michalik (Eds.), *Probleme und Perspektiven des Sachunterrichts: Vol. 25. Bildung im und durch Sachunterricht* (pp. 189–196). Klinkhardt Julius.
- Hofheinz, V. (2010). Das Wesen der Naturwissenschaften: Was die Naturwissenschaften ausmacht. *Unterricht Chemie*, 21(118/119), 8–13.
- Hugener, I., Pauli, C., & Klieme, E. (Eds.). (2006). *Materialien zur Bildungsforschung: Vol. 15. Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie: "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis". 3. Videoanalysen*. GfPF. <http://nbn-resolving.de/urn:nbn:de:0111-opus-31304>
- Jackson, T. (2013). Philosophical Rules of Engagement. In S. Goering, N. J. Shudak, & T. E. Wartenberg (Eds.), *Routledge Studies in Contemporary Philosophy: Vol. 47. Philosophy in schools: An introduction for philosophers and teachers* (pp. 99–109). Routledge.
- Kennedy, D. (2004). The role of a facilitator in a community of philosophical inquiry. *Metaphilosophy*, 35(5), 744–765.
- Kennedy, D. (2013). Developing Philosophical Facilitation: A Toolbox of Philosophical Moves. In S. Goering, N. J. Shudak, & T. E. Wartenberg (Eds.), *Routledge Studies in Contemporary Philosophy: Vol. 47. Philosophy in schools: An introduction for philosophers and teachers* (pp. 110–118). Routledge.
- Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., Mooij, S. M. M. de, Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience*, 33, 99–117. <https://doi.org/10.1016/j.dcn.2017.11.007>





- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Krüger, B., & Schick, D. (2012). *Wie wollen wir leben? Kinder philosophieren über Nachhaltigkeit*. oekom.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Waxmann.
- Lederman, N. G., Bell, R. L., Schwartz, R. S., & Abd-El-Khalick, F. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, 39(6), 497–521. <https://doi.org/10.1002/tea.10034>
- Lipman, M. (1988). *Philosophy goes to school*. Temple Univ. Press.
- Lipman, M. (2003). *Thinking in education* (2nd ed). Cambridge University Press.
- Lipman, M. (2009). Philosophy for Children: Some Assumptions and Implications. In E. Marsal (Ed.), *Hodos - Wege bildungsbezogener Ethikforschung in Philosophie und Theologie: Bd. 9. Children philosophize worldwide: Theoretical and practical concepts* (pp. 23–46). Lang.
- Lotz, M., Gabriel, K., & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung. *Zeitschrift Für Pädagogik*, 59(3), 357–380.
- Marsal, E. (2009). A German Framework for Philosophyzing with Children: The Five-Finger Model of Ekkehard Martens. In E. Marsal (Ed.), *Hodos - Wege bildungsbezogener Ethikforschung in Philosophie und Theologie: Bd. 9. Children philosophize worldwide: Theoretical and practical concepts* (pp. 503–512). Lang.
- Marsal, E. (2014). Socratic Philosophizing with the Five Finger Model: The Theoretical Approach of Ekkehard Martens. *Analytic Teaching and Philosophical Praxis*, 35(1), 39–49.
- Martens, E. (2004). Philosophieren mit Kindern als elementare Kulturtechnik. In H.-J. Müller & S. Pfeiffer (Eds.), *Denken als didaktische Zielkompetenz: Philosophieren mit Kindern in der Grundschule; Tagung "Philosophieren mit Kindern in der Grundschule" 23./24. April in Oldenburg* (pp. 7–18). Schneider-Verl. Hohengehren.
- Martens, E. (2010). Wozu Philosophie in der Schule? In K. Meyer (Ed.), *Universal-Bibliothek: Vol. 18723. Texte zur Didaktik der Philosophie* (pp. 156–172). Reclam.
- Martens, E. (2003). *Methodik des Ethik- und Philosophieunterrichts: Philosophieren als elementare Kulturtechnik*. Siebert.
- Mathis, C., & Conrad, S.-J. (2015). Kompetent mit Kindern philosophieren lernen. In H.-J. Fischer, H. Giest, & K. Michalik (Eds.), *Probleme und Perspektiven des Sachunterrichts: Vol. 25. Bildung im und durch Sachunterricht* (pp. 121–128). Klinkhardt Julius.
- McArdle, J. J., & Grimm, K. J. (2010). Five steps in Latent Curve and Latent Change Score Modeling with Longitudinal Data. In K. van Montfort, J. H. Oud, & A. Satorra (Eds.), *Longitudinal Research with Latent Variables* (pp. 245–274). Springer Berlin Heidelberg.
- McComas, W., & Olson, J. K. (1998). The Nature of Science in International Science Education Standards Documents. In W. McComas (Ed.), *Science & Technology Education Library: Vol. 5. The Nature of Science in Science Education: Rationales and Strategies* (pp. 41–52). Kluwer Academic Publishers.
- Meerwaldt, D., Borromeo, R., & Nevers, P. (2018). Philosophizing with Children in the Course of Solving Modeling Problems in a Sixth Grade Mathematics Classroom. *Analytic Teaching and Philosophical Praxis*, 34(1), 80–92.
- Michalik, K. (2009). Philosophieren mit Kindern als Unterrichtsprinzip und die Förderung von Wissenschaftsverständnis im Sachunterricht. In K. Michalik (Ed.), *Grundschulpädagogik interdisziplinär: Vol. 2. Philosophie als Bestandteil wissenschaftlicher Grundbildung? Möglichkeiten der Förderung des*

- Wissenschaftsverständnisses in der Grundschule durch das Philosophieren mit Kindern* (pp. 27–42). Lit-Verl.
- Michalik, K. (2013). Philosophieren mit Kindern als Unterrichtsprinzip: Bildungstheoretische Begründungen und empirische Fundierungen. *Pädagogische Rundschau*, 67, 635–649.
- Michalik, K. (2016). Staunen, Fragen, Nachdenken: Philosophieren. *Die Grundschulzeitschrift*, 30(297), 42–45.
- Michalik, K. (2004a). "Man kann nur ein bisschen wissen, nicht alles": Gespräche im Sachunterricht. *Grundschule*(5), 40–43.
- Michalik, K. (2004b). Denken wir eigentlich immer im Unterricht? In H.-J. Müller & S. Pfeiffer (Eds.), *Denken als didaktische Zielkompetenz: Philosophieren mit Kindern in der Grundschule; Tagung "Philosophieren mit Kindern in der Grundschule" 23./24. April in Oldenburg* (pp. 49–59). Schneider-Verl. Hohengehren.
- Michalik, K. (2008). Denken dürfen: Philosophieren mit Kindern als Kern des Grundschulunterrichts. *Grundschule*(12), 7–11.
- Michalik, K. (2016). Staunen, Fragen, Nachdenken: Philosophieren. *Die Grundschulzeitschrift*, 30(297), 42–45.
- Moser, M., Zimmermann, M., Pauli, C., Reusser, K., & Wischgoll, A. (2022). Student's vocal participation trajectories in whole-class discussions during teacher professional development. *Learning, Culture and Social Interaction*, 34, 100633. <https://doi.org/10.1016/j.lcsi.2022.100633>
- Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics*, 2(1), 13–43. [http://quantpsy.org/pubs/preacher\\_maccallum\\_2003.pdf](http://quantpsy.org/pubs/preacher_maccallum_2003.pdf)
- Rakoczy, K., & Pauli, C. (2006). Hoch inferentes Rating: Beurteilung der Qualität unterrichtlicher Prozesse. In I. Hugener, C. Pauli, & E. Klieme (Eds.), *Materialien zur Bildungsforschung: Vol. 15. Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie: "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis"*. 3. Videoanalysen (pp. 206–233). GPPF.
- Santi, M. (1993). Philosophizing and Learning to Think: Some Proposals for a Qualitative Evaluation. *Thinking: The Journal of Philosophy for Children*, 10(3), 16–22. <https://doi.org/10.5840/thinking199310318>
- Säre, E., Luik, P., & Tulviste, T. (2016). Improving pre-schoolers' reasoning skills sing the philosophy for children programme. *Trames. Journal of the Humanities and Social Sciences*, 20(3), 273. <https://doi.org/10.3176/tr.2016.3.03>
- Schleifer, M., & Courtemanche, L. (1996). The Effect of Philosophy for Children on Language Ability. *Thinking: The Journal of Philosophy for Children*, 12(4), 30–31. <https://doi.org/10.5840/thinking19961246>
- Seidel, T. (2005). Video analysis strategies of the IPN Video Study - a methodological overview. In T. Seidel, M. Prenzel, & M. Kobarg (Eds.), *How to run a video study: Technical report of the IPN Video Study* (pp. 70–78). Waxmann.
- Sharp, A. M. (2009). The Community of Inquiry as Ritual Participation. In E. Marsal (Ed.), *Hodos - Wege bildungsbezogener Ethikforschung in Philosophie und Theologie: Bd. 9. Children philosophize worldwide: Theoretical and practical concepts* (pp. 301–306). Lang.
- Sharp. (2018). A letter to a novice teacher: Teaching Stottlemeier's Discovery. In A. M. Sharp, M. Laverty, & M. R. Gregory (Hrsg.), *In community of inquiry with Ann Margaret Sharp: Childhood, philosophy and education* (S. 88–96). Routledge, Taylor & Francis Group.
- Siddiqui, N., Gorard, S., & See, B. H. (2019). Can programmes like Philosophy for Children help schools to look beyond academic attainment? *Educational Review*, 71(2), 146–165. <https://doi.org/10.1080/00131911.2017.1400948>





- Siddiqui, N., Gorard, S., & See, B. H. (2022). Can We Teach Non-Cognitive Outcomes? A Quasi-Experimental Study of Philosophy for Children. *Education Sciences*, 12(5), 322. <https://doi.org/10.3390/educsci12050322>
- Walker, C. M., Wartenberg, T. E., & Winner, E. (2013). Examining the Effects of Philosophy Classes on the Early Development of Argumentation Skills. In S. Goering, N. J. Shudak, & T. E. Wartenberg (Eds.), *Routledge Studies in Contemporary Philosophy: Vol. 47. Philosophy in schools: An introduction for philosophers and teachers* (pp. 277–287). Routledge.
- Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Hogrefe Verl. für Psychologie.
- Worley, E., & Worley, P. (2019). teaching critical thinking and metacognitive skills through philosophical enquiry. A practitioner's report on experiments in the classroom. *Childhood & Philosophy*, 15, 01–34. <https://doi.org/10.12957/childphilo.2019.46229>
- Wartenberg, T. E. (2013). Elementary School Philosophy. In S. Goering, N. J. Shudak, & T. E. Wartenberg (Eds.), *Routledge Studies in Contemporary Philosophy: Vol. 47. Philosophy in schools: An introduction for philosophers and teachers* (pp. 34–41). Routledge.
- Yan, S., Walters, L. M., Wang, Z., & Wang, C.-C. (2018). Meta-Analysis of the Effectiveness of Philosophy for Children Programs on Students' Cognitive Outcomes. *Analytic Teaching and Philosophical Praxis*, 39(1), 13–33.
- Zoller, E. (2015). *Selber denken macht schlau: Philosophieren mit Kindern und Jugendlichen; Anregungen für Schule und Elternhaus* (3. Auflage). Zytglogge Verlag.

### *appendix: supplementary File 1*

#### ***High-Inference Rating Instrument***

##### ***Philosophical richness***

###### *Basic idea:*

The criterion of philosophical richness captures the extent to which different questions and contributions are illuminated in a differentiated manner using philosophical methods (Martens, 2003). The process of capturing this is supported by questions based on the maieutic method, which is used by the teacher or possibly the students. Depending on their objective, these questions can be divided into four types (Helbling, 2018), on which the formulation of the first four indicators is based: describing facts and clarifying terms, working out differences and similarities, justifying opinions and reflecting upon evaluations, and creating hypotheses and considering consequences (thought experiments). Accordingly, in philosophical dialogue, concepts as well as phenomena and situations should be carefully considered and described. Similarities, differences and connections should be

reflected on and formulated. Another aspect of philosophical dialogue is the assessment of consequences and speculative thinking. Opinions expressed should always be justified and supported with arguments and examples. In order to illuminate a philosophical question in a differentiated manner, participants must express diverse and substantial thoughts. Thus, this category captures, on the one hand, the diversity and substance of the statements and, on the other hand, the differentiation of the philosophical reflection, jointly manifesting thorough thinking about these ideas.

*Indicators:*

- Terms, phenomena and situations are described and clarified.
- Similarities, differences and connections are formulated.
- Justifications or examples supporting opinions are expressed and evaluations are reflected upon.
- Hypotheses are formulated, consequences assessed and speculations expressed.
- Diverse and substantial thoughts are expressed.

*Answer:* overall impression

*Annotation:*

A very high level (4) means that at least the first three indicators and the fifth indicator are met in a highly differentiated way. A very high level is possible even without incorporating the fourth indicator. In addition to the number of indicators, the frequency and quality of conversational behaviour are decisive for the level of this characteristic.

***Co-construction***

*Basic idea:*

The co-construction criterion captures the extent to which the teacher's and students' statements and questions relate to and build on each other. Co-construction can be promoted by appropriate moderation questions from the teacher or possibly the students. A dialogue is considered co-constructive if the people involved refer to previous statements in their statements and rephrase,

supplement, question or contrast with them. The views of students should therefore not be expressed merely additively. Any ambiguities should be clarified by (follow-up) questions.

*Indicators:*

- Statements refer to previous statements by repeating, rephrasing, supplementing or differentiating them.
- Statements relate to previous statements by questioning or contrasting with them.
- Unclear statements are taken up and clarified by (follow-up) questions.
- Progress is evident in successive statements. Statements are not primarily made in an additive manner.

*Answer:* overall impression

*Annotation:*

A very high level (4) of this characteristic requires the contributions of all indicators. A very high level (4) is possible only if students sometimes behave co-constructively and converse with each other without the support of the teacher. Ping-pong conversations (teacher asks a question, students answer, teacher asks the next question, without any visible build-up or thinking on the part of the students) correspond to a low proficiency (1).

***Focus***

*Basic idea:*

The focus criterion captures the extent to which the statements and questions from the teacher and students focus on answering the philosophical question. The question should be examined from different perspectives, but the focus should not be lost. The contributions of the students should support the answering of the underlying question. By choosing appropriate questions, the teacher should also contribute to the focus on the original philosophical question. Summaries, transitions and questions that focus on the underlying issue should create a common thread in the conversation. Disturbances in the course of the dialogue due to non-observance of the rules of conversation and behaviour have a negative effect

on attention and focus. A particular challenge for the teacher is to recognise statements with special potential for the topic and to deepen them in conversation.

*Indicators:*

- Questions and statements contribute to answering the underlying question. The conversation does not digress in content.
- Summaries, transitions and questions that focus on the underlying issue create a common thread in the conversation.
- The underlying philosophical question is explicitly taken up again during or at the end of the dialogue.
- Basic dialogue rules are followed so that the dialogue is not disturbed or interrupted.
- The teacher recognises statements with potential for the topic and deepens them with specific questions.

*Answer:* overall impression

*Annotation:*

For a very high level (4) of this characteristic, all indicators must be taken into account. In addition to the number of indicators, the frequency and quality with which a conversational behaviour is shown are decisive for the level of the characteristic.

***Restrained facilitation***

*Basic idea:*

The art of restrained facilitation is to promote co-constructive dialogue and philosophical richness in a topic-focused manner without assuming too dominant a role. The moderator, in this case the teacher, behaves with restraint in philosophical dialogue and does not express his or her own views. Whenever possible, the teacher follows the students' thoughts and deepens them. The teacher, therefore, does not follow his or her own plan of dialogue but adapts the questions to the students' statements. A teacher-student ping-pong should be avoided; accordingly, the teacher must give the participants enough space to talk to each



other. The students should also be allowed to develop their trains of thought without being too quickly restricted by the teacher due to the focus of the topic.

*Indicators:*

- The teacher refrains from expressing personal opinions and evaluations and does not provide any answers.
- The teacher follows the students' trains of thought and deepens them by adapting questions flexibly instead of leading the dialogue toward a specific goal.
- The teacher creates space so that the students can talk to each other and limits her or his share of the speaking.
- The teacher creates space so that the students can develop their own trains of thought without being too quickly restricted by the focus of the topic.

*Answer:* overall impression

*Annotation:*

A very high level (4) in this characteristic means that all indicators are met at a high level. A high level (3, 4) of the characteristic is possible only if the students' share of the speaking is high. If the teacher behaves with restraint, but the conversation is more like a teacher-student ping-pong, the result is a rating of 2. If the teacher does not promote, or barely promotes, co-constructive dialogue and philosophical richness in a topic-focused manner, the result is a low rating (1, 2). If the teacher expresses his or her own opinion several times and provides answers, the result is a very low level (1).

*Theoretical foundation*

The categories and indicators described were created based on the literature. This theoretical foundation is documented in the following table.

Philosophical richness	
Terms, phenomena and situations are described and clarified.	Brüning, 2003, 2004; Cam & Beck, 1996; Helbling, 2018; Krüger & Schick, 2012; Martens, 2004, 2010; Michalik, 2008, 2016; Lipman, 1988

Similarities, differences and connections are formulated.	Cam & Beck, 1996; de Boer, 2015; Fisher, 2007; Haynes, 2008; Helbling, 2018; Kennedy, 2013; Krüger & Schick, 2012; Lipman, 1988
Justifications or examples supporting opinions are expressed and evaluations reflected upon.	Brüning, 2003, 2016; Fisher, 2007; Haynes, 2008; Helbling, 2018; Kennedy, 2013; Krüger & Schick, 2012; Martens, 2004, 2010; Michalik, 2008, 2016; Lipman, 1988
Hypotheses are formulated, consequences assessed and speculations expressed.	Brüning, 2003, 2004, 2016; Cam & Beck, 1996; Fisher, 2007; Haynes, 2008; Helbling, 2018; Martens, 2004, 2010, 2012; Michalik, 2008, 2016; Lipman, 1988
Diverse and substantial thoughts are expressed.	Michalik, 2016; Martens, 2004
<b>Co-construction</b>	
Statements refer to previous statements by repeating, rephrasing, supplementing or differentiating them.	Daniel et al., 2005; Daniel, 2008; Fisher, 2007; Haynes, 2008; Helzel & Michalik, 2015; Kennedy, 2013; Michalik, 2016; Sharp, 2009
Statements relate to previous statements by questioning or contrasting with them.	Daniel et al., 2005; Daniel, 2008; Fisher, 2007; Haynes, 2008; Helzel & Michalik, 2015; Kennedy, 2013; Krüger & Schick, 2012; Michalik, 2016; Sharp, 2009
Unclear statements are taken up and clarified by (follow-up) questions.	Cam & Beck, 1996; Kennedy, 2013; Michalik, 2016; Zoller, 2015; Sharp, 2009
Progress is evident in successive statements. Statements are not primarily made in an additive manner.	Daniel, 2008; Michalik, 2004b; Michalik, 2008; Michalik, 2016; Daniel et al., 2005; Fisher, 2007
<b>Focus</b>	
Questions and statements contribute to answering the underlying question. The conversation does not digress in content.	Cam & Beck, 1996; Jackson, 2013; Michalik, 2016; Zoller, 2015
Summaries, transitions and questions that focus on the underlying issue create a common thread in the conversation.	Brüning, 2003; Cam & Beck, 1996; de Boer, 2015; Krüger & Schick, 2012; Michalik, 2016; Zoller, 2015
The underlying philosophical question is explicitly taken up again during or at the end of the dialogue.	Brüning, 2003; Cam & Beck, 1996; de Boer, 2015; Michalik, 2016
Basic dialogue rules are followed so that the dialogue is not disturbed or interrupted.	Brüning, 2003; Cam & Beck, 1996; Fisher, 2007; Haynes, 2008; Jackson, 2013; Krüger & Schick, 2012; Michalik, 2016
The teacher recognises statements with potential for the topic and deepens them with specific questions.	Camhy, 2013; Krüger & Schick, 2012; Wartenberg, 2013
<b>Restrained facilitation</b>	



The teacher refrains from expressing personal opinions or evaluations and does not provide any answers.	Haynes, 2008; Helbling, 2018; Krüger & Schick, 2012; Michalik, 2016; Wartenberg, 2013
The teacher follows the students' trains of thought and deepens them by adapting questions flexibly, rather than leading the dialogue toward a specific goal.	de Boer, 2015; Jackson, 2013; Michalik, 2004a, 2016
The teacher creates space so that the students can talk to each other and limits her or his share of the speaking.	de Boer, 2015; Helbling, 2018; Michalik, 2004a, 2016
The teacher creates space so that the students can develop their own trains of thought without being too quickly restricted by the focus of the topic.	de Boer, 2015; Haynes, 2008; Jackson, 2013; Krüger & Schick, 2012; Michalik, 2004a, 2016

*submitted: 26.11.2023*

*approved: 13.06.2024*