sustainability

MDPI

Article

Inquiry-Based Facilitation, Diagnostic Reasoning, and Reflective Intelligence in Saudi Pre-Service Early Childhood Teachers: Associations and Implications for Sustainable STEM Teacher Education

Adel R. Althubyani

Department of Curriculum and Educational Technology, Taif University, Taif 21944, Saudi Arabia; adel@tu.edu.sa

Abstract

Sustainable STEM education is linked to teachers who can reflect on practice, diagnose student understanding, and facilitate inquiry-based learning, yet these competencies are often studied in isolation. Grounded in reflective intelligence, diagnostic competence, and constructivist theory, this cross-sectional self-report study examined associations among inquiry-based pedagogical facilitation (IBPF), diagnostic reasoning and pedagogical adaptation (DRPA), and reflective intelligence (RI) among 179 Saudi female pre-service early childhood teachers, testing whether DRPA statistically mediates the IBPF–RI association. A bilingual, expert-reviewed questionnaire was analyzed with PLS-SEM. Reliability and convergent validity were acceptable (Cronbach's $\alpha = 0.827$–0.945; AVE = 0.691–0.743), but discriminant validity was not conclusively established (HTMT 0.863–0.895). IBPF was associated with DRPA ($\beta = 0.794$, $f^2 = 1.710$), and DRPA with RI ($\beta = 0.539$, $f^2 = 0.364$). IBPF was directly associated with RI ($\beta = 0.345$), with a significant indirect path through DRPA ($\beta = 0.428$), consistent with partial mediation (all $p < 0.001$). The model explained 63.1% of DRPA and 70.6% of RI variance in-sample. As the design is cross-sectional, the associations are consistent with, but do not demonstrate, a practice-embedded account; longitudinal work should test whether reflection precedes or follows practice. Findings may conceptually inform sustainable teacher education under Vision 2030.

Keywords: reflective intelligence; diagnostic reasoning; inquiry-based facilitation; sustainable education; teacher competence; pre-service teachers; early childhood education; PLS-SEM; Saudi Arabia

Check for updates

Academic Editor: Vasiliki Brinia

Received: 19 September 2026
Revised: 22 November 2026
Accepted: 24 November 2026
Published: 27 November 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Teachers' ability to reflect on their teaching, to identify the learning needs of students, and to make corresponding adjustments to their teaching is central to the quality of STEM teaching [1,2]. This is particularly challenging in early childhood classrooms, as teachers are required to be metacognitively conscious, continuously assess student learning, and guide students in inquiry learning to achieve deep understanding [3]. The challenge is exacerbated in settings in transition to educational change, where student-centered, inquiry-oriented practices are supplanting traditional practices and requiring skills in addition to content knowledge [4]. The Vision 2030 initiative of Saudi Arabia underscores the importance of educating its citizens to become the best in the world at STEM, and the cognitive and metacognitive skills needed for good STEM teaching are a central focus of this reform, but empirical evidence on these skills in this context is lacking [5]. In this

study, effective STEM teaching is understood as a teaching and learning process in which teachers draw on professional competencies, namely the capacities to reflect on their practice, to diagnose what students understand, and to facilitate inquiry-based learning, so that students can actively construct knowledge. These three capacities are treated as pedagogical competencies, that is, knowledge-based and practice-based capacities that teachers use to plan, deliver, and adjust their teaching. For pre-service teachers, these competencies are still developing, and building them is a central aim of teacher preparation.

While teacher reflection, diagnostic reasoning and inquiry-based facilitation have all been identified as key to effective teaching and learning in STEM, they have traditionally been studied in isolation, and their interconnections and combined influence on the teaching process remain relatively under-researched [6,7]. Recently, reflective intelligence [8], defined as a recursive ability to reflect, integrate, and evolve across cognitive, emotional, and behavioral aspects, has been suggested as a possible unifying framework; however, to the author's knowledge, it has not been further developed or examined in early childhood teacher education or in relation to specific educational practices. In the field of biology, diagnostic reasoning has been seen as a key component of effective teaching, but, to the author's knowledge, its links with reflective capacity and its significance in facilitating inquiry have not been systematically explored. This lack of understanding has practical implications for teacher preparation programs in Saudi Arabia, as novice teachers are often constrained by limited reflective practice opportunities, tight time schedules, and the need to balance traditional and progressive methods of teaching [6]; teacher education and professional development programs are not evidence-based if it is not understood how these competencies relate.

Previous scholarship indicates that reflective practice is associated with professional competence [9] and that pedagogical content knowledge supports diagnostic competence [1]. However, these findings come from separate research streams, and they have not been brought together in an integrated empirical model, that is, a single model that specifies how reflective, diagnostic and facilitative capacities relate to one another within the same group of teachers. The link between reflective intelligence and diagnostic reasoning is theoretically plausible but, to the author's knowledge, has not been examined empirically, particularly in STEM contexts where teachers must develop their own metacognitive awareness while encouraging metacognition in young learners [10]. Inquiry-based facilitation is widely discussed as an essential part of meaningful STEM learning [11]; however, the cognitive and metacognitive capacities that accompany effective facilitation have not been clearly identified and tested.

A further open question concerns the ordering of these capacities. A common assumption is that reflection precedes and enables good practice. Similar questions of ordering have been raised for other emotional competencies: in a sample of Pakistani undergraduates, learners' attitudes and engagement were associated with trait emotional intelligence, which positions it as an outcome of learning experiences and not only as a predictor [12]. Practice-embedded accounts, by contrast, hold that diagnostic skill is built through repeated experience of situations in which expected and observed student responses differ [13], and that facilitative pedagogies such as inquiry- and problem-based learning place continuous diagnostic demands on teachers [14]. On this view, teachers who report more frequent inquiry-based facilitation would also report stronger diagnostic reasoning and pedagogical adaptation, and the evidence generated by diagnosing and adapting could in turn feed the metacognitive, self-regulating processes that characterize reflective intelligence [3,8]. Under this reading, diagnostic reasoning and pedagogical adaptation are candidate mediators between inquiry-based facilitation and reflective intelligence. This theoretical specification is tested as a pattern of associations. The ordering is presented as a theory-informed specifi-

cation and not as a confirmed sequence. Because the data are cross-sectional, the reverse ordering, in which reflective intelligence supports practice, cannot be ruled out, and it is revisited in the Discussion. The mediating role of diagnostic reasoning is a notable gap, especially in non-Western settings such as Saudi Arabia, where cultural and institutional conditions may differ from those of Western research settings [15] and, to the author's knowledge, contextual studies are rare.

The purpose of this study is to address these gaps by examining the associations among inquiry-based pedagogical facilitation (IBPF), diagnostic reasoning and pedagogical adaptation (DRPA), and reflective intelligence (RI) among Saudi female pre-service early childhood teachers using structural equation modeling. The focus is on pre-service teachers preparing to teach early childhood classrooms, because teacher preparation is the stage at which reflective, diagnostic, and facilitative capacities are first developed and at which teacher education programs can most directly act. The study specifies a model in which IBPF is associated with DRPA, DRPA is associated with RI, and IBPF is associated with RI both directly and indirectly through DRPA. The study builds on the growing literature on reflective intelligence in education by providing empirical evidence on how reflection, diagnosis, and facilitation are interrelated in STEM teaching, and it offers evidence that can inform integrated professional development, rather than programs that treat reflection, diagnosis, and facilitation as separate skill sets. To the author's knowledge, few, if any, quantitative studies have examined these relationships in early childhood STEM teacher education in Saudi Arabia. The findings may inform teacher preparation, professional learning and policy under Vision 2030, as Saudi Arabia reforms its education system toward more student-centered, inquiry-oriented teaching, and they contribute to the wider debate on whether teachers' reflective capacity precedes or follows their classroom practice. Because the design is cross-sectional and based on self-report, all relationships are interpreted as statistical associations and not as causal or developmental effects.

This focus on teacher capacity is also relevant to sustainability. Teachers are central to the quality of education, which is the focus of Sustainable Development Goal 4, and constructivist, active-learning approaches to STEM have been linked to sustainable education [14]. In Saudi Arabia, Vision 2030 positions education and STEM excellence as foundations for long-term human-capital development. Understanding how teachers' inquiry-based facilitation, diagnostic reasoning, and reflective capacity are related can therefore help to design professional development that aims to build an adaptable and resilient teaching workforce, which can be argued to be a precondition for sustainable education systems. This link to sustainability is conceptual, as the study does not measure sustainability outcomes or student learning, and its empirical findings concern associations among teacher capacities.

1.1. Research Purpose and Objectives

The main goal of this study is to examine the structural associations among inquiry-based pedagogical facilitation, diagnostic reasoning and pedagogical adaptation, and reflective intelligence among female pre-service early childhood teachers in Saudi Arabia. The study has the following purposes: (1) to estimate a structural model in which inquiry-based pedagogical facilitation is associated with reflective intelligence directly and indirectly through diagnostic reasoning and pedagogical adaptation; (2) to examine the direct, indirect and total associations among the three constructs; (3) to evaluate the reliability and validity of the measurement model in the Saudi educational context; (4) to assess the explanatory power of the model for diagnostic reasoning and pedagogical adaptation and for reflective intelligence; and (5) to discuss implications for teacher development in light of the Saudi educational reform agenda in Vision 2030, based on the findings of the study.

1.2. Research Questions

1. To what extent is inquiry-based pedagogical facilitation associated with diagnostic reasoning and pedagogical adaptation among Saudi female pre-service early childhood teachers?
2. To what extent are diagnostic reasoning and pedagogical adaptation associated with reflective intelligence?
3. To what extent is inquiry-based pedagogical facilitation directly associated with reflective intelligence, when diagnostic reasoning and pedagogical adaptation are taken into account?
4. Do diagnostic reasoning and pedagogical adaptation statistically mediate the association between inquiry-based pedagogical facilitation and reflective intelligence?

1.3. Hypotheses

H1. *Inquiry-based pedagogical facilitation is positively associated with diagnostic reasoning and pedagogical adaptation among Saudi female pre-service early childhood teachers.*

H2. *Diagnostic reasoning and pedagogical adaptation are positively associated with reflective intelligence among Saudi female pre-service early childhood teachers.*

H3. *Inquiry-based pedagogical facilitation is positively associated with reflective intelligence among Saudi female pre-service early childhood teachers.*

H4. *Diagnostic reasoning and pedagogical adaptation statistically mediate the association between inquiry-based pedagogical facilitation and reflective intelligence.*

2. Literature Review

2.1. Theoretical Framework

This research builds on three related theory streams. First, the study uses the concept of reflective intelligence of Markov and Slavova [8], a system-level cognitive function defined as transitive self-reflection and metacognitive awareness, building on Dewey's foundation, in which reflection is not linear but recursive, and, following Merkebu et al. [4], as higher-level analytical thinking, which is necessary for self-regulated learning and adaptive behavior. Second, it relies on the concept of diagnostic competencies developed by Kramer et al. [1], which links professional knowledge (pedagogical content knowledge, content knowledge and pedagogical knowledge) to diagnostic activities (evidence generation, evaluation and drawing conclusions) and to diagnostic accuracy, thereby treating diagnosis as an observable activity and not only as a cognitive trait, which makes it amenable to the type of self-report measurement used in the current study. Third, it places the inquiry-based pedagogical facilitation in the context of constructivist learning theory, including Vygotsky's social constructivism and scaffolding within students' zone of proximal development [11]. These frameworks coalesce in their conception of effective teaching as a dynamic interaction among teachers' metacognitive awareness, their dynamic interpretation of student learning, and their ability to provide contexts conducive to the active construction of knowledge, a convergence that warrants treating the three constructs as a structural system and not as standalone indicators of teaching quality.

2.2. Reflective Intelligence in Educational Contexts

Reflective intelligence involves self-aware, metacognitive, and outcome evaluation skills, which allow for changing beliefs and actions [3,8]. In educational contexts, this appears especially in teachers' capacity for critical reflection on their pedagogical choices,

understanding of student reactions, and systematic adaptation of their teaching methods [2]. Heckemann et al. [9] identified three levels of reflection, the third of which includes emotions and values in the analysis of the process. Empirical evidence supports the transformative potential of reflection: structured collaborative reflection on videotaped lessons was a significant means of enhancing pre-service teachers' analytical abilities [2], and action research has been shown to promote reflective thinking in pre-service preschool teachers [7].

Implementing reflective practice, however, is not without its problems, especially in STEM settings. Aslam et al. [6] found three main factors that hinder novice teachers: lack of classroom-management experience, limited understanding of how to integrate STEM, and time constraints. Similarly, Kausar [5] examined anatomy education in Saudi Arabia and found that opportunities for reflection were reduced by curriculum content, language barriers, and faculty shortages, indicating the need for both the development of reflective intelligence and supportive organizational structures. Despite these constraints, the relationship between teachers' reflective practices and students' self-regulated learning (SRL) is theoretically significant [10].

In this study, reflective intelligence refers to teachers' recursive capacity to examine their own teaching experience, integrate the resulting insight, and adapt their subsequent practice, with cognitive, emotional, and behavioral components [3,8]. It overlaps with, but is not identical to, several established constructs. Reflective practice and teacher reflection usually denote professional activities or processes, often structured and episodic, such as reflecting on a videotaped lesson [2,9]. Reflective thinking is typically examined as a set of skills or dispositions developed through specific tasks such as action research [7] or as a component of self-regulated learning [10]. Metacognition concerns knowledge about and monitoring of one's own cognition, and metacognitive reflection is one of its forms [4], while self-regulated professional learning emphasizes goal setting, monitoring, and strategy adjustment [11]. Reflective intelligence, as used here, is intended as an integrative, teacher-level capacity that spans these elements and is distinguished by the movement from reflection to changed action: awareness of outcomes, reconsideration of teaching methods, and learning from experience.

Reflective intelligence is used as an integrative label because reflective practice, metacognition, and self-regulated learning are usually studied separately, whereas the reflective intelligence literature [3,8] describes a unified, self-organizing capacity that responds to feedback. It is recognized, however, that there is a risk of relabeling existing constructs; the empirical distinctiveness of reflective intelligence from the other constructs in the model is therefore examined through discriminant validity tests (Section 4.1) and should be interpreted cautiously. The nine-item reflective intelligence scale was developed for this study (Section 3.3). Its items include reflecting on the outcomes of teaching activities (for example, whether learning objectives were achieved), adapting teaching in response to students' reactions, and learning from experience (for example, keeping notes for later use).

2.3. Diagnostic Reasoning and Pedagogical Adaptation

Diagnostic reasoning involves teachers' skills in evaluating students' understanding, pinpointing learning problems, and uncovering the root causes of comprehension difficulties [1]. In competence-oriented models of teachers' assessment, competencies are defined as context-specific, learnable cognitive dispositions that are needed to cope with specific situations [16], and a meta-analysis in medical and teacher education found that diagnostic competencies can be facilitated effectively through problem-solving, with scaffolding adapted to learners' knowledge base [17]. Pedagogical content knowledge is the knowledge most strongly related to diagnostic activities ($\beta = 0.29$, $p < 0.001$) and diagnostic accuracy ($\beta = 0.23$, $p = 0.002$), accounting for 17% of the variance in diagnostic activities,

which indicates that diagnostic competence is not limited to content knowledge or general pedagogical knowledge alone but is also related to pedagogical content knowledge [1]. Heitzmann et al. [18] developed an operationalization of diagnosis that supplements the structural model of Kramer et al. [1] with eight activities: Problem Identification, Questioning, Hypothesis Generation, Construction of Artifacts, Generation of Evidence, Evaluation of Evidence, Drawing of Conclusions, and Communication.

Diagnostic insight is followed by pedagogical adaptation, meaning that teachers adjust their teaching approaches and support to the identified student needs [13]; such translation, however, is neither automatic nor straightforward. Novice STEM teachers studied by Aslam et al. [6] reported difficulties with managing classrooms, developing curricula, and recognizing practical constraints. Kausar [5] found that reduced teaching time and inadequate anatomy-teaching experience also hindered the development of diagnostic skills in Saudi medical education, with 88.4% of students aware of the clinical importance of anatomy compared with only 19.9% of educators who felt that students were well prepared for diagnostic skills, highlighting a persistent gap between perceived importance and preparation. Taken together, this literature reinforces the idea that diagnostic reasoning and pedagogical adaptation are not a single action but a series of interrelated actions (noticing, interpreting, and acting) that are all subject to institutional and contextual constraints, even when a teacher has the necessary professional knowledge.

2.4. Inquiry-Based Pedagogical Facilitation in STEM Education

Inquiry-based pedagogical facilitation is an approach to teaching that moves beyond teacher-centered delivery and focuses on students' inquiry process, which involves questioning, exploring, and building knowledge [2]. The approach is based on constructivist epistemology, which places teachers in a facilitator role that creates an environment for active discovery instead of passive reception [11], and is related to the SRL approach, as inquiry involves students in setting goals, monitoring progress, and adjusting strategies [10]. In STEM education, developmentally appropriate inquiry balances structure and autonomy, as Coutts et al. [19] illustrated with exercise physiology students undertaking clinically based scenarios, where scaffolded inquiry supported professional competency development across levels of schooling. Inquiry is increasingly recognized as an essential component of STEM pedagogy, as a systematic review of 27 integrated-STEM framework papers revealed that inquiry is one of only five principles that appeared in the majority of such papers, along with integration, real-world problems, design, and teamwork [20]. Inquiry-driven, real-world problem engagement is also seen as a characteristic of integrated STEM education, and the way the problem is framed can affect motivational outcomes; Roehrig et al. [21] noted that a communally oriented framing is more motivating for a wider range of students than a narrowly technical one.

However, implementation issues limit inquiry-based teaching: time and coverage pressures conflict with exploratory learning, classroom management differs from that in conventional pedagogy, and student readiness varies, all of which may affect facilitation [11]. Although there is theoretical discussion about the link between reflective capacity and inquiry effectiveness, empirical research into these connections is still limited, and there is little evidence on how inquiry-based facilitation, teachers' diagnostic reasoning, and their reflective capacity are related. The gap is notable, since inquiry-based facilitation is student-led and therefore arguably requires teachers to make more continuous assessments than didactic instruction does. Building on this gap, Portillo-Blanco et al. [21] found that although there is a high degree of consensus on the concept of inquiry, there is less consensus and convergence on its operational characteristics and measurement across studies, which

is a caution for how inquiry-based facilitation is operationalized and measured in the current study.

2.5. Interrelationships Among Constructs and Proposed Mediation

Theoretically, reflective intelligence, diagnostic reasoning, and the facilitation of inquiry are not independent but are interrelated processes. The model tested in this study (see Figure 1) specifies inquiry-based pedagogical facilitation as the starting point of a sequence of associations. Inquiry-based facilitation is student-led, so teachers must continuously notice, interpret, and respond to students' thinking. Practice-embedded accounts of diagnostic skill [13] and work on facilitative pedagogies [14] suggest that such situations place sustained demands on teachers' diagnostic reasoning and pedagogical adaptation. Inquiry-based facilitation is therefore expected to be positively associated with diagnostic reasoning and pedagogical adaptation (H1).

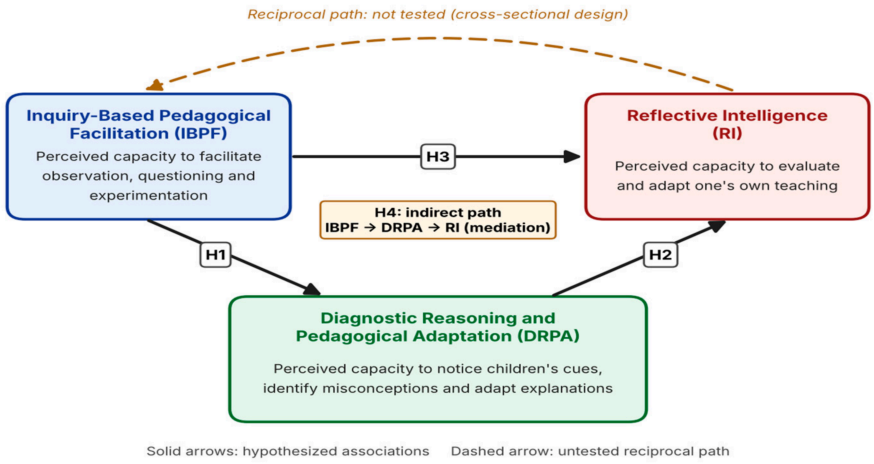

Figure 1. Conceptual model and hypotheses. Arrows represent hypothesized statistical associations and do not indicate temporal order or causation. The dashed arrow shows a reciprocal path that is discussed in Section 5.6 but was not tested.

Diagnostic reasoning and pedagogical adaptation, in turn, provide teachers with concrete evidence about what students understand and about how their own pedagogical decisions worked. This evidence is the material on which reflective intelligence operates: questioning one's interpretations, exploring alternative interpretations, and revisiting one's diagnostic schemas as more information is gathered [2,8]. Reflective practice has been linked to improved teaching in higher education [15], and structured reflection on classroom episodes has been linked to pre-service teachers' analytical abilities [2]. Diagnostic reasoning and pedagogical adaptation are therefore expected to be positively associated with reflective intelligence (H2).

The mediation hypothesis rests on the view that engaging in inquiry-based facilitation does not by itself translate into reflective capacity unless teachers interpret what students' responses mean. Reflection supported by AI-based emotion recognition has been reported to improve metacognitive abilities [22]. A teacher who facilitates inquiry on scientific argumentation, for instance, may notice through diagnostic reasoning that students have specific difficulties in evaluating evidence, adapt the inquiry activities accordingly, and use this experience as the basis for reflecting on their own practice. In this sequence, diagnostic reasoning and pedagogical adaptation link facilitation to reflective intelligence, so it is expected to statistically mediate the association between the two (H4), alongside a direct association between inquiry-based facilitation and reflective intelligence (H3). A reciprocal path in which reflective intelligence, once developed, shapes later facilitation is

conceivable, but it cannot be tested with cross-sectional data and is treated as a direction for future research.

2.6. Research Gap and Study Contribution

Although there is theoretical consistency, there are few empirical studies of these relationships. Previous studies document the effectiveness of reflective practice [9], diagnostic reasoning [1], and inquiry-based learning [11] individually, but few studies test the structural relationships among them. This is indicative of the field as a whole: a recent bibliometric mapping study of STEM education research over the past 22 years (2003–2024) reported that scholarly output has surged since 2018, but that descriptive and thematic-mapping studies remain the dominant approach, and empirical testing of the mechanisms linking teacher-level constructs is comparatively rare [23]. Few studies have quantitatively examined how inquiry-based facilitation, diagnostic reasoning, and reflective intelligence are related in early childhood STEM teaching, whether diagnostic reasoning statistically mediates the association between facilitation and reflective intelligence, or how these associations appear in non-Western contexts. Evidence on emotional competencies suggests that associations reported elsewhere cannot be assumed to hold across contexts: in Pakistani learners, some associations of trait emotional intelligence differed from global models and depended on the instructional context [12,24]. In Saudi Arabia, where educational transformation and STEM excellence are part of Vision 2030, the evidence base for developing teachers' competencies is still scarce [5]. In teacher education, the application of reflective intelligence, which stems largely from research on AI and cognition [3,8], is an emerging theory that needs empirical investigation. This study is designed to fill these gaps by using structural equation modeling to provide cross-sectional evidence for an integrated model of teacher cognition and practice that can inform theory and intervention design. In doing so, it also responds to calls in the wider teacher cognition literature for designs that examine mediated pathways, rather than the isolated bivariate relationships that have traditionally been studied in the teacher reflection and diagnostic competence literature.

3. Materials and Methods

The study is based on structural equation modeling (SEM), a family of multivariate techniques that combine a measurement model, which links observed questionnaire items to latent constructs, with a structural model, which specifies the relationships among those constructs [25,26]. The study used partial least squares SEM (PLS-SEM), a variance-based approach that estimates structural relationships by maximizing the explained variance of the endogenous constructs and that is commonly used for theory development with complex models that include mediation [25,27]. In this approach, the measurement model is evaluated first for reliability and validity, and the structural model is evaluated second [25,26]. In the present study, the measurement model links the questionnaire items to three latent constructs (IBPF, DRPA, and RI), and the structural model specifies the tested paths shown in Figure 1: IBPF → DRPA, DRPA → RI and IBPF → RI, with DRPA as a statistical mediator. The remainder of this section describes the research design, participants, questionnaire, and ethical considerations, followed by the data analysis procedure (Section 3.5).

3.1. Research Design

This study used a quantitative, cross-sectional survey design, with data collected through a self-report questionnaire, to examine how inquiry-based facilitation and diagnostic reasoning and pedagogical adaptation are associated with teachers' reflective intelligence, and to test diagnostic reasoning and pedagogical adaptation as a statistical mediator of the association between inquiry-based pedagogical facilitation and reflective intelligence.

3.2. Research Participants and Educational Context

The study recruited 179 female pre-service early childhood teachers enrolled in the undergraduate Early Childhood Education program at Taif University, located in Taif, Makkah Province, Saudi Arabia, using purposive (criterion-based) sampling, as described below. The program adopts a comprehensive, multidisciplinary approach to teacher preparation, equipping pre-service teachers with the pedagogical knowledge, instructional competencies, and practical experiences necessary to facilitate children's learning and development across multiple educational domains rather than specializing in a single academic discipline.

Within the Saudi teacher education context, teacher preparation generally follows either subject-specific pathways, such as physics, mathematics, and biology, or comprehensive programs designed to prepare generalist, early childhood teachers. Unlike subject-specific programs, early childhood teacher preparation integrates disciplinary knowledge and pedagogical approaches across multiple learning areas. The curriculum encompasses child development, developmental psychology, early childhood pedagogy, instructional strategies, and subject-specific teaching methods. STEM-related content is incorporated through courses in the program's 2023 specification, such as Developing Scientific and Mathematical Concepts in Kindergarten Children, Natural Sciences I and II, Mathematics I and II, Methods of Teaching Science, Methods of Teaching Mathematics in the Early Primary Grades, Educational Technology, and Emerging Technologies. Engineering-related concepts are embedded within science, mathematics, and interdisciplinary learning activities rather than offered as a separate discipline.

The program also incorporates supervised teaching practicums that provide opportunities to translate theoretical and pedagogical knowledge into authentic classroom experiences. In this study, practicum refers to the kindergarten field experiences of the program plan (Kindergarten Field Experience I and II, offered in the fifth and sixth semesters at four credit hours each; the English course titles are working translations of the Arabic titles in the program plan); the field experiences in the early primary grades were not counted. At the time of data collection, third-year students had completed the first of these practicums, and fourth-year students had completed both, whereas first- and second-year students had not yet undertaken one. During these placements, pre-service teachers plan, deliver, and evaluate developmentally appropriate learning activities across multiple educational domains, including science and mathematics. These experiences provide opportunities to engage children in inquiry-oriented activities, recognize emerging understandings and misconceptions, adapt instructional approaches, and reflect on teaching practices.

Although the program is not formally designated as a specialized STEM teacher preparation program, its multidisciplinary curriculum incorporates foundational knowledge and pedagogical competencies relevant to STEM education. The program prepares teachers for preschool through grade 6, who typically teach several subjects, including science and mathematics, for which foundational STEM knowledge is essential. To the author's knowledge, no dedicated STEM teacher preparation program exists for these grades in Saudi Arabia. This educational context provides a rationale for examining three interrelated constructs: reflective intelligence (RI), diagnostic reasoning and pedagogical adaptation (DRPA), and inquiry-based pedagogical facilitation (IBPF). Specifically, RI concerns pre-service teachers' perceived capacity to critically reflect on instructional decisions and learning experiences; DRPA addresses their perceived ability to identify children's learning needs, recognize misconceptions, and adapt teaching approaches; and IBPF concerns their perceived capacity to facilitate questioning, exploration, investigation, and knowledge construction through developmentally appropriate inquiry activities. The study therefore focuses on these cognitive and pedagogical dimensions of pre-service generalist teachers' approaches

to STEM-related instruction rather than evaluating program effectiveness or specialized STEM teaching expertise.

Participants were drawn from four academic levels: the second semester (first year), fourth semester (second year), sixth semester (third year), and eighth semester (fourth year). Their ages ranged from 18 to 23 years. The program admits approximately 250–300 students each year, so approximately 1000 students are enrolled across the four years of study. From this population, program instructors explained the aims and topic of the research to the pre-service teachers and identified, at each academic level, students who were willing to take part and who had studied, or were scheduled to study, the science-related courses of the program plan (Developing Scientific and Mathematical Concepts in Kindergarten Children, Natural Sciences I and II, and Methods of Teaching Science in the Early Primary Grades). Because no STEM-specialized courses are offered in early childhood programs in Saudi Arabia, these courses prepare generalist teachers to teach science in the primary grades. This gave a pool of 280 students. From this pool, 30 students (about 7–8 from each academic level) were first chosen at random for the pilot study (Section 3.3). All of the remaining 250 students were invited verbally to take part in the main study; the pilot participants were therefore outside the 250 invited and were not included in the main sample. The instructors also shared the link to the online questionnaire with the invited students; participation was voluntary, and no interviews were conducted. Of the 250 students invited, 186 returned the questionnaire (186/250 = 74.4%). Seven responses (7/250 = 2.8% of those invited; 7/186 = 3.8% of those returned) were excluded after screening for incomplete or invalid responses (responses were treated as invalid when the same response option was selected for every item), and the remaining 179 valid responses (179/250 = 71.6% of those invited) were retained for analysis. The pilot participants are counted in neither figure. The inclusion of participants from different stages of teacher preparation reflects varying levels of academic preparation and potential exposure to classroom teaching experiences, although the analysis treats them as a single sample. Table 1 summarizes the characteristics of the participants.

Table 1. Participant characteristics (N = 179 valid responses).

Characteristic	Category	*n*	%	Mean (SD)
Gender	Female	179	100	
Institution	Taif University, Taif, Makkah Province	179	100	
Program	Undergraduate Early Childhood Education	179	100	
Year of study (semester)				
	First year (second semester)	47	26.3	
	Second year (fourth semester)	46	25.7	
	Third year (sixth semester)	45	25.1	
	Fourth year (eighth semester)	41	22.9	
	Year of study (coded 1–4)			2.45 (1.11)
Practicum experience				
	None yet (second and fourth semesters)	93	52.0	

Table 1. *Cont.*

Characteristic	Category	*n*	%	Mean (SD)
	One practicum (sixth semester)	45	25.1	
	Two practicums (eighth semester)	41	22.9	
	Number of practicums (coded 0–2)			0.71 (0.82)
Response				
	Students invited	250	100	
	Responses received	186	74.4	
	Excluded (incomplete or invalid)	7	2.8	
	Valid responses analyzed	179	71.6	

Note. n = number of participants; % = percentage of the 179 valid responses, except in the Response section, where percentages are of the 250 students invited; M = mean; SD = standard deviation (n − 1). Year of study and the number of practicums is ordinal categories coded numerically, so their means and *SD*s are descriptive only. Practicum experience refers to the kindergarten field experiences completed at the time of data collection.

An a priori power analysis was not conducted, which is acknowledged as a limitation (Section 5.5). Sample adequacy was therefore assessed post hoc with the inverse square root method [28], under which the minimum sample size at 80% power and $\alpha = 0.05$ is $n > (2.486/|\beta_{min}|)^2$. The smallest structural coefficient in the model was IBPF → RI ($\beta = 0.345$), which requires approximately 52 observations. The coefficients for DRPA → RI ($\beta = 0.539$) and IBPF → DRPA ($\beta = 0.794$) require approximately 22 and 10 observations, respectively. The sample of 179 valid responses exceeds all of these thresholds.

3.3. Data Collection Tool

The study developed a bilingual (English and Arabic) questionnaire to measure reflective intelligence, diagnostic reasoning, pedagogical adaptation, and inquiry-based pedagogical facilitation, with the Arabic translation carried out by a native translator and then back-translated into English to resolve any ambiguity. The items were written by the author based on the theoretical concepts and then reviewed by two STEM and two psychology experts. A pilot study was then conducted to examine reliability and validity before the questionnaire was finalized and circulated via Google Forms from 10 December 2025, to 10 January 2026. Providing the questionnaire in both English and Arabic was intended to reduce language-related measurement error and to ensure that item wording captured the intended construct meaning for pre-service teachers whose working language is Arabic. The questionnaire was reviewed by experts for content appropriateness, which is not equivalent to a full validation; its reliability and validity were evaluated empirically through the PLS-SEM measurement model, first in the pilot study and then in the main sample (Section 4.1).

All items were newly written by the author from theory. The diagnostic reasoning and pedagogical adaptation scale was initially drafted with 10 candidate items. Before expert review, the author examined the pool and removed two items: one that overlapped in content with other items and one that stated an attitude rather than a frequency of practice. The remaining eight items were rated by the expert panel, pilot-tested and used in the main study, and no items were removed after expert review, the pilot study, or the main analysis; all eight items had outer loadings above 0.708 in the main sample (0.790–0.866; Figure 2). The relationship between the initial and final items is shown in Table S4 of the Supple-

mentary Materials. The reflective intelligence items operationalize the definition used in this study, namely a recursive capacity to examine one's teaching experience, integrate the resulting insight, and adapt practice [3,8], including reflection on practice [4,9]. The diagnostic reasoning and pedagogical adaptation items operationalize diagnostic activities and adaptation, such as noticing children's cues, probing their thinking, anticipating and correcting misconceptions, and adapting explanations [1,13,16,18]. The inquiry-based facilitation items operationalize the principles of inquiry-based learning, namely encouraging children to observe, question, experiment, and interpret [14,21,22]. Although all items use a frequency rating, the inquiry-based facilitation items and DRPA1 are written as short classroom scenarios, and the other items as concrete statements of practice, so that responses reflect how pre-service teachers would act in specific situations and not only their general agreement with abstract statements. Because 52.0% of respondents had not yet completed a practicum (Table 1), their responses are interpreted as self-reported or anticipated practice, based on coursework and, for those who had completed one or two practicums, possibly also on classroom experience, and not as observed teaching performance. How respondents without a practicum understood the frequency items was not examined directly (for example, through cognitive interviews), so this interpretation rests on the design of the items and on the coursework described in Section 3.2. The constructs are accordingly interpreted as pre-service teachers' perceived professional capacities and dispositions. A formal comparison by practicum experience was not undertaken, because practicum status is nested within year of study (Table 1), so any contrast would confound practicum with coursework exposure and academic level; the three-item IBPF scale and the need to establish measurement invariance first further limit the interpretability of such a test (Section 5.5). Four professors (two in STEM and two in psychology) rated each item as appropriate or not appropriate for the construct it was intended to measure and for the clarity of its wording. The item-level content validity index (I-CVI) was 1.00 for 14 items and 0.75 for six items (IBPF3, DRPA2, DRPA5, RI2, RI5, and RI9), each of which one professor judged not appropriate. The scale-level index was S-CVI/Ave = 0.925, above the benchmark of 0.90 [29], and S-CVI/UA was 0.70. For a panel of four experts, an I-CVI of 0.75 corresponds to a modified kappa of 0.67, which is regarded as good [30]. Because 92.5% of all ratings were "appropriate", chance-corrected coefficients such as Fleiss' kappa are unstable here ($\kappa = -0.08$), so the content validity indices are treated as the primary evidence. Since three of four professors judged each of these six items appropriate, they were retained without change. This evidence should nonetheless be read cautiously: with a panel of only four experts, stricter conventions for small panels expect unanimous agreement on each item, and S-CVI/UA (0.70) is below the 0.80 value commonly used as a benchmark [29,30]. Content validity is therefore best regarded as adequate at the scale level (S-CVI/Ave) but only moderately supported at the item level, and the six items with an I-CVI of 0.75 should be revisited in future validation with a larger panel.

The pilot study involved 30 female pre-service early childhood teachers (about 7–8 from each academic level) from the same program, who were not included in the main sample. In the pilot, DRPA (Cronbach's alpha = 0.925; AVE = 0.658) and RI (Cronbach's alpha = 0.952; AVE = 0.727) showed satisfactory convergent validity. IBPF showed a Cronbach's alpha of 0.690 (composite reliability = 0.827; AVE = 0.616), and the HTMT between IBPF and DRPA exceeded 1 (1.056). Because the pilot sample was small, which limits the stability of these estimates, and because the items were considered theoretically appropriate, the questionnaire was retained without modification for the main study. Reliability and discriminant validity were then re-evaluated in the main sample, in which IBPF showed a Cronbach's alpha of 0.827 and the IBPF–DRPA HTMT was 0.895 (Section 4.1). The final

questionnaire, after measuring content validity and conducting a pilot study, is provided as Supplementary Material in both English and Arabic.

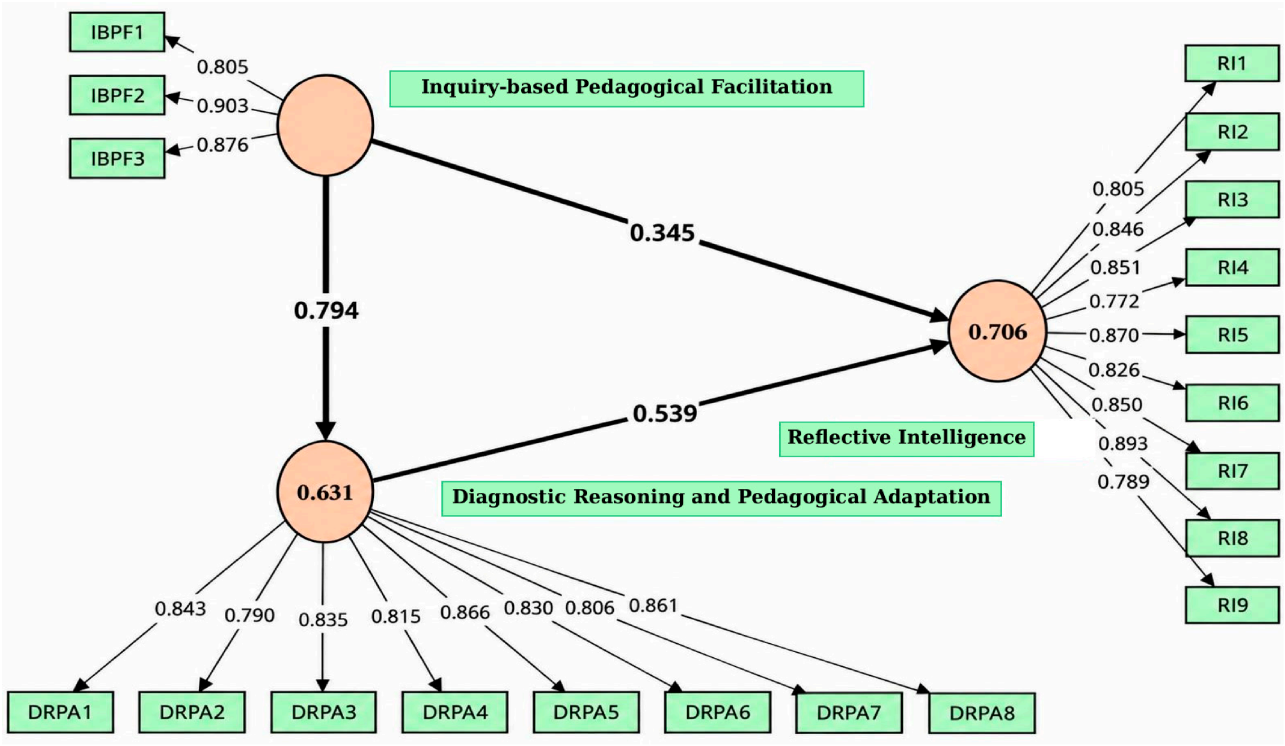

Figure 2. Structural and measurement model. Values on the arrows from the constructs to the indicators are outer loadings, values on the paths between constructs are standardized path coefficients, and values inside the circles are R^2.

The questionnaire comprised three constructs. Inquiry-based pedagogical facilitation (three items) assessed how often pre-service teachers report guiding children to observe, compare, and test ideas, to propose solutions to problems, and to try alternative designs in classroom situations, for example, encouraging children to observe and experiment with how leaf color changes. Diagnostic reasoning and pedagogical adaptation (eight items) assessed how often pre-service teachers report noticing children's thinking and misconceptions and responding to them diagnostically, including items such as "I notice the children's cues that indicate a misunderstanding of the scientific concept", "I ask questions that help me discover how children think", and "I change my way of explaining when I notice a misunderstanding". Reflective intelligence (nine items) assessed how often pre-service teachers report reflecting on their teaching and adapting it based on experience, including items such as "After each STEM activity, I think about the extent to which it has achieved the learning objectives", "I reconsider my teaching methods when I notice a weak response from the children", and "I keep notes about my activities so that I can learn from them later". All items were rated on a five-point frequency scale (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always).

3.4. Ethical Considerations

The research followed ethical protocols throughout data collection and was approved by the Scientific Research Ethics Committee (see the Institutional Review Board Statement). The study's purpose was stated in the questionnaire's introduction, and informed consent was obtained via a mandatory opt-in question before the questionnaire could be completed. Participating pre-service teachers were assured of anonymity and protection of their per-

sonal information. Several procedural steps were taken to reduce common method bias. First, the questionnaire was divided into three sections, which limited participants' ability to connect the independent and dependent variables. Second, the item statements were kept simple and concise to avoid ambiguity. Third, participants were informed of anonymity so that they could express their opinions freely. Fourth, they were assured that there were no right or wrong answers and were asked to give honest responses. The questionnaire was administered only to students enrolled in the Early Childhood Education program, and reminders were sent to encourage completion. These measures may have reduced common method bias, but they could not eliminate it.

3.5. Data Analysis Procedure

The study used PLS-SEM to test the proposed model in SmartPLS 4.1.1.7, examining the direct associations of inquiry-based pedagogical facilitation and diagnostic reasoning and pedagogical adaptation with reflective intelligence, and the mediating role of diagnostic reasoning and pedagogical adaptation in the association between inquiry-based facilitation and reflective intelligence. PLS-SEM was chosen for three reasons. First, the study is explanatory and theory-building in nature and uses a newly developed scale for which no validated measurement existed previously. Second, the model includes a mediating construct, and PLS-SEM is well suited to explaining variance in the endogenous constructs. Third, PLS-SEM is generally less demanding than covariance-based SEM (CB-SEM) with respect to sample size [25,26], which suited the sample of 179. The analysis proceeded in two stages: measurement model evaluation (convergent validity, discriminant validity including bootstrapped HTMT confidence intervals, collinearity, and model fit) and structural model evaluation (hypothesis testing, specific indirect effects, total effects, R^2, f^2, and predictive performance using PLSpredict [31]). The model was estimated with the path weighting scheme, a maximum of 3000 iterations, and a stop criterion of 10^{-7}. Hypotheses were tested with 5000 bootstrap subsamples drawn from the original sample ($n = 179$), using percentile bootstrap 95% confidence intervals and two-tailed tests at $\alpha = 0.05$. Paths were considered significant if the confidence interval excluded zero. Because the data are cross-sectional and self-reported, path coefficients, including those labeled direct, indirect, and total effects, are interpreted as model-based statistical associations and not as evidence of causal or temporal ordering.

4. Results

4.1. Measurement Model

Convergent validity was established: Cronbach's alpha exceeded 0.7 for all constructs (highest for reflective intelligence, 0.945), composite reliability exceeded the 0.7 threshold [25] for all constructs (highest for reflective intelligence, 0.946), and average variance extracted exceeded the 0.5 threshold [26] for all constructs (highest for inquiry-based pedagogical facilitation, 0.743) (Table 2).

Table 2. Convergent validity.

Construct	Cronbach's Alpha	Composite Reliability (rho_a)	Composite Reliability (rho_c)	AVE
Diagnostic Reasoning and Pedagogical Adaptation	0.936	0.937	0.947	0.691
Inquiry-based Pedagogical Facilitation	0.827	0.839	0.896	0.743
Reflective Intelligence	0.945	0.946	0.954	0.696

Discriminant validity was assessed using the Heterotrait–Monotrait Ratio (HTMT) and the Fornell–Larcker Criterion [32]. All HTMT point estimates fell below 0.90 but above the more conservative 0.85 threshold [33], and all Fornell–Larcker diagonal values exceeded their corresponding off-diagonal values, providing initial support for discriminant validity (Tables 3 and 4).

Table 3. Discriminant validity (HTMT).

Construct	DRPA	IBPF	RI
Diagnostic Reasoning and Pedagogical Adaptation (DRPA)	-		
Inquiry-based Pedagogical Facilitation (IBPF)	0.895	-	
Reflective Intelligence (RI)	0.863	0.874	-

Table 4. Discriminant validity (Fornell–Larcker criterion).

Construct	DRPA	IBPF	RI
Diagnostic Reasoning and Pedagogical Adaptation (DRPA)	0.831		
Inquiry-based Pedagogical Facilitation (IBPF)	0.794	0.862	
Reflective Intelligence (RI)	0.813	0.774	0.834

To assess discriminant validity more rigorously, bootstrapped HTMT confidence intervals were examined (5000 subsamples; percentile 95% intervals). The HTMT value for DRPA–IBPF was 0.895 (95% CI [0.820, 0.961]), for DRPA–RI 0.863 (95% CI [0.796, 0.920]), and for IBPF–RI 0.874 (95% CI [0.791, 0.949]) (Table 5). In all three cases, the upper bound lies below 1.00, so the constructs are statistically separable under the least stringent HTMT inference criterion [33]. However, the upper bounds exceed 0.90, the liberal cutoff recommended for conceptually similar constructs, and all three point estimates (0.895, 0.863, and 0.874) exceed the more conservative 0.85 threshold [33]. Discriminant validity is therefore not conclusively established: the constructs overlap substantially, and the structural results are interpreted as associations among closely related constructs, as also noted in Section 5.5. Two considerations motivate measuring the constructs separately. First, the Fornell–Larcker criterion was met for all three constructs, although the margins are small (for example, the square root of the AVE for DRPA was 0.831, against a correlation of 0.794 with IBPF). Second, the scales refer to different things and were measured differently: the inquiry-based facilitation items describe classroom situations in which the teacher guides children's investigation, the diagnostic reasoning items concern noticing and responding to children's thinking, and the reflective intelligence items concern the teacher's own evaluation of her teaching after an activity. The separation of inquiry-based facilitation from diagnostic reasoning in particular should not be regarded as settled, because their HTMT exceeded 1 in the small pilot sample (Section 3.3).

Table 5. Bootstrapped confidence intervals for HTMT.

Relationship	HTMT (Original Sample)	Sample Mean (M)	2.5%	97.5%
IBPF ↔ DRPA	0.895	0.896	0.820	0.961
RI ↔ DRPA	0.863	0.862	0.796	0.920
RI ↔ IBPF	0.874	0.875	0.791	0.949

Indicator-level collinearity was acceptable: all outer VIF values were below 5 [27] (Table 6). These values describe collinearity among indicators and do not test common method variance, which cannot be ruled out and is discussed in Section 5.5. All eight DRPA items (DRPA1–DRPA8) were retained; no indicators were removed during measurement model evaluation. Indicators were retained when their outer loadings exceeded the recommended threshold of 0.708 [25]; all 20 outer loadings were between 0.772 and 0.903 (Figure 2).

Table 6. Collinearity (outer VIF). IBPF = inquiry-based pedagogical facilitation; DRPA = diagnostic reasoning and pedagogical adaptation; RI = reflective intelligence.

Indicator	VIF
DRPA1	2.896
DRPA2	2.537
DRPA3	2.871
DRPA4	2.942
DRPA5	3.532
DRPA6	2.979
DRPA7	2.560
DRPA8	3.151
IBPF1	1.629
IBPF2	2.290
IBPF3	2.070
RI1	2.684
RI2	3.052
RI3	3.354
RI4	2.220
RI5	3.378
RI6	2.625
RI7	3.169
RI8	4.236
RI9	2.453

Model fit was judged acceptable on the basis of an SRMR of 0.053, below the 0.08 threshold [34]. The NFI was 0.831, which is modest, so the fit is interpreted with caution (Table 7). Because the structural model is saturated, the SRMR and other fit indices are identical for the saturated and estimated models and do not test the structural paths. Alternative orderings of the three constructs would yield identical fit indices, so the tested ordering rests on theory and not on the data.

Table 7. Model fitness.

Fit Index	Saturated Model	Estimated Model
SRMR	0.053	0.053
d_ULS	0.599	0.599

Table 7. *Cont.*

Fit Index	Saturated Model	Estimated Model
d_G	0.585	0.585
Chi-square	564.573	564.573
NFI	0.831	0.831

4.2. Structural Model

4.2.1. Hypothesis Testing

Hypothesis testing showed that diagnostic reasoning and pedagogical adaptation was positively associated with reflective intelligence ($\beta = 0.539$, $t = 6.429$, $p < 0.001$) more strongly than inquiry-based pedagogical facilitation was ($\beta = 0.345$, $t = 3.760$, $p < 0.001$), while the strongest direct path overall was from inquiry-based pedagogical facilitation to diagnostic reasoning and pedagogical adaptation ($\beta = 0.794$, $t = 24.542$, $p < 0.001$). The 95% bootstrap confidence intervals of all structural paths excluded zero (Table 8).

Table 8. Hypothesis testing, specific indirect effects, and total effects. β = original sample estimate (path coefficient or effect); M = sample mean; STDEV = bootstrap standard deviation; 2.5% and 97.5% = limits of the 95% percentile bootstrap confidence interval. RI = reflective intelligence.

Path	β	M	STDEV	T Stat	*p*	2.5%	97.5%
Hypotheses Testing							
DRPA → RI	0.539	0.542	0.084	6.429	<0.001	0.380	0.707
IBPF → DRPA	0.794	0.797	0.032	24.542	<0.001	0.729	0.856
IBPF → RI	0.345	0.343	0.092	3.760	<0.001	0.160	0.517
Specific Indirect Effects							
IBPF → DRPA → RI	0.428	0.433	0.073	5.848	<0.001	0.296	0.582
Total Effects							
DRPA → RI	0.539	0.542	0.084	6.429	<0.001	0.380	0.707
IBPF → DRPA	0.794	0.797	0.032	24.542	<0.001	0.729	0.856
IBPF → RI	0.774	0.776	0.038	20.573	<0.001	0.697	0.844

4.2.2. Mediation Analysis

Mediation analysis indicated that diagnostic reasoning and pedagogical adaptation statistically mediated the association between inquiry-based pedagogical facilitation and reflective intelligence ($\beta = 0.428$, $t = 5.848$, $p < 0.001$).

4.2.3. Total Effects

The total effect of inquiry-based pedagogical facilitation on reflective intelligence ($\beta = 0.774$, $t = 20.573$, $p < 0.001$) was statistically significant. All findings are presented in Table 8. The structural and measurement models are shown in Figure 2, and hypothesis testing is shown in Figure 3.

The model has moderate-to-substantial explanatory power, with R^2 values of 0.631 (63.1%) for DRPA and 0.706 (70.6%) for reflective intelligence, per the thresholds set by Hair et al. [25] (Table 9). These R^2 values describe variance explained within the sample. Predictive performance is assessed in Section 4.3.

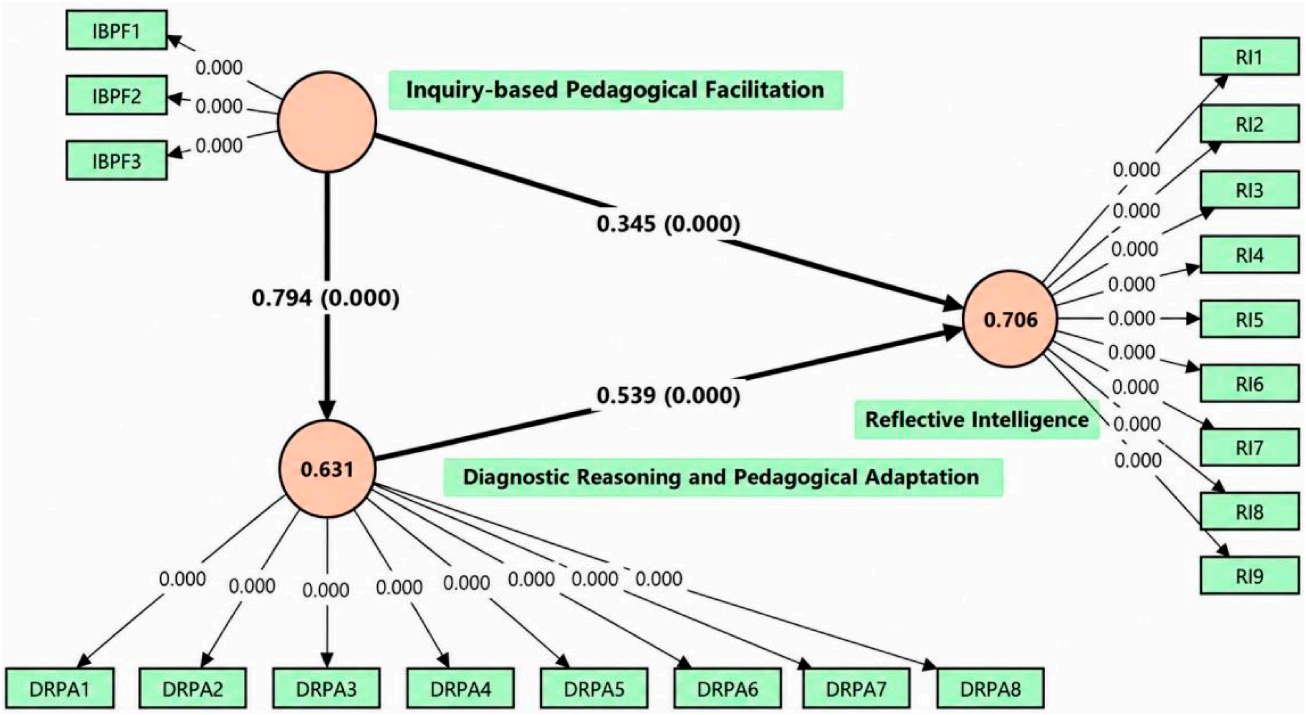

Figure 3. Hypothesis testing. Values on the paths are path coefficients with *p*-values in parentheses, and values on the arrows to the indicators are *p*-values; 0.000 indicates $p < 0.001$.

Table 9. Explanatory power (R-square).

Construct	R-Square	R-Square Adjusted
Diagnostic Reasoning and Pedagogical Adaptation	0.631	0.629
Reflective Intelligence	0.706	0.702

Effect sizes ranged from moderate to very high [25]: the effect of inquiry-based pedagogical facilitation on reflective intelligence was moderate ($f^2 = 0.149$), the effect of diagnostic reasoning and pedagogical adaptation on reflective intelligence was high ($f^2 = 0.364$), and the effect of inquiry-based pedagogical facilitation on diagnostic reasoning and pedagogical adaptation was very high ($f^2 = 1.710$) (Table 10).

Table 10. Effect size.

Path	f-Square
Diagnostic Reasoning and Pedagogical Adaptation → Reflective Intelligence	0.364
Inquiry-based Pedagogical Facilitation → Diagnostic Reasoning and Pedagogical Adaptation	1.710
Inquiry-based Pedagogical Facilitation → Reflective Intelligence	0.149

4.3. Predictive Performance (PLSpredict)

Predictive performance was assessed with PLSpredict [31] using 10 folds and 10 repetitions. Q^2predict was greater than zero for all indicators and for both endogenous constructs, indicating predictive relevance (Table 11). For the indicators, the PLS-SEM prediction error (RMSE) was lower than that of the linear model (LM) benchmark for 14 of the 17 indicators; the exceptions were DRPA4, DRPA5, and RI6. Because most, but not all, indicators outperformed the benchmark, the model shows medium out-of-sample predictive power

according to the guideline of Shmueli et al. [31]. For the latent variables, Q^2predict was 0.623 for DRPA and 0.590 for RI.

Table 11. PLSpredict results for the indicators and latent variables. IBPF = inquiry-based pedagogical facilitation; DRPA = diagnostic reasoning and pedagogical adaptation; RI = reflective intelligence.

Indicator	Q^2predict	PLS-SEM RMSE	LM RMSE
DRPA1	0.466	0.839	0.848
DRPA2	0.420	0.862	0.865
DRPA3	0.416	0.838	0.859
DRPA4	0.377	0.940	0.922
DRPA5	0.432	0.915	0.913
DRPA6	0.389	0.913	0.926
DRPA7	0.412	0.857	0.873
DRPA8	0.506	0.759	0.766
RI1	0.404	0.892	0.911
RI2	0.422	0.923	0.946
RI3	0.421	0.832	0.835
RI4	0.366	0.952	0.954
RI5	0.424	0.812	0.825
RI6	0.417	0.847	0.813
RI7	0.407	0.922	0.943
RI8	0.470	0.850	0.866
RI9	0.347	0.897	0.915
Latent variable			
Construct	Q^2predict	RMSE	MAE
DRPA	0.623	0.623	0.461
RI	0.590	0.649	0.455

5. Discussion

5.1. Interpretation of Key Findings

The present study examined the structural relationships among inquiry-based pedagogical facilitation (IBPF), diagnostic reasoning and pedagogical adaptation (DRPA), and reflective intelligence (RI) among Saudi female pre-service early childhood teachers, yielding four consistent and theoretically interpretable findings. The structural model showed moderate-to-substantial explanatory power in the sample, with IBPF and DRPA jointly accounting for 70.6% of the variance in RI and IBPF alone accounting for 63.1% of the variance in DRPA, both exceeding the moderate threshold (0.50), with the variance explained in RI approaching the substantial threshold (0.75) recommended by Hair et al. [25]. PLSpredict indicated medium out-of-sample predictive power for the indicators (Section 4.3). The results are consistent with a practice-embedded account in which self-reported inquiry-based facilitation is closely associated with both self-reported diagnostic reasoning and reflective intelligence, and they are compatible with, though they cannot be distinguished from, a reverse or reciprocal ordering of reflection and practice. Figure 1 shows the conceptual model, and Figures 2 and 3 show the estimated model; the reciprocal path discussed below was not tested. All four hypothesized paths were statistically significant, which lends confidence

that the model reflects a consistent pattern and not simply a collection of loosely related associations. Because the data are cross-sectional and self-reported, however, the paths are interpreted as associations; they do not establish temporal order or causation. Wording in this section and in Section 5.4 that refers to the development or acquisition of diagnostic and reflective competence reflects the theoretical accounts cited and is offered as a plausible interpretation; these data do not demonstrate such developmental processes.

The direct path from IBPF to DRPA is very large [25] ($\beta = 0.794$, $t = 24.542$, $p < 0.001$, $f^2 = 1.710$); that is, pre-service teachers who report more frequent inquiry-based facilitation also report markedly stronger diagnostic reasoning and pedagogical adaptation. This is consistent with the view of Reuker and Künzell [13] that diagnostic skill acquisition is action embedded: diagnostic skill is thought to be built incrementally through teachers' repeated interactions with actual pedagogical situations where predicted and observed outcomes vary, and not through abstract thinking. In a learning-lab structure in which learners observe, question, and experiment, the STEM teacher has to read students' responses, detect misconceptions, and adjust the explanation as the lesson proceeds, which is the kind of situation that practice-embedded accounts associate with diagnostic skill. The present finding is compatible with this theoretically proposed mechanism and applies it beyond the realm of physical education to STEM education in a non-Western context.

The substantial direct path from DRPA to RI ($\beta = 0.539$, $t = 6.429$, $p < 0.001$, $f^2 = 0.364$) indicates that pre-service teachers who report more often noticing student cues, probing children's thinking, and adapting explanations also report a more sophisticated ability to reflect on their teaching. This aligns with Kramer et al.'s [1] demonstration that diagnostic competence involves not only evidence generation and hypothesis formation but also the articulation of interpretive conclusions, a process that inherently requires reflection on the adequacy of one's own knowledge and observational frameworks, and with Heckemann et al.'s [9] account of deep, third-level reflection that integrates affective and evaluative dimensions. Diagnostic activities, particularly those that reveal unexpected student thinking, may create cognitive dissonance that motivates deeper reflective engagement, an interpretation that longitudinal or experimental work would need to test.

The direct association between IBPF and RI ($\beta = 0.345$, $t = 3.760$, $p < 0.001$), alongside the significant indirect path through DRPA ($\beta = 0.428$, $t = 5.848$, $p < 0.001$), is consistent with partial mediation: self-reported inquiry-based practice is associated with reflective intelligence both directly and indirectly via diagnostic reasoning and pedagogical adaptation. That the indirect path accounts for a larger share of the total effect ($\beta = 0.774$) than the direct path suggests that much of the association between inquiry-based facilitation and reflective intelligence is statistically accounted for by the indirect path through diagnostic reasoning and adaptation; the present data cannot show why the indirect path is the larger one.

5.2. Comparison with Prior Literature

The findings add cross-sectional evidence to prior scholarship on reflective practice, diagnostic competence, and inquiry-based facilitation without resolving the direction of the relationships that this scholarship posits. The prevailing theoretical logic positions reflection as an antecedent to effective diagnostic and facilitative practice: Rothe and Göbel [2] showed that structured collaborative reflection on videotaped teaching enhanced pre-service teachers' analytical capabilities, and Firat and Dinçer [7] found that action-research cycles developed reflective thinking among pre-service preschool teachers. These findings are consistent with the view that reflection is valuable for pre-service teachers' learning but do not preclude a reverse or reciprocal relationship, in which engagement with the demands of inquiry-based facilitation is associated with pre-service teachers' reflective intelligence. This two-way conceptualization is supported theoretically by Merkebu et al. [4],

who stated that reflective intelligence is a recursive capacity that is generated in cycles of engagement, not a capacity that exists in isolation before an action, and by Markov and Slavova [8], who described reflective intelligence as a product of contextual embeddedness and experiential feedback. The present study therefore offers a quantitative perspective that is consistent with the view that inquiry-based pedagogical contexts may be associated with reflection, and that reflective intelligence may be related to self-reported inquiry-based practice in both directions. Because the present data cannot establish direction, both orderings remain plausible.

The finding that IBPF is very strongly associated with DRPA is consistent with Reuker and Künzell's [13] action-theory perspective, which positions diagnostic skill development as contingent on the situation-action-effect associations teachers accumulate through practice. In theory, inquiry-based facilitation presents a richer repertoire of pedagogical situations than traditional didactic instruction, including moments where anticipated and actual student responses diverge, which are the occasions from which differentiated diagnostic noticing is proposed to emerge. The very large effect size observed ($f^2 = 1.710$) could reflect such a relationship, but it may equally reflect construct overlap (the HTMT between the two constructs was 0.895, with an upper confidence limit of 0.961; Section 4.1), shared method variance, or a tendency for pre-service teachers who report stronger diagnostic reasoning to report more frequent inquiry-based facilitation. Bou Saad et al. [14] similarly identified inquiry- and problem-based learning as demanding continuous student-centered assessment from teachers, reinforcing that facilitative pedagogies place ongoing diagnostic pressure on teachers. This pattern is directionally consistent with recent evidence from pre-service STEM teacher education: Payoungkiattikun et al. [35] found that an 18-week project-based learning intervention was followed by significant, large gains in pre-service science teachers' self-assessed metacognitive skills across three STEM disciplines, with project-based engagement emerging as the strongest predictor of metacognitive growth ($\beta = 0.45$, $p < 0.0001$) alongside academic level, together explaining 65% of the variance, loosely paralleling the present finding that self-reported inquiry-based practice is strongly associated with cognitive and metacognitive capacities. Notably, however, Payoungkiattikun et al. [35] could not rule out maturation effects given the absence of a control group, a limitation that the present design shares.

The DRPA–RI association extends Kramer et al.'s [1] finding that pedagogical content knowledge is the strongest predictor of diagnostic activities ($\beta = 0.29$, $p < 0.001$) and accuracy ($\beta = 0.23$, $p = 0.002$), suggesting an additional pathway to consider: diagnostic engagement may itself be linked to the reflective awareness that constitutes reflective intelligence, consistent with Heitzmann et al.'s [18] eight-step diagnostic model, in which hypothesis generation, evidence evaluation, and conclusion communication intrinsically require examining one's own knowledge and interpretive frameworks. Gebremariam and Sisay [15] similarly reported that structured diagnostic and evaluative activities in Ethiopian classrooms were associated with improved teacher pedagogical quality through enhanced self-awareness.

The Saudi-specific findings also merit comparison with literature on diagnostic challenges in similar settings. Kausar [5] found that reduced teaching hours, language barriers, and faculty shortages severely constrain diagnostic skill development in Saudi medical education. The present study suggests that where pre-service teachers report more frequent inquiry-based facilitation, they also tend to report stronger diagnostic and reflective capacities. This mirrors, in reverse, Aslam et al.'s [6] finding that novice STEM teachers' reflective and diagnostic difficulties stem partly from limited experience with student-centered methodologies. Read alongside one another, these studies suggest, as a possibility that the present data cannot test, that the institutional constraints documented across Saudi

education (time pressure, resource limitations, and limited exposure to student-centered methods) may be relevant to the amount of inquiry-based practice that teachers report, with which diagnostic and reflective capacities appear to be associated.

Not all recent evidence is consistent with this practice-embedded account of reflection, however, and the discrepancy is instructive rather than merely contradictory. Li [36] tested an analogous bridge hypothesis among 377 Chinese pre-service teachers and their supervisors, examining whether practical teaching competence and lifelong-learning/reflective disposition co-vary as reflective-practice theory would predict; of 20 correlation coefficients across five competence domains and four lifelong-learning attributes, 19 were statistically non-significant or practically trivial, and the study's explicit test of the reflective bridge hypothesis was not supported. Li [36] attributes this null result to participants' own reports of limited hands-on teaching opportunities, weak mentoring, and insufficient structured reflection time, conditions under which practice was too thin and unstructured to be accompanied by the hypothesized reflective gains. This is consistent with, rather than contradictory to, the present findings: where Li's [36] pre-service sample was asked about practical teaching competence in general and reported limited hands-on opportunities, the present study asked Saudi pre-service early childhood teachers about one specific facilitative pedagogy, inquiry-based facilitation, in their STEM teaching activities, not generic "practical teaching competence". Taken together, the two studies suggest that it may be the particular character and intensity of inquiry-based practice, as reported by respondents, and not classroom experience in the abstract, that is associated with whether a reflective or diagnostic bridge emerges, a boundary condition that future research should test directly. Self-regulated learning research outside the field of teacher education revealed a parallel warning: de Vreugd et al. [37] tested a learning analytics dashboard and found that when these dashboards were used to encourage university students to reflect on their own learning activities, it did lead to new awareness of their learning, but this awareness did not necessarily result in behavior change, as students often reported that they had not yet tried, had forgotten, or believed they already did what was suggested and were unable to do otherwise. The need for the observational and longitudinal follow-up discussed below is relevant to the interpretation of the present study's self-report structural model: high reflective intelligence scores reflect teachers' perceptions of reflecting and adapting, but, as in the study by de Vreugd et al. [37], such self-reports do not necessarily mean that these pre-service teachers are actually reflective and adaptable in the classroom.

Outside of teacher education, a similar mediation relation has recently been documented: Gürefe and Evrensel Yumrutepe [38] showed that critical thinking tendency (CTT) played a key role in the relationship between students' mathematics-related affect and their reflective problem-solving in middle school, and the direct effect of affect decreased significantly when CTT was added as the mediator. While that study investigated students' mathematical cognition, it is instructive that in both studies reflection is positioned after engagement in the statistical model, which reflects the models' specification and not a demonstrated temporal order: in that study, critical thinking tendency intervened between mathematics-related affect and reflective problem-solving, whereas in the present study, diagnostic reasoning and pedagogical adaptation are the intervening constructs between inquiry-based facilitation and reflective intelligence.

5.3. Theoretical Contributions

This study makes four contributions to understanding teacher cognition among pre-service early childhood teachers. First, it provides quantitative evidence that is consistent with Reuker and Künzell's [13] theoretical proposal that diagnostic skill is built through repeated engagement in inquiry-based practice, a proposal that these cross-sectional data

can neither test nor confirm, offering an empirical association that is compatible with a mechanism that had been proposed theoretically, and it extends the diagnostic competence literature, which has emphasized professional knowledge as the driver of diagnostic quality [1], by pointing to self-reported inquiry-based facilitation as a further correlate of self-reported diagnostic reasoning, whose practical relevance intervention studies would need to test. Second, it positions reflective intelligence within a practice-related model: rather than treating it solely as a pre-existing trait [8] or a product of deliberately structured reflection [2,7], the findings suggest that it is also associated with the metacognitive demands of inquiry-based facilitation, extending Li et al.'s [3] view of reflective intelligence as a self-organizing system responsive to environmental feedback.

Third, the partial mediation of DRPA advances understanding of how practice may be related to metacognitive capacity: inquiry-based facilitation is associated with reflective intelligence partly through its association with diagnostic reasoning and adaptation, which offers a conceptual architecture in which practice, diagnosis, and reflection are linked as a set of associations, consistent with constructivist accounts of professional learning [11]. Finally, few, if any, SEM-based studies have examined these relationships in Saudi STEM education, and this study brings reflective intelligence theory from the cognitive science and artificial intelligence fields [3,8] to early childhood teacher education in a non-Western, Gulf-region context. The questionnaire showed acceptable reliability and convergent validity (Cronbach's alpha: 0.827–0.945; average variance extracted: 0.691–0.743) and acceptable model fit (SRMR: 0.053), which supports reliable operationalization of the scales in this Arabic-speaking sample, although discriminant validity was not conclusively established (Section 4.1) and further validation is needed before cross-cultural comparative use.

5.4. Practical Implications

The results have possible implications for teacher education, professional learning and policy. These should be read as hypotheses for future testing and not as demonstrated consequences, because the study is cross-sectional, relies on self-report, and did not measure student outcomes, teachers' professional growth over time, or the effects of any intervention. The strong association between IBPF and DRPA suggests that pre-service teachers who report more frequent inquiry-based facilitation also report stronger diagnostic reasoning. Teacher education programs could therefore examine whether continuous, mentored immersion in inquiry-based facilitation, for example through structured placements during the practicum phase, the three-stop technique for video analysis [13], and structured debriefing, supports the development of diagnostic and reflective competence together, a question that intervention studies could test. Regarding reflective intelligence, the results raise the possibility that explicit diagnostic training, for instance through the video-based reflection approach of Rothe and Göbel [2], may be a useful complement to reflection journaling, a comparison that the present data cannot address.

In terms of supporting inquiry teaching, granting pre-service teachers time, flexibility during field experience, and organized opportunities for peer observation and mentoring may be a worthwhile way to support teachers' diagnostic and reflective learning rather than a pedagogical luxury; this may be especially relevant in Saudi Arabia, where schools are gender-separated and where such learning communities might be particularly effective. Time constraints are a well-documented problem in implementing inquiry-based pedagogy and in teacher reflection, and they could be considered during curriculum reform to leave room for authentic inquiry facilitation [5,6]. Policy makers and teacher education programs might also consider whether inquiry-based facilitation competence should be included in teacher preparation standards and evaluation frameworks, but the present findings do not show that doing so would improve learning outcomes or teacher professionalism; such

claims require studies that measure student outcomes and teachers' professional growth over time.

These possible implications also connect to sustainable education. First, teachers are central to the quality of education that Sustainable Development Goal 4 aims for, and the capacity to diagnose learners' needs and adapt instruction is part of sustaining quality teaching as curricula and pedagogies change. Second, in Saudi Arabia, Vision 2030 treats education and STEM excellence as foundations for long-term human-capital development, and teacher competencies that are associated with one another may support this aim. Third, reflective and diagnostic capacities may contribute to the resilience of teachers and schools during reform, when student-centered practices replace traditional ones [4]. Constructivist, active-learning approaches to STEM have likewise been linked to sustainable education [14]. These links are conceptual: the present study did not measure student outcomes, long-term changes in teacher competence, long-term human-capital indicators, or resilience, and they should be tested directly.

5.5. Limitations

This study has several limitations. First, because of the cross-sectional design, it is not possible to conclude that the associations between inquiry-based facilitation and diagnostic reasoning and reflective intelligence are causal or that they result from inquiry-based facilitation. The associations could arise because pre-service teachers who already report higher levels of these capacities are more likely to facilitate inquiry, or because these capacities are stable characteristics of particular teachers; reverse and reciprocal relationships are equally possible. Competing orderings were not estimated, and they cannot be distinguished empirically in a saturated cross-sectional model. Second, all constructs were measured with the same self-report questionnaire, at one time point, from the same respondents. Social desirability may have inflated scores on normatively desirable constructs such as reflection and inquiry facilitation, and self-reports are an imperfect measure of actual teaching behavior in classrooms. Common method variance may also have inflated the associations among the constructs. A VIF-based collinearity check did not indicate a problem [27], but it does not test common method bias. Several procedural remedies were applied (Section 3.4), but others, such as separating the measurement of the constructs in time or collecting data from multiple sources, were not used.

Third, the sample consisted of female pre-service early childhood teachers from a single university and region, recruited through purposive sampling of students whom their instructors identified as having studied, or being scheduled to study, the science-related courses of the program plan (280 identified, of whom 250 were invited to the main study, from approximately 1000 enrolled students), which limits generalizability across gender, institutions, regions, in-service teachers, and other educational stages. Selection bias is also possible, because pre-service teachers who are more interested in reflective and inquiry-based practice may have been more willing to respond. Because the instructors identified students who were willing to take part, the sample may over-represent students who are more interested in STEM-related content, and instructor selection may have introduced additional selection bias. Of the 250 students invited, 179 provided valid responses (71.6%), so non-response bias cannot be excluded. The program is a generalist early childhood program and not a specialized STEM teacher preparation program, so the findings concern pre-service generalist teachers' approaches to STEM-related instruction and cannot be assumed to hold for subject-specific STEM teacher preparation or for male teachers, particularly because schools in Saudi Arabia are gender-separated. Their applicability to other teacher-training programs requires further investigation. Fourth, although participants came from four stages of the program with different amounts of practicum experience, the

analysis did not examine whether the relationships differ by year of study or practicum experience, so it is unclear whether they are consistent across stages of training. In addition, because 52.0% of respondents had not completed any practicum, their responses to frequency-based items probably reflect self-reported or anticipated practice rather than actual teaching, so the findings should be read as associations among perceived capacities and not among observed classroom behaviors. They should not be read as direct evidence of teaching effectiveness or actual classroom behavior. A formal comparison by practicum experience was not undertaken because practicum status is nested within year of study (Table 1), so any contrast would confound practicum experience with coursework exposure and academic level. The three-item IBPF scale and the need to establish measurement invariance first further limit the interpretability of such a test. Because more than half of the respondents had not yet taught in a practicum, the associations may also partly reflect shared self-perception or common method variance rather than observed practice. The findings are therefore interpreted as associations among perceived capacities in a mixed-stage sample, and multi-group analysis, including tests of measurement invariance, is recommended for future research.

Fifth, the inquiry-based pedagogical facilitation scale has only three items, which may have limited the breadth of measurement. Its reliability ($\alpha = 0.827$) and AVE (0.743) were acceptable, but a short scale cannot cover the facets of questioning, scaffolding, and formative assessment, may limit content validity, and allows only limited examination of dimensionality; a more extensive operationalization may yield more differentiated relationships. This is consistent with the systematic review of Portillo-Blanco et al. [21], which found that inquiry, one of the few STEM education principles with broad conceptual consensus, is still highly inconsistent in the literature with respect to its operational characteristics; the present three-item measure can be understood as one defensible but partial operationalization of an inquiry construct that the field has not yet standardized. In addition, the expert panel was small (four professors), six items had an I-CVI of 0.75, and S-CVI/UA was 0.70 (Section 3.3), so the content validity of the instrument is only moderately supported and should be confirmed with a larger panel. Sixth, the three constructs overlap conceptually and are strongly correlated; the upper confidence limits of the HTMT exceeded 0.90 for all three pairs (Table 4) and all three point estimates exceeded the more conservative 0.85 threshold, so their empirical distinctiveness should be re-examined in independent samples. Seventh, an a priori power analysis was not conducted, and sample adequacy was assessed post hoc (Section 3.2). The pilot sample ($n = 30$) was also too small to evaluate the measurement model reliably: in the pilot, the reliability of the inquiry-based facilitation scale (Cronbach's alpha = 0.690) and its discriminant validity from diagnostic reasoning were weak, but both improved in the main sample, although the discriminant validity results remain subject to the cautions noted above. Eighth, the study did not assess sustainability outcomes or student learning, so its link to sustainable education is conceptual and should be kept distinct from the empirical findings. Finally, the Saudi STEM context, which is a positive aspect in terms of contextual relevance, restricts generalizability to other national and cultural contexts.

5.6. Directions for Future Research

The results indicate some areas for further research. Panel studies tracking Saudi pre-service teachers from teacher preparation into their first years of teaching with time-lagged SEM or growth curve modeling would be able to test the proposed reciprocal relationship, while pre-post measurement designs in intervention studies that introduce training in IBPF would help clarify the direction of the association between IBPF and DRPA. Classroom observations, stimulated recall, and think-aloud protocols could be integrated with a mixed-methods design, using the three-stop video analysis approach developed by

Reuker and Künzell [13] in early childhood STEM classrooms in Saudi Arabia to deepen the understanding of the associations between self-reported inquiry-based facilitation and diagnostic reasoning.

Additionally, moderating factors like teacher experience, qualifications, and type of school should be explored in future studies, and multi-group SEM could be useful in this regard, since Aslam et al. [6] showed that novice STEM teachers' challenges in diagnosis and reflection may differ qualitatively from those of experienced teachers. The model also should be expanded to student outcome variables such as achievement, inquiry skills, and self-regulated learning, and tested for measurable student benefits, especially as Radović et al. [10] identified reflective thinking as a weak point in students' self-regulated learning.

Finally, comparative research could be conducted in other GCC countries, with female-only or co-educational teaching models, and in contexts outside the GCC, such as the Ethiopian context studied by Gebremariam and Sisay [15], to determine whether these relationships are similar or different and which elements of the model are common across countries and which are specific to the Saudi context, thereby promoting culturally responsive professional development aligned with Vision 2030.

6. Conclusions

The aim of this study was to explore the structural relationships among the three constructs of inquiry-based pedagogical facilitation (IBPF), diagnostic reasoning and pedagogical adaptation (DRPA), and reflective intelligence (RI), within the context of Saudi female pre-service early childhood teachers, who constitute a future teaching workforce at the heart of Saudi Arabia's educational transformation but have been relatively understudied. Using PLS-SEM on 179 valid responses collected via a purpose-built, bilingual, expert-reviewed questionnaire, the study tested direct effects, a mediated pathway, and the explanatory power of a three-construct model. All four hypotheses were supported, and the model showed moderate-to-substantial explanatory power in the sample, accounting for 63.1% of the variance in DRPA and 70.6% in RI.

The principal findings form a theoretically interpretable pattern. IBPF was very strongly associated with DRPA ($\beta = 0.794$, $p < 0.001$, $f^2 = 1.710$), consistent with the theoretical view that the student-responsive demands of inquiry-based practice may offer a rich context for diagnostic competence, which these data cannot confirm. DRPA, in turn, was positively associated with RI ($\beta = 0.539$, $p < 0.001$, $f^2 = 0.364$), consistent with the idea that self-reported noticing of student cues, testing of diagnostic hypotheses, and adaptation of instruction relate to pre-service teachers' metacognitive self-awareness. IBPF was also directly associated with RI ($\beta = 0.345$, $p < 0.001$), and the significant indirect path through DRPA ($\beta = 0.428$, $p < 0.001$) was consistent with partial mediation, indicating that self-reported inquiry-based practice is statistically associated with reflective intelligence both directly and through the indirect path via diagnostic reasoning and pedagogical adaptation. These findings are compatible with reflective intelligence being closely linked to the frequency of self-reported inquiry-based practice, but they are equally compatible with reflection preceding practice, and the present design cannot settle the order.

Theoretically, the study provides quantitative evidence consistent with Reuker and Künzell's [13] action-theoretic model of diagnostic skill acquisition in a population of pre-service early childhood teachers, extends reflective intelligence from its origins in cognitive science and AI research into early childhood teacher education, and proposes a practice-embedded account, to be tested in longitudinal work, that enriches existing frameworks [1,8]. Practically, teacher educators and professional development designers in Saudi Arabia and comparable reform contexts may wish to consider sustained, mentored immersion in inquiry-based facilitation as a promising context in which diagnostic and

reflective competence could be developed together, subject to testing in longitudinal and intervention studies; policy frameworks aligned with Vision 2030 could consider whether inquiry-based approaches in STEM classrooms are accompanied by gains in student learning quality and teacher professional growth, which would require studies that measure student outcomes and follow teachers over time.

These conclusions are bounded by the study's cross-sectional design, which precludes causal attribution; its reliance on self-report measures, which may introduce social desirability bias; its restriction to female Saudi pre-service teachers, which limits generalizability across gender, to in-service teachers, and across national contexts; and the narrow operationalization of IBPF using three items. Because the three constructs overlap substantially (Section 4.1), the results also do not show that inquiry-based facilitation, diagnostic reasoning, and reflective intelligence are clearly separable competencies, and they do not support claims about independent developmental mechanisms. The findings concern relationships among self-reported capacities and not teaching effectiveness or classroom behavior. Future research should address these boundaries through longitudinal or experimental designs, mixed-methods protocols incorporating classroom observation and video-based stimulated recall, moderation analyses, and cross-cultural comparative studies extending the model to other GCC and non-Western educational settings. For Saudi Arabia and comparable education systems pursuing Vision 2030-aligned reform, this raises the possibility that investment in inquiry-based pedagogy may be accompanied by stronger self-reported diagnostic and reflective capacities among pre-service teachers, an implication that remains to be tested.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su182312099/s1, Supplementary Material (final questionnaire items in English and Arabic) containing Tables S1–S3 (Table S1. Reflective intelligence (RI) items, Table S2. Inquiry-based pedagogical facilitation (IBPF) items, Table S3. Diagnostic reasoning and pedagogical adaptation (DRPA) items) and Table S4 (relationship between the initial draft items and the final diagnostic reasoning and pedagogical adaptation items).

Funding: The author would like to acknowledge the Deanship of Graduate Studies and Scientific Research, Taif University, for funding this work.

Institutional Review Board Statement: IRB approval obtained (approved by the Scientific Research Ethics Committee of Taif University (47-024) on [28 September 2025] as part of a broader research project encompassing multiple studies, including the current study. Subsequently, a separate ethical approval specific to the current study was obtained from the same committee (48-036) on [17 September 2026).

Informed Consent Statement: Informed consent for participation was obtained from all subjects involved in the study.

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

Acknowledgments: Anthropic's Claude Sonnet 5.5 and OpenAI's GPT-6 were used for language editing, for the graphical presentation of figures based on the author's own model and results, and for formatting the manuscript according to the journal's style requirements. The statistical analysis was conducted by the author in SmartPLS 4.1.1.7; these tools were not used to analyze data or to generate findings. All data, findings, interpretations, analysis, and academic content, including means, standard deviations, item scores, and all narrative descriptions, were produced exclusively by the author through original empirical research. The author reviewed all edited text and takes full responsibility for the content of the publication. I am thankful to all the participants who voluntarily participated in this research and consented to the acknowledgement.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Kramer, M.; Förtsch, C.; Boone, W.J.; Seidel, T.; Neuhaus, B.J. Investigating pre-service biology teachers' diagnostic competences: Relationships between professional knowledge, diagnostic activities, and diagnostic accuracy. *Educ. Sci.* **2021**, *11*, 89. [CrossRef]
2. Rothe, L.; Göbel, K. Preservice teachers' reflection processes when collaboratively reflecting on videotaped classroom teaching. *Educ. Sci.* **2024**, *14*, 1357. [CrossRef]
3. Li, L.; Wang, Y.; Zhao, H.; Kong, S.; Teng, Y.; Li, C.; Wang, Y. Reflection-Bench: Probing AI intelligence with reflection. *arXiv* **2025**, arXiv:2410.16270. [CrossRef]
4. Merkebu, J.; Veen, M.; Hosseini, S.; Varpio, L. The case for metacognitive reflection: A theory integrative review with implications for medical education. *Adv. Health Sci. Educ.* **2024**, *29*, 1481–1500. [CrossRef] [PubMed]
5. Kausar, N. Anatomy teaching in problem-based learning curriculum: Challenges and opportunities in the Kingdom of Saudi Arabia. *Front. Educ.* **2025**, *10*, 1686355. [CrossRef]
6. Aslam, S.; Alghamdi, A.A.; Abid, N.; Kumar, T. Challenges in implementing STEM education: Insights from novice STEM teachers in developing countries. *Sustainability* **2023**, *15*, 14455. [CrossRef]
7. Firat, Z.S.; Dinçer, F.Ç. An Action Research on the Development of the Reflective Thinking Skills of Preservice Preschool Teachers. *Syst. Pract. Action Res.* **2024**, *37*, 985–1013. [CrossRef]
8. Markov, K.; Slavova, V. Transitive self-reflection, A fundamental criterion for detecting intelligence. *Proceedings* **2025**, *126*, 8. [CrossRef]
9. Heckemann, B.; Schols, J.M.; Halfens, R.J. A reflective framework to foster emotionally intelligent leadership in nursing. *J. Nurs. Manag.* **2014**, *23*, 744–753. [CrossRef] [PubMed]
10. Arvatz, A.; Peretz, R.; Dori, Y.J. Self-Regulated Learning and Reflection: A Tool for Teachers and Students. *Metacogn. Learn.* **2025**, *20*, 15. [CrossRef]
11. Ortube, A.F.; Panadero, E.; Dignath, C. Self-regulated learning interventions for pre-service teachers: A systematic review. *Educ. Psychol. Rev.* **2024**, *36*, 113. [CrossRef]
12. Malik, S.; Rashid, S.; Haider, A. Trait emotional intelligence as a dynamic learning outcome: Predictive roles of attitude and engagement in EFL pedagogy. *Lang. Teach. Res.* **2026**, 1–34. [CrossRef]
13. Reuker, S.; Künzell, S. Learning diagnostic skills for adaptive teaching: A theoretical foundation. *Cogent Educ.* **2021**, *8*, 1887432. [CrossRef]
14. Bou Saad, R.; Llorens Garcia, A.; Cabre Garcia, J.M. Mapping constructivist active learning for STEM: Toward sustainable education. *Sustainability* **2025**, *17*, 6225. [CrossRef]
15. Gebremariam, H.T.; Sisay, A.B. Reflective practice in higher education institutions: Exploring teachers' reflective experiences and challenges in Ethiopian public universities. *Interchange* **2024**, *55*, 405–428. [CrossRef]
16. Herppich, S.; Praetorius, A.-K.; Förster, N.; Glogger-Frey, I.; Karst, K.; Leutner, D.; Behrmann, L.; Böhmer, M.; Ufer, S.; Klug, J.; et al. Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teach. Teach. Educ.* **2018**, *76*, 181–193. [CrossRef]
17. Chernikova, O.; Heitzmann, N.; Fink, M.C.; Timothy, V.; Seidel, T.; Fischer, F. Facilitating diagnostic competences in higher education—A meta-analysis in medical and teacher education. *Educ. Psychol. Rev.* **2020**, *32*, 157–196. [CrossRef]
18. Heitzmann, N.; Seidel, T.; Opitz, A.; Hetmanek, A.; Wecker, C.; Fischer, M.; Ufer, S.; Schmidmaier, R.; Neuhaus, B.; Siebeck, M.; et al. Facilitating diagnostic competences in simulations in higher education. *Frontline Learn. Res.* **2019**, *7*, 1–24. [CrossRef]
19. Coutts, R.; Irwin, P.; Del Vecchio, L. A reflective practice learning experience with higher education exercise physiology students. *Reflective Pract.* **2024**, *25*, 251–266. [CrossRef]
20. Portillo-Blanco, A.; Deprez, H.; De Cock, M.; Guisasola, J.; Zuza, K. A systematic literature review of integrated STEM education: Uncovering consensus and diversity in principles and characteristics. *Educ. Sci.* **2024**, *14*, 1028. [CrossRef]
21. Roehrig, G.H.; Dare, E.A.; Ellis, J.A.; Ring-Whalen, E. Beyond the basics: A detailed conceptual framework of integrated STEM. *Discip. Interdiscip. Sci. Educ. Res.* **2021**, *3*, 11. [CrossRef]
22. Chen, S.; Cheng, H.; Huang, Y. Emotion recognition in self-regulated learning: Advancing metacognition through AI-assisted reflections. In *Trust and Inclusion in AI-Mediated Education: Where Human Learning Meets Learning Machines*; Kourkoulou, D., Tzirides, A.-O., Cope, B., Kalantzis, M., Eds.; Springer: Cham, Switzerland, 2024; pp. 185–212. [CrossRef]
23. Laksita, G.D.; Pusporini, W.; Mutiara, E.; Zafrullah, Z.; Zamzami, Z. Implementation of STEM education in schools: A bibliometric analysis (2003–2024). *J. Penelit. Pendidik. IPA* **2024**, *10*, 944–959. [CrossRef]
24. Rashid, S.; Malik, S. Beyond universal models: Predicting trait emotional intelligence's context-contingent effects on EFL learners' attitudes, motivation, anxiety, and engagement. *Educ. Sci.* **2025**, *15*, 1137. [CrossRef]
25. Hair, J.F.; Hult, G.T.M.; Ringle, C.M.; Sarstedt, M. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 3rd ed.; Sage: Thousand Oaks, CA, USA, 2021.
26. Sarstedt, M.; Ringle, C.M.; Hair, J.F. Partial least squares structural equation modeling. In *Handbook of Market Research*; Homburg, C., Klarmann, M., Vomberg, A., Eds.; Springer: Cham, Switzerland, 2022; pp. 587–632.

27. Hair, J.F.; Alamer, A. Partial least squares structural equation modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Res. Methods Appl. Linguist.* **2022**, *1*, 100027. [CrossRef]
28. Kock, N.; Hadaya, P. Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Inf. Syst. J.* **2018**, *28*, 227–261. [CrossRef]
29. Polit, D.F.; Beck, C.T. The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Res. Nurs. Health* **2006**, *29*, 489–497. [CrossRef] [PubMed]
30. Polit, D.F.; Beck, C.T.; Owen, S.V. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res. Nurs. Health* **2007**, *30*, 459–467. [CrossRef] [PubMed]
31. Shmueli, G.; Sarstedt, M.; Hair, J.F.; Cheah, J.-H.; Ting, H.; Vaithilingam, S.; Ringle, C.M. Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *Eur. J. Mark.* **2019**, *53*, 2322–2347. [CrossRef]
32. Fornell, C.; Larcker, D.F. Structural equation models with unobservable variables and measurement error: Algebra and statistics. *J. Mark. Res.* **1981**, *18*, 382–388. [CrossRef]
33. Henseler, J.; Ringle, C.M.; Sarstedt, M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Mark. Sci.* **2015**, *43*, 115–135. [CrossRef]
34. Hu, L.; Bentler, P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model.* **1999**, *6*, 1–55. [CrossRef]
35. Payoungkiattikun, W.; Hemtasin, C.; Intanin, A.; Thongsuk, T. Project-based learning as a catalyst for fostering metacognitive skills in preservice science teachers. *Eur. J. Educ. Res.* **2025**, *14*, 453–470. [CrossRef]
36. Li, Z. Practical teaching competence and lifelong learning attributes of pre-service teachers in local normal universities in China. *Front. Educ.* **2026**, *11*, 1861123. [CrossRef]
37. de Vreugd, L.; van Leeuwen, A.; Jansen, R.; van der Schaaf, M. Learning analytics dashboard design and evaluation to support student self-regulation of study behaviour. *J. Learn. Anal.* **2024**, *11*, 249–262. [CrossRef]
38. Gürefe, N.; Evrensel Yumrutepe, N. Affective entry characteristics towards mathematics and reflective thinking skills towards problem solving: The mediating role of critical thinking. *J. Intell.* **2026**, *14*, 149. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.