Journal of Turkish Science Education, 2024, *21*(3), 448-466.
DOI no: 10.36681/tused.2024.024

The effect of changes in teaching methods on pupils' academic performance in biology

Irena Labak[1], Ivan Kujundžić[2], Branko Bognar[3]

[1]*J. J. Strossmayer University of Osijek, Croatia, Corresponding author, ilabak@biologija.unios.hr, ORCID ID: 0000-0001-9394-5131*

[2]*J. J. Strossmayer University of Osijek, Croatia, ORCID ID: 0009-0005-4542-4389*

[3]*Faculty of Humanities and Social Sciences, J. J. Strossmayer University of Osijek, Croatia, ORCID ID: 0000-0002-8037-6555*

ABSTRACT

Biology teachers received professional development to effect instructional changes that ensure student cognitive engagement and knowledge acquisition at higher cognitive levels. We asked the following questions: 1) What are the initial needs of teachers to promote active learner engagement and knowledge acquisition at higher cognitive levels? 2) What changes in teaching practice does each form of support trigger? 3) Do supportive and reflection-based professional development succeed in improving pupil academic achievement? Teachers received support in the form of interactive lectures and ready-made examples in the form of a written lesson plan to develop teaching practices that promote cognitive engagement and knowledge acquisition at higher cognitive levels. Throughout the professional development programme, they reflected on the success of their teaching practices derived from the lectures and implemented according to the prepared plans based on feedback. The analysis of video recordings of lessons enabled the collection of feedback, while learning communities facilitated critical discussions. Changes in teaching were monitored and identified through (self-)evaluation of recorded lessons using the Teaching Observation Form (TOF). The impact of the training on students' academic performance was determined using knowledge tests administered before and after the teacher training. Although teachers made positive changes in their teaching, these did not lead to an improvement in students' academic performance.

RESEARCH ARTICLE

ARTICLE INFORMATION
Received:
16.11.2023
Accepted:
14.02.2024

KEYWORDS:
Active learning, learning communities, professional learning, recorded lesson analysis, reflection.

To cite this article: Labak, I., Kujundžić, I., & Bognar, B. (2024). The effect of changes in teaching methods on pupils' academic performance in biology. *Journal of Turkish Science Education, 21*(3), 448-466. DOI no: 10.36681/tused.2024.024

Introduction

The intent of this paper is to support the critical discourse on professional development and perspectives on the complex construct of professional development (Mooney Simmie et al., 2024), where what matters is not only whether certain changes are elicited in teachers, but also whether these changes have a significant impact on pupil learning (Desimone, 2023). In today's world of uncertainty, "teachers need to engage in development processes throughout their lifetime, to be accountable for their practices, and to use and draw from research, counter-intuitive knowledge, and evidence as

valuable contributions to their thinking, to their practices and to their professional development" (Mooney Simmie, 2023, p. 917).

It is not easy to give a clear answer to the question of what constitutes effective professional development. Several pieces of research (e.g. Desimone, 2009; Darling-Hammond et al., 2017; Dunst et al., 2015) have highlighted a number of characteristics of effective teacher professional development: focusing on learners' understanding of subject matter and how they learn that content, facilitating active engagement, encouraging collaboration with other teachers, aligning with established curricula and school policies, and delivering learning sessions of appropriate duration to allow for practice and feedback. Recently, such "lists" of characteristics of effective professional development have been criticised. Research on the effectiveness of professional development that includes the aforementioned characteristics provides conflicting results, suggesting that the education field does not have a coherent, cohesive vision of what makes professional development programmes effective (Desimone, 2023). Furthermore, Asterhan & Lefstein (2024) note that the consensus on the key features of effective professional development is not based on solid evidence from large-scale, replicated and rigorously controlled research studies. Hill et al. (2022) criticise the empirical research on which the consensus on a core set of characteristics is based. Yang et al. (2020) were unable to confirm the effectiveness of the established characteristics of effective professional development on pupil learning in their study. Asterhan & Lefstein (2024) caution against the ambition to identify general characteristics of effective professional development at all. They believe that it is unrealistic to expect a universal answer to the question of which approach to professional development is effective. It is necessary to focus on understanding the conditions under which change occurs in the classroom (Hayes et al., 2024). The effectiveness of professional development programmes and their implementation depends on a variety of environmental factors, such as teachers' working conditions, instructional materials and other resources, school leadership, and informal processes of teacher learning (Asterhan & Lefstein, 2024). Some successful professional development programmes include peer support and the sharing of experiences (Hill & Papay, 2022). Teachers need to be supported in examining their existing pedagogical beliefs and how they manifest these in the classroom culture (Hayes et al., 2024). Desimone (2023) argues that professional development programmes should focus on helping teachers become experts who can make decisions about what each of their pupils needs. It is important to provide teachers with ongoing mentoring and collegial support to encourage the development of specific professional skills and knowledge and to maintain habits of mind (Graham et al., 2020).

Literature Review

Active learning as a feature of high-quality teaching (Baumert et al., 2010; Förtsch et al., 2016) positively affects learner performance (Dogani, 2023; Neumann et al., 2012;) and is necessary to acquire knowledge at higher cognitive levels. According to Bloom's taxonomy, higher cognitive levels include the processes of understanding, applying, analysing, evaluating and creating according to revised Bloom's taxonomy (Anderson & Krathwohl, 2001). Cognitive activation implies the planned involvement of learners in the teaching and assessment processes, enabling them to master the process of independent learning and self-assessment. It is reflected in the interaction between teachers and pupils by asking questions that promote higher-level cognitive processes. Teachers' questions and feedback positively affect learners' performance (Kyriakides et al., 2013) and contribute to an in-depth understanding of the content by activating their prior knowledge and promoting discussion about the content being learned (Förtsch et al., 2016; Praetorius et al., 2014). In addition to the questions the teacher asks, the questions posed by learners are also important, as is their free expression of ideas and hypotheses and the oral or written expression of understanding of the content being learned. Encouraging learners to engage in the above activities develops the skills of analysis, assessment, and creation, which are higher-level cognitive skills (Aisyah et al., 2018) and include various forms of thinking such as critical, logical, and creative (Mainali, 2012). Metacognitive higher-level thinking processes, which are simultaneously part of the cognitive system (Ristić Dedić, 2019), control the

above-mentioned cognitive activities. The described interaction encourages learners to monitor their work and progress and to use metacognitive knowledge and skills, which ultimately helps to practice and master the process of self-assessment. This process is inseparable from the learning process.

More cognitive processes in teaching can be stimulated by active learning methods such as flipped classroom and inquiry-based learning. The flipped classroom is an active learning approach implemented to improve the quality of learning in school (Ozdamli & Asiksoy, 2016). It includes homework in the sense of acquiring the information needed to solve tasks at higher cognitive levels in the classroom (Bergmann & Sams, 2012). At home, learners acquire knowledge at the level of reproduction using materials selected and designed by the teacher, which is then expanded to higher cognitive levels in the classroom. According to Candaş & Altun (2023) and Kurnianto et al. (2019) the flipped classroom improves critical thinking skills and learning outcomes in science and positively impacts motivation for critical thinking.

Inquiry-based learning is a very appropriate strategy for teaching science (Constantinou et al., 2018; Ladachart et al., 2022) and thus biology. It takes place in phases that correspond to the scientific methodology. In the conceptualisation phase, preceded by the orientation phase, pupils pose research questions and hypotheses, then test the hypotheses and draw conclusions (Pedaste et al., 2015). It requires high cognitive engagement and indicates metacognitive skill development (Nunaki et al., 2019). Using these strategies, learners acquire biological factual, conceptual, and procedural knowledge at higher cognitive levels determined by the outcomes of the prescribed subject curriculum.

Rationale, Objectives, Research Questions and Research Design

The teachers involved in our study received support in the form of interactive lectures (first line of support) and ready-made examples in the form of a written lesson plan (second line of support) to develop teaching practices that promote students' cognitive engagement and knowledge acquisition at higher cognitive levels. In addition, they reflected throughout the professional development program based on feedback on the success of their teaching practices derived from the lectures and implemented according to the prepared plans. An analysis of the video recordings of the lessons enabled feedback to be gathered, while learning communities facilitated critical discussions. The described imitates the model of reflective learning, which according to Vizek-Vidović & Vlahović Štetić (2007) comprises several phases. The first phase is the reflection that takes place during the planning and implementation of a particular activity (reflection in action – first level loop). The second phase is a reflection on what has been done and the identification of possible improvements (reflection on action – second level loop). Finally, there is a critical review and reflection on the reflection itself, which is the third level loop, and then the process circles back to the first level (Vizek-Vidović & Vlahović Štetić, 2007). Reflection enables teachers to take an active role in their own professional development, as individuals who monitor, supervise and guide their own professional growth (Labak, 2020).

Fostering pupils' cognitively active engagement during classes should become the central goal of teachers' professional development. In our study, teachers underwent active professional development to effect changes in teaching that ensured learners' cognitive engagement and knowledge acquisition at higher cognitive levels. To guide our research, we asked the following questions:

1. What are the initial needs of teachers to promote active pupil engagement and gain knowledge at higher cognitive levels?
2. What changes in teaching practice does each form of support trigger?
3. Do supportive and reflection-based professional development succeed in improving pupil academic achievement?

It is known that high-quality learning experiences that relate directly to the curriculum and instruction and involve active learning, practice and feedback can lead to changes in teachers'

classroom practice. However, it remains uncertain whether these innovative ideas and practices have an impact on pupil learning (Desimone, 2023). Improving pupil learning is the most distant variable in a long causal chain of effects (Asterhan & Lefstein, 2024). Teachers who teach professionally experience changes in their knowledge and teaching skills, which are reflected in the introduction of changes to their teaching methods. During the introduction of these changes, teachers evaluate their effectiveness and modify their teaching practices accordingly. Active pupil engagement and knowledge acquisition at higher cognitive levels require complex learner-teacher interactions that often involve the introduction of major innovations in teaching practices. For new learning strategies to be effective, learners must learn to use them in class, discuss them, reflect on them, and explain under what conditions they are effective (Česi & Ivančić, 2019). While teachers are introducing innovations into their practice, students are at the stage of adopting and practising these innovations and are not yet using them as their typical learning methods. Therefore, in relation to the research questions posed, our hypothesis is that teachers will change their teaching practices. However, we expect that these changes in teaching will not result in improved pupil achievement while professional development is ongoing.

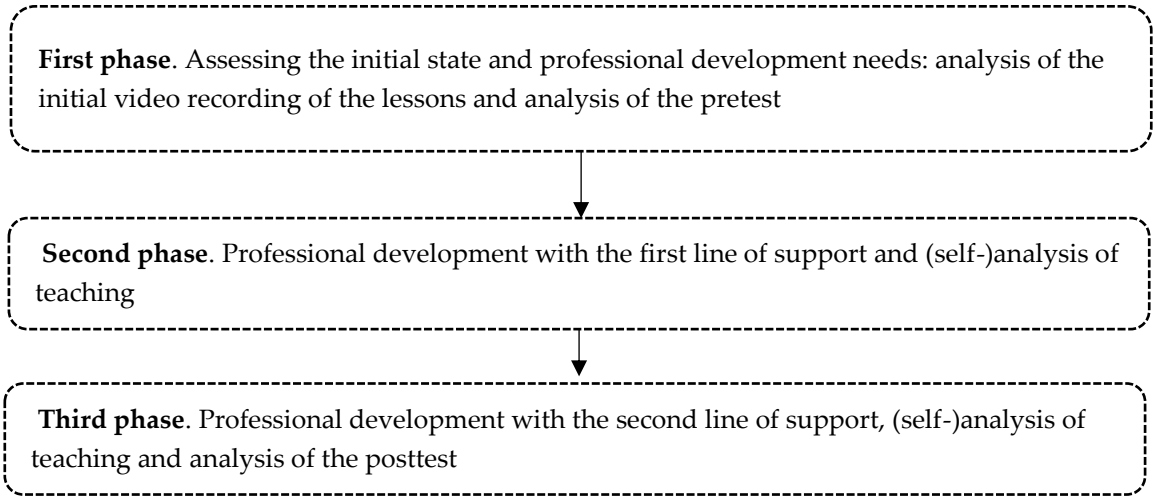
To verify the set hypotheses, we conducted a pretest and posttest before and after the professional development. The results of the posttest were compared with the results of the pretest. It created an experiment with a group where the success of a treatment is determined by comparing the pretest and posttest (Mills & Gay, 2019).

Methods

The research included four biology teachers and their 8th-grade pupils (mean age = 14.1; N = 134). Teachers underwent professional development during the second semester of the 2021/2022 academic year (from February to the end of May). We designed a professional development program that included the three phases listed in Figure 1.

Figure 1

Phases of the professional development program



First Phase

Before the implementation of professional development, one biology lesson was recorded for each teacher. These were regular lessons that were scheduled in the curriculum at the time of recording. The video recordings were analysed using the Teaching Observation Form (TOF). The TOF (Bezinović et al., 2012) assesses the presence of the teaching features that fall into six categories:

classroom atmosphere, the structure of the lesson, involvement and motivation, individualisation and differentiation of teaching, teaching metacognitive skills and learning strategies, feedback, and formative assessment. This research selected some of the TOF categories, i.e., features essential for the active involvement of learners in the teaching process which help them attain higher-level cognitive engagement (Table 1 and Table 2 in the Results, referring to the level of questions asked during teaching to attain student understanding and self-assessment). The lessons were analysed independently by two raters involved in the project and by the teachers themselves to determine the representation of certain features listed in the TOF. If a feature was not present, it was given a score of zero; if it was present but to an insufficient degree, it was given a score of 0.5; and if a feature was sufficiently present, it was given a score of 1. Table 1 shows the ratings of each teacher's teaching features (indicated by numbers from 1 to 4), and the numerical values in the columns marked RA indicate the average of the assessed feature from two raters. Inter-rater reliability was determined using Cohen's kappa coefficient, whose values can range from 0 (no agreement between raters) to 1 (excellent agreement between raters), with values below 0.20 indicating poor agreement, from 0.21 to 0.40 fair, from 0.41 to 0.60 moderate, from 0.61 to 0.80 good, and from 0.81 to 1.00 very good agreement (Landis & Koch, 1977). Table 1 also shows the average presence of teaching features, which was calculated by adding the self-assessment scores and the raters' average scores separately and dividing them by the number of teachers. If the average of a feature assessed by the raters was 0.5 or less, it was marked with a ↑ sign, which meant that it needed improvement. In this step, the pretest was conducted and evaluated to determine the students' prior knowledge. The pretest and the analysis of the lesson were the grounds for planning professional development.

Second Phase

The training programme began with an initial online learning community in which teachers critically reflected on the feedback they had gained from analysing the initial video recordings of all teachers' lessons. Following the learning community, the teachers took part in an interactive lecture. The outcome of the lecture was: *Discuss the didactic and methodological design of the lesson in which pupils respond to higher cognitive level questions, formulate observations and conclusions independently and make a self-assessment of the learning process and progress in relation to the objectives of the lesson*. The outcome of the lecture was determined based on the analysis of the recordings of the initial lessons and the pretest (see the arrows in Table 1 in the AZU column). After the learning community, the teachers applied what they had learned. Specifically, they independently planned a lesson that was recorded and analysed (by themselves) using the TOF. This independent lesson planning based on what was learned was the first line of support. In the second phase, there were two learning communities, and two lessons were recorded and analysed. The changes prompted by the first line of support were recorded and analysed following the same procedure as for the initial lesson The changes are presented as the average of the presence of the individual features of all teachers (Table 3).

Third Phase

After recording and analysing two lessons, teachers received a second line of support, viz ready-made examples in the form of a written lesson plan. The lesson plans included flipped classroom and inquiry-based learning, teaching approaches that can enhance all features of the lessons that were found to need improvement during the analysis of the recordings (Table 3, progress column compared to the initial recording). The lessons held after the second line of support were recorded as before and analysed in learning communities. All changes in relation to the first line of support are shown in Table 3 as the average of features for all teachers. Two lessons were recorded and three learning communities that analysed the recordings were held as the second line of support.

Statistical Analysis

For this study, we designed a pretest and posttest. Each test consisted of two questions to determine the first cognitive level (remember) and nine to examine higher cognitive levels (understanding, applying, analysing, evaluating and creating) determined according to a revision of Bloom's taxonomy of educational objectives (Anderson & Krathwohl 2001).

Before implementing the research instruments, we conducted a pilot study on a sample of students who did not participate in this research to examine their measurement characteristics. The pilot study for the pretest test was conducted on a sample of 288 students, and for the posttest test on a sample of 136 students. We calculated Cronbach's alpha coefficient as a measure of the reliability of both tests. For the pretest, the coefficient was 0.97, and for the posttest, it was 0.94. Considering these values, both tests are highly reliable (Bukvić, 1982). In addition, we calculated an item difficulty index (p) and a discrimination index (D) for each question in both tests. The item difficulty index indicates how easy or difficult a question is, while the discrimination index indicates how effective a question is in measuring differences between students (Cohen et al., 2007; Danuwijaya, 2018). Difficult and unacceptable questions were excluded from the test before the implementation with the pupils whose results we present in this study.

In both tests, the maximum score on the first-level cognitive questions was 2.5, while the maximum score on the higher-level questions was 22.5. Because the professional development programme was conducted with the expectation that the resultant teaching would elicit higher cognitive levels of learning, the tests evaluated mainly higher level of knowledge. In the tests, pupils were given the possible score next to each question and asked to estimate how many points they expected to score on each question. The ratio between the achieved and the expected score served as an assessment of the acquired self-assessment skills (Pavlin-Bernardić & Vlahović-Štetić, 2019).

The normality of the data distribution was calculated using the Kolmogorov-Smirnov test, while the homogeneity of variances was determined using the Levan homogeneity test. The overall performance in each written test is presented by descriptive statistics. Differences in the pretest scores between pupils from each teacher were determined by the ANOVA test, while differences in scores on the pretest and posttest were determined by the paired samples t test. Statistical tests were performed using the statistical software package Statistika 12 (Quest Software Inc., Aliso Viejo, CA, USA) with a significance level of $\alpha = 0.05$.

Findings

Initial Teacher Needs

The initial needs of the teacher were identified through the analysis of the first lesson and the pretest. The analysis of the initial lesson revealed that three out of five observed teaching features related to asking questions needed improvement. Two teachers needed to improve asking questions ($\bar{x}1$ indicated by bold numbers). For these two teachers, two of the five features are absent in teaching, while the other two teachers exhibited these features to a sufficient or insufficient degree. In relation to the teacher's willingness to respond to the pupils' questions and their free expression of ideas and asking questions, the teachers' self-assessment (SA) and the raters' assessment (RA) did not coincide. The teachers assessed features as being present when pupils asked about the rules for completing tasks and assignments, while raters assessed only questions and answers related to understanding the content being taught. All three observed teaching features related to content understanding need improvement and all four teachers needed to improve in this regard. The self-assessment and the raters' assessment in this latter part of the teaching observation correspond. The greatest discrepancy is in the average presence of the features related to independent notetaking. Teachers assessed copying from the board or presentation as independent notetaking, while raters rated only notes students made individually. The analysis of the initial recordings also revealed that all the observed

features related to self-assessment needed to be improved. Regarding self-assessment features, teachers' and raters' assessments are largely similar, as shown in Table 1.

Table 1

Distribution of Students by Gender

Teaching features	SA1	RA1	SA2	RA2	SA3	RA3	SA4	RA4	$\overline{x}$ SA	$\overline{x}$ RA	AZU
Teaching features related to asking questions											
The teacher willingly answers the pupils' questions.	1	0	1	0.5	1	0.5	1	0.5	1	0.4	↑
Pupils are free to express their ideas or ask for clarifications.	0.5	0	1	1	1	1	1	0	0.9	0.5	↑
The teacher allows pupils enough time to answer questions	1	1	1	0.5	1	1	1	1	1	0.9	
The class is interactive (lots of questions and answers).	0.5	1	1	1	1	1	1	1	0.9	1	
The teacher asks thought-provoking questions (which stimulate higher-level cognitive processes).	0.5	0.5	0.5	1	0.5	0.5	0	0	0.4	0.5	↑
$\overline{x1}$	0.7	0.5	0.9	0.8	0.9	0.8	0.8	0.5			
Teaching features related to the understanding of the content being taught											
The teacher emphasises understanding and not just memorising concepts.	0.5	1	1	0	1	1	0	0	0.6	0.5	↑
The teacher encourages pupils to express in their own words how they understood the content being taught.	0.5	0	0	1	0.5	0.5	0	0	0.3	0.4	↑
The teacher encou-rages pupils to independently take notes and organise the content (e.g., by highlighting key ideas and concepts or making simple mind maps).	0.5	0.5	1	0	0	0	0	0	0.4	0.1	↑
$\overline{x1}$	0.5	0.5	0.6	0.3	0.5	0.5	0	0			
Teaching features related to self-assessment											
The teacher clearly states the objectives of the lesson (learning outcomes).	0	0	0	0	1	1	0	0	0.3	0.3	↑
The teacher encourages pupils to monitor and review their work (e.g., to identify and correct errors, to verify the solution they have reached).	0	0	1	0	0	0	0	0.5	0.3	0.1	↑
The teacher asks pupils to evaluate their own work and progress.	0	0.5	0	0	0	0	0.5	0	0.1	0.1	↑

Teaching features	SA1	RA1	SA2	RA2	SA3	RA3	SA4	RA4	$\overline{x}$ SA	$\overline{x}$ RA	AZU
$\overline{x1}$	0	0.2	0.3	0	0.3	0.3	0.2	0.2			

*Note.*SA – self-assessment of teaching features, RA – raters' assessment of teaching features, $\overline{x}$SA – mean representation of teaching features of all teachers – self-assessment, $\overline{x}$RA – mean representation of teaching features of all teachers – raters' assessment, AZU – aspect that needs to be improved; ↑ improvement necessary, $\overline{x1}$– mean representation of teaching features of an individual teacher; For Teacher 1, 1 Fleiss' kappa showed that there was good agreement between the raters (κ=.709 (95% CI, .695 to .722), p < .0005); for Teacher 2 moderate agreement (κ=.557 (95% CI, .543 to .571), p < .0005; for Teacher 3 moderate agreement κ=.585 (95% CI, .571 to .598), p < .0005, and for Teacher 4 very good agreement κ=.827 (95% CI, .812 to .841), p < .0005.

Figure 2

Pupils' scores on questions of different cognitive levels of the pretest

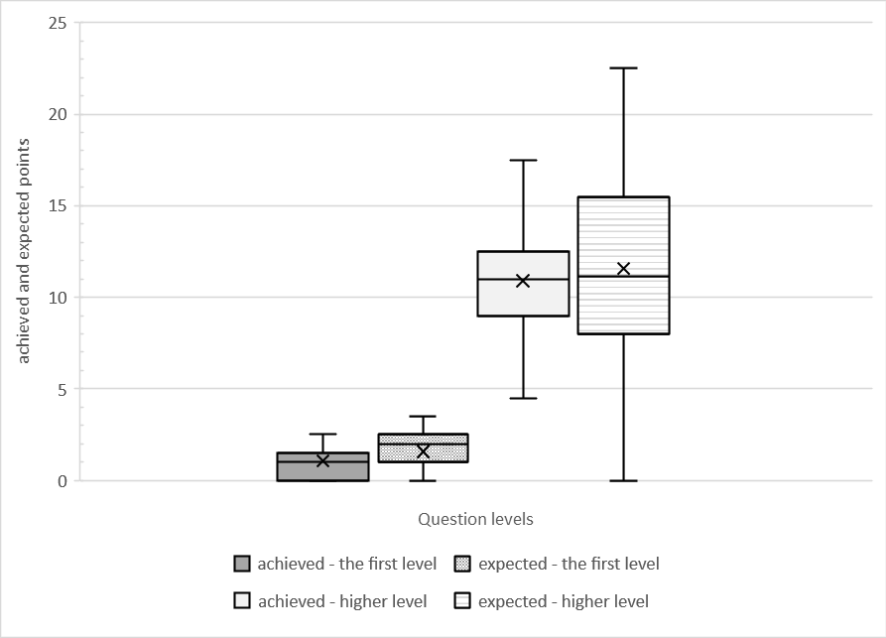


Figure 2 shows that pupils scored an average of 1.1 on the first-level questions, while the self-assessed average was 1.6. For the first-level questions, they could score a maximum of 2.5 points. The achieved score ranges from 0 to 2.5 points, with 25% of them scoring 0 points and 25% scoring 1.5 points or more. They overestimated themselves in the self-assessment. At the same time, 25% expected 1 point or less, while 25% expected 2.5 points or more. For the higher-level questions, the highest possible score was 22.5. The average score was 10.9 points, while the expected average score was 11.6 points. Pupils scored between 4.5 and 17.5 points while expecting 0 to 22.5. At the same time, 25% scored 9 points or less, and the same number expected 8 points or less. An equal number of students (25%) scored 12.5 points or more and expected 15.5 points or more. The students of the teacher with the number 3 achieve the best average results in the questions from both levels examined, but the difference to the students of the other teachers is not statistically significant.

Table 2

The scores of pupils in solving questions of different cognitive levels in the pretest of the knowledge of individual teachers

Question levels	Teachers	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min.	Max.
						Lower Bound	Upper Bound		
First- level	1.00	41	0.85	0.85	0.13	0.58	1.12	0.00	2.50
	2.00	39	1.24	0.91	0.15	0.95	1.53	0.00	2.50
	3.00	14	1.29	1.05	0.28	0.68	1.89	0.00	2.50
	4.00	40	1.05	0.85	0.13	0.78	1.32	0.00	2.50
Higher-level	1.00	41	10.60	2.57	0.40	9.79	11.41	4.50	16.00
	2.00	39	10.66	3.22	0.52	9.62	11.71	4.50	17.50
	3.00	14	11.50	3.05	0.82	9.74	13.26	6.50	16.50
	4.00	40	11.26	3.16	0.50	10.25	12.27	5.00	19.50

Changes in Teaching as a Result of Professional Development and Their Impact on Pupil Academic Performance

Two lines of support incited the changes – independent didactic-methodical design of lessons based on what was learned in the learning communities (first line of support) and the examples of lessons' didactic-methodical plan (second line of support). Table 3 shows the analysis of the third lesson, i.e., the application of the first line of support; and the analysis of the fifth lesson, i.e., the application of the second line of support. After that, the teachers completed their professional development programme.

Table 3

Analysis of the final recording regarding teaching features and changes as a result of teachers' professional development

Teaching features	1st line of support		Progress compared to the initial recording	2nd line of support		Progress compared to the 1st line of support
	$\overline{x}$ SA	$\overline{x}$ RA		$\overline{x}$ SA	$\overline{x}$ RA	
Teaching features related to asking questions						
The teacher willingly answers the pupils' questions.	0.8	0.2	↙	0.6	0.6	↗
Pupils are free to express their ideas or ask for clarifications.	0.7	0.7	↗	0.8	1	↗
The teacher allows pupils enough time to answer the questions he or she poses.	1	1	↔	1	1	↔
The class is interactive (lots of questions and answers).	1	1	↔	1	1	↔
The teacher asks thought-provoking questions (which stimulate higher-level cognitive processes).	0.7	1	↗	0.6	1	↔
Teaching features related to the understanding of the content being taught						
The teacher emphasises understanding and not just memorising concepts.	0.8	0.8	↗	0.8	0.8	↔
The teacher encourages pupils to express in their own words how they understood the content being taught.	0.6	0.7	↗	0.7	0.9	↗

The teacher encourages pupils to independently take notes and organise the content (e.g., by highlighting key ideas and concepts or making simple mind maps).	0.3	0.3	↔	0.9	0.9	↗
Teaching features related to self-assessment						
The teacher clearly states the objectives of the lesson (learning outcomes).	0.5	0.5	↔	0.7	0.6	↗
The teacher encourages pupils to monitor and review their work (e.g., to identify and correct errors, to verify the solution they have reached).	0.8	0.6	↗	0.7	0.7	↗
The teacher asks pupils to evaluate their own work and progress.	0.5	0.5	↔	0.4	0.5	↔

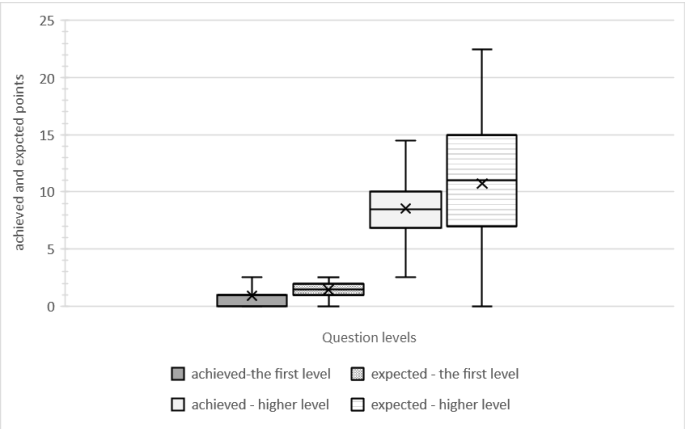
Note. $\bar{x}$ SA – mean representation of teaching features of all teachers – self-assessment, $\bar{x}$t RA – mean representation of teaching features of all teachers – raters' assessment, ↔ no progress, ↗ progress, ↙ decline.

On average, the first line of support led to progress on all teaching features that needed improvement, except for the teacher's willingness to answer pupils' questions, where there was a decline. For this feature, there are differences between the teachers' self-assessment and the raters' assessment that did not even out until after the second line of support, when this feature was present to a greater extent in the classroom. The first line of support resulted in progress in understanding. Only the feature related to pupils taking notes independently required further support. Clearly stating the objective and outcome of the lesson, which is a function of (self-)assessment, was partially represented after the first line of support, and a slight increase was observed after the second line of support. The incentive for learners to monitor and assess their own work and progress did not improve after the second line of support. In summary, on average, teachers made progress on all observed features by the end of their professional development programme. Pupil self-assessment could still be improved.

Although teachers made positive changes in teaching, these were not reflected in pupil performance. Comparing the posttest (Figure 3) to the pretest (Figure 2), they achieved a lower average score on first-level questions ($t(129)=2.03$; $p=0.045$, effect size Cohen's $d = 0.18$) and a lower average score on higher-level questions in the posttest ($t(129)=6.71$, $p < 0.001$; effect size Cohen's $d = 0.59$). In the first-level questions, they scored 0.9 points on average, while they expected to score 1.4. For higher-level questions, they scored an average of 8.6 points, while the self-assessed average was 10.7. The scores on higher-level questions in the posttest range from 2.5 to 14.5 points, which is less than the pretest. A quarter of the pupils scored 6.9 or less, and the same number scored 10 or more. Although the average of achieved and expected score is the same for the higher-level questions, the differences in expected distribution are visible, so they expected a range from 0 to 22.5, which is also the maximum score. A quarter of the pupils expected 15 points or more.

Figure 3

Students' score in solving questions of different cognitive levels in the posttest



Discussion

In this study, we observed the changes in teaching practice that promote pupils' active engagement and acquisition of knowledge at higher cognitive levels as an outcome of professional development for biology teachers. To bring about these changes, support was provided in the form of interactive lectures (first line of support) and ready-made lesson plans (second line of support). Throughout the training programme, teachers reflected on the success of the teaching practice learned in the lectures and implemented using the lesson plans provided, based on feedback. An analysis of the video recordings of the lessons enabled feedback to be gathered, while learning communities facilitated critical discussions. The design of the applied professional development programme tested the hypothesis that teachers would exhibit positive changes in instructional practice but that these positive changes would not result in improved pupil achievement while professional development is ongoing. After the first line of support, progress in lesson design and increased pupil activity were observed compared to the first lesson. With the introduction of the second line of support, almost all aspects that were not improved by the first line of support were improved. Despite the positive changes in teaching, the analysis of the post-test and its comparison with the pre-test showed no positive effects of the changes in teaching on the pupils' performance.

According to Asterhan & Lefstein (2024), it is unrealistic to expect to find a clear answer to the question of which approach to professional development is the most effective and to what extent. The approach must correspond to the desired change that one wants to bring about. For example, is it subject-specific knowledge, pedagogical knowledge or the development of teaching skills? Or is the aim to develop a teacher who carries out a self-evaluation of their teaching and takes action based on this? In this case, improving professional judgment may be best achieved through hands-on simulations or collaborative planning of practice presentations with like-minded colleagues (Horn & Garner, 2022). The effectiveness of the design depends critically on how it is implemented (Patfield et al., 2021). Whether something is effective depends on the professional and school environment in which teachers work, the professional knowledge, skills, judgment, and wisdom of leaders and teachers, and how all these factors interact (Asterhan & Lefstein, 2024). Therefore, in designing our professional development programme, we first identified the needs of the teachers involved in our study. By evaluating the pretest and analysing the video recordings of the first lesson, we gained partial insight into their usual teaching practices, and all of our future interventions were based on the aspects of teaching that we identified as needing improvement through these analyses. The pretest examined the pupils' prior knowledge that they had acquired before teachers' professional development program. The content included in the pretest was studied during the school year in which the pretest was administered and the year before when the pupils attended seventh grade. The relatively low pretest score suggests that many of them had forgotten some facts that are critical not only for answering first-level questions but also for application in solving tasks of higher cognitive level. The obtained scores can also be explained by the way teachers teach. Considering that the observation of the classes showed that, on average, pupils do not ask enough questions that contribute to the understanding of the content taught, that they do not express their ideas freely enough and do not ask enough for clarifications, and that teachers do not ask enough questions that stimulate thinking (Table 1), the lower pretest score was expected. On average, pupils' active participation is also underrepresented or absent in expressing their understanding of what is taught or summarising and organizing notes independently. Self-assessment, essential for higher-level learning, is also poor on average. Pupils are not encouraged to monitor and review their work or self-assess, and there is no clear setting of the objective at the beginning of the lesson so that they can self-assess. The self-assessment results suggest that pupils are not accustomed to self-assessing their work and results, which may explain the differences in self-assessed and actual scores.

Although some features important for active engagement and knowledge acquisition at higher cognitive levels are present in the initial lesson such as interactive teaching in terms of alternating questions and answers, it is evident from the results of the pretest that they are not

sufficient to contribute to better student learning outcomes. To achieve better understanding, instruction that promotes higher-level cognitive engagement of learners is required, which, according to Lee et al. (2019) and Mayer (2004), leads to better performance and deeper conceptual understanding. In such teaching, the presence and interaction of all observed features are essential. Indeed, the indeterminate difference in the scores achieved by students on the pretest (Table 2) suggests that all teachers provide similar teaching (Table 1) that leads to the same learning outcomes. According to Table 1 all teachers, except Teachers 2 and 3, need to improve teaching features related to asking questions, while all four teachers need to improve teaching features related to understanding and self-assessment. According to an equal score achieved on higher-level questions on the pretest it can be assumed that acquiring knowledge at higher cognitive levels requires teaching that involves the interaction of all features, not just some.

In the pretest, there was a discrepancy between the self-assessed score and the achieved score (Figure 2), indicating that the pupils had not developed self-assessment skills. Self-assessment is based on learners' metacognition. Unfortunately, teachers lack the knowledge to develop and implement it in the classroom (Ben-David & Orion, 2013; Labak, 2022; Seraphin et al., 2012). In addition to the barely present features that relate to self-assessment, there are also partially present features functioning as self-assessment (e.g., the learner seeking clarification and answering questions that stimulate thinking). These features are crucial for elaborating one of the strategies of independent learning and contributing to knowledge at higher cognitive levels (Pavlin-Bernardić & Vlahović-Štetić, 2019). They are also a tool for immediate feedback on pupil understanding and assessment of learning progress (Hattie, 2008). Independent notetaking and organising of acquired content are also strategies for monitoring lessons, which (like other student activities observed in this study) should be guided by metacognitive processes. The absence of the above teaching features makes it impossible to engage learners' metacognitive processes systematically and explicitly in learning. If teachers do not teach them metacognitive strategies that help them monitor their progress and take control of their learning, it will not ultimately lead to better educational outcomes (Ristić Dedić, 2019).

The Impact of Professional Development on Teaching Practice and Its Influence on Pupils' Academic Achievement

The training programme implemented in our study, which consisted of support and continuous reflection, led to positive changes in teaching. At the end of the programme, teachers showed all the characteristics during their teaching practice that contribute to active engagement and higher level learning through their interactions. The support we gave them could be likened to coaching and mentoring as described by Jin et al (2021). Contemporary training models have evolved into coaching and mentoring models that function like an apprenticeship where individuals observe and learn with and from others who are recognised as experts (Jin et al., 2021). Eshchar-Netz & Vedder-Weiss (2021) argue that novice teachers may refrain from sharing their work with others or seeking guidance from more experienced colleagues, while experienced teachers may be reluctant to disclose the challenges they face. Despite the sharing of teaching tips among teachers, it may prove difficult to establish constructive professional dialog for reflective collaborative inquiry (Eshchar-Netz & Vedder-Weiss, 2021). In our study, we fostered a culture of reflection and learning community where collaborative and friendly relationships were cultivated that allowed teachers to voice their observations, limitations and challenges. They were able to clearly articulate which form of support was of greater benefit to them and why. For example, they mentioned that the second form of support was helpful, but they found that a pre-packaged lesson plan did not quite suit their pupils (e.g. flipped classroom and inquiry-based learning, which also affected the final test score).

The first line of support for teachers contributed to the realisation of changes in their teaching, which related to almost all observed features (Table 3). The most progress (in the sense that all teachers exhibited a feature) was made in asking questions that stimulate thinking (Table 1 and Table 3). To ask such questions, teachers must possess adequate content knowledge. The progress achieved

indicates the improvement in pedagogical content knowledge developed during professional development. Baumert et al. (2010) found a relationship between maths teachers' pedagogical content knowledge and pupils' cognitive involvement. The improvement in pedagogical content knowledge is also related to progress in changing the focus from teaching focused on memorising concepts to teaching focused on understanding as indicated by the recorded progress on the first two features related to understanding (Table 3). Only at the end of professional development programme, after the second line of support, did pupils begin to ask more questions about understanding the content they were learning (Table 3, first feature). It indicates that teachers needed clear instructions on how to create situations in which learners would ask questions, but also that learners needed time to adjust to the new teaching conditions To clearly articulate goals, teachers needed a different form of support, which resulted in a relatively small change in teaching practice. Even in the final recorded lesson, teachers only partially encouraged pupils to self-assess concerning the defined objective. Self-assessment is part of formative evaluation, which according to Vingsle (2015), is a complex process that is difficult to integrate into teaching practice and requires psychological and practical support (Yan et al., 2021). Encouraging learners to take notes independently and organise the content they have learned has been significantly improved through the second line of support. In our study, as researchers, we were able to tailor the programme to the teachers' needs, which corresponds to the model of professional development described in (Labak, 2020), through reflective discussions in learning communities. During the reflection, teachers recognised the challenge of applying what they had learned in the interactive lectures to their classroom practice. Therefore, we decided to introduce a second line of support for them. A similar result, especially among university teachers was also found in the study by Labak & Blažetić (2023).

According to Deibl et al. (2018), lesson planning is not an intuitive process, especially when we introduce innovations that require deep thought, observationand reflection. The adequate way to reflect is through analysis of teaching videos, as it provides an understanding of how pupils learn specific content (Grossman, 2014), which according to Boston & Smith (2009), is necessary for teachers so they could effectively address identified learners' educational interests. The professional development program designed for this study included learning community meetings with interactive lectures, implementation of what was learned in class, analysis of teaching videos and reflection, and implementation of a ready-made didactic-methodical teaching practices. Islami et al. (2022) refer to these approaches as something teachers commonly use during their professional development. There is a growing consensus that professional development needs to take a collaborative approach that goes beyond the traditional boundaries of training to include observational placements, skills enhancement, proactive planning, achievement of learning objectives and adaptive expertise. There is a newfound urgency to support teacher experimentation with inquiry-based learning, drawing on evidence, integrating reflection, embracing the concept of learning from mistakes, and extending professional development programmes over time (Blackmore & O'Mara, 2022; Mooney Simmie et al., 2024). The positive changes teachers experienced in this research should be seen as a product of the interplay of all the approaches used. Support and reflection were the backbone of teachers' professional learning and allowed them active participation in their learning process. Darling-Hammond et al. (2017) describe active learning as one including collaboration, coaching, feedback, reflection, and models and modelling. In our study experts and teachers were equally involved in learning communities, although their roles differed. Experts provided professional support to teachers in their presentations, the design of teaching examples, and feedback on the effectiveness of implementation, which Nugent et al. (2016) reported had a positive impact on teachers' confidence. Teachers supported each other in the form of discussion about how best to adapt the planned changes to students and specific teaching conditions.

Professional development for teachers should help them to understand and improve their teaching practice with the aim of improving pupil learning (Lozano Cabezas et al., 2022). Guskey & Yoon (2009) believe that teachers' professional development leads to better learner performance because it allows them to better understand what they teach and how learners learn. According to

Desimone et al. (2002), changes in teaching can help to improve learning. However, there is evidence that this does not necessarily lead to improved student academic achievement (Yang et al. 2020; Yoon et al., 2007). These results are not surprising given that improving learning performance is the most distant variable in the impact chain (Asterhan & Lefstein, 2024). The professional development programme is expected to have an impact on teachers' skills, beliefs, and/or knowledge, resulting in improved instructional practices that influence the cognitive, motivational, and/or affective aspects of learner engagement and ultimately translate into individual learner test scores (Kennedy, 2016). The assumption of our study, which we confirmed, is that positive changes in teaching do not necessarily result in better academic performance while teachers are in the professional development phase. Pupils scored lower on average on both cognitive level questions on the posttest than on the pretest. This setback may be related to the fact that they were exposed to the acquisition of two types of knowledge arising from their teachers' professional learning: content knowledge, determined by the biology curriculum outcomes, and metacognitive knowledge. Intending to improve the observed characteristics, teachers introduced changes such as inquiry learning and flipped classrooms, which were new to them and their students.

During a short period (from February to the end of May), the pupils were exposed to various changes in the classroom. During this time, they learned various biological content conceptually based on evolution, which according to Ross et al. (2010), is a concept of abstract but grounded ideas for understanding numerous processes and phenomena in biology. They adopted this concept at higher cognitive levels dictated by the subject curriculum, using strategies that enabled reaching higher cognitive levels and ensuring improvement in the teaching practices of the teachers involved in the study. Most of these strategies were new to learners. They were simultaneously experiencing new teaching strategies and learning biology content. Because the assessment of the effects of teachers' professional development on their academic performance occurred immediately after the implemented changes, learners did not have sufficient time to internalise the new learning practices. We hypothesise that effective teaching intervention needs more time to show positive effect. In other words, we might expect changes in teaching and educational outcomes would become more visible when everything teachers learned during their professional development became a well-established and common teaching practice and a way of learning students are familiar with. The lack of knowledge retention testing due to pupils' transition to higher levels of education is both a limitation of our study and an opportunity for further research.

Conclusion and Implications

The focus of our research has been on two aspects: first, we wanted to find out whether a developed professional development programme brings about positive changes in teaching and under what conditions these changes occur. Secondly, we wanted to gain insight into the conditions under which changes in teaching lead to improvements in student learning.

The implemented professional development programme appeared to be effective in changing teaching practices but did not show improved pupil academic performance. During professional development, teachers analysed the effectiveness of their teaching concerning the innovations introduced, and pupils were at the stage of accepting these innovations as a new way of learning. The effect of professional development on learners' academic performance would be good to test after teachers and learners have internalised the new learning approaches. Knowledge retention tests are likely to be a more effective and reliable means of testing the interdependence of teaching changes and the effectiveness of learning, and they are essential for critically considering cause-and-effect relationships, especially when introducing teaching changes that involve higher learner (meta)cognitive engagement and the acquisition of knowledge at higher cognitive levels. Unfortunately, it was impossible to achieve this in our study because the participants were pupils in the final grades of primary school who continued their education in high schools after the completion of this study. The applied design of the study included not only the question of how well professional

development programmes work, but also why they (do not) work. Therefore, the results of our research support the critical discourse on professional development and offer perspectives on the complex construct of professional development.

Acknowledgements

This work was supported by the Croatian Science Foundation under the project IP-2018-01-8363.

References

Aisyah, A., Salehuddin, K., Aman, I., Yasin, R. M., & Mimiko, N. (2018). Eliciting elements of higher order thinking skills in the higher secondary examination question structure in Japan and Malaysia. In M. Y. Mohamad Noor, B. E. Ahmad, M. R. Ismail, H. Hashim, & M. A. Abdullah Baharum (Eds.), *Proceedings of the Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)* (pp. 455–464). Springer. https://doi.org/10.1007/978-981-13-0203-9_42

Anderson, L. W., & Krathwohl, D. R. (Eds.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of educational objectives*. Longman.

Asterhan, C. S. C., & Lefstein, A. (2024). The search for evidence-based features of effective teacher professional development: A critical analysis of the literature. *Professional Development in Education, 50*(1), 11–23. https://doi.org/10.1080/19415257.2023.2283437

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Ben-David, A., & Orion, N. (2013). Teachers' voices on integrating metacognition into science education. *International Journal of Science Education, 35*(18), 3161–3193. https://doi.org/10.1080/09500693.2012.697208

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. International Society for Technology in Education.

Bezinović, P., Marušić, I., & Ristić Dedić, Z. (2012). *Opažanje i unapređivanje školske nastave* [Observing and improving school teaching]. Agencija za odgoj i obrazovanje [Education and Teacher Training Agency].

Blackmore, J., & O'Mara, J. (2022). Putting professional learning practice first in innovative learning environments. *Studies in Continuing Education, 44*(2), 232–246. https://doi.org/10.1080/0158037X.2021.1977921

Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education*, *40*(2), 119–156. http://www.jstor.org/stable/40539329

Bukvić, A. (1982). *Načela izrade psiholoških testova* [Principles of designing psycholgical tests]. Zavod za udžbenike i nastavna sredstva.

Candaş, B., & Altun, T. (2023). Investigating flipped laboratory practices of science students teachers. *Journal of Turkish Science Education, 20*(1), 173-188. https://doi.org/10.36681/tused.2023.010

Cohen, L., Manion, L., & Morrison, K. (2007). *Metode istraživanja u obrazovanju* [Research methods in education]. Naklada Slap.

Constantinou, C. P., Tsivitanidou, O. E., & Rybska, E. (2018). What is inquiry-based science teaching and learning? In O. E. Tsivitanidou, P. Gray, E. Rybska, L. Louca, & C. P. Constantinou (Eds.), *Professional development for inquiry-based science teaching and learning* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-319-91406-0_1

Česi, M., & Ivančić, Đ. (2019). *Izazovi i umijeća učenja i poučavanja: Hrvatski jezik i inkluzivni pristup* [Challenges and skills of learning and teaching: Croatian language and inclusive approach]. Naklada Ljevak.

Danuwijaya, A. (2018). Item analysis of reading comprehension test for post-graduate students. *English Review: Journal of English Education*, *7*(1). 29-40. https://doi.org/10.25134/erjee.v7i1.1493

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://doi.org/10.54300/122.311

Deibl, I., Zumbach, J., Geiger, V. M., & Neuner, C. M. (2018). Constructive alignment in the field of educational psychology: Development and application of a questionnaire for assessing constructive alignment. *Psychology Learning & Teaching*, *17*(3), 293–307. https://doi.org/10.1177/1475725718791050

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Desimone, L. M. (2023). Rethinking teacher PD: A focus on how to improve student learning. *Professional Development in Education, 49*(1), 1–3. https://doi.org/10.1080/19415257.2023.2162746

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, *24*(2), 81–112. https://doi.org/10.3102/01623737024002081

Dogani, B. (2023). Active learning and effective teaching strategies. *International Journal of Advanced Natural Sciences and Engineering Researches, 7*(4), 136–142. https://doi.org/10.59287/ijanser.578

Dunst, C. J., Bruder, M. B., & Hamby, D. W. (2015). Metasynthesis of in-service professional development research: Features associated with positive educator and student outcomes. *Educational Research and Reviews, 10*(12), 1731–1744. https://doi.org/10.5897/ERR2015.2306

Eshchar-Netz, L., & Vedder-Weiss, D. (2021). Teacher learning in communities of practice: The affordances of co-planning for novice and veteran teachers' learning. *Journal of Research in Science Teaching, 58*(3), 366–391. https://doi.org/10.1002/tea.21663

Förtsch, C., Werner, S., Dorfner, T., von Kotzebue, L., & Neuhaus, B. J. (2016). Effects of cognitive activation in biology lessons on students' situational interest and achievement. *Research in Science Education, 47*(3), 559–578. https://doi.org/10.1007/s11165-016-9517-y

Graham, L. J., White, S. L. J., Cologon, K., & Pianta, R. C. (2020). Do teachers' years of experience make a difference in the quality of teaching? *Teaching and Teacher Education, 96*, 103190. https://doi.org/10.1016/j.tate.2020.103190

Grossman, P. (2014). *Collecting evidence of instruction with video and observation data.* National Academy of Education. https://naeducation.org/wp-content/uploads/2016/10/grossman-nces-longitudinal-surveys.pdf

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *90*(7), 495–500. https://doi.org/10.1177/003172170909000709

Hattie, J. (2008). *Visible learning: A Synthesis of over 800 meta-analyses relating to achievement*. Routledge. https://doi.org/10.4324/9780203887332

Hayes, K. N., Preminger, L., & Bae, C. L. (2024). Why does teacher learning vary in professional development? Accounting for organisational conditions. *Professional Development in Education, 50*(1), 108–128. https://doi.org/10.1080/19415257.2023.2283433

Hill, H. C., Papay, J. P., & Schwartz, N. (2022). *Dispelling the myths: What the research says about teacher professional learning*. Research Partnership for Professional Learning. https://annenberg.brown.edu/sites/default/files/rppl-dispelling-myths.pdf

Hill, H. C., & Papay, J. P. (2022). *Building better PL: How to strengthen Teacher learning*. Research Partnership for Professional Learning, Annenberg Institute.

Horn, I., & Garner, B. (2022). *Teacher learning of ambitious and equitable mathematics instruction: A sociocultural approach*. Routledge. https://doi.org/10.4324/9781003182214

Islami, R. A. Z. E., Anantanukulwong, R., & Faikhamta, C. (2022). Trends of teacher professional development strategies: A systematic review. *Shanlax International Journal of Education, 10*(2), 1-8. https://doi.org/10.34293/education.v10i2.4628

Jin, X., Li, T., Meirink, J., van der Want, A., & Admiraal, W. (2021). Learning from novice–expert interaction in teachers' continuing professional development. *Professional Development in Education, 47*(5), 745–762. https://doi.org/10.1080/19415257.2019.1651752

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980. https://doi.org/10.3102/0034654315626800

Kurnianto, B., Wiyanto, W., & Haryani, S. (2019). Critical thinking skills and learning outcomes by improving motivation in the model of flipped classroom. *Journal of Primary Education*, *8*(6), 282-291. https://journal.unnes.ac.id/sju/jpe/article/view/27783

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, *36*, 143–152. https://doi.org/10.1016/j.tate.2013.07.010

Labak, I. (2020). Upravljanje razvojem pedagoških kompetencija [Managing the development of pedagogical competences]. *Školski vjesnik: Journal of Pedagogic Theory and Practice, 69*(2), 461–480. https://hrcak.srce.hr/249030

Labak, I. (2022). Unaprjeđivanje metakognitivne dimenzije kompetencije učiti kako učiti kod učitelja. [Improving the metacognitive dimension of competence *to learn how to learn* in students and teachers]. *Napredak: Journal of Interdisciplinary Research in Education, 163*(1–2), 181–199. https://hrcak.srce.hr/279091

Labak, I., & Blažetić, S. (2023). A successful learning environment for biology teachers in higher education: Needs assessment. *Journal of Education and E-Learning Research, 10*(3), 530–538. https://doi.org/10.20448/jeelr.v10i3.4919

Ladachart, L., Phothong, W., Phornprasert, W., Suaklay, N., & Ladachart, L. (2022). Influence of an inquiry-based professional development on science teachers' orientations to teaching science. *Journal of Turkish Science Education, 19*(3), 979-996. https://doi.org/10.36681/tused.2022.159

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lee, J., Song, H.-D., & Hong, A. J. (2019). Exploring factors, and indicators for measuring students' sustainable engagement in e-learning. *Sustainability, 11*(4), 985. https://doi.org/10.3390/su11040985

Lozano Cabezas, I., Iglesias Martínez, M. J., Arroyo Salgueira, S., Camús Ferri, M. D. M., & Giner Gomis, A. (2022). What teaching models do pre-service teachers learn during placements? *Cogent Education, 9*(1), 2034393. https://doi.org/10.1080/2331186X.2022.2034393

Mainali, B. P. (2012). Higher order thinking in education. *Academic Voices: A Multidisciplinary Journal*, *2*(1), 5–10. https://doi.org/10.3126/av.v2i1.8277

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *The American Psychologist*, *59*(1), 14–19. https://doi.org/10.1037/0003-066X.59.1.14

Mills, G. E., & Gay, L. R. (2019). *Educational research: competencies for analysis and applications* (12th ed.). Pearson.

Mooney Simmie, G. M. (2023). Teacher professional learning: A holistic and cultural endeavour imbued with transformative possibility. *Educational Review, 75*(5), 916–931. https://doi.org/10.1080/00131911.2021.1978398

Mooney Simmie, G., O'Meara, N., Forster, A., Ryan, V., & Ryan, T. (2024). Towards a model of teachers' continuing professional development (CPD): A border crossing journey with embedded contradictions, ethical dilemmas and transformative possibilities. *Professional Development in Education, 50*(1), 46–58. https://doi.org/10.1080/19415257.2023.2283420

Neumann, K., Kauertz, A., & Fischer, H. E. (2012). Quality of instruction in science education. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 247–258). Springer. https://doi.org/10.1007/978-1-4020-9041-7_18

Nugent, G., Kunz, G., Houston, J., Kalutskaya, I., Wu, C., Pedersen, J., Lee, S. C., DeChenne-Peters, S. E., Luo, L. L., & Berry, B. (2016). *The effectiveness of technology-delivered science instructional coaching in middle and high school*. National Center for Research on Rural Education. https://eric.ed.gov/?id=ED571809

Nunaki, J. H., Damopolii, I., Kandowangko, N. Y., & Nusantari, E. (2019). The effectiveness of inquiry-based learning to train the students' metacognitive skills based on gender differences. *International Journal of Instruction, 12*(2), 505-516. https://doi.org/10.29333/iji.2019.12232a

Ozdamli, F., & Asiksoy, G. (2016). Flipped classroom approach. *World Journal on Educational Technology: Current Issues*, *8*(2), 98–105. https://doi.org/10.18844/wjet.v8i2.640

Patfield, S., Gore, J., & Harris, J. (2021). Shifting the focus of research on effective professional development: Insights from a case study of implementation. *Journal of Educational Change, 24*(2), 345–363. https://doi.org/10.1007/s10833-021-09446-y

Pavlin-Bernardić, N., & Vlahović-Štetić, V. (2019). Kognitivni aspekti samoregulacije učenja [Cognitive aspects of self-regulated learning]. In V. Vizek Vidović, & I. Marušić (Eds.), *Kompetencija učiti kako učiti: Teorijske osnove i istraživanja u hrvatskom kontekstu* [The learning to learn competence: theoretical foundations and research in the Croatian context] (pp. 89–110). Institut za društvena istraživanja u Zagrebu.

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, *31*(1), 2-12. https://doi.org/10.1016/j.learninstruc.2013.12.002

Ristić Dedić, Z. (2019). Metakognitivni aspekti samoregulacije učenja [Metacognitive aspects of self-regulated learning]. In V. Vizek Vidović & I. Marušić (Eds.), *Kompetencija učiti kako učiti: Teorijske osnove i istraživanja u hrvatskom kontekstu* [The learning to learn competence: theoretical foundations and research in the Croatian context] (pp. 89–110). Institut za društvena istraživanja u Zagrebu [Institute for Social Research in Zagreb].

Ross, P. M., Taylor, C. E., Hughes, C., Whitaker, N., Lutze-Mann, L., Kofod, M., & Tzioumis, V. (2010). Threshold concepts in learning biology and evolution. *International Journal of Biology*, *47*, 47–54. https://iubs.org/wp-content/uploads/2021/12/BI_Numero_47.pdf#page=49

Seraphin, K. D., Philippoff, J., Kaupp, L., & Vallin, L. M. (2012). Metacognition as means to increase the effectiveness of inquiry-based science education. *Science Education International*, *23*(4), 366–382.

Vingsle, C. (2015). Formative assessment: Teacher knowledge and skills to make it happen. In K. Krainer, & N. Vondrová (Eds.), *CERME 9—Ninth Congress of the European Society for Research in Mathematics Education* (pp. 3172–3173). Charles University in Prague, Faculty of Education and ERME. https://hal.archives-ouvertes.fr/hal-01289837

Vizek-Vidović, V., & Vlahović Štetić, V. (2007). Modeli učenja odraslih i profesionalni razvoj [Models of adult learning and professional development]. *The Annual of Social Work*, *14*(2), 283–310. https://hrcak.srce.hr/14421

Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, *28*(3), 228–260. https://doi.org/10.1080/0969594X.2021.1884042

Yang, R., Porter, A. C., Massey, C. M., Merlino, J. F., & Desimone, L. M. (2020). Curriculum-based teacher professional development in middle school science: A comparison of training focused on cognitive science principles versus content knowledge. *Journal of Research in Science Teaching*, *57*(4), 536–566. https://doi.org/10.1002/tea.21605

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, REL 2007–No. 033). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. https://eric.ed.gov/?id=eD498548