education sciences

MDPI

Article

Measuring Reflective Inquiry in Professional Learning Networks: A Conceptual Framework

Ariadne Warmoes [1,*], Chris Brown [2], Iris Decabooter [3] and Els Consuegra [1]

[1] BRITE, Vrije Universiteit Brussel, 1050 Brussel, Belgium; els.consuegra@vub.be
[2] Southampton Education School, University of Southampton, Southampton SO17 1BJ, UK; christopher.brown@soton.ac.uk
[3] U-Rise Research Group, Hasselt University, 3590 Diepenbeek, Belgium; iris.decabooter@uhasselt.be
* Correspondence: ariadne.warmoes@vub.be

Abstract: This study introduces a conceptual framework to measure reflective inquiry in professional learning networks (PLNs). The framework represents a continuum of three components: critical dialog, the use of multiple data sources, and depth of reflection. It is validated through 8 h of PLN conversations (2195 contributions). The results show that the majority of dialog within PLNs consisted of free-flowing discussions (64.37%), using personal experiences as data source (92.57%), and descriptive reflection with external attribution (R1: 79.82%). High-level reflective inquiry was rare (0.18%). Findings highlight the need for skilled facilitation and strategies to foster reflective inquiry.

Keywords: reflective inquiry; professional learning networks; measurement instrument; professional development; in-service teachers

1. Introduction

For education to stimulate and support positive societal change, schools and education systems must be populated by the most effective educators. This means a workforce of lifelong learners who engage in continuous learning to ensure that their knowledge and practice responds to the changing needs of society and students (Schleicher, 2012). Collaborative approaches, in which social capital resources (e.g., knowledge and support) are exchanged with colleagues, can serve as an effective way of facilitating such learning (and such approaches are more effective than other forms of professional development (OECD, 2015). In this context, professional learning networks (PLNs)—defined as groups of educators who engage in collaborative learning with the aim of improving teaching and student outcomes—provide a promising means through which to achieve this type of collaborative approach at scale (Stoll, 2015). However, if PLNs are to deliver meaningful change through collaboration, reflective inquiry is essential as it encourages educators to embrace new ideas and work together effectively toward improvement (Katz & Earl, 2010; Poortman et al., 2022; Rincón-Gallardo & Fullan, 2016; Warmoes et al., 2025). This process relies on a reflective, collaborative, and dialogic approach, which helps participants deepen and change their understanding of both themselves and their teaching. Research further confirms that when educators engage in collaborative reflective practices, the likelihood of learning significantly increases (Alzayed & Alabdulkareem, 2020; Mitchell & Sackney, 2011; Vanblaere & Devos, 2016).

Despite ongoing research into PLNs, there remains ambiguity surrounding the concept, processes, and stimulation of reflective inquiry (Poortman et al., 2022). Research gaps persist in understanding how to effectively enhance educators' reflective inquiry,

Academic Editors: Kira Elena Weber and Hendrik Lohse-Bossenz

Received: 31 January 2025
Revised: 27 February 2025
Accepted: 5 March 2025
Published: 8 March 2025

Citation: Warmoes, A., Brown, C., Decabooter, I., & Consuegra, E. (2025). Measuring Reflective Inquiry in Professional Learning Networks: A Conceptual Framework. *Education Sciences*, *15*(3), 333. https://doi.org/10.3390/educsci15030333

measure the success of such approaches, and ascertain their impact on teaching practice and student outcomes (Poortman et al., 2022). Currently, approaches used to measure reflective inquiry are challenging. They often rely on subjective tools, such as self-reports, or labor-intensive methods, like recording and analyzing discourse, limiting studies to a small scale (Brown et al., 2021). Consequently, more effective approaches associated with measuring reflective inquiry need to be established. In response, Brown et al. (2021) previously proposed an AI approach involving the use of AI to analyze and identify aspects of reflective and non-reflective inquiry. Such an approach offers several benefits. AI can take just minutes to process hours of recorded conversation and produce an objective report to indicate whether educators in a PLN are engaging in reflective practices. An AI approach could potentially provide an objective, large-scale, quantitative analysis of qualitative data to assess the success of any new approaches to facilitating reflective inquiry.

Correspondingly, this paper aims to measure reflective inquiry objectively and apply such an approach to existing data to gain an initial understanding of how reflective inquiry translates into conversation, while also exploring the potential of AI to help measure reflective inquiry within PLNs. The research questions it seeks to address are as follows:

- How can reflective inquiry in conversations within PLNs be measured?
- How can AI be used to support this measurement approach?
- What insight does this measurement approach provide about the different types of reflective inquiry occurring in PLNs?

2. Theoretical Framework

2.1. Reflective Inquiry in Professional Learning Networks

The potential of collective reflection is further demonstrated in research on PLNs, where reflective inquiry has been identified as a key enactment process influencing positive outcomes (Katz & Earl, 2010; Poortman et al., 2022; Rincón-Gallardo & Fullan, 2016; Warmoes et al., 2025). Research on professional learning communities (PLCs)—which differ from PLNs in terms of their scope and membership (see Brown & Poortman, 2018)—has reported similar findings (e.g., Doğan & Adams, 2018; Vangrieken et al., 2015). Furthermore, multilevel analyses by Vanblaere and Devos (2016) show that, among the various characteristics of PLCs, only reflective dialog significantly contributes to perceived changes in practices.

Reflective inquiry is a concept that encompasses various terms and definitions, such as reflective professional inquiry (e.g., Poortman et al., 2022), critical colleagueship (e.g., Kintz et al., 2015), collaborative inquiry (e.g., Katz & Earl, 2010), and learning conversations (e.g., Earl & Timperley, 2008). While these definitions share common ground, they highlight different aspects that need to be considered. A process-oriented dimension is often central to these definitions, emphasizing the methods through which collective learning occurs. Next to this, elements such as sharing critical opinions, giving feedback, introducing new perspectives, and confronting conflicts are highlighted as essential for fostering learning (Katz & Dack, 2013). Furthermore, some scholars incorporate an explicit layer of inquiry, emphasizing systematic investigation and the application of new ideas (e.g., Earl & Timperley, 2008). Building on these perspectives, Brown et al. (2021) provide a comprehensive definition of reflective inquiry as "a collaborative, dialogic process in which educators both consider and aim to address pressing educational issues or problems. Such a process involves the collective generation and testing of ideas linked to enhancing their own practice, with these ideas based on evidence in the form of literature and/or data and displaying internal attribution" (p. 9).

2.2. Observing Reflective Inquiry

Based on this definition, a measurement instrument should focus on (1) the degree of collaborative and dialogic process (C), (2) the degree to which multiple data sources (D) are used to generate and/or test ideas, and (3) the degree of reflection (R), as evidenced by internal attribution (see Table 1 for the summary). In the next paragraph, we will delve into these three components in greater detail, drawing on the literature to outline each one and proposing a continuum for each component, where the highest levels across all three indicate the presence of reflective inquiry.

Table 1. Summary of the focus and evaluation of reflective inquiry.

Aspect	Collective Dialogue (C)	Data Sources (D)	Reflection (R)
Focus	Interaction	Use of multiple data sources	Individual depth of reflection
Observation	Degree of questioning and engagement	Degree of integration and triangulation of sources	Depth of reasoning and attribution

2.2.1. The Degree of Collaborative and Dialogic Process (C)

Educational change requires us to foster a critical stance toward teaching using confronting traditional practices and adopt an inquiry-driven mindset (Earl & Timperley, 2008). Through talking, new realizations and greater insights come about. However, when a group of people engages in free-flowing discussion, they often settle on topics where there is already a high level of agreement among members, leading to a lack of variability in perspectives (Katz et al., 2009). Earl and Timperley (2008) acknowledge that improvement involves participants taking the time to understand each other's viewpoints, but that the ultimate purpose is to challenge each other's reasoning, upon which the different viewpoints are based.

Therefore, a key aspect of the dialogic process is the intentional disruption of biased thinking (Katz & Dack, 2013). This requires participants to challenge each other's viewpoints, sharing critical opinions, giving feedback, and encouraging critical reflection (Charteris & Smardon, 2014; De Keijzer et al., 2020; Katz & Dack, 2013). Probing questions are important for intentional disruption, as they encourage reflection, synthesis, and critical thinking, helping individuals challenge their assumptions (Kintz et al., 2015). In contrast, clarifying questions focus on gathering information. Kintz et al. (2015) found that communities with deeper discussions used probing questions more often, while shallower discussions relied on clarifying ones. Based on these insights, we propose the following levels of collaborative and dialogic process within reflective inquiry:

C1: The first is free-flowing conversation. At this level, conversations are characterized by free-flowing dialog, often settling on topics where there is already high agreement among members (Katz et al., 2009). Initiating speakers focus on sharing descriptive information or recounting experiences, providing a starting point for discussion, without inviting critical engagement. Responding speakers agree with the initiator's input, often offering their own experiences or perspectives.

C2: The second is engagement through clarifying questions. At this stage, speakers take the time to understand each other's viewpoints (Earl & Timperley, 2008). Initiating speakers frame their input to invite further discussion. Responding speakers expand on the initiator's ideas by asking clarifying questions that help gather information and facilitate understanding (Kintz et al., 2015). Feedback is framed as suggestions, but stays rather practical.

C3: The third is intentional interruption. At this level, initiating speakers introduce topics with probing or thought-provoking questions that challenge existing prac-

tices and assumptions (Kintz et al., 2015) Responding speakers actively challenge each other's viewpoints, share critical opinions, and give feedback (Charteris & Smardon, 2014; De Keijzer et al., 2020; Katz & Dack, 2013).

2.2.2. The Degree of Generating and/or Testing Ideas Based on Multiple Data Sources (D)

The second component of the framework emphasizes the importance of using multiple sources—such as data, evidence, and practical knowledge—to generate or assess ideas for improving practice. While educators often rely on local, anecdotal knowledge because it feels personal and contextually relevant, this knowledge is frequently based on untested assumptions (Earl & Timperley, 2008). However, change requires going beyond intuition by integrating additional data sources and testing assumptions. Triangulating formal data (e.g., classroom observations, assessments), informal data (e.g., professional judgment), and research evidence strengthens decision-making and ensures more robust outcomes (Farley-Ripple et al., 2018; Schildkamp & Datnow, 2020). Rather than prioritizing one type of data over another, it is the thoughtful linking of these sources that creates actionable insights (Kintz et al., 2015). Schildkamp and Datnow (2020) found that teams engaging in higher levels of inquiry integrate diverse data sources to develop new knowledge and focus on actionable plans to improve teaching and learning. Conversely, teams with lower levels of inquiry rely predominantly on informal data, often engaging in storytelling or description without triangulating.

Based on these insights, we propose the following levels of the use of multiple data sources in reflective inquiry:

D1: The first is personal experiences. This level reflects Earl and Timperley's (2008) observation that educators often rely on their rich personal encyclopedias of experiences and actions, which, while valuable, remain untested. Contributions center on storytelling or the re-telling of personal experiences (Schildkamp & Datnow, 2020). There is no reference to formal data or research evidence to support their points. Responding speakers at this level rely on informal data as well, often agreeing or expanding on the initiating speaker's observations without introducing or referencing other sources.

D2: The second is personal experience with reference to another source. At this level, initiating speakers' contributions still rely on personal experiences, but they begin incorporating references to additional sources. These sources can help expand, assess, or compare personal experience and may include formal data (e.g., classroom observations, standardized tests, curriculum and audit reports) and research evidence (Schildkamp & Datnow, 2020). This reflects an effort to move beyond anecdotal knowledge by linking it to other data sources (Farley-Ripple et al., 2018). However, the use of sources remains limited, typically referencing only one additional source, and the connection is made in a thoughtful but not yet fully integrated manner (Kintz et al., 2015).

D3: The third is the balanced use of multiple data sources. At this highest level, initiating speakers engage in the deliberate and balanced integration of multiple data sources to test assumptions and ground their reflections or action plans (Farley-Ripple et al., 2018; Schildkamp & Datnow, 2020). At this level, responding speakers build on the initiating speaker's contributions by offering suggestions informed by evidence, triangulating across multiple sources themselves.

2.2.3. Degree of Reflection and Internal Attribution (R)

Dewey (1910) originally described reflection as the "active, persistent, and careful consideration of any belief or supposed form of knowledge in light of the grounds that support it". Since then, scholars have expanded on this concept, examining its attributes, development, and measurement (e.g., Schön, 1987; Van Manen, 1977). Reflection is not

a binary concept but occurs across levels, from non-reflective description to critical reflection (Van Manen, 1977). Reflection often begins with describing a situation, before progressing toward more complex reflective processes, such as questioning assumptions, comparing different perspectives, and making judgments (Jay & Johnson, 2002; Liu et al., 2017; Walkington et al., 2001). This is in line with other models and protocols of reflection, where structured reflection begins with description, as individuals first need to articulate what they perceive as a problem or challenge before deeper analysis can take place (Brookfield, 1995; Schön, 1983). In higher levels of reflection, individuals recognize the subjectivity of knowledge, critically examine their assumptions, and engage with multiple perspectives (El-Dib, 2006). From this perspective, merely describing a situation does not represent reflection in itself, but it is a necessary step in the reflective process. Reflection cannot occur without first describing the problem or situation being analyzed, as the emergence of initial understandings is essential before moving toward deeper analysis.

In recent years, collective reflection has gained attention for its significant advantages over individual reflection. When educators engage in critical dialog, they can bring presuppositions, ideas, and beliefs to the surface, making these explicit and available for exploration (Hakkarainen et al., 2004). As Nonaka and Takeuchi (1995) emphasize, making tacit knowledge explicit is a key part of professional growth—this involves revealing subjective insights, intuitions, and beliefs that are otherwise difficult to articulate and subjecting them to collaborative critique and refinement. Such dialog requires participants to articulate their perspectives, explain their reasoning, and engage with competing viewpoints, thereby challenging assumptions in the process.

A key component of reflection, particularly in groups, is internal attribution, where educators examine how their own teaching practices influence student outcomes rather than attributing success or failure to external factors like socioeconomic status or family background (Brown et al., 2021; Earl & Timperley, 2008; Farley-Ripple et al., 2018). Teams that move beyond external attribution and adopt internal accountability are more likely to achieve meaningful professional growth and positively impact student outcomes (Mitchell & Sackney, 2011; Schildkamp & Datnow, 2020).

R1: The first level is description. At this level, initiating speakers recount experiences or provide descriptive information (Jay & Johnson, 2002; Liu et al., 2017; Van Manen, 1977; Walkington et al., 2001). Responding speakers acknowledge or expand on the descriptive information shared by the initiator but do not analyze or question it. Such interactions do not make tacit knowledge explicit (Nonaka & Takeuchi, 1995).

R2: The second level is analysis with external attribution. At this level, initiating speakers identify problems or express a need for change, offering explanations, ideas, and beliefs that are articulated and made explicit for exploration (Jay & Johnson, 2002; Hakkarainen et al., 2004; Liu et al., 2017; Walkington et al., 2001). They make tacit knowledge explicit, but explanations are often rooted in external circumstances (Earl & Timperley, 2008). Responding speakers contribute by analyzing the initiator's problem or idea, offering additional perspectives or expanding on the initiator's logic.

R3: The third level is critique with internal attribution. At the highest level, initiating speakers engage in critical reflection by examining their own assumptions, beliefs, and actions (El-Dib, 2006) and expressing personal accountability for their role in the situation (Mitchell & Sackney, 2011; Schildkamp & Datnow, 2020). Initiators move beyond simply describing or analyzing to explicitly question established practices and explore how their own actions contribute to the situation they observe (Brown et al., 2021; Earl & Timperley, 2008; Farley-Ripple et al., 2018). At this level, responding speakers engage critically with the initiator's ideas, evaluating their reasoning and encouraging deeper reflection through internal attribution.

2.2.4. Towards a Framework for Reflective Inquiry

Table 2 summarizes the three aspects of reflective inquiry and the three levels each aspect encompasses. When a conversation is labeled with C3, D3, and R3—indicating a high degree of critical dialog, the use of multiple data sources to generate and/or test ideas, and a high degree of reflection through internal attribution—we classify it as reflective inquiry. However, this framework does not account for actual behavior changes. While these conversations demonstrate the highest levels of reflective inquiry, we cannot establish a direct link between reflective inquiry in conversations and meaningful changes in practice.

Table 2. Framework for reflective inquiry.

	Speaker	Degree of Collective Dialogue (C)	Degree of Use of Multiple Data Sources (D)	Depth of Reflection (R)
1	Initiating	The speaker shares contributions without inviting further engagement.	The speaker relies on personal experiences.	The speakers' contribution consists of descriptive information.
	Responding	The speaker acknowledges others' contributions or shares their own contribution.	The speaker agrees or expands on others' contributions with personal experiences.	The speaker acknowledges or expands on others' contributions with descriptive information.
2	Initiating	The speaker shares contributions and invites input.	The speaker relies mainly on personal experiences but references an additional data source.	The speakers' contribution offers explanations or rationale but relies mostly on external attribution.
	Responding	The speaker engages with others' contributions by asking clarifying questions and/or offering suggestions.	The speaker expands on others' contributions by referencing an additional data source.	The speaker offers reasons or explanations for others' contributions.
3	Initiating	The speaker shares contributions while asking probing questions to provoke critical dialogue and explore assumptions or actions collectively.	The speaker integrates a balanced use of multiple data sources.	The speaker critically reflects on their contribution, offering internally attributed explanations and questioning their role in a situation.
	Responding	The speaker builds on others' contributions with probing questions, critiques, or challenges assumptions.	The speaker expands on others' contributions by triangulating multiple sources.	The speaker critically reflects on others' contributions, offering explanations with internal attributions questioning others' role in a situation.

3. Methodology

To address the first research question—how can reflective inquiry in conversations within PLNs be measured?—we validated the model by analyzing conversational data from three PLNs, observed between January 2023 and May 2024. A total of 19 PLN sessions were recorded, yielding approximately 78 h of discussion. Each network exhibited a distinct design and focus area. All PLNs were organized and led by one or two facilitators, and participants varied in age and experience. Network 1 consisted of school leaders from eight elementary schools, concentrating on implementing change processes to stimulate more goal and data-driven practices. Network 2 included first-grade teachers from approximately 20 secondary schools, focusing on promoting reading comprehension and motivation.

Network 3 was composed of first-grade math teachers from four secondary schools, with an emphasis on foundational math literacy.

For the validation process, we used theoretical sampling. Theoretical sampling is a purposive, iterative sampling method aimed at developing and refining theoretical constructs. It involves selecting data based on their potential to contribute to the creation and testing of emerging theories. Researchers collect an initial sample, analyze the data, and then choose subsequent samples based on emerging theoretical insights. Theoretical saturation occurs when all of the main variations in the phenomenon have been identified and incorporated into the emerging theory (Ritchie & Lewis, 2003). In this approach, the researcher deliberately searches for extreme variations in each concept in the theory to exhaustion (Hennink et al., 2016). Their study on saturation in qualitative research (in field methods) highlights that six samples (around 10%) often suffice for understanding themes and testing coding frameworks, provided the data are rich (Guest et al., 2005). Therefore, we selected ±10% of the data, divided over the four PLNs, to obtain a variety of analyzed segments of data where educators engaged in conversation (see Table 3). We applied the coding framework to these conversations to look for examples of each code and to tweak the framework.

Table 3. Overview of selected data.

PLN	Segment	# Minutes
PLN 1	Segment 1	41
	Segment 2	42
	Segment 3	51
PLN 2	Segment 4	80
	Segment 5	68
	Segment 6	69
PLN 3	Segment 7	30
	Segment 8	30
	Segment 9	119
Total	n = 9	530 min (8.84 h)

Note: PLN = Professional Learning Network.

The recorded conversations were transcribed using Trint software 1.0.655, after which all transcriptions were manually checked by the first author for accuracy. The verified transcripts were then downloaded in Excel, where each speaker's contribution was represented as a separate row (n = 2195). Every contribution was manually coded with a code to assess the degree of the collaborative and dialogic process (C1, C2, or C3), the use of multiple data sources to generate and/or test ideas (D1, D2, or D3), and the degree of reflection (R1, R2, or R3) (see coding example in Table 4).

In order to ensure objectivity and accuracy in the analysis, we measured interrater reliability. This was assessed by the third author by independently coding 10% of the dataset and calculating Krippendorff's Alpha. The initial agreement yielded a Krippendorff's Alpha of 0.489. Upon reviewing discrepancies, we found that most coding inconsistencies occurred in the categories D2 (the use of multiple data sources) and R3 (critique with internal attribution). To enhance precision, we refined the definitions of these categories, clarifying what constitutes "another data source" and specifying criteria for when reflection is critical with internal attribution. Following these refinements, a second round of coding was conducted, resulting in a Krippendorff's Alpha of 0.904. This demonstrated a substantial increase in reliability.

Table 4. Example of coding.

Speaker	Contribution	C	D	R	Explanation
I	Yeah, the big problem with those kids is that many of them don't even have a book. They may have seen a book from a distance once. Before the last lesson, I even brought one of my own books because I felt like, if I use books. . . <deleted personal information> Uh, I had brought some books from her kindergarten class, and even those were difficult for them to read. So let alone if I were to use the chemistry book.	1	1	1	The speaker describes a personal experience descriptively without asking questions to facilitate the conversation (C1). The contribution relies on personal experience ("They may have seen a book from a distance once"). No other data sources are referenced (D1). The reflection is descriptive, without analyzing underlying reasons or implications ("I had brought some books from her kindergarten class, and even those were difficult for them to read") (R1).
R	It's a challenging audience. But even so, we shouldn't let ourselves be guided by the level they're currently at. We really need to aim higher, and that can actually spark their interest as well.	2	1	2	The responder engages with the initiator's ideas by adding a different perspective ("We shouldn't let ourselves be guided by the level they're currently at"). The perspective is suggestive (C2) rather than confrontational (C3). The statement is based on a general principle ('we really need to aim higher'). They do not reference other data sources (D1). The responder acknowledges a need for change and provides a rationale ("that can actually spark their interest as well") (R2).
I	But that's exactly it, yeah. So I hope to see some progress there, but I'm already getting the reaction, 'Is it reading again?' But we're working on that.	1	1	1	The contribution acknowledges the responder's comment but there are no questions to facilitate the conversation (C1). The speaker uses personal experiences, such as student reactions ("is it reading again?") without referencing additional sources (D1). The contribution is descriptive, without analyzing or critiquing their approach (R1).
R	But how did you approach all the reading lessons? Have you really already done that by reading aloud according to the theoretical principles? Or was it just free reading?	2	2	1	The responder asks clarifying questions to gather more context ("But how did you approach all the reading lessons?") (C2). The responder references theoretical principles as a source, suggesting a more structured approach to reading (D2). However, they do not triangulate across multiple sources or provide specific evidence (D3). The contribution gives not an explicit rationale or reasoning for the questions (R1).
I	Read freely, and they had to read that for themselves and then read out loud.	1	1	1	This statement is an answer to the question but does not invite further dialogue (C1). The contribution relies only on personal observations (D1). The statement is descriptive, recounting what was done during the lesson ("Read freely, and they had to read that for themselves") (R1).

Table 4. *Cont.*

Speaker	Contribution	C	D	R	Explanation
R	I think that might still be a step too far for them because they don't enjoy reading themselves yet. If you enjoy reading, you'll naturally share that enthusiasm, and you saw <anonymized> do that earlier—it's a completely different approach. So I really think you need to focus on implementing <reading programme> because there's an entire pedagogy behind it that changes how you approach reading. It's much more than just saying, 'We have to read for fifteen minutes,' and it takes a completely different approach.	3	2	3	The responder challenges the practice ("I think that might still be a step too far for them") (C3) emphasizing the need to focus on implementing <reading programme> and contrasting it with a surface-level approach ("It's much more than just saying, 'We have to read for fifteen minutes.'"). The responder references an evidence-based reading programme. However, it is a standalone source (D2), without integration of other sources (D3). The statement gives rationale ("because they don't enjoy reading themselves yet.") and critiques the current practice ("It's much more than just saying, 'We have to read for fifteen minutes'") with internal attribution (R3), stimulating the speaker to change their own method and practice ("So I really think you need to focus on implementing <reading programme>").

Note: I = Initiating speaker; R = responding speaker; C = degree of collective dialogue; D = degree of use of multiple sources; R = depth of reflection.

To address our second research question—how AI can be used to support the measurement of reflective inquiry—we had ChatGPT code 10% of the data using the optimized framework, and afterwards we assessed intercoder reliability. Based on the recommendations of Theelen et al. (2024), we refined the prompts to ensure accuracy in coding by making them clear, specific, and closely aligned with the predefined coding categories. This included leveraging ChatGPT's own definitions to enhance its understanding of complex concepts and coding tasks. Additionally, we provided contextual information to improve the model's performance, such as theoretical frameworks and sample annotations, which allowed ChatGPT to apply codes more effectively. To further optimize its output, we utilized structured datasets, as these were found to improve coding reliability by offering explicit cues for analysis, thereby minimizing ambiguity. We gave 10% of the data to ChatGPT 4.0. After ChatGPT coded the CSV file, intercoder reliability was measured using Krippendorff's Alpha Calculator (Marzi et al., 2024), resulting in a Krippendorff's Alpha of 0.503, which is considered too low. Even after additional training and refined prompts, the alpha remained below acceptable thresholds. The free version of ChatGPT was not useful; when we input data into the chat (three contributions at once), the model began hallucinating its own framework during the coding process.

Afterward, to address our third research question—what insight does this measurement approach provide about the different types of reflective inquiry occurring in PLNs?—we used the built-in functions in Excel to calculate the frequency and percentage of each individual code to determine how often each code and code combination was applied. Specifically, we used the following formula to count the occurrences of a specific code (e.g., C1): "=COUNTIF(Column_C, "C1")". In this code, Column_C represents the column containing the coded data C1-C3 and "C1" specifies which code being counted. Similarly, we applied this formula to the other codes.

Additionally, we examined the distribution of all 27 possible code combinations (C1-D1-R1, C2-D1-R1, C3-D1-R1, . . . C3-D3-R3). To systematically analyze these combinations, we first created a new tab where we manually listed all possible combination identifications

(combi_IDs) and their corresponding merged code strings (e.g., combi_ID 1 = C1D1R1, combi_ID 2 = C2D1R1, combi_ID 27 = C3D3R3). Next, in our coded dataset, we added a new column to automatically assign the correct combination ID to each row using the VLOOKUP function "=VLOOKUP(CONCATENATE(Column_C, Column_D, Column_R), Lookup_Table_Range, Column_Index, FALSE)". Column_C, Column_D, and Column_R refer to the columns containing the individual codes (C, D, and R) in the main dataset. Lookup_Table_Range refers to the range in the separate tab where the predefined combination IDs are stored. Column_Index specifies the column number in the lookup table that contains the combination ID. FALSE ensures an exact match.

Once each row had an assigned combi_ID, we calculated the frequency of each unique combination using the COUNTIF function: =COUNTIF(Column_combi_ID, Specific_Combi_ID). In his function, Column_combi_ID represents the column containing the assigned combination IDs and Specific_Combi_ID (1–27) refers to the specific combination being counted.

This approach provided a structured and systematic way to identify patterns and trends in reflective inquiry, ensuring that we could quantitatively analyze the interactions between different levels of collaborative dialog, the use of multiple data sources, and the degree of reflection.

4. Results

Overall, the analysis of 2195 contributions revealed the predominance of basic interaction and reliance on informal data. Table 5 shows the relative contribution of each code to the total contributions made by both participants and facilitators. Participants accounted for the majority of contributions (74%), with facilitators contributing 26%. Most conversation fell into C1, free-flowing conversation (64.37%), followed by C2, engagement through clarifying questions (32.76%), and C3, intentional interruption (2.87%). Similarly, the degree of data use was largely concentrated in D1, reliance on personal experience (92.57%), with limited contributions incorporating an additional data source (D2: 6.97%) or achieving the balanced use of multiple sources (D3: 0.46%). Reflection depth followed a similar trend, with R1, describing, accounting for 79.82% of contributions, while deeper reflection levels—R2, explaining (18.31%), and R3, critiquing (1.87%)—were less frequent.

Table 5. Overview of general distribution of codes between participants and facilitators across the categories.

Category	*Code*	*Participant*	*%*	*Facilitator*	*%*	*Total*	*Total %*
Total		*n* = 1634	74%	*n* = 561	26%	*n* = 2195	100%
Degree of collectiveness in dialogue (C)	C1	1135	51.70%	278	12.67%	1413	64.37%
	C2	493	22.46%	226	10.3%	719	32.76%
	C3	6	0.27%	57	2.6%	63	2.87%
Degree of use of multiple data sources (D)	D1	1540	70.16%	492	22.41%	2032	92.57%
	D2	91	4.15%	62	2.82%	153	6.97%
	D3	3	0.14%	7	0.32%	10	0.46%
Depth of reflection (R)	R1	1342	61.14%	410	18.68%	1752	79.82%
	R2	268	12.21%	134	6.10%	402	18.31%
	R3	24	1.1%	17	0.77%	41	1.87%

Figure 1 further illustrates the internal distribution of codes between participants and facilitators across the three categories. Facilitators generally show a more balanced distribution across levels 1, 2, and 3, while participants are more skewed toward level 1. Facilitators demonstrate higher levels dialog (C2 and C3) compared to participants, while

the distribution for the use of data sources (D codes) is similar for both groups, with a clear predominance of D1 and minimal contributions at D3. In terms of reflection (R codes), both groups are primarily concentrated at R1, though participants exhibit slightly higher proportions of R2 and R3, suggesting progress toward deeper reflection.

Below, we examine the three levels separately, along with the occurrence of reflective inquiry.

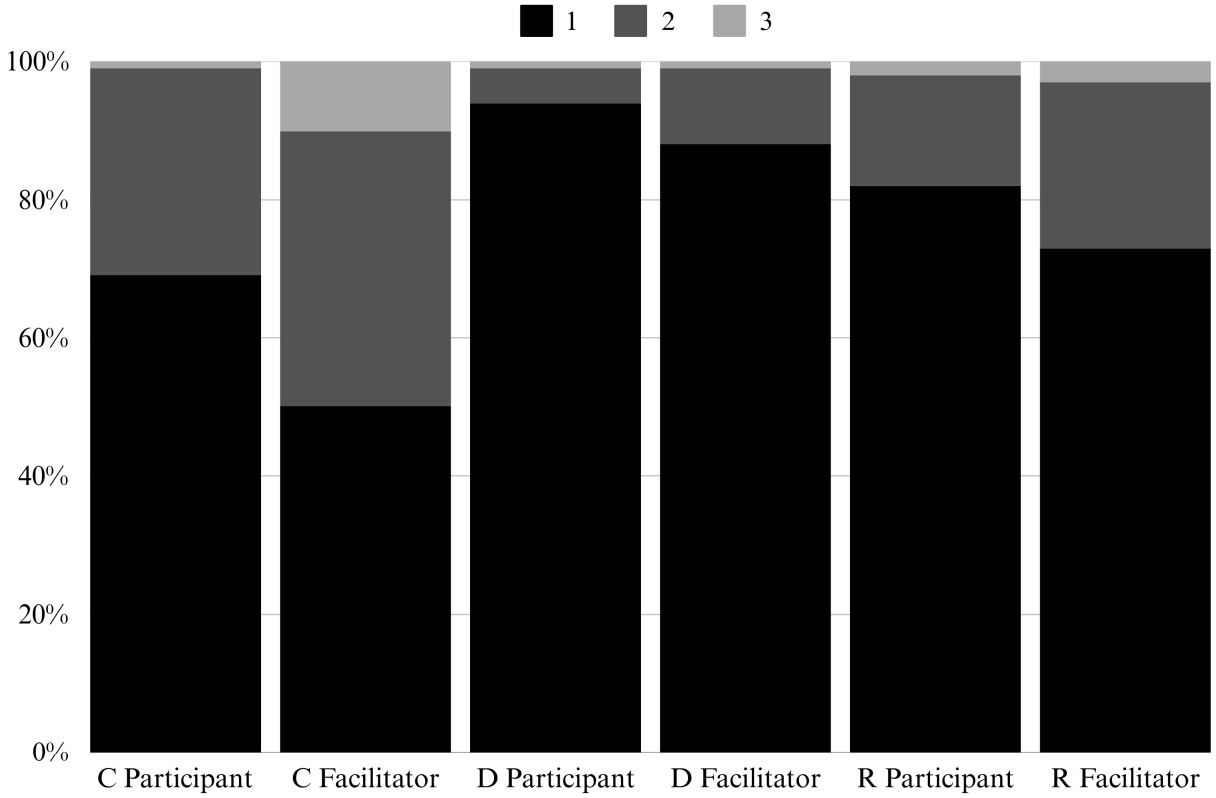

Figure 1. Internal distribution of codes between participants and facilitators across the three categories.

4.1. Degree of Collaborative Dialogue

The analysis of dialog indicated that free-flowing conversations (C1) were the most frequently observed, accounting for 64.37% of all contributions. An excerpt of a conversation, coded as C1, is provided below. These interactions were typically characterized by initiating speakers (I) sharing information without actively inviting further engagement through questions or requests for opinions. A response (R) often consisted of brief agreement (e.g., "that's really nice") or responders' personal experiences (e.g., "that's not how it works in our school cluster"), which shifted the conversation to another speaker or topic.

> **Speaker I:** The professionalization days are focused on that as well. Everyone comes together: all first-grade teachers, all second-grade teachers, and so on, to exchange ideas. (C1)
>
> **Speaker R1:** That's really nice. (C1)
>
> **Speaker R2:** That's not how it works in our school cluster. (C1)
>
> **Speaker R3:** Same here. (C1)
>
> **Speaker R2:** Even with sharing resources, there are still schools who say, "No, this is specific to our school". I think that's nonsense. (C1)

In contrast, the next level of engagement (C2) involved more engaged dialog through asking questions and offering suggestions. These interactions accounted for 32.76% of the

data. An instance of this type of dialog can be found below. Contributions at this level often included initiating speakers asking questions or seeking opinions (e.g., "Isn't that it? What do you think?") or responding speakers posing clarifying or exploratory questions to facilitate dialog (e.g., "Do you do that with all your mathematical problems too?"). C2 interactions were also characterized by responding speakers providing suggestions (e.g., "Maybe with a 'problem of the week', you could offer problems at varying levels of difficulty.").

> **Speaker I:** We've been working on it, and it gets a bit clearer every year, which is nice. But about mathematising, so that's turning something into a math problem, and demathematising, is taking it away from math. That's what I think those two words mean. Like turning Dutch into math or turning math into Dutch. Isn't that it? What do you think? (C2)
>
> **Speaker R1:** That's what it means. And do you do that with all your mathematical problems too? (C2)
>
> **Speaker I:** Yes, we look for solution strategies to solve certain real-life problems. (C1)
>
> **Speaker R1:** And those solution strategies, have you given them specific names? (C2)
>
> [. . .]
>
> **Speaker R2:** Maybe with a "problem of the week," you could offer problems at varying levels of difficulty. (C2)

The third level, C3 or intentional interruption, was the least frequent, representing only 2.87% of the dialog segments. Of the 63 contributions coded as C3, 6 were made by participants, while the remaining 57 came from facilitators. These contributions were marked by critical engagement and questioning assumptions to deepen discussions. In the C3 example below, a speaker responded to a suggestion to introduce core knowledge tests each trimester by challenging how the proposal would fail to evaluate core knowledge, instead arguing it would assess prerequisite knowledge.

> **Speaker R1:** [. . .] That's why I'd propose having a core knowledge test at the start or end of each trimester. A few per semester would clearly show where they stand. (C2)
>
> **Speaker R2:** You'd miss the point of core knowledge. For example, if it's about triangles, core knowledge only covers naming the triangles. Then it's no longer core knowledge—it's just studying. It's more useful to say, "With Pythagoras, you get a relationship, but what was it again?" It's meant to build on or recall what they've seen before. Otherwise, it's just a prerequisite knowledge test. (C3)

4.2. The Degree of Generating and/or Testing Ideas Based on Multiple Sources

The data revealed that conversations predominantly relied on personal experiences (D1), constituting 92.57% of the contributions. At this level, participants based their reflections on anecdotal evidence or personal observations. For example, when asked about consultation meetings, one speaker primarily used personal anecdotes and, in analyzing the situation, relied on personal observations (e.g., "I've told them, and I'm noticing a slight change").

> **Speaker R1:** You experience these difficulties during the meeting. Do you ask questions, or do you run into your own lack of motivation? (D1)

> **Speaker I:** I find the most tiring thing to keep convincing others about it. They are so quick to say: "I need special needs assistance, or there needs to be a diagnosis". They want to forget the phases before that and go straight to the third phase of IEP. (D1)
>
> **Speaker R2:** Or they rely on you as the special needs coordinator. (D1)
>
> **Speaker I:** I also got the question last week. I thought it was a good meeting in itself—it was a meeting with the second preschooler's teacher who has resistance, and I just clearly sent an email on the needed preparation. If nothing comes. . . I mainly bump into the teachers who have never had anything to say, but then the kids are in the third preschool, and it's just "they didn't see that in the previous year," so there was a gap in that. So I've started this school year by telling them very clearly that when it's time to consult, you get ready, you make an overview about the kids, and you tell me if there is a child you want to discuss at the meeting. I've told them, and I'm noticing a slight change. (D1)

D2 contributions represented 6.97% of the total contributions. Participants primarily referred to self-collected data, such as logbooks or tracking student scores, as well as standardized test scores, curriculum documents, and audit reports. Facilitators similarly referenced self-collected data but also mentioned research evidence more frequently, curriculum documents, and standardized test data. Less frequently, both groups cited sources like government data, professional development insights, and expert opinions.

Only 0.46% of contributions demonstrated the balanced use of multiple data sources (D3), with 3 out of 10 contributions made by participants. For instance, learning support coordinator systematically monitored their tasks and cross-referenced them with the evidence-based roles associated with their position to evaluate alignment and identify areas for improvement:

> I'm someone who also likes to do it for myself, too. As a teacher, you check the goals you want to achieve with your students, and as a special needs coordinator, I also want to check my goals for myself. I check my <evidence based> roles, I want to monitor myself with the number of hours I have, 32 out of 36, I personally think that's achievable, and that I should strive for that. So I created a template for myself with those <evidence based> roles and their explanations in it, and once a month, at the end of the month, I'll take my agenda and I'm going to tout for myself: that task includes that role, that task with that role. In such a way that I see: I did not uptake that role much at the moment, or I have not covered that much, that I can check why? Was there a question or is that something I've lost sight of?

4.3. Degree of Reflection and Internal Attribution

The depth of reflection was largely limited to R1, describing, which accounted for 79.82% of contributions. At this level, participants mainly described experiences, practices, or problems without deeper analysis or critique. For instance, a participant added that "I find it really difficult with students who don't have books. I used to give them a punishment task to teach them to be prepared, but honestly, that doesn't sit well with me".

Contributions, including reasoning for or explanations of practices (R2), represented 18.31% of the contributions. However, these reasons or explanations often highlighted external attributions for challenges or successes. For example, one contribution stated that "The worksheets and tests from <anonymized> are perceived as very challenging. But they say you should see it as a progress monitoring tool rather than a test. That's a difficult shift for us because my teachers want immediate results".

Finally, R3, critiquing, which involves critical self-reflection and internal attribution, only occurred in 1.87% of contributions. These instances demonstrated a willingness to question personal assumptions or practices. An illustrative example is an additional needs learning support coordinator, who reflects on her competences to ask teachers the right questions:

> Yes, I think for myself, because I've only been a special needs coordinator for a few months, it's still difficult to ask the right questions to them like that. Certain things that make me wonder, we have a mentor and when I hear her asking questions, when she afterwards discusses them with us, that I think yes, sure, those are the most logical questions. But I won't figure that out myself, I'm still early in my career, I still need to learn how to ask the right questions.

4.4. Occurrence of Reflective Inquiry

Out of 2195 contributions, the combination of C3, D3, and R3, defined as reflective inquiry, occurred only 4 times, making up 0.18% of all interactions. For an overview of the distribution for each code combination, see Table 6.

Table 6. Overview of the distribution of each code combination.

Combi_ID	R	D	C	n=	%
1	**C1**	**D1**	**R1**	**1135**	**51.71%**
2	**C2**	**D1**	**R1**	**551**	**25.10%**
3	C3	D1	R1	15	0.68%
4	C1	D2	R1	34	1.55%
5	C2	D2	R1	15	0.68%
6	C3	D2	R1	2	0.09%
7	C1	D3	R1	0	0.00%
8	C2	D3	R1	0	0.00%
9	C3	D3	R1	0	0.00%
10	**C1**	**D1**	**R2**	**187**	**8.52%**
11	C2	D1	R2	115	5.24%
12	C3	D1	R2	15	0.68%
13	C1	D2	R2	35	1.59%
14	C2	D2	R2	30	1.37%
15	C3	D2	R2	15	0.68%
16	C1	D3	R2	1	0.05%
17	C2	D3	R2	3	0.14%
18	C3	D3	R2	1	0.05%
19	C1	D1	R3	8	0.36%
20	C2	D1	R3	2	0.09%
21	C3	D1	R3	4	0.18%
22	C1	D2	R3	12	0.55%
23	C2	D2	R3	3	0.14%
24	C3	D2	R3	7	0.32%
25	C1	D3	R3	1	0.05%
26	C2	D3	R3	0	0.00%
27	**C3**	**D3**	**R3**	**4**	**0.18%**

Note: Combi_ID = combination identifier; C = degree of collective dialogue; D = degree of use of multiple sources; R = depth of reflection.

The most frequent combinations were number 1 (C1, D1, R1), number 2 (C2, D1, R1), and number 10 (R2, R1, R1). Number 1 (C1, D1, R1) accounted for 51.71% of contributions. These interactions were characterized by the descriptive sharing of experiences based on personal experiences, without inviting further engagement or exploration. Another common combination was C2, D1, and R1 (25.10%), where other speakers engaged with

clarifying questions or provided suggestions, but the reasoning remained descriptive. Free-flowing conversation, (C1) paired with personal experiences (D1) and analytical reasoning (R2), occurred in 8.52% of contributions. Here, speakers shared their rationale for certain practices but did not engage others with questions or asked for input.

True reflective inquiry (C3, D3, R3) only occurred in four instances. All four were contributed by facilitators. An excerpt of the conversation coded as C3, D3, and R3 can be found below. The facilitator demonstrated questioning established ways of working (e.g., "We won't convince teachers if we say, 'We read twice a week and have a few results, but they're not really convincing.' That creates a vicious cycle."), calling for change with internal attribution (e.g., "If you can convince people with data, you'll take a step forward. (. . .) so show them."), giving arguments (e.g., "Subject teachers are not language teachers. They think differently, work differently, and want numbers, and numbers can be persuasive. So show them."), and showing the balanced use of multiple data sources (e.g., collecting data on amount of books read, effect on math results and using evidence-based reading scan), while also facilitating critical dialog through probing questions (e.g., "How often are you reading? That's where we need to start.").

> **Speaker R:** That's also problematic for a school if no one can handle the data. You must be able to work with data literacy—you need to work with standard deviation, boxplots, and the information they provide. (C2, D2, R2)
>
> **Speaker I:** In <learning platform>, there's a possibility to work with data, but apparently, that wasn't purchased. And among the staff, there's no data analyst either. (C1, D1, R1)
>
> **Speaker R:** You should definitely look for colleagues willing to help. This will also support realizing this initiative. If you can convince people with data, you'll take a step forward. Who do we want to convince? We all want to win over subject teachers. Subject teachers are not language teachers. They think differently, work differently, and want numbers, and numbers can be persuasive. So show them. For example, "We started with zero books; students had never read a book before. Now it's 27 books". And this is reflected in math results. If you can demonstrate that, you've done excellent work, really. Let me check, I printed it out for you. This comes from 'The Reading Scan.' Did you also conduct this with your school? I believe you did some time ago. This aspect of the reading policy plan is critical for expanding it across the school. But the key question remains: How often are you reading? That's where we need to start. From there, we won't convince teachers if we say, "We read twice a week and have a few results, but they're not really convincing". That creates a vicious cycle. (C3, D3, R3).

5. Discussion

This study explored the measurement of reflective inquiry and its varying levels, as well as their prevalence within PLNs. The use of a coding framework to measure reflective inquiry provides a structured way of assessing dialog. This is reflected in the interrater reliability of Krippendorff's Alpha of 0.904.

Interrater reliability with ChatGPT was below the acceptable threshold, suggesting that the complexity of the framework exceeds the current capabilities of ChatGPT. This finding aligns with observations by Theelen et al. (2024), who noted that AI tools struggle when labels are not explicitly present in the text, as they lack the ability to accurately infer contextual cues. Similarly, Khademi (2023) demonstrated that AI tools underperform compared to human experts when tasked with evaluating complex constructs, particularly due to challenges in interpreting semantic and pragmatic nuances. Farrokhnia et al. (2023)

further highlighted that ChatGPT lacks a deep understanding of the meaning behind the words it processes. While it recognizes patterns and generates plausible responses, its inability to grasp underlying concepts limits its performance, especially in tasks requiring higher-order thinking skills such as critical and analytical reasoning (Rudolph et al., 2023). These limitations were likely compounded in our study, where distinguishing between nuanced levels of collaborative dialog, data use, and reflection required not only textual analysis but also an understanding of the broader context, reasoning, and intent behind the contributions.

The results indicate that conversations within PLNs are primarily characterized by free-flowing interactions (C1) and a preference for personal experiences (D1), with few examples of critical reflection (R3).

One plausible explanation for the predominance of free-flowing interactions (C1: 64.37%) is group dynamics. Katz and Dack (2013) observe that groups tend to settle on areas of agreement, which can limit the range of perspectives explored and reduce opportunities for challenging existing thinking. Similarly, Earl and Timperley (2008) found that while participants often demonstrate respect and support for one another, they may hesitate to challenge others' interpretations or actions. This highlights that psychological safety and trust are essential for participants to share challenges, reflect on failures, and question existing practices without fear of judgment, which is a prerequisite for critical and transformative dialog (Katz & Earl, 2010). Additionally, participants' professional backgrounds and work experience influence group dynamics. More experienced teachers may bring rich professional knowledge, but if discussions are hierarchical, they may limit open inquiry and mutual learning rather than encourage critical engagement (York-Barr et al., 2016). Beyond group dynamics, individual competencies also play a crucial role in shaping critical conversations. The depth of critical discussions depends on participants' ability to ask probing questions, analyze information, and engage in reflective dialog (Rodgers, 2002; Katz & Earl, 2010). This illustrates the complexity of fostering the critical dialog needed for transformational learning (Katz & Dack, 2013). Although low levels of critical dialog require more intentional facilitation to foster substantive changes in practice, they may still contribute positively by enhancing job satisfaction and motivation. The patterns of collaborative dialog observed in this study are consistent with Little's (1990) findings that interactions focused on mutual assistance and sharing may contribute to teacher satisfaction but are less likely to drive innovation or significant instructional changes. A recent review similarly highlights that while PLNs are associated with positive effects on PLN members' motivation and well-being, their impact on practice and school-level change can vary (Warmoes et al., 2025). It seems plausible that conversations characterized by lower levels of reflective inquiry may enhance job satisfaction but require more intentional facilitation to foster substantive changes in practice.

The frequent use of personal experiences (92.57%) aligns with Schildkamp and Datnow's (2020) findings that team conversations often center around storytelling and description, rather than around integrating diverse data sources. Earl and Timperley (2008) argue that incorporating evidence into conversations not only requires us to bring data into discussions but also foster a mindset that challenges ideas and promotes the creation of new knowledge. However, this process does not occur spontaneously—it requires explicit support to develop inquiry skills, such as the ability to analyze data and generate insights (Ciampa & Gallagher, 2016). Encouraging the use of multiple data sources and supporting teams in connecting diverse evidence can contribute to richer discussions and more actionable insights (Kintz et al., 2015; Schildkamp & Datnow, 2020).

Instances of R3-level reflections were relatively uncommon, reflecting the difficulty of fostering internal attribution and critical self-examination (El-Dib, 2006). Similarly, Earl and

Timperley (2008) noted how hard it is to penetrate tacit knowledge, make it explicit, and move learning forward on a new level. The capacity to engage in reflective dialog varies between participants based on their previous training, openness to critique, and ability to connect insights to practice (Alzayed & Alabdulkareem, 2020). Kuh (2015) also highlights that, without structured processes to guide discussions toward examining teaching and learning, conversations can shift toward general issues, missing opportunities for deeper critical reflection.

Reflective inquiry was only observed in 0.18% of contributions. While this may reflect the inherent challenges of developing an inquiry stance, as noted by some, it also underscores the importance of deliberate facilitation in supporting these processes (Alzayed & Alabdulkareem, 2020). Facilitators emerge as key contributors to deeper levels of reflective inquiry, with a disproportionate share of contributions at C3, D3, and R3 levels. This finding aligns with research highlighting the role of skilled facilitators in disrupting entrenched thinking, modeling reflective practices, and encouraging critical dialog (Fisher-Yoshida & Yoshida, 2022; Katz & Dack, 2013). Similarly, Murugaiah et al. (2012) emphasized the critical role of facilitators in advancing learning processes, particularly in resolution and critical thinking. These findings suggest that facilitators play an essential role in creating the conditions for reflective inquiry and supporting teams in achieving higher levels of engagement and learning.

6. Conclusions

This study examined how reflective inquiry can be measured and its prevalence within PLNs. The first research question asked "How can reflective inquiry in conversations within PLNs be measured?" Findings demonstrate that reflective inquiry can be measured using a coding framework that categorizes dialog into three dimensions: the degree of collaborative dialog (C1–C3), the degree of use of data sources (D1–D3), and the depth of reflection (R1–R3). This framework provides a structured method for analyzing dialog by identifying contributions ranging from basic-level interactions and reliance on informal data to those involving critical dialog, the balanced use of multiple data sources, and internal attribution. This is reflected in the interrater reliability of Krippendorff's Alpha of 0.904. For the second research question, "how can AI be used to support this measurement approach?", we examined ChatGPT's ability to independently code the data based on the coding framework. The interrater reliability with ChatGPT was below the acceptable threshold, suggesting that the complexity of the framework exceeds the current capabilities of ChatGPT.

The third research question asked "What insight does this measurement approach provide about the different types of reflective inquiry occurring in PLNs?". Findings indicate that the majority of dialog within PLNs consisted of free-flowing conversation (C1: 64.37%) and reliance on personal experience (D1: 92.57%) and descriptive reflection (R1: 79.82%). True reflective inquiry was observed in 0.18% of contributions and was exclusively associated with facilitators. This shows that facilitators play an important role in modeling and fostering higher levels of reflective inquiry.

7. Limitations and Suggestions for Future Research

This study provides insights into reflective inquiry within PLNs, but there are some limitations. First, the research was conducted in a natural setting without an intervention designed to promote reflective inquiry. While the PLNs aimed to foster learning, data use, and changes in practice, reflective inquiry was not explicitly stimulated. As a result, the findings reflect existing practices rather than the potential impact of targeted interventions.

Second, the study does not account for changes in behavior resulting from reflective inquiry. Although the conversations analyzed demonstrated instances of reflective inquiry, it is not possible to determine whether these interactions directly influenced meaningful changes in practice. The framework isolates the dialog itself and does not account for contextual factors that may mediate the translation of reflective inquiry into change.

Third, while the coding framework provides a structured method for analyzing conversation, the reliance on a purposive sample comprising 10% of the data, although consistent with theoretical sampling principles, may limit the generalizability of the findings.

Future research could build on these findings by expanding the dataset to include a larger and more diverse sample of PLNs, enabling a broader understanding of reflective inquiry across various contexts. Additionally, investigating how factors such as facilitation, group composition, activities, trust, collaboration and individual competencies influence the depth and nature of reflective inquiry would provide valuable insights into how to foster reflective inquiry. For example, different levels of experience among participants may affect their ability to challenge assumptions and integrate multiple data sources, while group cohesion and facilitation strategies might shape the extent to which probing questions and constructive discussions emerge. Moreover, examining the relationship between reflective inquiry in conversations and measurable changes in teaching practices, student outcomes, or organizational improvement would bridge the gap between reflective inquiry and outcomes.

Author Contributions: Conceptualization, A.W., C.B. and E.C.; methodology, A.W.; validation, I.D.; formal analysis, A.W.; investigation, A.W.; data curation, A.W.; writing—original draft preparation, A.W.; writing—review and editing, A.W., C.B. and E.C.; visualization, A.W.; supervision, E.C.; project administration, A.W.; funding acquisition, E.C. and A.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Education in Flanders (Belgium) and is part of a broader project entitled 'Steunpunt Centrale Toetsen in Onderwijs'. The collaboration between the VUB and University of Southampton was funded by FWO grant for a long stay abroad (V450424N).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by approved by the Human Sciences Ethics Committee of the Vrije Universiteit Brussel (Approval Code: ECHW_402.02) on the 2nd of February 2023.

Informed Consent Statement: Written informed consent has been obtained from the participants to publish this paper.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy reasons.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

PLN	professional learning network
PLC	professional learning community

References

Alzayed, Z. A., & Alabdulkareem, R. H. (2020). Enhancing cognitive presence in teachers' professional learning communities via reflective practice. *Journal of Education for Teaching International Research and Pedagogy*, *47*(1), 18–31. [CrossRef]

Brookfield, S. D. (1995). *Becoming a critically reflective teacher*. Jossey-Bass.

Brown, C., & Poortman, C. (2018). *Networks for Learning: Effective collaboration for teacher, school and system improvement* (1st ed.). Routledge.

Brown, C., Poortman, C., Gray, H., Ophoff, J. G., & Wharf, M. (2021). Facilitating collaborative reflective inquiry amongst teachers: What do we currently know? *International Journal of Educational Research*, *105*, 101695. [CrossRef]

Charteris, J., & Smardon, D. (2014). Dialogic peer coaching as teacher leadership for professional inquiry. *International Journal of Mentoring and Coaching in Education*, *3*(2), 108–124. [CrossRef]

Ciampa, K., & Gallagher, T. L. (2016). Teacher collaborative inquiry in the context of literacy education: Examining the effects on teacher self-efficacy, instructional and assessment practices. *Teachers and Teaching*, *22*(7), 858–878. [CrossRef]

De Keijzer, H., Jacobs, G., Van Swet, J., & Veugelers, W. (2020). Identifying coaching approaches that enable teachers' moral learning in professional learning communities. *International Journal of Mentoring and Coaching in Education*, *9*(4), 375–392. [CrossRef]

Dewey, J. (1910). *How we think*. Available online: https://pure.mpg.de/pubman/item/item_2316308_3/component/file_2316307/Dewey_1910_How_we_think.pdf (accessed on 4 March 2025).

Doğan, S., & Adams, A. (2018). Effect of professional learning communities on teachers and students: Reporting updated results and raising questions about research design. *School Effectiveness and School Improvement*, *29*(4), 634–659. [CrossRef]

Earl, L. M., & Timperley, H. (2008). Professional Learning Conversations: Challenges in using evidence for improvement. In *Springer eBooks*. Springer. [CrossRef]

El-Dib, M. a. B. (2006). Levels of reflection in action research. An overview and an assessment tool. *Teaching and Teacher Education*, *23*(1), 24–35. [CrossRef]

Farley-Ripple, E., May, H., Karpyn, A., Tilley, K., & McDonough, K. (2018). Rethinking connections between research and practice in education: A conceptual framework. *Educational Researcher*, *47*(4), 235–245. [CrossRef]

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, *61*(3), 460–474. [CrossRef]

Fisher-Yoshida, B., & Yoshida, R. (2022). Transformative learning and its relevance to coaching. In *Springer eBooks* (pp. 935–948). Springer. [CrossRef]

Guest, G., Bunce, A., & Johnson, L. (2005). How many interviews are enough? *Field Methods*, *18*(1), 59–82. [CrossRef]

Hakkarainen, K., Palonen, T., Paavola, S., & Lehtinen, E. (2004). *Communities of networked expertise: Professional and educational perspectives*. Available online: http://ci.nii.ac.jp/ncid/BB05229407 (accessed on 4 March 2025).

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2016). Code saturation versus meaning saturation. *Qualitative Health Research*, *27*(4), 591–608. [CrossRef] [PubMed]

Jay, J. K., & Johnson, K. L. (2002). Capturing complexity: A typology of reflective practice for teacher education. *Teaching And Teacher Education*, *18*(1), 73–85. [CrossRef]

Katz, S., & Dack, L. A. (2013). Towards a culture of inquiry for data use in schools: Breaking down professional learning barriers through intentional interruption. *Studies in Educational Evaluation*, *42*, 35–40. [CrossRef]

Katz, S., & Earl, L. (2010). Learning about networked learning communities. *School Effectiveness and School Improvement*, *21*(1), 27–51. [CrossRef]

Katz, S., Earl, L., & Jaafar, S. (2009). *Building and connecting learning communities: The power of networks for school improvement*. Corwin Press. [CrossRef]

Khademi, A. (2023). Can ChatGPT and bard generate aligned assessment items? A reliability analysis against human performance. *arXiv*, arXiv:2304.05372. [CrossRef]

Kintz, T., Lane, J., Gotwals, A., & Cisterna, D. (2015). Professional development at the local level: Necessary and sufficient conditions for critical colleagueship. *Teaching and Teacher Education*, *51*, 121–136. [CrossRef]

Kuh, L. P. (2015). Teachers talking about teaching and school: Collaboration and reflective practice via Critical Friends Groups. *Teachers and Teaching*, *22*(3), 293–314. [CrossRef]

Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record the Voice of Scholarship in Education*, *91*(4), 509–536. [CrossRef]

Liu, Q., Zhang, S., Wang, Q., & Chen, W. (2017). Mining online discussion data for understanding teachers reflective thinking. *IEEE Transactions On Learning Technologies*, *11*(2), 243–254. [CrossRef]

Marzi, G., Balzano, M., & Marchiori, D. (2024). K-Alpha Calculator–Krippendorff's Alpha Calculator: A user-friendly tool for computing Krippendorff's Alpha inter-rater reliability coefficient. *MethodsX*, *12*, 102545. [CrossRef]

Mitchell, C., & Sackney, L. (2011). *Profound improvement: Building capacity for a learning community*. Taylor & Francis.

Murugaiah, P., Azman, H., Thang, S. M., & Krish, P. (2012). Teacher learning via communities of practice: A Malaysian case study. *International Journal of Pedagogies and Learning*, *7*(2), 162–174. [CrossRef]

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company*. Oxford University Press. [CrossRef]

OECD. (2015). *Education policy outlook 2015: Making reforms happen*. OECD Publishing. [CrossRef]

Poortman, C. L., Brown, C., & Schildkamp, K. (2022). Professional learning networks: A conceptual model and research opportunities. *Educational Research*, *64*(1), 95–112. [CrossRef]

Rincón-Gallardo, S., & Fullan, M. (2016). Essential features of effective networks in education. *Journal of Professional Capital and Community*, *1*(1), 5–22. [CrossRef]

Ritchie, J., & Lewis, J. (2003). *Qualitative research practice: A guide for social science students and researchers*. SAGE.

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record The Voice of Scholarship in Education*, *104*(4), 842–866. [CrossRef]

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, *6*(1). [CrossRef]

Schildkamp, K., & Datnow, A. (2020). When data teams struggle: Learning from less successful data use efforts. *Leadership and Policy in Schools*, *21*(2), 147–166. [CrossRef]

Schleicher, A. (Ed.). (2012). *Preparing teachers and developing school leaders for the 21st century: Lessons from around the world*. International Summit on the Teaching Profession. OECD Publishing. [CrossRef]

Schön, D. A. (1983). *The Reflective Practitioner: How professionals think in action*. Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Available online: http://psycnet.apa.org/record/1987-97655-000 (accessed on 4 March 2025).

Stoll, L. (2015). Using evidence, learning and the role of professional learning communities. In C. Brown (Ed.), *Leading the use of research and evidence in schools* (1st ed., pp. 54–65). IOE Press.

Theelen, H., Vreuls, J., & Rutten, J. (2024). Doing research with help from ChatGPT: Promising examples for coding and Inter-Rater reliability. *International Journal of Technology in Education*, *7*(1), 1–18. [CrossRef]

Vanblaere, B., & Devos, G. (2016). Exploring the link between experienced teachers' learning outcomes and individual and professional learning community characteristics. *School Effectiveness and School Improvement*, *27*(2), 205–227. [CrossRef]

Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research Review*, *15*, 17–40. [CrossRef]

Van Manen, M. (1977). Linking ways of knowing with ways of being practical. *Curriculum Inquiry*, *6*(3), 205. [CrossRef]

Walkington, J., Christensen, H. P., & Kock, H. (2001). Developing critical reflection as a part of teaching training and teaching practice. *European Journal of Engineering Education*, *26*(4), 343–350. [CrossRef]

Warmoes, A., Decabooter, I., Struyven, K., & Consuegra, E. (2025). Exploring learning outcomes: The impact of professional learning networks on members, schools, and students. *School Effectiveness And School Improvement*, 1–35. [CrossRef]

York-Barr, J., Sommers, W. A., Ghere, G. S., & Montie, J. (2016). *Reflective practice for renewing schools: An action guide for educators*. Corwin Press.