

Evaluating middle school physics teachers' use of AI in lesson planning: A comparative study with three groups from Austria

Florian Budimaier^{1,2}, Matthias Fasching¹, Anna Reumann^{1,3} and Esmeralda Campos^{1,*}

¹*University of Vienna, Austrian Educational Competence Centre Physics,
Porzellangasse 4, 1090 Vienna, Austria*

²*University College of Teacher Education Vienna, Grenzackerstraße 18, 1100 Vienna, Austria*

³*University College of Teacher Education Burgenland,
Thomas-Alva-Edison-Straße 1, 7000 Eisenstadt, Austria*



(Received 9 January 2025; accepted 20 March 2026; published 24 April 2026)

Since 2022, artificial intelligence (AI) tools have become publicly available, thereby enabling physics teachers to employ them in a multitude of ways. The aim of this research is to evaluate the potential of AI in assisting physics teachers in lesson planning. The objective is to identify which cognitive demands of lesson planning can be carried out by the AI, therefore reducing the workload for teachers. To compare teachers with different backgrounds, we examine three groups: prospective, out-of-field, and experienced physics teachers. The participants planned a lesson on Newton's third law using an AI tool. To analyze the participants' lesson planning process, we employed evaluative qualitative content analysis. Each of the groups presented a different approach to lesson planning using AI. For example, out-of-field physics teachers focused on the physics content, while experienced physics teachers were more prone to asking how to adapt the learning materials to the students' cognitive level. One of the key findings was that the AI tools considered the clarity of the learning objectives and phrasing of the lessons without necessarily being prompted, in all three groups. Moreover, in general, the experienced physics teachers were more critical about using AI for lesson planning than the other two groups. The findings of this study might inform teacher professional development courses focused on the use of AI in physics education.

DOI: [10.1103/n611-h1mh](https://doi.org/10.1103/n611-h1mh)

I. INTRODUCTION

Since 2022, artificial intelligence (AI) tools have become publicly available, opening the door for physics teachers to use them in a variety of ways. In physics education research, the role of AI has mainly been studied by (i) testing its power to answer physics questions in well-known tests [1–3], (ii) analyzing its ability to demonstrate an understanding of physics discussion [4], and (iii) assessing its potential for aiding physics teachers in task development [5]. Kieser and others investigated ChatGPT's performance when answering the Force Concept Inventory (FCI) and found that, when prompted to answer with misconceptions, its answers approximate real human responses: the distribution of answers to the FCI resembled real students' distributions, and the accuracy of responses dropped considerably [1]. Polverini and Gregorcic investigated ChatGPT-4, which has the ability to process image data, on its performance when

answering the test of understanding graphs in kinematics [2]. They found that, although it provides good strategies for addressing each problem, it has limitations in the interpretation of the graphs, which leads to low performance. Yeadon and Hardy investigated ChatGPT-3.5 Turbo on its performance to answer over 1000 physics questions [3]. They found that it performed better on questions in early educational stages and worse on questions at the university level. Gregorcic and Pendrill analyzed the correctness of the physics knowledge of ChatGPT when answering a basic question of acceleration, finding that it does not master the physics concepts and often contradicts itself [4]. Küchemann *et al.* evaluated the quality of lesson planning when prospective physics teachers used either textbooks or AI to plan their lessons [5]. They found no significant difference in correctness, but the students using ChatGPT presented difficulties with output quality. The focus of this research follows from the need to critically assess the use of AI in task development, specifically in lesson planning, since it is a decisive activity for the teaching practice of educators.

Other areas of education research raise some implications of using AI in lesson planning. Powell and Courchesne found that ChatGPT was efficient at producing a lesson plan for first-grade students following the 5E model [6], but they often found problematic components, inaccuracies, and

*Contact author: esmeralda.campos@univie.ac.at

missing details [7]. Gonçalves Costa *et al.* conducted a study to compare the teaching sequences developed by a person, ChatGPT 3.5, and ChatGPT 4 [8]. They found that ChatGPT 4 outperforms ChatGPT 3.5, but they still exhibit limitations like shallow reasoning and sequences that could reinforce misconceptions. Davis and Lee found that ChatGPT designs stand-alone lessons, instead of integrating prior lessons or strategic preparation of future learning outcomes [9]. Even though using AI can help relieve some educational tasks and potentially prevent teacher burnout [10], the literature emphasizes that teachers need to critically analyze the outcomes of using AI for lesson planning and refine them [11]. Possible incorrect information and concerns about students' critical thinking were also the challenges most often reported by physics instructors in the study from Wattanakasiwich *et al.* [12]. Şimşek investigated teachers' and prospective teachers' prompts in 5E lesson planning and found no clear correlation between the number of prompts and the quality of the lesson plan [13]. Acquah *et al.* recognized social influence as the most important predictor of prospective teachers' intention to use AI in lesson planning [14].

As suggested by the literature, technological knowledge is an ever-evolving aspect that requires continuous interaction with technology. The integration of technological knowledge in pedagogy is a crucial aspect in educational frameworks like technological pedagogical content knowledge (TPCK) [15], where the knowledge of technology is not only considered for its pedagogical applications but also for its potential in aiding teachers to prepare their classes and evaluate their students. Since the emergence of AI and its applications to science education, it is crucial to investigate the intersections between AI and educational practice, not only while teaching, but also in the preparation tasks. It becomes necessary to learn more about AI's potential to help physics teachers in their task development and lesson planning. To this end, we aim to identify which cognitive demands of lesson planning [16] are met by Austrian middle-school physics teachers when being supported by an AI tool. We conducted a qualitative study comparing three different types of participants: prospective (P), out-of-field (O), and experienced (E) physics teachers, which describes the different backgrounds of middle-school physics teachers in Austria. The cognitive aspects, as well as teachers' reflections on their interaction with AI, are important to understand, given the intersection between technology and pedagogy. Therefore, our research questions are as follows:

RQ1: Which cognitive demands related to the lesson planning process are met by three groups of Austrian middle school physics teachers when using AI to aid in their lesson planning?

RQ2: How does the use of AI to aid in lesson planning compare between the prospective, out-of-field, and experienced physics teachers?

RQ3: How do the perceptions of using AI to aid in their lesson planning compare between the prospective, out-of-field, and experienced physics teachers?

This article is organized as follows: We first present the model of cognitive demands of lesson planning [16] as a theoretical framework in Sec. II. Section III presents the methodology, research design, and analytical approach. In Sec. IV, we present the results of the individual groups, as well as a comparison between the groups and teachers' perceptions of using AI, by classifying them into the cognitive demands proposed in the theoretical framework. In Sec. V, we will discuss the results. We conclude the article by summarizing the research and answering the research questions.

II. THEORETICAL FRAMEWORK

Lesson planning is a decisive activity for the teaching practice of physics educators. As part of the “pedagogical cycle of teaching,” planning is the link between reflecting on past and conducting future lessons [17]. Similar to Großmann and Krüger [18], we argue that high-quality lesson planning may positively influence the quality of physics teaching. This claim can be substantiated, as lesson planning often addresses domains for which positive effects on learning outcomes are evident, such as cognitive task analysis or considering students' prior abilities [19]. In line with this assumption, lesson planning is regarded as a crucial aspect for the professional development of physics teachers [17]. The process of lesson planning is closely connected to the acts of decision making and pedagogical reasoning. Sardo-Brown [20] defined lesson planning as the “instructional decisions made prior to the execution of plans during teaching” (p. 519). Those decisions are the result of educators' pedagogical reasoning, which is conducted each time anew for the specific content, group of pupils, and context [17]. Thereby, the rationales are based on teachers' individual professional knowledge, including, for example, knowledge about the physics content, curriculum, or assessment [17].

Since these decision-making processes are often complex and ill structured, König *et al.* developed a model called “Cognitive Demands of Lesson Planning” [16], which we use as a lens on lesson planning with AI in this study. We opted for this model in contrast to others (e.g., Refs. [21,22,18]), as (i) the demands are specifically focused on the lesson planning process instead of respective teacher dispositions or classroom performance; (ii) the model has a reliable foundation considering theoretical and empirical research on social systems, teacher expertise, as well as teacher competency; (iii) it allowed us to use well-operationalized demands for structuring the complex processes in lesson planning and limit the aspects to be considered in our research; and (iv) it is empirically tested through the analysis of lesson plans and provides respective

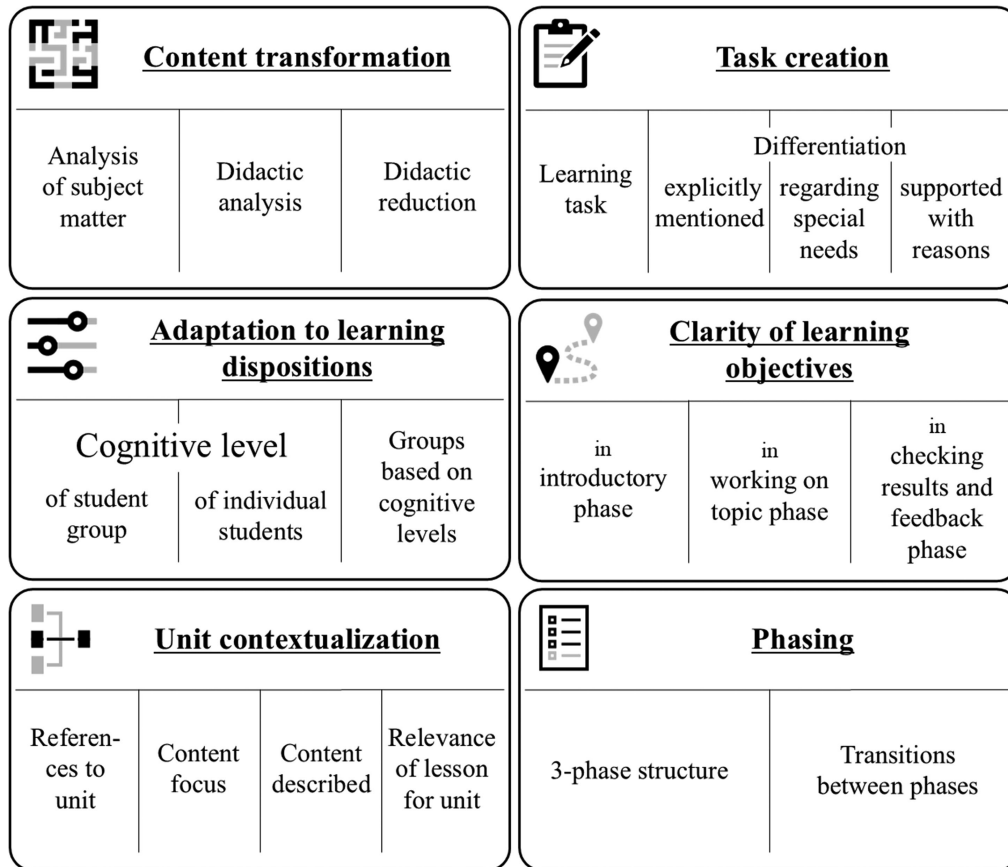


FIG. 1. “Cognitive Demands of Lesson Planning”: Each of the six demands consists of two to four subcategories, which are derived from the coding manual of König *et al.* [16].

coding manuals. We describe limitations of analyzing lesson planning via cognitive demands in Sec. V.

Cognitive demands of lesson planning conceptualizes classroom teaching as a social system within three key dimensions: content, social context, and time [16,23]. Lesson planning represents a process in which educators make decisions based on their professional knowledge and thereby orient themselves cognitively within this social system [16]. Furthermore, they shape upcoming lessons through their planning decisions and respective pedagogical reasoning. As the dimensions of content, social context, and time are quite abstract, König *et al.* [16] provided six concrete, albeit not completely clear-cut, cognitive demands for teachers in lesson planning (see Fig. 1). These can be placed within the three dimensions of social systems and will be presented in the light of physics teaching in the following paragraphs.

A. Content transformation

In the process of lesson planning, teachers transform content with the aim of providing it in forms that are accessible to their students. As posited by König *et al.* [16], whose explanations are based on the theoretical frameworks of Shulman [24] and Klafki [25], the cognitive

demand for content transformation encompasses the analysis and clarification of the lesson's subject matter. Based on this, teachers either explicitly or implicitly reason why it is transformed into a learning content for physics lessons. Furthermore, through didactic reduction of the learning content,¹ physics teachers decide on which specific key ideas students should learn and how they can be represented appropriately [16]. Hence, content transformation includes decisions on a specific learning path as well as the respective implementation through instructional materials, graphical representations, teaching methods, or social forms [27–29].

B. Task creation

With regard to Bromme [30] and Kang [31], König *et al.* [16] identify the creation and selection of learning tasks as a central aspect of the lesson planning process. Thereby, teachers develop or select tasks (i.e., for working sheets), and affect the extent to which learner’s prerequisites are considered. As these prerequisites vary regarding different

¹In contrast to König *et al.* [16], we have a broader understanding of this cognitive demand which connects to the concept of “educational reconstruction” [26]. However, due to reasons of consistency, we draw on the original term “didactic reduction” [16].

cognitive levels or special needs, teachers may use differentiation measures and support them by generic or subject-specific reasons in task creation. In addition, task creation encompasses implicit or explicit thoughts about links to preceding and subsequent activities within the lesson or unit [16]. In contrast to other cognitive demands, task creation demands physics teachers to think about their lesson at the most concrete level of the whole planning process [32].

C. Adaptation to student learning dispositions

According to König *et al.* [16], lesson planning may also encompass thoughts on how to acquire diagnostic information about students' cognitive learning dispositions, and how to take this information into account in the implementation of the lesson [16,32]. For example, physics teachers could design a short assessment tool that they use in their classroom or retrieve information from physics education research regarding students' prior knowledge. Meeting the demand of adapting to student learning dispositions in lesson planning may become apparent if discrepancies between students' conceptions and physical concepts are addressed in a way that the learning content lies within the students' "zone of proximal development" [17,33,34]. Therefore, the adaptation to students' learning dispositions can concern the whole group of learners (e.g., by following the most common prior conceptions) or individuals (e.g., by offering differentiated learning tasks).

D. Clarity of learning objectives

In the process of lesson planning, physics teachers must also meet the cognitive demand of setting appropriate learning objectives. Thereby, they gain clarity about the aims of each phase of the physics lesson, for example, concerning the learning content, methods, and activities [16]. Following König *et al.* [16] and Maulana *et al.* [35], teachers may engage with learning objectives at different levels, for example, ranging from an implicit consideration to the explicit communication to students.

E. Unit contextualization

According to König *et al.* [16], as well as Krepf and König [36], lesson planning comprises the cognitive demand of integrating a single lesson into the broader context of a whole unit. In turn, starting from the unit level, teachers may think about how such a unit can be appropriately divided into subthemes and single lessons [16]. Consequently, meeting the cognitive demand of unit contextualization becomes apparent if teachers explicitly clarify the lesson's references to the unit and describe its focus as well as the respective content. With regard to Arnold and Koch-Priewe [37], unit contextualization can be demonstrated, for instance, by elucidating the added value

and relevance of the individual lesson in relation to the overall objective of the unit [16].

F. Phasing

The final cognitive demand of lesson planning is the phasing of lessons into sequences with similar activities, methods, or social forms. König *et al.* [16] identify this as an "instrument for planning and managing lesson times" (p. 473), mostly divided into introductory, working, and feedback phases. As with the other cognitive demands, phasing assists physics teachers with orientation during lesson preparation. In addition, König *et al.* [16] emphasize, with regard to Muijs *et al.* [38], that teachers may also explicate links between lesson phases for minimizing interruptions on the students' learning path.

G. The cognitive demands for lesson planning combined with using AI

In light of our research topic, some of the six cognitive demands of lesson planning might be supplemented by the utilization of AI tools. Physics teachers may use these tools to retrieve knowledge on the subject matter or its transformation into suitable learning pathways, obtain inspirations for concrete tasks or representations, formulate learning objectives, and assess the alignment of different lesson phases. Additionally, teachers may employ these tools to sequence an entire unit within a single lesson or to structure specific lessons into several phases. Furthermore, prompts could also explicitly address pedagogical reasoning, thereby supporting decisions regarding the most appropriate teaching methodology. Therefore, it becomes necessary to investigate which learning demands can be supplemented with the use of AI.

For integrating AI in educational planning, the gap between the capabilities of AI systems and the skills of teachers in harnessing those capabilities should be considered. It should be noted that the capabilities of AI systems can also be questioned, as there are biases and limitations in the training models [39]. However, in this study, we focus on the bridge between the current AI capabilities and teachers' skills to use them. As noted by Welleck *et al.* [40], AI tools, when prompted in a naive or unsophisticated manner, tend to produce outputs that are highly probable but repetitive and not particularly useful for complex tasks like lesson planning (i.e., misconceptions are not taken into account, the output is a list of experiments). This is a core issue for physics teachers who may be new to AI technologies and unaware of the importance of crafting specific, nuanced prompts. Their lack of knowledge in this area may limit the effectiveness of AI-generated content, which can become an obstacle in the learning process of students. Furthermore, non-AI experts often struggle with prompt design due to their limited understanding of how large language models (LLMs) interpret and process input. Users tend to apply human-to-human

communication strategies when designing prompts for AI, which leads to suboptimal results [41,42]. Without systematic knowledge of prompting, teachers are unlikely to achieve the desired outputs, thereby reducing the effectiveness of AI tools in educational contexts. The *chain-of-thought prompting* (CoT) approach offers a potential solution [39,43]. This method enhances LLMs' ability to reason by breaking down complex tasks into smaller, intermediate steps. For education, this is particularly relevant as it provides a framework for structuring prompts in a way that mirrors logical reasoning processes.

III. METHODOLOGY

For this study, we applied a qualitative methodology to analyze the use of AI in lesson planning within three groups of physics teachers in Austria. The implementation involved using an AI tool as an aid to create a lesson plan. The participants provided the lesson plans, the AI chats, and a reflection on their experience. The lesson plans, the AI chats that emerged in the lesson planning process, and written reflections about that process were analyzed with qualitative content analysis [44].

A. Sample selection

In this study, we present three groups of Austrian middle school physics teachers examined in a comparative design: prospective, out-of-field, and experienced physics teachers. The three groups represent the diversity of middle school physics teachers in the Austrian educational context. Each group shares certain characteristics and the institutional context. Group P stands for prospective physics teachers ($N_P = 10$) studying at the University of Vienna. The participants in group P possess a physics background and are majoring in physics education. They usually aim to teach middle or high school level physics, but have no prior experience teaching. Group O stands for out-of-field physics teachers ($N_O = 13$) continuing their education at the University College of Teacher Education Burgenland. The participants in group O are currently teaching at the middle school level and have extensive teaching experience in other subjects but limited content knowledge in physics. Group E stands for experienced physics teachers ($N_E = 9$) participating in a professional development course offered by the University of Vienna to become mentors for prospective physics teachers. The participants in group E have over ten years of teaching experience at the middle school level and solid content knowledge. Informed consent was obtained from all participants.

B. Study design

The data were collected using the SoSci Survey [44] platform. The participants were provided with a link by the researchers, which directed them to an online survey containing all the requisite tasks. Additionally, the link

offered the possibility of uploading any documents created by the participants during the study. The researchers provided a concise introduction via Zoom but remained available in the online meeting to address any inquiries the participants might have had. The group of prospective and experienced physics teachers used a maximum of 1.5 h to complete the task, whereas the out-of-field physics teachers used a minimum of 2 h.

The objective for the participants was to develop a lesson plan on the topic of Newton's third law while utilizing an artificial intelligence tool of their choosing (e.g., ChatGPT and Google Gemini). The lesson plan should be designed for students in lower secondary school and should be of a duration of 50 min, in accordance with the standard lesson length in Austria. All participants were provided with a planning tool in tabular form. The initial column of the table outlined the requisite elements of the lesson plan, including the duration of each phase, the activities to be undertaken by both teachers and students during each phase, and the materials necessary for each phase. The number of phases was not specified by the researchers. In addition to describing the various phases of the lesson, the participants were also required to create worksheets or presentation slides if such materials were mentioned in the lesson plan. This step was not mandatory, but nearly all participants created some form of learning material. Once the lesson planning process was complete, the participants were required to upload all materials to the online platform, including the lesson plan and a copy of their chat with the AI.

The subsequent phase of the study required participants to compose a brief reflection, comprising 200–250 words, on their experiences utilizing AI for lesson planning. In accordance with the model proposed by König *et al.* [16], the participants were required to incorporate a minimum of three out of the six identified cognitive demands (see Sec. II) associated with lesson planning into their reflection. These cognitive demands were explicitly introduced before the reflection task. Even though this requirement may introduce a risk of bias, we decided to include it because otherwise, the reflections could have been too dispersed to be analyzed within this theoretical framework. We address this limitation in the discussion. In the final step, participants should indicate whether they utilized additional resources (e.g., physics textbooks, physics education literature, online resources, etc.) in their lesson planning process. The online survey concluded with participants indicating their consent for the utilization of all materials created by them for research purposes.

C. Data analysis

In accordance with the study design, a dataset of three documents was accessible for each participant: a tabular lesson plan, an AI-mediated chat, and a reflection. Different types of qualitative content analysis were applied to different types of data in three consecutive steps. First, the lesson plans

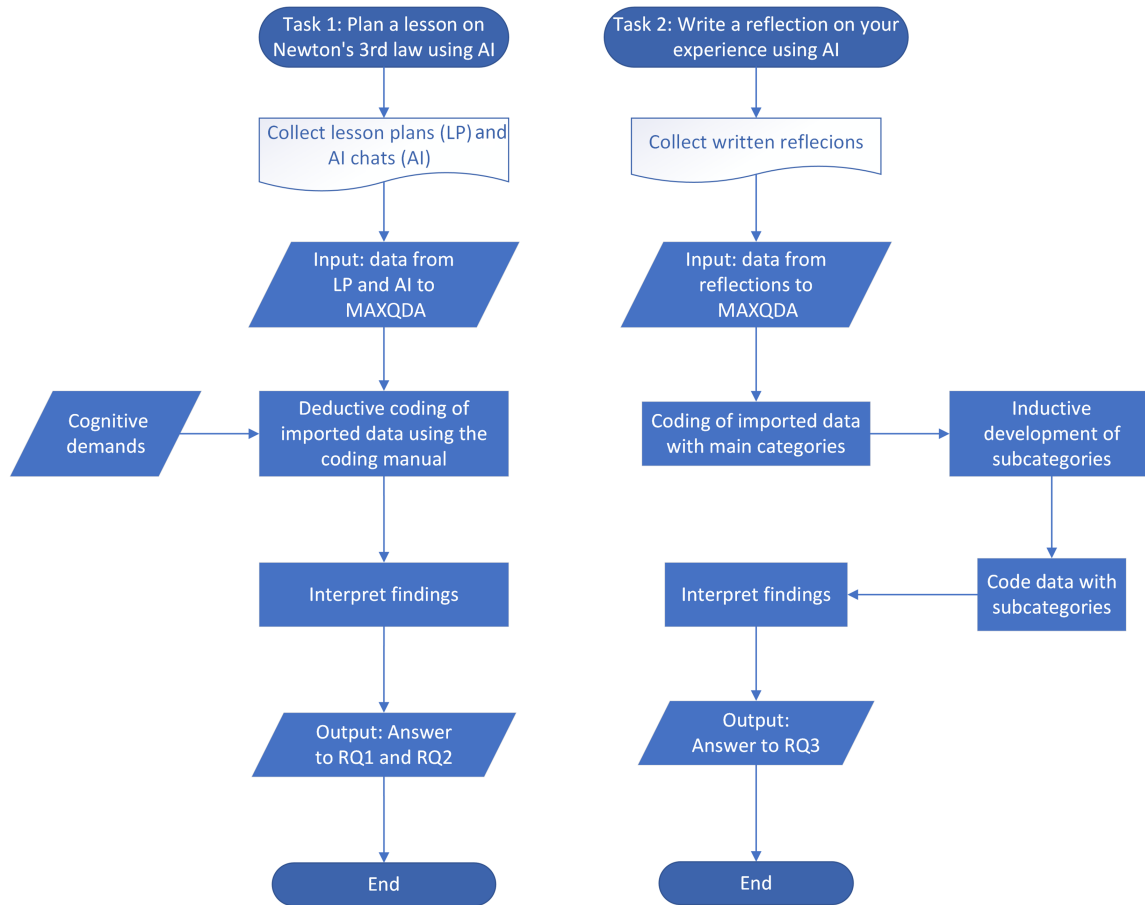


FIG. 2. Flowcharts that illustrate the data analysis process for the two different tasks: lesson plan creation and writing a reflection.

were analyzed through evaluative qualitative content analysis [44]. Second, using the same method, the AI chats were analyzed, separating between teachers' prompts and the answers given by the LLM. Third, the reflections were analyzed through structuring qualitative content analysis [44]. All the steps described were conducted using MAXQDA [45] software. Figure 2 provides an overview of the data collection and coding process for the two tasks.

1. Evaluative qualitative content analysis

This type of qualitative content analysis uses a deductive approach to assess and evaluate data. Rather than developing the categories through the analysis, coding is based on preexisting categories, which are derived from the literature. In our case, the cognitive demands for lesson planning by König *et al.* [16] served as categories for coding. In their study, König and colleagues present a comprehensive coding scheme for the analysis of cognitive demands in written lesson plans. The coding scheme comprises six categories for the respective cognitive demands, as well as three to four subcategories for each demand. In the original coding scheme, the subcodes for content transformation, task creation, adapting to learning dispositions, and clarity of learning objective were assigned

a value of 0 to 2 points, while the subcodes for unit contextualization and phrasing were given a value of either 0 or 1 point. However, we elected to refrain from assigning points for the various subcategories, instead focusing on the mere presence or absence of each subcategory within the dataset. The objective was not to evaluate the participants' lesson plans in terms of their relative merit, but rather to ascertain which of the identified cognitive demands were met by the teachers, with the assistance of AI. Prior to commencing the coding process, the coding scheme was translated into German, as the entirety of the data was in German, necessitating that the subsequent analysis be conducted in German as well.

Lesson plans were analyzed with evaluative qualitative content analysis to ascertain which of the cognitive demands proposed by the theoretical framework from König *et al.* were met by the teachers. Figure 3 presents an example of how a lesson plan can show which of the cognitive demands were met by the teacher, in this case, aided by an AI system.

AI chats were analyzed with evaluative qualitative content analysis as well to ascertain which of the cognitive demands proposed by the theoretical framework from König *et al.* were met by the teachers and the AI system

		Task creation: Learning task description	Content transformation: Analysis of subject matter		
	Clarity of objectives:	Time	Student activities	Teacher activities	Material
Phasing: 3-phase structure	Introductory phase	15 min	Listen attentively and take notes.	Explain the principle of interaction: Forces always act reciprocally between two bodies. If body A exerts a force on body B (actio), body B reacts with an equal but oppositely directed counterforce (reactio). Emphasize the importance of the opposing directions of the forces. Illustrate with diagrams or sketches on the board.	Blackboard and chalk M1 Concept+idea board picture
	Core working phase	15 min	Observe experiment and make notes of observations.	Show the two permanent magnets Explain the forces of attraction: When the magnets attract each other, they exert forces on each other. Emphasize the equivalence of the forces (actio and reactio).	Two permanent magnets M2Experimental setup
		15 min	Participate in the experiment Note down the observations	Prepare the skateboards and the rope Explain the experiment: The students pull on the skateboards, although only one is pulling on the rope. Emphasize that the movement occurs despite different pulling forces.	Skateboard, rope M3 Experiment set-up
	Feedback phase	5 min	Actively participate in the discussion Summarize the most important points	Summarize the learning objectives: Understand the principle of interaction Understand the importance of opposing forces Answering questions from students	

FIG. 3. Example of a lesson plan containing the codes of content transformation, task creation, clarity of learning objectives, and phasing by a prospective physics teacher (P59).

in the process of lesson planning. We aimed to differentiate between the teachers’ own ideas and the ideas given to them by AI. Consequently, the coding scheme was utilized twice: once for teachers’ prompts and once for the AI answers. To facilitate the differentiation between the codes, the coding scheme was replicated in MAXQDA. By comparing the codes given to the participants’ output and the codes

given to the AI’s output, it should be possible to draw conclusions about which cognitive demands were met by the AI on its own, which cognitive demands were only present when explicitly prompted and which cognitive demands were not taken into account by the AI (or the participants) at all. Furthermore, we could also ascertain the alignment of the AI responses with the actual instructional

To help you plan your lesson on Newton's 3rd axiom, I need the following information:

About your lesson:

Age of the students: The knowledge and interests of the students vary depending on their age.

Prior knowledge: What prior knowledge do your students already have about mechanics or forces?

Curriculum: What specific learning objectives are set out in the curriculum for this topic?

Time frame: How much time do you have available for the entire lesson?

Materials: What materials do you have available in the classroom (e.g., balls, springs, balances)?

About the 5E model:

Focus: Would you like to concentrate on certain aspects of Newton's 3rd axiom (e.g., everyday examples, mathematical formulation)?

Differentiation: Are there any pupils who need special support or should be particularly challenged?

With this information, I can help you with the following aspects of lesson planning:

Engage: How can you get students excited about the topic and pique their curiosity?

Explore: What experiments or activities are appropriate for students to discover Newton's 3rd axiom on their own?

Explain: How can you explain the axiom in a clear and understandable way without overloading it?

Elaborate (Elaborate): What tasks or projects can students complete to deepen their understanding?

Evaluate: How can you check whether the learning objectives have been achieved and the students' knowledge is secure?

Adaptation to learning dispositions

Task creation

FIG. 4. Example of an answer given by Google Gemini to the following prompt: “I would like to plan a lesson on Newton’s third law using the 5E model [6]. What information do you need for this?” (E79). The colored boxes indicate the respective cognitive demands, which were coded in relation to different parts of the answer.

The technical content is also not clearly explained, only after specific requests. Teaching and learning objectives are well formulated. The structure of the lesson is given in detail, but everything else is formulated very vaguely. It is assumed that the pupils have prior knowledge, but it is not specified what this should be. The worksheets were actually formulated quite well and it was also taken into account that there should be differentiated material.

AI does not provide sufficient content
learning objectives are clearly formulated
clear phasing
prior knowledge not taken into account
suggestions for differentiation

FIG. 5. Exemplary coding of the reflection from participant O26. Note that coloring refers to the different cognitive demands used in the coding system.

strategies and materials employed by the teachers in the lesson planning and classroom implementation phases. For instance, it is conceivable that the AI proposed a multitude of experiments for the lesson, yet the educators did not act upon this suggestion, preferring instead to prioritize content knowledge over the instruction of experimental techniques. An exemplary AI answer from one of the chats is shown in Fig. 4.

2. Structuring qualitative content analysis

Teachers' reflections were analyzed using structuring qualitative content analysis to ascertain what experiences the teachers had while using AI to help them plan a lesson. This approach was chosen, as it is more inductive in nature and aims to develop categories for coding from within the dataset. In the first step, the reflections were coded using the six cognitive demands presented in the theoretical framework. It is important to note that the coding scheme for the lesson plans and AI chats used several subcodes for every cognitive demand, which were not applied here. Instead, after highlighting where the cognitive demands applied, all segments related to the same cognitive demand were pooled together. Reading through these pooled segments, the researchers then extracted relevant information. For example, if several teachers complained about the AI not considering students' conceptions, a new inductive code "AI does not take students' conceptions into account" was created. Doing this with each of the six cognitive demands delivered a new set of codes, which was based on teachers' perceptions of AI. This new set of codes was then used to code all the reflections again. Figure 5 shows an example of how one of the reflections was coded.

The whole dataset was coded as described above by the first author. A smaller set of 4 out of 32 participants was also coded by the second author to investigate interrater reliability. Cohen's kappa [45] was then calculated to be $\kappa = 0.64$, representing a good interrater agreement [46].

IV. RESULTS

We present the results in five subsections. Sections IV A–IV C correspond to an analysis of each group: A for

prospective physics teachers, B for out-of-field physics teachers and C for experienced physics teachers. In these subsections, we analyzed each group separately by presenting how they approached the lesson planning and the cognitive demands that we found in their lesson plans and chats with AI (separated by prompts and AI answers). Section IV D presents a comparison of the three groups. Section IV E focuses on the reflections of the three types of participants with a comparative structure. Throughout the results section, we include examples from the different types of participants and documents. We identified these examples by a five digit code: (i) the group (P = prospective physics teachers, O = out-of-field physics teachers, E = experienced physics teachers), (ii) the two digit number that was given to each participant by SoSci Survey, (iii) the type of document (R = reflection, L = lesson plan, P = prompts, and AI = AI answers) and, (iv) the position (paragraph) in the document. The results represent the views of the three types of teachers within our sample, so they are not generalizable. Instead, they offer a qualitative comparison of the perception regarding the use of AI for lesson planning by three types of physics teachers in Austria, limited to the sample, as further explained in Sec. V C.

As we did not give any specific instructions to the participants on what AI assistant to use, we briefly report on their respective choices. Most of the teachers ($N = 23$) used *ChatGPT*, while four teachers used *Microsoft Copilot*, and three used *Google Gemini*. Two teachers also utilized AI software trained for specific tasks: *Teachino*, which is aimed at helping teachers with preparing their lessons, and *Slidesgo*, an AI for creating presentation slides. Based on the available dataset from each participant, we could not find any relation between which AI assistant was used and the resulting lesson plan. Therefore, going further, the type of AI assistant does not represent a unit of analysis.

A. Prospective physics teachers

Prospective physics teachers (PPTs) had a direct and practical approach to lesson planning using AI. Eight out of 10 PPTs started the conversation asking the AI tool to create a plan on the topic of Newton's third law and, in

some cases, provided more requirements like the expected length of the lesson and activities for the students and teachers. For example, “Plan a lesson on the interaction principle (Newton’s 3rd law) for secondary level I. Use the planning matrix, required time, student activity, teacher activity, and materials” (P59-P:1). The subsequent prompts were used to include activities, experiments, or more detailed explanations. For example, “Can you please tell me the activities of the pupils and the activities of the teacher during the individual phases of this planning?” (P61-P:49).

The other two PPTs had conversational approaches with the AI tool. These two participants did not prompt directly for a lesson plan and probably decided to do the lesson plan on their own, requiring only ideas from ChatGPT. One of them started by asking ChatGPT for an explanation: “How would you explain Newton’s 3rd law?” (P60-P:1). Then, the participant asked for an introduction, examples, experiments, and finally for a worksheet. The other participant asked directly for an experiment, “Which experiment can be used to teach the interaction principle to children?” (P67-P:1), and a quiz.

To answer the research question RQ1 for group P, we analyzed prospective physics teachers’ lesson plans, prompts, and the answers provided by the AI according to the cognitive demands present in the lesson planning process. We present the results in Table I. As the AI answers can be quite long and repetitive, we only include examples that are short, concise, and relevant to understanding the impact of the AI answers on the cognitive demand.

Looking at the results in Table I, it is apparent that some aspects of lesson planning in the framework of the cognitive demands were never explicitly mentioned by the PPT. For example, they did not prompt the AI to perform a didactic analysis of the content, nor did they write something about grouping students based on their different cognitive levels. Furthermore, the PPTs did not consider differentiation in their prompts and lesson plans, while the AI did consider explicit differentiation. The same can be observed regarding the cognitive demand of unit contextualization. While the AI described the content of the lesson, the PPT did not explicitly adapt that information.

B. Out-of-field physics teachers

Out-of-field physics teachers (OPTs) started their lesson planning process in a variety of ways. Six OPTs started with a short and concise prompt, like “Create a lesson plan for secondary level 1 on the topic of Newton’s 3rd axiom” (O40-P:1). Four OPTs created full prompts, for example, “Please create a 50-minute lesson on the principle of interaction (Newton’s 3rd law) for secondary level I. Divide the lesson into 3 phases: A start phase in which interest is aroused and attention is created. A main phase in which an experiment on the topic is planned, carried out,

documented, and evaluated. And a final phase in which the results of the experiment are summarized and saved together. For each phase, please explain which activity the students will carry out, which activity the teacher will carry out, which materials are required, and how long each phase will take” (O31-P:1). The subsequent prompts of these OPTs included the creation of worksheets (sometimes with their solutions), diagrams, and experiments. Out of these ten OPTs who started with short or full prompts, four asked in the subsequent prompts for explanations of the third law (e.g., “Explain Newton’s 3rd axiom to me,” O40-P:1).

The other three OPTs had an indirect approach, as they required information instead of the full lesson plan. For example, one of them started by prompting a picture of Newton’s laws, followed by an explanation (“Find a picture of Newton’s 3 axioms”/“explain the 3 law in more detail and more simply” O34-P:1) and worksheets. Another teacher started by asking learning objectives (“Hello, I need a formulation of objectives for my physics lesson on the topic of the principle of interaction/3rd Newton’s axiom”), followed by how to introduce the topic, a worksheet, and its solution. The other OPTs started by asking for a “simple explanation of Newton’s third law” (O52-P:1), followed by a worksheet and an experiment.

To answer research question RQ1 for group O, we analyzed out-of-field physics teachers’ lesson plans, prompts, and the answers provided by the AI according to the cognitive demands present in the lesson planning process. An exemplary lesson plan including the applied coding can be found in the Appendix (Figs. 6 and 7). We present the results of the analysis in Table II.

The various aspects of content transformation were more often prompted and part of the AI answer than they are explicitly mentioned in the lesson plan. However, it is possible that some of the information is still represented implicitly in the lesson plan, as the teachers used the analysis of the subject matter and the didactic reduction as a foundation for their lesson planning. Regarding task creation, 12 out of 13 OPTs included a learning task in their lesson plan, while only 2 considered explicit differentiation. It is noticeable that the AI proposed using explicit differentiation more often than the teachers prompted it. Regarding the cognitive demand of clarity of learning objectives, it can be observed that although the teachers rarely prompted that, it was included in the AI’s response and therefore also appeared in the lesson plan. The same is apparent regarding the three-phase structure. Some of the aspects related to the cognitive demands are not present in the data obtained from the OPT at all, e.g., describing the relevance of the lesson for the unit or giving a description of the transition between the different phases of a lesson.

C. Experienced physics teachers

Experienced physics teachers (EPTs) approached the lesson planning with AI in diverse ways. Five EPTs created

TABLE I. Number of codes related to the six cognitive demands from the prospective teachers ($N_p = 10$). Note that every code is only applied once per document. The column “LP” describes how often this code was applied to the lesson plan, “P” relates to the teachers’ prompts, and “AI” refers to the answer given by the AI. Comparing how often a certain code was applied to the different document types allows us to draw conclusions about how the teachers used AI for lesson planning. For example, a description of the objectives for every phase of the lesson is present in nearly all of the lesson plans and was also suggested by the AI. This leads to the conclusion that most of the teachers used AI to formulate clear learning objectives.

Cognitive demand	Coding structure (Example)	Document type		
		LP	Chat	
			P	AI
Content transformation	Analysis of subject matter (Teacher example: “How would you explain Newton’s 3rd law?” P60-P:1)	1	1	1
	Didactic analysis (AI example: “ <i>Ensuring learning success: The combination of practical experiments and theoretical development ensures that different types of learners are addressed.</i> ” P62-AI:45)	0	0	1
	Didactic reduction (Teacher example: “A few ideas and thoughts on how to explain and implement this principle in a nice and interesting way for pupils in lower secondary” P58:P1)	2	8	7
Task creation	Learning task (Teacher example: “Also create some worksheets and mark their use in the planning matrix” P56-P:4)	9	5	4
	Explicit differentiation (AI example: “Differentiated tasks: Additional tasks with different levels of difficulty to support students at different learning levels” P57-AI:153)	0	0	4
	Explicit differentiation (special needs)	0	0	0
	Differentiation supported with reasons	0	0	0
Adaptation to learning dispositions	Cognitive level of student group (Teacher example: “Make it more student-friendly” P67-P:1)	0	1	2
	Cognitive level of individual students	0	0	0
	Groups based on cognitive levels	0	0	0
Clarity of learning objectives	Introductory phase (Teacher example: “Write down what exactly the teacher should explain to the students in phase 1” P59-P:113)	8	2	7
	Working on topic phase (Teacher example: “Can you organize phase two so that the students come up with their own hypotheses and the teacher is more in the background?” P61-P:121)	10	1	7
	Checking results and feedback phase (Teacher example: “Create a handout on the topic as a learning basis for a performance assessment” P72-P:234)	10	1	7
Unit contextualization	References to unit (Teacher example: “A crossword reel, which is done at the end and has as a result the law of interaction, which is then a transition to the next unit” P60-L:1)	1	0	0
	Content focus (Teacher example: “Objective: Understand the principle of interaction qualitatively and apply it to various situations.” P57-L:1)	5	8	7
	Content described	0	0	6
	Relevance of lesson for unit	0	0	0
Phasing	Three-phase structure (Teacher example: “Can you please break down the materials by phase?” P61-P:121)	10	3	7
	Transition between phases	0	0	0

straightforward prompts with clear indications for the GPT. Three of them were full prompts, for example, “Plan a physics lesson (50 min) on the principle of interaction (i.e., Newton’s 3rd law) for lower secondary school. The pupils already know the concept of force and Newton’s first two

laws. When planning the lesson, consider students’ existing preconceptions and make references to their everyday lives” (E83-P:1). Two of them were simpler prompts, like “Create a teaching unit on the subject of the law of interaction” (E91:P1). The subsequent prompts were used

TABLE II. Number of codes related to the six cognitive demands from the out-of-field teachers ($N_O = 13$). Note that every code is only applied once per document. The column “LP” describes how often this code was applied to the lesson plan, “P” relates to the teachers’ prompts, and “AI” refers to the answer given by the AI. Comparing how often a certain code was applied to the different document types allows us to draw conclusions about how the teachers used AI for lesson planning. For example, a description of the objectives for every phase of the lesson is present in nearly all of the lesson plans and was also suggested by the AI. This leads to the conclusion that most of the teachers used AI to formulate clear learning objectives.

Cognitive demand	Coding structure (example)	Document		
		LP	Chat	
			P	AI
Content transformation	Analysis of subject matter (Teacher example: “Simple explanation of Newton’s third law” O52-P:1)	2	6	5
	Didactic analysis	0	0	1
	Didactic reduction (Teacher example: “What interesting teaching introduction could be used in physics lessons on the topic of the interaction principle/3rd law?” O30-P:1)	3	9	8
Task creation	Learning task (Teacher example: “Create a worksheet on the topic of the 3rd Newtonian axiom for secondary level 1 with pictures” O40-P:1)	12	9	9
	Explicit differentiation (Teacher example: “Create differentiated worksheets,” O26-P:4)	2	2	7
	Explicit differentiation (special needs)	0	0	0
	Differentiation supported with reasons	0	0	0
Adaptation to learning dispositions	Cognitive level of student group (Teacher example: “The pupils’ prior knowledge of the interaction principle (Newton’s 3rd law) is limited” O45-P:1)	0	1	1
	Cognitive level of individual students	0	0	0
	Groups based on cognitive levels	0	0	0
Clarity of learning objectives	Introductory phase (Teacher example: “How could you make the introduction with the repetition more creative so that the students participate more actively?” O35-P:1)	12	1	10
	Working on topic phase	12	0	10
	Checking results and feedback phase (Teacher example: “Write me a suitable repetition of Newton’s two laws, which have already been dealt with in class” O35-P:1)	12	1	10
Unit contextualization	References to unit (Teacher example: “Write me a suitable repetition of Newton’s two laws, which have already been dealt with in class” O35-P:1)	3	1	3
	Content focus (Teacher example: “Topic Principle of interaction—Newton’s 3rd law” O36-P:1)	8	9	9
	Content described	1	0	8
	Relevance of the lesson for the unit	0	0	0
Phasing	Three-phase structure (Teacher example: “Divide the lesson into three phases” O31-P:1)	13	1	8
	Transition between phases	0	0	0

to include additional elements like activities or experiments for the students, activities for the teachers, or more detailed explanations. For example, “After the welcome and motivation, plan an experiment for the students!” (E93-P:1).

Two of the EPTs started by prompting a lesson and asking the GPT for the information it needed to conduct the task: “Plan a lesson for me, what information do you need?” (E80-P:1). This created two different results: one of them provided all the information that the GPT needed

in only one prompt, resulting only in 4 subsequent prompts, whereas the other teacher provided an answer per prompt and required several modifications, resulting in 40 prompts. One of the teachers who had started with a simple prompt, moved to another platform, and followed the idea of asking the GPT what information it needed. So, even if it was not the first prompt, it was something that a third teacher decided to do after failed attempts with a simple prompt.

TABLE III. Number of codes related to the six cognitive demands from the experienced teachers ($N_E = 9$). Note that every code is only applied once per document. The column “LP” describes how often this code was applied to the lesson plan, “P” relates to the teachers’ prompts, and “AI” refers to the answer given by the AI. Comparing how often a certain code was applied to the different document types allows us to draw conclusions about how the teachers used AI for lesson planning. For example, a description of the objectives for every phase of the lesson is present in all of the lesson plans and was also suggested by the AI. This leads to the conclusion that most of the teachers used AI to formulate clear learning objectives.

Cognitive demand	Coding structure	Document type		
		LP	Chat	
			P	AI
Content transformation	Analysis of subject matter (Teacher example: “Explain Newton’s 3rd axiom: “If a body A exerts a force on a body B, an equal but opposite force of body B acts on body A.” E92-L:3)	3	0	1
	Didactic analysis	0	0	0
	Didactic reduction (Teacher example: “Are there any other demonstration experiments?” E91-P:37)	1	7	8
Task creation	Learning task (Teacher example: “Create a worksheet for the experiment with specific observation tasks” E80-P:1)	8	3	4
	Explicit differentiation (Teacher example: “Differentiate from the easy version” E79-P:1076)	0	1	2
	Explicit differentiation (special needs) (AI example: “Are there students with special needs that need to be considered?” E79-AI:120)	0	0	1
	Differentiation supported with reasons	0	0	0
Adaptation to learning dispositions	Cognitive level of student group (Teacher example: “When planning, take into account pupils’ existing preconceptions and make references to their everyday lives” E83-P:1)	0	6	5
	Cognitive level of individual students	0	0	0
	Groups based on cognitive levels	0	0	0
Clarity of learning objectives	Introductory phase (Teacher example: “Can you base the lesson on an exciting introductory example? For example, a superhero jumping off something” E83-P:2)	9	4	9
	Working on topic phase (Teacher example: “Plan an Explore phase. This should take about 20 min and include an experiment” E79-P:458)	9	1	9
	Checking results and feedback phase (Teacher example: “Create a reflection sheet for the final phase of this teaching unit” E96-P:1)	9	2	9
Unit contextualization	References to unit (Teacher example: “How do you repeat Newton’s first two axioms for a secondary level 1 class?” E91-P:171)	1	1	2
	Content focus (Teacher example: “ <i>Principle of interaction</i> ” E80-P:1)	2	8	6
	Content described	0	0	0
	Relevance of lesson for unit (Teacher example: “How would you plan the transition to the 3rd axiom after the repetition?” E91-P:194)	0	1	2
Phasing	Three-phase structure (Teacher example: “ <i>Use the 5E model</i> ” E94-P:1)	9	3	9
	Transition between phases	0	0	0

One of the EPTs used an AI (*Teachino*) that asked the teacher about their subject, to which the teacher responded: “Subject: Physics/Age of pupils: 13/Language of instruction: German/Curriculum available” (E85:P1). The AI subsequently asked for more details. After four rounds of questions by the AI and answers by the teacher, the AI gave

the first attempt at lesson planning. After that, the teacher prompted for modifications like, “Can you plan the lesson according to the Conceptual Change teaching model?” (E85:P2).

To answer the research question RQ1 for group E, we analyzed experienced physics teachers’ lesson plans,

TABLE IV. Comparison between the coding of PPTs ($N_P = 10$), OPTs ($N_O = 13$), and EPTs ($N_E = 9$). The percentages express how many of the participants in each group met the respective subcategory of the cognitive demand. This allows us to analyze differences and similarities between the groups. For example, nearly half of the OPTs used AI for the analysis of subject matter, while only 10% of the PPTs and none of the EPTs did.

Cognitive demand	Coding structure	Document type					
		Lesson plan			Chat		
		PPT (%)	OPT (%)	EPT (%)	PPT (%)	OPT (%)	EPT (%)
Content transformation	Analysis of subject matter	10	15	33	10	46	0
	Didactic analysis	0	0	0	0	0	0
	Didactic reduction	20	23	11	80	69	78
Task creation	Learning task	90	92	89	50	69	33
	Explicit differentiation	0	15	0	0	15	11
	Explicit differentiation (special needs)	0	0	0	0	0	0
	Differentiation supported with reasons	0	0	0	0	0	0
Adaptation to learning dispositions	Cognitive level of student group	0	0	0	10	8	67
	Cognitive level of individual students	0	0	0	0	0	0
	Groups based on cognitive levels	0	0	0	0	0	0
Clarity of learning objectives	Introductory phase	80	92	100	20	8	44
	Working on the topic phase	100	92	100	10	0	11
	Checking results and the feedback phase	100	92	100	10	8	22
Unit contextualization	References to the unit	10	23	11	0	8	11
	Content focus	50	62	22	80	69	89
	Content described	0	8	0	0	0	0
	Relevance of the lesson for the unit	0	0	0	0	0	11
Phasing	Three-phase structure	100	100	100	30	8	33
	Transition between phases	0	0	0	0	0	0

prompts, and the answers provided by the AI according to the cognitive demands present in the lesson planning process. We present the results of the analysis in Table III.

Regarding the cognitive demand of adaptations to learning dispositions, two-thirds of the EPTs consider the cognitive level of the student group when interacting with the AI. For example, they prompt the AI to take students' existing preconceptions into account. However, this aspect is not explicitly mentioned in their lesson plans. Furthermore, in relation to the clarity of learning objectives, several EPTs prompt the AI with specific ideas on how the introduction of the lesson should be designed, e.g., by using an exciting introductory example.

Other aspects of the cognitive demands were not explicitly prompted by the EPT; for example, they did not inquire about the subject matter. However, one-third of the lesson plans feature an analysis of the subject matter; therefore, it can be assumed that the EPT already possessed the relevant knowledge about Newton's third law. It is also striking that some of the aspects related to the cognitive demands are not present in the data at all, e.g., supporting differentiation with reasons or describing the content of the lesson.

D. Comparison between the groups

To answer RQ2, we examined differences in how often a specific code is present within each of the groups. Looking at these comparisons, Table IV shows that there are only a few relevant differences between the groups. For most of the subcodes the numbers align, if one considers that the three groups were not of the same size ($N_P = 10$, $N_O = 13$, and $N_E = 9$). It is also important to note that the agreement is not only visible if teachers met the specific cognitive demand but even more when they did not meet it. For example, although learning tasks were a part of nearly every lesson plan, none of the teachers gave reasons for how differentiation between different students' needs should be implemented. This aligns with the fact that none of the teachers differentiated between the different cognitive levels of individual students. These findings suggest that differentiation was not a relevant part of lesson planning for the teachers in our sample. However, this cannot be generalized as the teachers planned a fictional lesson, with no specific group of students in mind. It might be that in a more realistic setting, these teachers would have thought more carefully about their students' needs.

Despite the general agreement of the teachers in the three groups, some differences can also be observed. As one might expect, the OPT used AI to analyze the content of the lesson a lot more than the other two groups. An explanation for this could be that, as these teachers did not study physics, they feel the need to familiarize themselves with the content of the lesson before they can start planning. Therefore, they prompted the AI to give them a “simple explanation of Newton’s third law” (O52-P:1). Conversely, only one of the PPT and none of the EPTs asked the AI to explain Newton’s third law to them, as they studied physics and already knew enough about the topic. However, the lesson plans created by the EPT featured more explicit analysis of content than the other two groups. This strengthens the argument that these teachers had enough experience with the physics content and therefore did not need to rely on AI.

The OPT also had the highest number of codes regarding the creation of learning tasks. Ninety-six percent of them prompted the AI to help them with that cognitive demand, in most cases to “create a worksheet on the topic ‘law of interaction’” (O45-P:1). Again, the EPT used the AI least (33%), while still 89% of the lesson plans created by the EPT featured a learning task. Same as with the analysis of content, the EPTs relied more on their own experience as physics teachers than on the AI. The PPT did not use the AI as frequently as the OPT (50%), but still more often than the EPT, as they had physics knowledge but still lacked experience in teaching physics.

Regarding the cognitive demand adaptation to learning dispositions, the situation is exactly the opposite of that

with the two previous examples. Here, the EPT used the AI by far more frequently to find out more about “which ideas can students have about the law of interaction?” (E91-P:94). While 67% of the EPT inquired about students’ ideas, only one each of the PPT and OPT asked the AI a similar question. The presence of codes in our data suggests that only the EPT recognized the need to think about students’ ideas before planning the lesson and therefore used the AI accordingly. This association between the degree of experience in teaching physics and the usage of AI for dealing with students’ learning dispositions should be further examined with bigger and more representative samples.

The EPT also used the AI more often than the other two groups combined to get ideas on how the introductory phase of the lesson could be planned. One teacher, for example, wanted to have an “exciting introductory example” like “a superhero jumping off something” to “get all the pupils involved” (E83-P:2). Another teacher wanted “to start with a picture for activation” (E79-P:226). This suggests that the EPT had a greater awareness of the importance of getting students cognitively activated at the beginning of the lesson to motivate the topic. Again, having more experience in physics teaching might have been the reason for these differences.

E. Reflections

To answer RQ3, we also investigated teachers’ reflections about their experiences with using AI in lesson planning. Table V summarizes the results. The coding

TABLE V. Results from teachers’ reflections on their experiences with using AI for lesson planning. The coding structure was obtained inductively from the data, using structured qualitative content analysis. The percentages express how many of the participants in each group met the respective code. This allows to compare teachers’ perceptions of using AI for lesson planning between the PPT ($N_P = 10$), OPT ($N_O = 13$), and EPT ($N_E = 9$). For example, while half of the PPTs thought that AI created useful tasks, none of the EPTs thought that way.

Cognitive demand	Coding structure	Group		
		PPT (%)	OPT (%)	EPT (%)
Content transformation	AI delivers useful content transformation	40	31	0
	AI does not deliver useful content transformation	20	8	67
	AI helps teachers to repeat content	0	23	0
Task creation	AI creates useful tasks	50	31	0
	AI does not create useful tasks	0	8	78
	AI gives suggestions for differentiation	0	23	0
Adaptation to learning dispositions	AI adapts the lesson plan to learning dispositions	20	38	11
	AI does not adapt the lesson plan to learning dispositions	50	31	44
Clarity of learning objectives	AI gives clear learning objectives	40	62	44
	AI does not give clear learning objectives	10	0	44
Unit contextualization	AI does not offer unit contextualization	0	0	33
Phasing	AI gives a clear structure	50	69	67
	The lesson is structured by the 5E model	10	0	11

structure was created inductively according to the method of structuring qualitative content analysis [47]. It is apparent that teachers had different viewpoints on the AI abilities for lesson planning. Therefore, the coding structure for most of the cognitive demands features opposing pairs of codes. Furthermore, comparing groups reveals different perceptions of AI within distinct groups of teachers.

Regarding content transformation, the PPT and OPT had the tendency to perceive AI as useful to give them relevant explanations and examples for Newton's third law. About a third of the participants from these two groups shared that viewpoint, with only very few participants stating the opposite. Additionally, 23% of the OPT mentioned that the AI delivered them with a helpful "repetition of the physics background" (O35-R:2). Conversely, 67% of the teachers from group E argued that the AI tool does not deliver appropriate content transformation. Teachers mentioned, for example, that "the physics content in particular must be viewed very critically" (E79-R:2) and that the AI features "unclear and confusing tasks" (E80-R:2).

The same differences between the three groups, which were just described for content transformation, also apply to the cognitive demand of task creation. While most of the PPT and OPT were satisfied with the learning tasks created by the AI, none of the EPT found these tasks to be useful. Instead, they described them as "banal and not expedient" (E85-R:2) or "very shallow" (E87-R:2). Furthermore, the aspect of differentiation was only mentioned by OPT. They highlighted that the AI did come up with ideas for student differentiation by itself, without being specially prompted to do so.

In contrast to the last two cognitive demands, the differences between the three groups were not as clear in the context of adaptation to learning dispositions. Here, the whole group of participants is split between two opposing opinions. On the one hand, 38% of the OPTs, 20% of the PPTs and 11% of the EPTs shared the view that the "consideration of the students' prior knowledge worked well" (O45-R:2). On the other hand, more teachers argued that there was "no consideration of prior knowledge and students' ideas" (P65-R:2) (50% of OPT, 31% of PPT, 44% of EPT). This suggests that while both views were present in all three groups, the critical viewpoint was more prevalent in groups P and E, while the teachers in group O had a more positive perception of AI's abilities to take students' ideas into account.

Concerning the clarity of learning objectives, the teachers showed more consensus compared to the other cognitive demands described above. Between 40% and 62% of the participants in the three groups argued that AI can give clear learning objectives. One teacher specifically liked that "the learning objectives are clearly formulated, realistic and harmonize with the lesson planning" (O43-R:2). However, the EPT again were more critical, as another 44% of them mentioned that AI does not adapt the lesson to learning

dispositions. One teacher, for example, criticized that "the objectives are not based on the curriculum" (E80-R:2).

The participants showed the highest consensus regarding the cognitive demand phasing. Here, 50% of the PPTs, 69% of OPTs, and 67% of the EPTs agreed that AI gives a clear structure in the lesson plan. One teacher specifically noted that "the basic structure of the lesson is definitely one that can also be used in practice" (P59-R:2). Furthermore, two teachers argued that the lesson plan was structured by the 5E model.

Contrasting the three groups shows a general tendency for EPT to be the most critical users of AI, while OPTs and PPTs seem to perceive it as more helpful and positive. One argument from the EPT was that for them "the workload is longer with AI than without it" (E87-R:2). These results suggest that experience in teaching the subject is a key factor in teachers' perception of the usefulness of AI for lesson planning. Although the OPT and the EPT both had experience in teaching in general, the experience in teaching physics made the EPT much more critical when using AI to plan a physics lesson, while the OPT saw the AI as a welcome opportunity to help them when having to teach unfamiliar content.

V. DISCUSSION

The aim of this study was to investigate how three groups of Austrian middle-school physics teachers use AI in their lesson planning. As suggested by previous research, AI proved efficient in producing a lesson plan [7]. Similar to Powell and Courchesne [7], at least one of the participants found that ChatGPT knows about the 5E model of instruction. Even without specifically prompting, the lesson plans were, to some extent, aligned with that model. Although improvements can be made, all the lesson plans might be used as a template for teaching Newton's third law.

A. Summary of key findings

To analyze the lesson plans created by the teachers with the help of AI, we investigated the extent to which the cognitive demands for lesson planning proposed by König *et al.* [16] have been met. By applying that method, we were able to describe a profile of the teachers' use of AI in lesson planning for three different groups of participants in this study.

The prospective physics teachers chose a very direct and practical approach to using AI in lesson planning. Most of the introductory prompts just asked for a lesson on Newton's third law, giving few additional requirements. Most of the following prompts were focused on the didactic reduction, asking how to explain the content for students. The out-of-field physics teachers applied different approaches in the use of AI. While some of the teachers applied short and concise prompts, similar to the PPTs, others made use of more

detailed prompting strategies, giving the AI a lot more information as a basis for lesson planning. Furthermore, three of the teachers in this group did not start their interaction with AI by prompting it to plan a lesson, but by asking for information they would use to plan the lesson themselves. The experienced physics teachers again applied different prompting strategies. While some of the teachers used direct prompts and provided enough information to AI, two of them used a reverse approach. Instead of prompting the AI on what to do right at the beginning, these teachers asked the AI what information it would need to plan a lesson on Newton's third law. Then the AI gave a detailed response, listing all the necessary information, which was subsequently provided by the teachers.

When analyzing the results in Tables I–III, we also found interesting patterns for all groups regarding prompting. In some codes, there is a mismatch between their frequency in the prompts and how they show in the lesson plan. For example, regarding the cognitive demands of content transformation, the subcode didactic reduction was prompted by 24 participants but only appeared in 5 lesson plans. On the other end of this mismatch, we found that the participants did not heavily prompt the clarity of learning objectives and phasing, but both aspects are present in the lesson plans.

There were only some differences between the lesson planning of the three distinct groups of participants. The number of codes in Table V is similar for most of the categories, given that there were different numbers of participants in each group. One visible exception is task creation. The out-of-field physics teachers met this cognitive demand more often than the prospective and experienced physics teachers. On the contrary, regarding content analysis, we observed greater experience of the prospective and experienced physics teachers. A plausible reason for this is that out-of-field teachers needed more information on the content because physics is not their area of expertise. As mentioned before, teachers did not sufficiently meet the cognitive demand of adaptation to learning dispositions, with the exception of the experienced physics teachers. These teachers might be more aware of the importance of taking learning dispositions into account due to their experience in teaching physics.

B. Implications regarding the use of AI in lesson planning

There were also some questionable components of the AI's suggestions for the lesson plans. For example, the AI once suggested the usage of a firing gun to demonstrate Newton's third law. However, firing a gun or cannon was also one of the most common examples to explain Newton's third law mentioned by the AI (only swimming appeared more often). The fact that it was also suggested as an experiment for the physics classroom shows that the AI cannot really "understand" the difference between a

theoretical example and a practical experiment the way that humans do. Although this is highly unlikely, there might be grave consequences if a teacher were to act on the AI's suggestions.

Furthermore, the AI tools tended to create stand-alone lessons, with only very few connections to other areas of mechanics. Similar results were reported by Davis and Lee [9]. The AI also did not take misconceptions into account and, in line with the results by Gonçalves Costa *et al.* [8], did not give reasons for how it designed the lessons. However, as the teachers did not explicitly ask for this, it remains unclear to what extent reasoning by the AI would be possible if prompted. Therefore, we agree with van den Berg and Du Plessis [11] that teachers need to critically analyze the lesson plans produced by AI. They need to be able to critically scrutinize and enhance these lesson plans, especially regarding the selection of differentiative measures, the adaptation to learning dispositions, the justification of the lesson's relevance for the unit, and the development of the transition between the phases, as these aspects occur least frequently in the lesson plans examined.

The observed patterns described in the previous subsection also reflect fundamental characteristics of large language models (LLMs). As Welleck *et al.* [40] noted, LLMs tend to produce highly probable responses based on training data, which can result in generic and sometimes superficial content when prompted in an unsophisticated manner. It is likely that the training data of the LLM contained the cognitive demands that later appeared in the AI chats with a high number of encodings, even if the teachers had not asked for them directly (e.g., clarity of the learning objectives). It also appears that the LLMs were trained to segment lessons into phases. This presents a challenge for nonexpert AI users, such as teachers, who may lack knowledge of how to craft precise, specific, and effective prompts. Without such skills, AI-generated outputs can lack depth and fail to address complex instructional needs, such as addressing students' conceptions and reasoning about differentiation between students with diverse needs. Training teachers to apply prompt engineering techniques could enhance the quality of AI-generated lesson plans by promoting structured reasoning and incorporating essential didactic elements. This approach improves AI output and empowers teachers to better understand and guide the AI's reasoning process, increasing the practical value of chatbots in educational contexts.

C. Limitations

We acknowledge the limitations of this study at various levels. At the theoretical level, it is important to acknowledge that, by choosing the cognitive demands of lesson planning as a theoretical framework, our research has a strong focus on the cognitive aspects of lesson planning, while emotional or affective aspects are not represented [21]. Furthermore, there may be an overlap of the cognitive

demands, and there could be other relevant cognitive demands that were not considered in this study [16]. The possibility of overlapping might have impacted the interrater reliability of the data analysis, as it was found to be good, but not excellent. An inductive approach, such as grounded theory or phenomenography, could yield an exhaustive set of cognitive demands.

At the methodological level, we acknowledge the limited sample size. As our methodological approach was qualitative, we focused on an in-depth analysis of data from these three groups. Although this approach allowed us to present detailed results for our sample, we cannot generalize our results. Using a much greater sample size in combination with quantitative analysis would have allowed us to check for any significant differences between distinct groups of teachers or various aspects of lesson planning. Future research might employ such an approach to generalize the findings. Moreover, we collected the data in the summer of 2024. With the rate at which AI is being developed, the quality of the responses given by AI and the teachers' prompting skills might have already improved. For this reason, the reproducibility of the results is low.

Regarding the study design, a limitation might stem from the fact that we asked the participants to plan a lesson on Newton's third law for lower secondary school. However, for some of the participants, this lesson planning remained solely fictional as they had never taught this specific topic. This is especially relevant for the prospective and out-of-field teachers. Planning a lesson for a fictional class instead of a real one might have resulted in teachers creating a more general lesson plan, which is not tailored to the needs of specific students. This might be a possible reason for the lack of differentiation in the lesson plans.

Furthermore, to answer RQ3, we asked teachers to reflect on their experiences with AI regarding at least three of the cognitive demands for lesson planning. We consider this approach necessary for conducting the analysis within our chosen theoretical framework. However, giving the teachers specific aspects to reflect on might have skewed the way they described their perception of using AI for lesson planning. We cannot rule out the possibility that the participants would focus on quite different aspects of lesson planning in their reflection if we had not given them the cognitive demands as a guideline. For example, when experienced teachers were prompted to think about task creation, they reported that the AI does not create useful tasks. If they had not been asked specifically, they maybe would not have reported on this aspect and focused more on what AI actually could do.

We also acknowledge that our interpretation of the results was based on the hypothesis that if a specific code is in the lesson plan of a participant, but was not prompted by the same participant, the AI considered this cognitive demand by itself. This hypothesis was based on the fact that

the data used to train the AI resulted in different weights for the various aspects of lesson planning. Nevertheless, we cannot say for sure that if a specific demand is not part of the lesson plan, the teacher did not think about the demand, or if they just did not prompt it because it was already clear to them.

VI. CONCLUSION

In this study, we asked prospective, out-of-field, and experienced physics teachers to plan a lesson on Newton's third law using AI to learn more about AI's potential to help physics teachers in their task development and lesson planning. The objective of the study was to identify how these distinct groups of physics teachers can use AI for lesson planning and what their perceptions about the role of AI for lesson planning are. We approached this objective through a qualitative methodology and analyzed the results based on the cognitive demands of lesson planning.

RQ1 focuses on the cognitive demands present in the lesson planning of middle school physics teachers with three groups: prospective, out-of-field, and experienced physics teachers. To answer RQ1, we found that the most frequent cognitive demands that were explicitly addressed by the participants when chatting with the AI were content transformation and unit contextualization. Task creation and clarity of learning objectives were present in fewer numbers, while adaptation to learning dispositions and phasing were only considered by a few teachers. On the other hand, clarity of learning objectives was the most frequent cognitive demand met by the AI tools, followed by phasing. Based on this finding, we argue that teachers should be clear about the learning objectives and phasing when prompting AI to make a lesson plan. Even though the AI tools already consider learning objectives and phasing when creating lesson plans, teachers should have the learning objectives and phases clearly defined beforehand to evaluate the quality of the lesson plans provided by the AI tools before using them and modify them when necessary.

To answer RQ2, we compared the results of prospective, out-of-field, and experienced physics teachers. We found similar patterns in the emergence of most categories. One category where we found notable differences was task creation, which was more prevalent with the out-of-field physics teachers. Another difference was the adaptation of learning dispositions, which experienced physics teachers included mostly in their prompting. We discussed that this difference may be explained by the experienced physics teachers being more aware of students' learning dispositions.

Concerning RQ3, we described teachers' perceptions of using AI in lesson planning, based on written reflections. Following the inductive qualitative analysis of their answers, we found that the teachers would discuss the

advantages of using AI, as well as critically examine their content. In general, the EPTs were most critical in the use of AI, especially regarding content transformation and task creation. The adaptation to learning dispositions was seen critically by a lot of teachers from all three groups. Positive remarks were made regarding the clarity of learning objectives and phrasing.

We acknowledge some educational implications of this study. One of the most important aspects of effective use of AI tools for physics education is prompting. By teaching educators how to use effective prompting techniques, universities, schools, or other teacher-training entities can enable them to create more structured and effective lesson plans (i.e., they can provide a structure that includes important points of physics education, such as students' conceptions). This technique not only improves the quality of AI-generated outputs but also empowers teachers to better understand and control the AI's reasoning process, making it a valuable tool in modern education. Targeted training in these areas could increase teachers' TPACK and therefore their ability to improve their teaching by making use of innovative technologies. For prospective physics teachers, the awareness of adapting the lesson plan to learning dispositions could be enhanced, so that they prompt the AI tools to effectively cover this aspect. For out-of-field physics teachers, the AI tools could aid in the content knowledge of physics as a starting point when creating their lessons, while also being aware of the learning dispositions. Given the fact that AI tends to produce superficial and sometimes even questionable results, teachers also need to have the ability to critically analyze the AI output. As the less experienced physics teachers tended to have a very positive view regarding AI, teacher training must focus on fostering reflective competencies of prospective teachers.

To conclude this study, we would like to present avenues for future research. It is important to consider that, even though the technological development of AI is rapidly

evolving, research on its applications in physics education remains limited. An important area of research is the risks related to AI use. Some of the risks found in early studies, mainly between 2022 and 2023 during the LLMs' first versions, are no longer a cause for concern due to improvements in recent versions of the LLMs and the use of plugins. Some risks remain, for example, the risk of AI hallucination is still highly relevant, and physics teachers should be aware of their responsibility in ensuring the quality of AI outputs before using them. Future research should focus on why physics teachers use AI, not only on how they use it. Will AI provide solutions for physics teachers? If so, what kind of solutions (time management, administrative, and pedagogical)? Are physics teachers using AI meaningfully to improve their practice or adopting an AI-first approach without reflection? A question that remains open in physics education research is whether we will be able to keep up with the rapid evolution of LLMs, as the research data can be outdated by the time of publication. We must, however, persist in this task because with every technological development, new questions arise.

ACKNOWLEDGMENTS

The authors thank the Austrian Educational Competence Centre physics education research group for providing access to the study's participants, and the participants for agreeing to take part. Open Access funding provided by the University of Vienna.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX

Unit contextualization		Students understand the concept of the principle of interaction (Newton's 3rd law) and can apply it in simple examples. “action = reaction”			
Clarity of objectives:		Time	Student activities	Teacher activities	Material
Phasing: 3-phase structure	Introductory phase	Intro 5–7 min	- Pupils greet the teacher - Pupils answer the instructor's question - Pupils listen and think in EA about what happens when a ball is thrown against a wall. - Ideas are collected in the plenary - Pupils listen and ask any questions on the topic	- Welcoming the pupils - Instructor asks what the pupils can remember from the last unit - Instructor summarizes the content of the last unit and asks the question “What happens if we throw a ball against a wall?” - Gives the pupils time to collect ideas (~1–2 min) - Instructor writes on the blackboard/smartboard what pupils answer - Instructor repeats/summarizes the results and leads to the topic of the principle of interaction.	
	Core working phase	Work 30–35 min		- Instructor repeats/summarizes the results and leads on to the topic of the principle of interaction. The motto of the unit is “Forces always occur in pairs and are of equal magnitude, but they are directed in opposite directions. ‘action = reaction’” - Instructor gives an example from everyday life and shows an experiment (throwing a ball - initial question of the unit) Table tennis: “When you hit a table tennis ball with a bat, the bat exerts a force on the ball (action). This force pushes the ball forward. At the same time, the ball exerts an equal but opposite force on the bat (reaction). This force causes you to feel the club back”). Swimming: “You paddle in the water with your arms and legs, exerting a force on the water. (Action) The water pushes you back with an equal force, causing you to move forward in the water. (Reaction)”. - Instructor now asks the pupils to go together in pairs or fours and work on worksheet M1 Alternative: Groups of 3 - Instructor goes around and asks for support, clarifies any questions etc.	M1
	Content transformation: Elementarization				
	Task creation: Learning task description		- Pupils now go together in pairs or fours and work on M1 Alternative: Groups of 3 - Pupils may need help/work in a concentrated manner/help each other etc.	- Instructor now asks the pupils to go together in pairs or fours and work on worksheet M1 Alternative: Groups of 3 - Instructor goes around and asks for support, clarifies any questions etc.	
	Feedback phase	Close 10 min	Pupils present their results in plenary	- Instructor adds any information on the pupils' contribution, summarizes the results/findings of the unit	

FIG. 6. Example of coding structure in a lesson plan from the out-of-field physics teachers.

Unit contextualization

Objective: Principle of interaction (Newton's 3rd law) for secondary level I

Clarity of objectives:	Time	Student activities	Teacher activities	Material
Introductory phase	10 min	Pupils name hypotheses, which are noted on the board. e.g., "I will roll backwards," "Nothing happens," "The ball only moves forwards" Pupils' observations are requested and hypotheses are discussed.	Pupils are asked to imagine the following thought experiment: "Imagine you are sitting on a wheelchair in the physics room and have a medicine ball in your hand. What happens if you throw the ball forcefully away from you?" Pupils are then asked to name hypotheses (min. 3 to max. 5 hypotheses) Instructor takes a vote All pupils must vote on at least one hypothesis. Thought experiment (wheelchair and medicine ball) is demonstrated. Introduction: Today we are learning about the principle of interaction, a basic rule that applies to all forces.	Medicine ball Rolling chair
Core working phase	20 min	Pupils receive worksheet and carry out experiment independently.	Explain the experiment with the wheelchairs and ropes. Discuss the three variants: One pupil pulls on the rope. The other student pulls. Both pull at the same time. Addition: Different masses Divide the class into small groups - 6 small groups Hand out the worksheet with observation tasks: Task 1: Note the movements when only one person is pulling. Task 2: Describe what happens when both people pull at the same time. Task 3: What changes if the two people have different masses?	Rolling chair (available in the PH hall) 6x ropes Worksheet
Task creation: Learning task description				
Content transformation: Analysis of subject matter	7 min	Pupils share their observations and conclusions. Pupils note the interaction principle on their worksheet.	The observations are compared with each other and discussed if necessary. Reflect on the results of the experiments and the principle of interaction is introduced and entered in the memory box on the worksheet. If body A exerts a force F_A (actio) on body B, body B exerts the opposing force F_B (reactio) on body A. The force and the opposing force are equal. The force and counterforce are equal in magnitude but directed in exactly opposite directions. Please note: Interaction forces always act on two different bodies!	
	7 min	Ask pupils to identify the force and counterforce	Consolidation: Pictures from everyday life are shown (e.g. recoil when walking or rowing) Force and counterforce are drawn on the board	Pictures
Feedback phase	5 min	Pupils do a quiz at the end Quiz (M2)	Discuss the answers together	

FIG. 7. Example of coding structure in a lesson plan from the experienced physics teachers.

- [1] F. Kieser, P. Wulff, J. Kuhn, and S. Küchemann, Educational data augmentation in physics education research using ChatGPT, *Phys. Rev. Phys. Educ. Res.* **19**, 020150 (2023).
- [2] G. Polverini and B. Gregorcic, Performance of ChatGPT on the test of understanding graphs in kinematics, *Phys. Rev. Phys. Educ. Res.* **20**, 010109 (2024).
- [3] W. Yeadon and T. Hardy, The impact of AI in physics education: A comprehensive review from GCSE to university levels, *Phys. Educ.* **59**, 025010 (2024).
- [4] B. Gregorcic and A.-M. Pendrill, ChatGPT and the frustrated Socrates, *Phys. Educ.* **58**, 035021 (2023).
- [5] S. Küchemann, S. Steinert, N. Revenga, M. Schweinberger, Y. Dinc, K. E. Avila, and J. Kuhn, Can ChatGPT support prospective teachers in physics task development?, *Phys. Rev. Phys. Educ. Res.* **19**, 020128 (2023).
- [6] R. W. Bybee and N. M. Landes, Science for life & living: An elementary school science program from biological sciences curriculum study, *Am. Biol. Teach.* **52**, 92 (1990).
- [7] W. Powell and S. Courchesne, Opportunities and risks involved in using ChatGPT to create first grade science lesson plans, *PLoS One* **19**, e0305337 (2024).
- [8] G. Gonçalves Costa, W. J. D. Nascimento Júnior, M. N. Mombelli, and G. Girotto Júnior, Revisiting a teaching sequence on the topic of electrolysis: A comparative study with the use of artificial intelligence, *J. Chem. Educ.* **101**, 3255 (2024).
- [9] R. O. Davis and Y. J. Lee, Prompt: ChatGPT, Create My Course, Please!, *Educ. Sci.* **14**, 24 (2024).
- [10] R. Hashem, N. Ali, F. El Zein, P. Fidalgo, and O. Abu Khurma, AI to the rescue: Exploring the potential of ChatGPT as a teacher ally for workload relief and burnout prevention, *Res. Pract. Technol. Enhanced Learn.* **19**, 023 (2024).
- [11] G. van den Berg and E. Du Plessis, ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education, *Educ. Sci.* **13**, 998 (2023).
- [12] P. Wattanakasiwich, K. Kaewkhong, and D. Katwibun, Physics instructors' acceptance and implementation of generative AI, *Phys. Rev. Phys. Educ. Res.* **21**, 010155 (2025).
- [13] N. Şimşek, Integration of ChatGPT in mathematical story-focused 5E lesson planning: Teachers, and pre-service teachers' interactions with ChatGPT, *Educ. Inf. Technol.* **30**, 11391 (2025).
- [14] B. Y. S. Acquah, F. Arthur, I. Salifu, E. Quayson, and S. A. Nortey, Preservice teachers' behavioural intention to use artificial intelligence in lesson planning: A dual-staged PLS-SEM-ANN approach, *Comput. Educ.* **7**, 100307 (2024).
- [15] P. Mishra and M. J. Koehler, Technological pedagogical content knowledge: A framework for teacher knowledge, *Teach. Coll. Rec.* **108**, 1017 (2006).
- [16] J. König, M. Krepf, A. Bremerich-Vos, and C. Buchholtz, Meeting cognitive demands of lesson planning: Introducing the CODE-PLAN model to describe and analyze teachers' planning competence, *Teach. Educ.* **56**, 466 (2021).
- [17] J. Carlson *et al.*, The refined consensus model of pedagogical content knowledge in science education, in *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*, edited by A. Hume, R. Cooper, and A. Borowski (Springer, Singapore, 2019), pp. 77–94.
- [18] L. Großmann and D. Krüger, Assessing the quality of science teachers' lesson plans: Evaluation and application of a novel instrument, *Sci. Educ.* **108**, 153 (2024).
- [19] J. Hattie and G. C. R. Yates, *Visible Learning and the Science of How We Learn* (Routledge, Oxon and New York, 2013).
- [20] D. Sardo-Brown, A longitudinal study of novice secondary teachers' planning: Year two, *Teach. Teach. Educ.* **12**, 519 (1996).
- [21] M. Cevikbas, J. König, and M. Rothland, Empirical research on teacher competence in mathematics lesson planning: recent developments, *ZDM Int. J. Math. Educ.* **56**, 101 (2024).
- [22] L. Großmann, M. Koberstein-Schwarz, D. Scholl, D. Krüger, and A. Meisert, Establishing common ground in empirical research on science teachers' lesson planning competence: A scoping review, *Stud. Sci. Educ.* **61**, 329 (2024).
- [23] N. Luhmann, *Soziale Systeme* (Suhrkamp, Frankfurt am Main, 1984), Vol. 666.
- [24] L. Shulman, Knowledge and teaching: Foundations of the new reform, *Harv. Educ. Rev.* **57**, 1 (1987).
- [25] W. Klafki, Didactic analysis as the core of preparation of instruction (Didaktische Analyse als Kern der Unterrichtsvorbereitung), *J. Curric. Stud.* **27**, 13 (1995).
- [26] R. Duit, H. Gropengießer, U. Kattmann, M. Komorek, and I. Parchmann, The model of educational reconstruction—A framework for improving teaching and learning science, in *Science Education Research and Practice in Europe*, edited by D. Jorde and J. Dillon (SensePublishers, Rotterdam, 2012), pp. 13–37.
- [27] F. K. Oser and F. J. Baeriswyl, Choreographies of teaching: Bridging instruction to learning, in *Handbook of Research on Teaching*, edited by V. Richardson, 4th ed. (American Educational Research Association, Washington DC, 2001), pp. 1031–1065.
- [28] R. Duschl, S. Maeng, and A. Sezen, Learning progressions and teaching sequences: A review and analysis, *Stud. Sci. Educ.* **47**, 123 (2011).
- [29] H. Krabbe and H. E. Fischer, Instructional design, in *Physics Education*, edited by H. E. Fischer and R. Girwidz (Springer International Publishing, Cham, 2021), pp. 83–112.
- [30] R. Bromme, *Das Denken von Lehrern bei der Unterrichtsvorbereitung* (Beltz, Weinheim; Basel, 1981).
- [31] H. Kang, Preservice teachers' learning to plan intellectually challenging tasks, *J. Teach. Educ.* **68**, 55 (2017).
- [32] J. König, A. Bremerich-Vos, C. Buchholtz, I. Fladung, and N. Glutsch, Pre-service teachers' generic and subject-specific lesson-planning skills: On learning adaptive teaching during initial teacher education, *Eur. J. Teach. Educ.* **43**, 131 (2020).
- [33] D. Kaltakci-Gurel, Assessment of conceptual understanding in physics, in *The International Handbook of Physics*

- Education Research: Teaching Physics*, edited by M. F. Taşarand P. R. L. Heron (AIP Publishing LLC, Melville, New York, 2023), pp. 19-1–19-20.
- [34] *Mind in Society*, edited by L. S. Vygotskij (Harvard University Press, Cambridge, MA; London, England, 1978).
- [35] R. Maulana, M. Helms-Lorenz, and W. van de Grift, Validating a model of effective teaching behaviour of pre-service teachers, *Teach. Teach. Educ.* **23**, 471 (2016).
- [36] M. Krepf and J. König, Structuring the lesson: An empirical investigation of pre-service teacher decision-making during the planning of a demonstration lesson, *J. Educ. Teach.* **49**, 911 (2023).
- [37] K.-H. Arnold and B. Koch-Priewe, The merging and the future of classical German traditions in general didactics: A comprehensive framework for lesson planning, in *Beyond Fragmentation. Didactics, Learning and Teaching in Europe*, edited by B. Hudson and M. A. Meyer (Budrich, Leverkusen, 2011), pp. 252–264.
- [38] D. Muijs, L. Kyriakides, G. van der Werf, B. Creemers, H. Timperley, and L. Earl, State of the art—teacher effectiveness and professional learning, *Sch. Eff. Sch. Improv.* **25**, 231 (2014).
- [39] G. Polverini and B. Gregorcic, How understanding large language models can inform the use of ChatGPT in physics education, *Eur. J. Phys.* **45**, 025701 (2024).
- [40] S. Welleck, I. Kulikov, S. Roller, E. Dinan, K. Cho, and J. Weston, Neural text generation with unlikelihood training, [arXiv:1908.04319](https://arxiv.org/abs/1908.04319).
- [41] J. D. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, and Q. Yang, *Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts*, edited by A. Schmidt, K. Väänänen, T. Goyal, P. O. Kristensson, A. Peters, S. Mueller, J. R. Williamson, and M. L. Wilson (ACM, Hamburg, Germany, 2023), pp. 1–21.
- [42] S. Schulhoff *et al.*, The prompt report: A systematic survey of prompting techniques, [arXiv:2406.06608](https://arxiv.org/abs/2406.06608).
- [43] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, Chain-of-thought prompting elicits reasoning in large language models, [arXiv:2201.11903](https://arxiv.org/abs/2201.11903).
- [44] D. J. Leiner, Software SoSci Survey, SoSci Survey GmbH, München (2024), <https://www.sosicisurvey.de/>.
- [45] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Taylor and Francis, Hoboken, 1988).
- [46] J. R. Dettori and D. C. Norvell, Kappa and beyond: Is there agreement?, *Global Spine J.* **10**, 499 (2020).
- [47] U. Kuckartz and S. Rädiker, *Qualitative Content Analysis: Methods, Practice and Software*, 2nd ed. (Sage, London and Thousand Oaks and New Delhi and Singapore, 2023).