

ARTIFICIAL INTELLIGENCE AS A PRACTICE PARTNER: THE POTENTIAL OF GENERATIVE SIMULATIONS IN INITIAL TEACHER EDUCATION

Flavius-Cristian MĂRCĂU

Lecturer PhD., Faculty of Educational Sciences, Law and Public Administration,
Constantin Brâncuși University of Târgu Jiu, Romania
flavius.marcau@utgjiu.ro

Elena PALIȚĂ

Associate professor PhD., Faculty of Educational Sciences, Law and Public Administration,
Constantin Brâncuși University of Târgu Jiu, Romania
elena.palita@utgjiu.ro

Abstract

Initial teacher education faces a persistent paradox: students are expected to develop complex professional competences – diagnosing a pupil’s misconception, defusing a conflict, conducting a difficult conversation with a parent – while opportunities to practise them under supervision remain few, uneven and largely left to chance. This article examines the potential of generative artificial intelligence to narrow this gap through simulated practice. Methodologically, we combine a critical review of the literature on simulation-based learning, feedback and reflective practice with an instrumental case study of Practicum, a web platform we designed for students preparing for the teaching profession in Romania. We analysed the platform’s initial catalogue of 20 role-play scenarios (extensible by trainers), its 60 assessment criteria and the instructions governing four distinct roles assigned to the language model: role-play partner, formative evaluator, source-grounded tutor and reflective mentor. The scenario categories cover five of the seven competence domains of the Romanian teacher professional standards, and two thirds of the criteria target the relational, emotional and ethical dimensions of teaching – precisely those that written case studies cannot exercise. Building on these findings, we propose a five-stage model of AI-mediated simulated practice (preparation, rehearsal, mirroring, reflection, transfer) in which the human mentor remains indispensable. We argue that the main potential of AI lies in repeatable, adaptive and immediately debriefed rehearsal, while its main risks concern the authenticity of simulated characters, the reliability of automated evaluation and the delegation of professional judgement.

Keywords: artificial intelligence; simulated practice; initial teacher education; role-play simulation; formative feedback; reflective practice

1. Introduction

The emergence of large language models (LLMs) and their unprecedented accessibility, from late 2022 onwards, prompted a wide range of responses in higher education, from enthusiasm among those who regard them as a universal assistant to concern among those who fear for academic integrity. Between these two extremes, the debate has remained focused largely on how students write, read or are assessed. Far less attention has been paid to another area, perhaps the most sensitive in teacher education: the point at which the student moves from *knowing* to *doing*, namely teaching practice. This observation led us to examine the place that

new technologies might occupy in practising the professional competences of future teachers and, above all, the conditions under which such a place is legitimate.

The gap between theory and practice is not a new problem. More than two decades ago, Fred Korthagen and Jos Kessels pointed out that the theoretical foundations taught at university are difficult to transfer into classroom action when they are not constructed from students' concrete experiences (Korthagen and Kessels, 1999). Ken Zeichner (2010) described the divide between university courses and experiences in schools, while Linda Darling-Hammond (2006) showed that successful teacher education programmes are distinguished precisely by the coherent integration of clinical experience into the theoretical pathway. In Romania, Romița Iucu (2004) analysed systems, policies and strategies for teacher education, and the practical component remains a recurring issue in the debate on the quality of initial teacher education.

Teaching practice, as it is currently organised, often resembles a lottery of situations. A student may complete the entire programme without ever encountering a parent who publicly challenges a grade, a child who refuses to participate in a lesson or a colleague who suggests, in a friendly tone, that they should 'round up' an average mark. It is precisely in these rare but decisive situations that the beginning teacher is most vulnerable. When such situations do arise, the student is usually an observer rather than the main participant or, if they are the main participant, their mistake carries a cost borne by real children.

Let us conduct a thought experiment and put ourselves in the position of a second-year student enrolled in the Primary and Pre-school Education programme. They have studied educational psychology, the didactics of mathematics and communication theory. In principle, they know that feedback should be descriptive, that a child's emotions should be validated and that a concerned parent first needs to be heard. Yet they will most likely have their first real conversation with a dissatisfied parent only after entering employment, alone, without any prior rehearsal and without anyone to tell them afterwards what went well and what did not. Between what they know and what they will have to do lies a space that initial teacher education covers only partially.

Simulation is a long-established response to this problem. Microteaching, proposed by Dwight Allen and Kevin Ryan (Allen and Ryan, 1969), clinical simulations involving trained actors, described by Benjamin Dotger (2013), and mixed-reality environments in which students interact with avatars controlled by human operators (Dieker et al., 2014) have demonstrated that practice in a protected setting can prepare students for encounters with the real classroom. However, all these forms entail substantial costs in time, personnel and infrastructure, which is why they remain exceptions rather than routine practices at most universities. Large language models change this equation: a single system can portray a seven-year-old child who makes mistakes in subtraction, a concerned parent or a demanding mentor, respond to what the student says and then comment on the student's intervention. Whether this promise withstands careful analysis remains to be seen.

The urgency of the topic is heightened by the regulatory context. Law No. 198/2023 on Pre-University Education and Law No. 199/2023 on Higher Education reshaped the architecture of initial teacher education by introducing the teaching master's programme, while Order of the Minister of Education No. 7,386/2024 approved the Profile and Professional Standards for Teachers, organised into seven competence domains. For the first time, practical training has an explicit frame of reference against which any practice tool, including one based on artificial intelligence (AI), can and must be evaluated.

To continue our inquiry, we need to answer the following set of questions: **(H1)** What does generative AI add to simulated practice compared with traditional forms of simulation? **(H2)** Under what conditions can this potential be transformed into professional learning rather than mere mechanical practice or the delegation of professional judgement to a machine? **(H3)** What are the pedagogical, ethical and legal limitations that a university must consider when integrating such tools into the education of future teachers?

With regard to the research objectives, we sought to:

- critically analyse the literature on simulation-based learning, formative feedback and the use of generative AI in teacher education;
- examine, as an instrumental case study, the Practicum platform, designed for students preparing for the teaching profession, in relation to national professional standards;
- develop a model of AI-mediated simulated practice that specifies the place of each human and artificial component in the professional learning cycle;
- identify risks and formulate an agenda for empirical validation.

The novelty of our approach lies in the fact that we do not discuss the potential of AI in the abstract, but on the basis of a tool that has actually been built in Romanian, aligned with national professional standards and subjected to explicit critical analysis. To the best of our knowledge, this paper is among the first analyses of its kind in Romania.

2. Literature review

2.1. *The theory–practice gap and pedagogies of practice*

The theoretical foundations of practice in teacher education derive from three complementary traditions. The first is experiential learning. David Kolb (1984) describes learning as a cycle in which concrete experience is followed by reflective observation, abstract conceptualisation and active experimentation. The absence of any one of these stages leaves experience unprocessed or theory unapplied. The second tradition is associated with Donald Schön (1983), who distinguishes between *reflection-in-action* and *reflection-on-action* and argues that competent professionals do not apply rules mechanically, but construct solutions in uncertain, unique and often conflicting situations. The third tradition concerns the nature of professional knowledge: Lee Shulman (1987) developed the concept of pedagogical content knowledge, that specific blend of subject-matter mastery and knowledge of how to teach it, later extended by Punya Mishra and Matthew Koehler through the TPACK framework, which also integrates the technological dimension (Mishra and Koehler, 2006).

On this basis, Pam Grossman and colleagues, comparing teacher education with the education of clergy and clinical psychologists, proposed three concepts that describe how relational professions prepare their novices: *representations* of practice (case studies, video recordings, examples), the *decomposition* of practice into components that can be learnt separately, and *approximations* of practice, namely opportunities to practise components of professional work under simplified, low-stakes conditions (Grossman et al., 2009). We regard this distinction as essential to our inquiry because AI-assisted simulation is, by definition, an approximation of practice, and its value depends on how successfully it preserves the relevant complexity of the real situation while eliminating risks to children.

A further point must be added. Research on expertise has shown that superior performance is not the result of simply accumulating experience, but of *deliberate practice*: targeted practice of difficult components, with immediate feedback and the opportunity for repetition (Ericsson et al., 1993). Traditional teaching practice provides experience, but rarely

deliberate practice: a student cannot repeat the same conversation with the same parent three times, trying a different strategy each time.

2.2. Simulation in teacher education

The history of simulation in teacher education begins, in its modern form, with microteaching developed at Stanford University in the 1960s: short lessons delivered to a small group, recorded and subsequently analysed (Allen and Ryan, 1969). This was followed by clinical simulations, in which trained actors play the roles of parents or pupils (Dotger, 2013), and, more recently, by mixed-reality environments in which students interact with a class of avatars animated in real time by a human operator (Dieker et al., 2014). A review by David Kaufman and Alice Ireland highlighted the potential of simulations to provide repeated practice, in a safe environment, for skills that are otherwise difficult to acquire (Kaufman and Ireland, 2016).

Although heterogeneous, the empirical evidence generally supports this direction. The systematic review conducted by Hanneke Theelen, Antoine van den Beemt and Perry den Brok, based on 15 studies, shows that computer-based classroom simulations had positive effects on classroom management and teaching skills in general, while four of the studies also reported positive effects on future teachers' self-efficacy. Specific effects on interpersonal competence remained insufficiently investigated (Theelen et al., 2019). The meta-analysis conducted by Olga Chernikova and colleagues across 145 empirical studies in higher education, including teacher education, reports a large overall effect of simulation-based learning, $g = 0.85$ (Chernikova et al., 2020). More importantly for our argument, the same meta-analysis shows that the effect depends on the support provided: learners with substantial prior knowledge benefited more from reflection phases, whereas those with limited knowledge learned better when they received examples. In other words, simulation does not produce learning merely by existing, but through the pedagogical architecture surrounding it.

Along similar lines, the experimental study led by Julie Cohen, involving 105 randomly assigned future teachers, showed that *coaching* provided between mixed-reality simulation sessions produced significant and substantial improvements in skills compared with a condition in which students merely reflected independently on their own performance (Cohen et al., 2020). The finding is important: solitary reflection is not sufficient, and external, specific and timely feedback makes the difference. In health professions education, where simulation has a longer tradition, the meta-analysis led by David Cook showed large effects of technology-enhanced simulation compared with no intervention (Cook et al., 2011), although these conclusions cannot be transferred automatically to teacher education because pedagogical situations are less amenable to standardisation than clinical ones.

2.3. Feedback and professional confidence

If simulation provides the opportunity to practise, feedback is what transforms practice into learning. John Hattie and Helen Timperley showed that feedback is among the most powerful influences on learning, but that its effects vary considerably: it becomes effective when it answers three questions—Where am I going?, How am I going? and Where to next?—and when it addresses the task, the process and self-regulation rather than the person (Hattie and Timperley, 2007). General praise or global criticism, by contrast, has weak or even negative effects. As we shall see, this distinction underpinned the design of the evaluator used by the platform under analysis.

A second, often neglected dimension is professional confidence. Albert Bandura (1997) showed that self-efficacy is built primarily through mastery experiences, while Megan Tschannen-Moran and Anita Woolfolk Hoy adapted this concept to the teaching profession, demonstrating that teachers’ self-efficacy influences how they organise their work and persist when faced with difficulties (Tschannen-Moran and Woolfolk Hoy, 2001). From this perspective, simulation could serve a dual role: not only enabling the practice of techniques, but also building confidence before students encounter real situations. We emphasise that this remains, for the time being, a hypothesis requiring empirical verification.

2.4. Generative AI and simulated practice

Reviews of AI in higher education showed, even before the emergence of generative models, that the field was dominated by technical perspectives and that education specialists were poorly represented in both the design and evaluation of systems (Zawacki-Richter et al., 2019). Wayne Holmes, Maya Bialik and Charles Fadel cautioned that the relevant question is not what technology can do, but what educational purpose it serves (Holmes et al., 2019). After 2022, Enkelejda Kasneci and colleagues systematised the opportunities offered by language models in education—personalisation, material generation and feedback—as well as their challenges: accuracy, bias, user dependency, academic integrity and data protection (Kasneci et al., 2023).

The use of LLMs as simulation partners is a very recent direction. Julia Markel and colleagues developed GPTeach, a tool in which future teaching assistants practise interactions with students simulated by a GPT model; the tool was subsequently used in the education of more than 800 novice teaching staff (Markel et al., 2023). Ethan Mollick and colleagues proposed a multi-agent AI architecture in which distinct roles—mentor, role-play partner and evaluator—work together to provide personalised instruction, practice and feedback, opening up, as the title of their paper suggests, the prospect of ‘simulated practice at scale’ (Mollick et al., 2024). In teacher education, Jieun Lim and colleagues reported improvements in future teachers’ problem-solving skills with both a rule-based simulation and one enhanced by generative AI, compared with scenarios presented exclusively in text format. They found no significant differences between the two types of simulation, but realism and motivation were perceived to be higher in the generative AI condition (Lim et al., 2025).

However, the same literature also identifies limitations. The case study conducted by Longwei Zheng and colleagues on the experiences of 12 future teachers in a classroom simulation with LLM-generated pupils identified, alongside perceived benefits, a series of barriers to authenticity that hinder what the authors call the ‘suspension of disbelief’—the temporary acceptance of fiction without which simulation loses its force (Zheng et al., 2025). In a systematic review of the literature on simulated learners, Tanja Käser and Giora Alexandron show that, in almost half of the studies analysed, their validity—the extent to which they behave like real learners—is unsupported by any evidence (Käser and Alexandron, 2024). These concerns are compounded by the general risks of language models: their tendency to produce plausible but false content (Ji et al., 2023), their reproduction of biases present in training data (Bender et al., 2021), and, when models are used as evaluators, a series of documented biases—a preference for longer responses, sensitivity to presentation order and a tendency to favour their own formulations—even though their overall agreement with human evaluators may be high (Zheng et al., 2023). A widely used strategy for limiting fabricated information is *retrieval-augmented generation*, in which the model formulates its answer on the basis of passages retrieved from a controlled collection of documents (Lewis et al., 2020).

2.5. The regulatory and ethical framework

In Romania, the Profile and Professional Standards for Teachers, approved by Order of the Minister of Education No. 7,386/2024, organise teachers' competences into seven domains: (1) knowledge of pupils and how they learn; (2) curriculum and teaching; (3) learning environment and well-being; (4) pupil guidance and collaboration with families/guardians and the community; (5) ethics, reflexivity and continuous learning; (6) institutional development; and (7) digital technology in education. The standards are differentiated by career stage and level of education and provide a reference point for both initial and continuing teacher education.

At European level, Regulation (EU) 2024/1689 on artificial intelligence classifies, among others, systems intended to evaluate learning outcomes in education and vocational training institutions as high-risk systems (Annex III, point 3), with direct consequences for any tool that assigns scores to students. UNESCO's guidance on generative AI in education and research (Miao and Holmes, 2023) and the European Commission's ethical guidelines for educators (European Commission, 2022) likewise emphasise human agency, transparency and teacher oversight.

Three findings emerge from the literature review. First, simulation is a validated form of approximation of practice, but its effectiveness depends on the pedagogical support accompanying it—examples, feedback, reflection and coaching. Second, generative AI can substantially reduce the cost of simulation, but the authenticity of simulated characters and the reliability of automated evaluation remain open questions. Third, analyses linking these tools to national professional standards and to the realities of initial teacher education in Romania are almost entirely absent. This is the gap addressed by our inquiry.

3. Methodology

3.1. Research design

The research methodology was qualitative and analytical-constructive. We chose an instrumental case study in the sense used by Robert Stake (1995): the case, the Practicum platform, is not studied for its own sake, but in order to understand a broader phenomenon, namely the potential and limitations of generative AI in the simulated practice of future teachers. Because the authors are also the platform's designers, the inquiry also follows the logic of *design science research*, in which an artefact is constructed to address an identified problem and is then described, theoretically justified and critically evaluated (Hevner et al., 2004). We clarify at the outset that the paper does not report data collected from users; the empirical evaluation of effects on learning is the subject of a subsequent stage, for which we propose a protocol in Section 4.7.

3.2. The case analysed

Practicum is a web application intended for students enrolled in psycho-pedagogical teacher education programmes (DPPD, Levels I and II) and those in the Primary and Pre-school Education programme (PIPP) at Constantin Brâncuși University of Târgu Jiu. The platform brings together five components: (a) a role-play simulation module, in which the language model portrays the characters in a professional situation while the student plays the role of the teacher; (b) an evaluator that produces a structured report at the end of each simulation; (c) a tutor that answers questions exclusively on the basis of course materials uploaded by the institution; (d) a reflective teaching practice journal accompanied by formative feedback; and (e) supplementary tools for lesson planning and self-evaluation. Access is differentiated between students and

administrators, with the latter able to upload course materials, deactivate certain functions and, most importantly, manage the simulation catalogue.

The scenario catalogue is not closed. Teacher educators with administrator privileges can create any number of new scenarios directly through the platform interface, without any changes to the code, specifying for each one the category, initial context, age level or class, number of stages (between two and six), assessment criteria (between two and six) and the study programmes for which it is intended. A new scenario can be published immediately or kept as a draft until it has been validated. The platform is currently in the testing phase, which is why the catalogue presently contains four scenarios in each category.

From a technical perspective, the platform accesses a commercial language model through the provider’s application programming interface (in the configuration analysed, GPT-4o mini, developed by OpenAI), while searches across course materials are conducted using vector representations of the text. The outputs that populate the interface—assessment reports, feedback on journal entries and tests—are requested from the model in a structured format and automatically validated before being displayed. We do not elaborate on these details because they are not the object of the analysis; we mention them only because they explain how the pedagogical constraints were translated into the operation of the system.

3.3. Corpus and analysis procedures

The corpus analysed comprised: (1) the platform’s initial catalogue of 20 scenarios developed for the testing phase, including the description of the initial situation, level of education, number of stages and associated assessment criteria (60 criteria in total); (2) the system instructions that define the model’s behaviour in the four roles it performs; (3) the structure of the assessment reports and reflective feedback; and (4) the text of the Profile and Professional Standards for Teachers, approved by Order No. 7,386/2024. Because the catalogue can be extended at any time, we did not treat it as a definitive inventory, but as a sample illustrating the tool’s design logic and the type of practice it enables.

We used four complementary methods. Through *content analysis* of the catalogue, each scenario was coded according to the competence domain in the professional standards, the level of education, the type of interlocutor and the number of stages; the 60 assessment criteria were then grouped inductively into thematic categories through constant comparison. Through *document analysis*, we examined the system instructions, identifying the pedagogical constraints and safety measures they impose on the model. The *comparative method* enabled us to relate the adopted solutions to traditional forms of simulation and to the tools described in the international literature. Finally, through *theoretical synthesis*, we integrated the findings into a model of AI-mediated simulated practice.

3.4. Ethical considerations and reflexivity

The research did not involve human participants, and no real student conversations were used in the analysis; consequently, ethics committee approval was not required. We are aware that our dual position as researchers and designers of the tool may encourage an overly favourable interpretation. To limit this risk, we reported the catalogue data in a quantitative, verifiable form, made the criteria for analysis explicit, and devoted at least as much space to risks and limitations as to potential benefits.

4. Results and discussion

4.1. What does the student practise? Analysis of the scenario catalogue

The first finding concerns coverage. The 20 scenarios in the initial catalogue are distributed equally—four in each of the platform’s five categories—with each category corresponding to one of the first five competence domains in the professional standards (Table 1). The domains of ‘institutional development’ and ‘digital technology in education’ are not represented. We note that, unlike the number of scenarios, which is unlimited, the categories are fixed within the platform architecture; this delimitation therefore does not reflect the testing phase, but a design decision: both domains concern individual or organisational activities rather than interpersonal interactions that can be practised through role-play. The category names used on the platform are adapted for students (for example, ‘Knowledge of pupils and diagnosis’ or ‘Learning environment and classroom management’), but they correspond directly to the official domains.

Table 1. Distribution of scenarios by competence domain and level of education

Competence domain (Order of the Minister of Education No. 7,386/2024)	Scenarios	Pre-school	Primary	Professional relationships	Student interventions
D1. Knowledge of pupils and how they learn	4	1	3	–	13
D2. Curriculum and teaching	4	1	3	–	15
D3. Learning environment and well-being	4	2	2	–	15
D4. Pupil guidance and collaboration with families/guardians and the community	4	1	2	1	14
D5. Ethics, reflexivity and continuous learning	4	–	–	4	14
Total	20	5	10	5	71

Note: ‘Professional relationships’ refers to scenarios involving interactions between adults (colleagues, teaching practice mentors), without reference to a particular group or class. The final column totals the number of stages in the scenarios, with each stage corresponding to one student intervention. Source: the author’s analysis of the platform’s initial catalogue (testing phase).

In terms of level of education, half of the scenarios (10) take place in primary education, one quarter (5) in pre-school education, and the remaining five concern professional relationships with colleagues and the teaching practice mentor. In 12 scenarios (60%), the simulated interlocutors are children, while in 8 (40%) they are adults (parents, colleagues or mentors). A simulation involves three or four student interventions (3.55 on average). The short duration is a deliberate choice: the scenarios in the initial catalogue are designed as brief exercises that can be repeated, although teacher educators can also create more extensive simulations of up to six stages.

The second finding, and perhaps the most relevant to the discussion of AI’s potential, concerns what is assessed. The 60 criteria associated with the scenarios correspond to 35 distinct formulations, which we grouped into four categories (Table 2).

Table 2. Categories of assessment criteria associated with the scenarios

Category	Distinct criteria	Occurrences	% of occurrences	Examples of criteria
Cognitive and didactic	9	20	33.3	Clarity of explanation; Correct diagnosis of the error; Checking understanding
Socio-emotional (relationship with children)	10	16	26.7	Empathy; Conflict management; Constructive firmness
Professional communication with adults	8	14	23.3	Empathic communication; Professional clarity; Managing tension
Ethical and reflective	8	10	16.7	Authentic reflection; Ethical conduct; Professional integrity
Total	35	60	100.0	

Source: the content analysis conducted by the author on the initial catalogue; inductive coding through constant comparison.

We find that two thirds of the criterion occurrences (66.7%) concern the relational, emotional and ethical dimensions of the profession, while only one third concern the cognitive and didactic dimension in a narrow sense. We consider this finding not to be a mere accident of design, but the expression of a particular logic: these are precisely the dimensions that are most difficult to practise through traditional forms of university-based education. A written case study can ask a student to describe what they would do when faced with an angry parent, but it cannot ask them to do it under the pressure of an immediate response that changes depending on the tone they use. This, we believe, is the core of AI's potential in simulated practice: not the transmission of knowledge, which other tools can also provide, but the practice of professional conversation.

The analysis also reveals a weakness. Of the 35 distinct criteria, 21 (60%) appear in only one scenario, and some formulations are semantically similar ('Ethical conduct', 'Professional conduct' and 'Professional integrity'; 'Managing tension' and 'Managing tension within the group'; 'Empathy' and 'Empathic communication'). For a student who completes a single scenario, this diversity is not a problem. For monitoring progress over time, however, it becomes an obstacle because scores obtained in different scenarios are not comparable. The issue becomes increasingly important because the catalogue is open-ended: each teacher educator who creates a new scenario formulates the criteria freely, so their dispersion will increase as the catalogue expands. The resulting recommendation is to consolidate the criteria into a common rubric linked directly to the descriptors in the professional standards, from which teacher educators would select criteria for new scenarios rather than formulating them each time. In this way, every simulation would contribute to a coherent competence profile.

4.2. The four roles of AI and the architecture of constraints

Analysis of the system instructions shows that the same language model is used in four distinct roles, each with a pedagogical function and its own set of constraints (Table 3). We observe that designing each role involved specifying what the model is *not* permitted to do rather than what it must do.

Table 3. Roles assigned to the language model, their functions and constraints

AI role	Pedagogical function	Main constraints	Theoretical foundations
Role-play partner	Approximation of practice; repeated practice	Remains in role throughout; does not offer advice or evaluate during the scene; does not resolve the situation for the student; characters change only if the intervention justifies it; age-appropriate language	Grossman et al. (2009); Ericsson et al. (1993)
Formative evaluator	Immediate, criteria-based feedback	Evaluates only the student's responses; relies on observable evidence; score of 1–5 for each criterion and overall score of 0–100; strengths and specific recommendations	Hattie and Timperley (2007); Cohen et al. (2020)
Source-grounded tutor	Activation of theoretical knowledge	Answers only from the institution's materials; cites the source; explicitly states when information is unavailable	Shulman (1987); Lewis et al. (2020)
Reflective mentor	Reflection-on-action	Separates facts from interpretations; asks probing questions; does not award marks or make diagnoses; refers to institutional procedures in situations involving risk to children	Schön (1983); Korthagen and Kessels (1999)

Source: documentary analysis of the platform's system instructions.

The role-play partner is the most delicate of the four roles. The instructions require the model to portray only the characters in the scenario (children, parents, colleagues or mentors), not to utter or anticipate the teacher's responses and, above all, not to resolve the situation for the student: the characters calm down, cooperate or understand only to the extent that the student's intervention realistically justifies such a change. Each category has additional instructions. In diagnostic scenarios, the simulated child uses reasoning that is incorrect but plausible for their age and does not discover the error independently. In microteaching scenarios, the class reacts in varied ways (some pupils follow the explanation, another loses concentration, another asks an unexpected question). In ethics scenarios, the simulated mentor or colleague maintains their initial position and moderates their tone only if the student's responses demonstrate authentic reflection. The simulation unfolds in stages: the model opens the scene, responds to each intervention and brings it to a natural conclusion in the final stage, without introducing new problems.

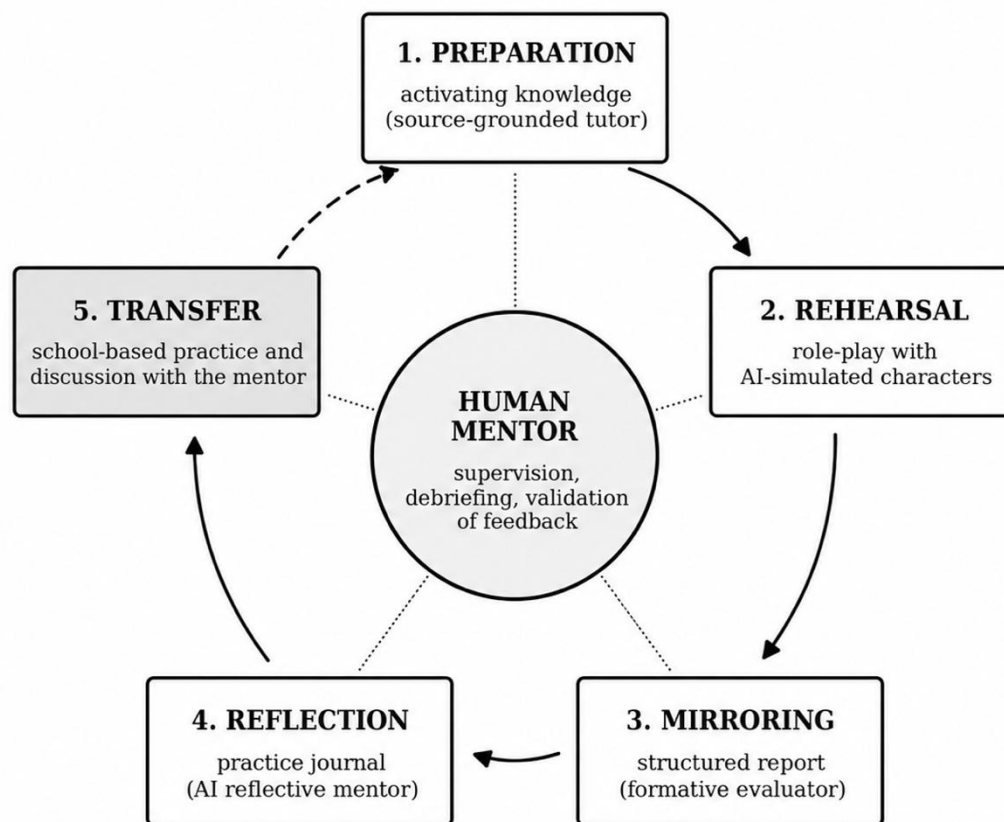
We consider this ‘pedagogy of restraint’ to be the fundamental condition for formative value. Left unconstrained, a language model tends to be cooperative, conciliatory and explanatory—the very opposite of an upset six-year-old child or a defensive parent. If the character gives way too quickly, the student learns that any intervention works; if the character begins to offer advice, the simulation becomes a lesson and the student once again becomes a listener.

The evaluator intervenes only after the scene has ended and receives the scenario criteria in addition to the transcript. The instructions require it to evaluate only the student's responses, rely on observable evidence, refrain from inferring intentions, and maintain consistency between justifications and scores. Recommendations must be specific, and vague formulations such as ‘be better prepared’ are explicitly excluded. The report structure operationalises almost literally the three questions proposed by Hattie and Timperley (2007): scores answer ‘How am I going?’, the justifications explain why, and the recommendations indicate the next step. If the model fails to produce a valid report, the system displays a fallback report, explicitly marked as indicative, rather than an invented result.

The tutor and reflective mentor complete the cycle. The former answers questions exclusively on the basis of passages retrieved from the materials uploaded by the institution, identifies the source document and must state when the information is absent. The latter analyses the student’s teaching practice journal (context, observed evidence, successes, principal difficulty and next step) and responds with probing questions and specific steps, without turning feedback into a mark. The instructions prohibit it from making psychological diagnoses or labelling children and, if the text suggests a risk to a child’s safety, require it to recommend following the institution’s procedures and discussing the matter with the mentor immediately.

4.3. A model of AI-mediated simulated practice

Building on the findings above and the literature reviewed, we propose a five-stage model of AI-mediated simulated practice (Figure 1). In addition to the three components described by Grossman et al. (2009)—representation, decomposition and approximation of practice—our model adds two stages that, in our view, are decisive if simulation is to become learning: mirroring and transfer.



solid line: the course of a simulation; dashed line: restarting the cycle after real-world practice

Figure 1. Model of AI-mediated simulated practice: five stages and the central role of the human mentor (developed by the author)

- (1) **Preparation:** the student activates relevant theoretical knowledge (for example, typical errors in subtraction with borrowing or communication with parents), supported by the tutor grounded in the course materials. This stage responds to Chernikova et al.'s (2020) finding that students with limited prior knowledge benefit more from examples and support before practice.
- (2) **Rehearsal:** the student assumes the role of the teacher and interacts with the simulated characters under low-stakes conditions, with the opportunity to repeat the scenario.
- (3) **Mirroring:** the evaluator presents the student with a structured image of their own intervention: what they did, what effect it had and what they could change. We chose the term *mirroring* to emphasise that the report is not a verdict, but a reflection to be interpreted.
- (4) **Reflection:** the student reinterprets the experience in the teaching practice journal, while the reflective mentor challenges them through questions; this is the point at which, in Schön's (1983) terms, reflection-on-action prepares future reflection-in-action.
- (5) **Transfer:** the simulated experience is compared with real practice in school and discussed with the human mentor, who validates, nuances or contradicts the automatically generated feedback. At the same time, real situations encountered during teaching practice can be transformed by teacher educators into new scenarios that restart the cycle for other students.

The stages depend on one another, and bypassing any of them diminishes the value of the whole. Rehearsal without mirroring becomes mere repetition. Mirroring without reflection remains a score that the student accepts or rejects without understanding it, while reflection without transfer risks turning the simulation into a self-contained game. We placed the human mentor at the centre of the model not to comply with convention, but because the findings of Cohen et al. (2020) show that coaching provided by a teacher educator between simulation sessions produces effects that solitary reflection does not.

To illustrate how the model works, let us return to the thought experiment in the introduction. Our student selects the scenario in which a child in the preparatory class consistently writes 'b' instead of 'd'. Before beginning, the student asks the tutor what the course materials say about letter confusions at the onset of writing. They then enter the simulation: the child answers honestly, according to their six-year-old logic, without naming the error. If the student corrects the child directly, the child may become unsure; if the student asks how the child reasoned, the child explains their thinking. At the end, the report may show, for example, that the diagnosis was correct but that the explanation was too abstract for the child's age. In the journal, the student notes what they would change, while the reflective mentor's questions require them to distinguish between what they observed and what they inferred. The following week, during teaching practice in school, they discuss a similar real case with the teacher mentor. Nothing in this process replaces the real child; each stage does, however, make the student better prepared to encounter that child.

4.4. What does AI add? Dimensions of its potential

Returning to the first research question, we consider that the potential of AI in simulated practice can be described through five dimensions.

Repeatability without victims. The student can afford to make mistakes, repeat the exercise and try another strategy—precisely the kind of deliberate practice (Ericsson et al., 1993)

that school-based teaching practice cannot provide. A mistake that has a cost for a child in the real classroom becomes a source of learning in the simulation.

Equal access to rare situations. Simulation transforms the lottery of situations into a curriculum of situations. Conflict, a challenged grade or an ethical dilemma no longer depends on the luck of a placement, but can be practised by every student. In our view, this represents equity in teacher education. Moreover, the curriculum of situations is not fixed once and for all: it can grow with the institution’s experience through scenarios added by teacher educators.

Adaptability. Unlike a written case study or a traditional branching scenario, which offers a finite number of pathways, the generated character responds to the student’s specific wording. Two runs through the same scenario are never identical, preventing the memorisation of a ‘correct answer’ and requiring adaptation.

Immediacy of feedback. The gap between action and feedback, often measured in days or weeks in real-world practice, is reduced to a few seconds, and feedback is available at any time, including outside seminar hours.

The record of practice. Each simulation leaves behind a transcript, a report and, potentially, a reflection. These can become materials for discussion with the mentor and for the teaching practice portfolio, giving the teacher educator access to the student’s thinking process rather than only to the final product.

We add a sixth dimension, with the status of a hypothesis: *building confidence*. If self-efficacy develops through mastery experiences (Bandura, 1997), then a series of successful simulation experiences before the real encounter might reduce the anxiety associated with beginning practice. We emphasise, however, that we do not yet have data to confirm this hypothesis.

4.5. Limitations of the potential

The potential described above cannot be separated from its limitations. We have grouped them into three paradoxes to emphasise that, in each case, the same feature of the technology is simultaneously a source of value and risk.

The authenticity paradox. The more realistic the simulated character, the greater the formative value of the simulation, but also the greater the risk of producing inappropriate content. At the same time, language models tend to be coherent, cooperative and articulate, whereas real children are not always so. The barriers to authenticity identified by Zheng et al. (2025) and the difficulty of validating simulated learners highlighted by Käser and Alexandron (2024) show that the problem is not specific to the platform analysed, but applies to the entire category of tools. Instructions preventing characters from giving way too quickly reduce this risk, but do not eliminate it.

The evaluation paradox. The automated evaluator provides what a human mentor can rarely provide—immediate feedback for every student and every attempt—but its reliability has yet to be established. The documented biases of models used as evaluators (Zheng et al., 2023) have particular significance in a pedagogical context: a preference for longer answers could favour lengthy interventions, even though a brief and clear intervention is often the most appropriate when addressing a five-year-old child. We therefore consider that AI-generated scores should be treated as hypotheses for discussion rather than measurements, at least until they have been calibrated against human evaluators.

The accessibility paradox. The ease with which students can turn to AI is simultaneously its main advantage and its main danger. A permanently available assistant can encourage the

delegation of professional judgement: a student may be tempted to request a ready-made lesson plan or to ‘learn’ the formulations that produce high scores instead of developing their own reasoning. The platform analysed limits this risk—the characters do not offer solutions, and the automatic generation of lesson plans can be deactivated by the administrator—but the problem remains essentially one of pedagogical culture rather than technical configuration.

These paradoxes are compounded by legal and ethical constraints. If assessment reports were used to award marks or determine whether a student had passed teaching practice, the system would very likely fall within the provisions on high-risk AI systems in Regulation (EU) 2024/1689, with the corresponding obligations regarding documentation, transparency and human oversight. Moreover, the wording of the annex, which concerns the evaluation of learning outcomes even when used to steer the learning process, suggests that strictly formative use is not automatically excluded from its scope. A dedicated legal analysis is therefore necessary. Even in a formative setting, students’ texts are transmitted to an external provider, which requires users to be informed, compliance with data protection legislation, and the avoidance of any personal data relating to children in placement schools. Finally, the quality of the Romanian language produced by the model depends on the linguistic resources on which it was trained, while the gender, social or cultural biases that the characters might reproduce (Bender et al., 2021) have not yet been systematically tested.

4.6. The role of the mentor and implications for initial teacher education

The discussion above supports a conclusion that we regard as central: AI does not abolish the mentor’s role, but changes their place in the teacher education cycle. In the traditional model, the mentor was the source of the first response to the student’s intervention; in the proposed model, the mentor becomes the interpreter and validator of that response, discussing the automatically generated report, challenging it where appropriate and relating it to observations from the real classroom. This is a shift not in status but in function, and the function cannot be automated because it requires knowledge of the context, the student and the real children. A new role is also added: that of scenario author. The platform’s capacity for unlimited simulation creation turns teacher educators from users of a tool into designers of their own curriculum of situations: a conflict observed during a placement, a difficult conversation reported by a teacher mentor or an ethical dilemma arising in school can become, within minutes, exercises for future generations of students. The value of this openness nevertheless depends on the quality of the scenarios, which requires a minimum set of common rules for formulating the context and assessment criteria.

For initial teacher education programmes—whether DPPD programmes or the teaching master’s programme introduced by Law No. 199/2023—the implications are practical. First, simulations should be integrated into teaching practice seminars and followed by *debriefing* discussions, rather than being left to students as an independent and unsupervised activity. Second, assessment criteria should be explicitly aligned with the descriptors in the professional standards so that simulation contributes to a competence profile that is comparable over time. Third, teacher educators themselves need preparation to interpret AI-generated reports critically and to teach students to do the same (Miao and Holmes, 2023), a competence that, moreover, corresponds to the ‘digital technology in education’ domain of the professional standards.

4.7. An agenda for empirical validation

At present, the potential described in this paper has the status of a theoretically grounded hypothesis. Its validation requires, in our view, at least four lines of research. The first is a quasi-

experimental study comparing a group of students who complete a fixed number of simulations and reflections over one semester with a group undertaking the usual teaching practice seminar activities. Effects should be measured using self-efficacy scales administered before and after the intervention (Tschannen-Moran and Woolfolk Hoy, 2001) and a standardised performance task independently assessed by two human evaluators. The second line of research is evaluator calibration: comparing AI-generated scores with those assigned by experienced teacher educators to the same set of transcripts and estimating agreement between them. The third concerns authenticity: the extent to which students and experienced teachers perceive the simulated characters as plausible for their age and context. The fourth, no less important, concerns safety: analysing a sample of anonymised transcripts to identify inappropriate content, stereotypes and factual errors.

5. Research limitations

Although our inquiry has, we believe, several strengths—it is among the first analyses of its kind in Romania and is based on a tool that has actually been built—it also has a number of limitations that must be acknowledged.

The first and most important limitation is the absence of empirical data on effects on learning. The paper describes and analyses a tool and formulates a model, but does not demonstrate that its use leads to better professional competences. The second limitation arises from the author’s dual role as researcher and designer of the platform. Although we have attempted to mitigate this risk by reporting verifiable data and explicitly analysing the risks, independent validation is necessary.

The third limitation concerns the scenario catalogue analysed. The platform is in the testing phase, and the 20 scenarios constitute an initial catalogue focused predominantly on pre-school and primary education. Although the architecture allows any number of new scenarios to be added, including scenarios for DPPD students in legal, economic, technical or medical disciplines who will teach in lower-secondary, upper-secondary or vocational education, the content analysis findings describe the current state of the catalogue and will change as it expands. The fourth limitation is technological: the platform’s behaviour depends on a commercial language model, updates to which may alter results without the user being informed, so strict reproducibility of interactions is not guaranteed.

Finally, the interaction is exclusively text-based. Tone of voice, eye contact, gesture, the noise of a class of 25 children and time pressure—defining elements of a real pedagogical situation—are absent from the simulation, which limits transfer, particularly in classroom management.

6. Conclusions

Having reached the end of our inquiry, after analysing the literature and a tool built for the simulated practice of future teachers, we are now in a position to answer the questions formulated in the introduction.

What does generative AI add to simulated practice? Our answer is that, first and foremost, it provides a professional conversation partner that responds to what the student says, is available at any time and costs incomparably less than actors or mixed-reality environments. Analysis of the initial scenario catalogue showed that two thirds of the assessment criteria address the relational, emotional and ethical dimensions of the profession—the very components that conventional university-based education can describe but cannot provide opportunities to

practise. Repeatability without victims, equal access to rare situations, adaptability, immediacy of feedback and the record of practice are, in our view, the five dimensions of this potential.

Under what conditions does this potential become professional learning? Only when simulation is integrated into a complete cycle—preparation, rehearsal, mirroring, reflection and transfer—in which the human mentor remains indispensable, not only as an interpreter of feedback, but also as an author of scenarios capable of transforming real situations from schools into exercises for future generations of students. The analysis showed that the formative value of AI depends less on what the model can do and more on what it is prohibited from doing: resolving the situation for the student, offering solutions during the scene, awarding marks or inventing sources.

What are the limitations? The authenticity of the characters, the reliability of automated evaluation and the risk of delegating professional judgement are the principal pedagogical limitations; the legal classification of automated evaluation, data protection and potential model biases are the principal ethical and legal limitations. None of these, we believe, is a reason to reject the use of AI in teacher education, but each is a reason to introduce it cautiously and under supervision.

In conclusion, we consider that generative AI can become a genuine practice partner for future teachers, provided that it remains what it is: a partner, not a substitute. It cannot replace the real child, the real classroom or the mentor who knows the child, the class and the student alike. It can, however, ensure that when a student stands before a class or a dissatisfied parent for the first time, they do not do so without ever having practised. For a profession in which children often bear the cost of beginners' mistakes, this is by no means insignificant.

Declarations

Use of artificial intelligence tools. The platform analysed uses generative AI models, which are the object of the study. An AI-based assistant (Claude, Anthropic) was used in preparing the manuscript to analyse the platform's source code. The final text is the authors' own work, and they accept full responsibility for its content.

Conflict of interest. Flavius Cristian Mărcău is the designer of the platform analysed, and the implications of this position are discussed in Sections 3.4 and 5.

REFERENCES

- Allen, D. W., Ryan, K., *Microteaching*, Addison-Wesley, Reading, MA, 1969.
- Bandura, A., *Self-efficacy: The exercise of control*, W. H. Freeman, New York, 1997.
- Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S., On the dangers of stochastic parrots: Can language models be too big?, *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, ACM, New York, pp. 610–623, 2021. <https://doi.org/10.1145/3442188.3445922>
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., Fischer, F., *Simulation-based learning in higher education: A meta-analysis*, *Review of Educational Research*, 90(4), pp. 499–541, 2020. <https://doi.org/10.3102/0034654320933544>
- Cohen, J., Wong, V., Krishnamachari, A., Berlin, R., *Teacher coaching in a simulated environment*, *Educational Evaluation and Policy Analysis*, 42(2), pp. 208–231, 2020. <https://doi.org/10.3102/0162373720906217>
- European Commission, Directorate-General for Education, Youth, Sport and Culture, *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*, Publications Office of the European Union, Luxembourg, 2022. <https://doi.org/10.2766/153756>
- Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., Hamstra, S. J., *Technology-enhanced simulation for health professions education: A systematic review and meta-analysis*, *JAMA*, 306(9), pp. 978–988, 2011. <https://doi.org/10.1001/jama.2011.1234>
- Darling-Hammond, L., *Constructing 21st-century teacher education*, *Journal of Teacher Education*, 57(3), pp. 300–314, 2006. <https://doi.org/10.1177/0022487105285962>
- Dieker, L. A., Rodriguez, J. A., Lignugaris/Kraft, B., Hynes, M. C., Hughes, C. E., *The potential of simulated environments in teacher education: Current and future possibilities*, *Teacher Education and Special Education*, 37(1), pp. 21–33, 2014. <https://doi.org/10.1177/0888406413512683>
- Dotger, B. H., *“I had no idea”: Clinical simulations for teacher development*, Information Age Publishing, Charlotte, NC, 2013.
- Ericsson, K. A., Krampe, R. T., Tesch-Römer, C., *The role of deliberate practice in the acquisition of expert performance*, *Psychological Review*, 100(3), pp. 363–406, 1993. <https://doi.org/10.1037/0033-295X.100.3.363>
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., Williamson, P. W., *Teaching practice: A cross-professional perspective*, *Teachers College Record*, 111(9), pp. 2055–2100, 2009. <https://doi.org/10.1177/016146810911100905>
- Hattie, J., Timperley, H., *The power of feedback*, *Review of Educational Research*, 77(1), pp. 81–112, 2007. <https://doi.org/10.3102/003465430298487>
- Hevner, A. R., March, S. T., Park, J., Ram, S., *Design science in information systems research*, *MIS Quarterly*, 28(1), pp. 75–105, 2004. <https://doi.org/10.2307/25148625>
- Holmes, W., Bialik, M., Fadel, C., *Artificial intelligence in education: Promises and implications for teaching and learning*, Center for Curriculum Redesign, Boston, MA, 2019.
- Iucu, R. B., *Teacher education: Systems, policies, strategies*, Humanitas Educațional, Bucharest, 2004.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A., Fung, P., *Survey of hallucination in natural language generation*, *ACM Computing Surveys*, 55(12), art. 248, 2023. <https://doi.org/10.1145/3571730>
- Käser, T., Alexandron, G., *Simulated learners in educational technology: A systematic literature review and a Turing-like test*, *International Journal of Artificial Intelligence in Education*, 34(2), pp. 545–585, 2024. <https://doi.org/10.1007/s40593-023-00337-2>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al., *ChatGPT for good? On opportunities and challenges of large language models for education*, *Learning and Individual Differences*, 103, art. 102274, 2023. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kaufman, D., Ireland, A., *Enhancing teacher education with simulations*, *TechTrends*, 60(3), pp. 260–267, 2016. <https://doi.org/10.1007/s11528-016-0049-0>
- Kolb, D. A., *Experiential learning: Experience as the source of learning and development*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
- Korthagen, F. A. J., Kessels, J. P. A. M., *Linking theory and practice: Changing the pedagogy of teacher education*, *Educational Researcher*, 28(4), pp. 4–17, 1999. <https://doi.org/10.3102/0013189X028004004>
- Law No. 198/2023 on Pre-University Education, Official Gazette of Romania, Part I, No. 613 of 5 July 2023.

- Law No. 199/2023 on Higher Education, Official Gazette of Romania, Part I, No. 614 of 5 July 2023.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., et al., Retrieval-augmented generation for knowledge-intensive NLP tasks, *Advances in Neural Information Processing Systems*, 33, pp. 9459–9474, 2020.
- Lim, J., Lee, U., Koh, J., Jeong, Y., Lee, Y., Byun, G., Jung, H., Jang, Y., Lee, S., Moon, J., Development and implementation of a generative artificial intelligence-enhanced simulation to enhance problem-solving skills for pre-service teachers, *Computers & Education*, 232, art. 105306, 2025. <https://doi.org/10.1016/j.compedu.2025.105306>
- Markel, J. M., Opferman, S. G., Landay, J. A., Piech, C., GPTeach: Interactive TA training with GPT-based students, *Proceedings of the Tenth ACM Conference on Learning @ Scale*, ACM, New York, pp. 226–236, 2023. <https://doi.org/10.1145/3573051.3593393>
- Miao, F., Holmes, W., Guidance for generative AI in education and research, UNESCO, Paris, 2023. <https://doi.org/10.54675/EWZM9535>
- Ministry of Education, Order No. 7,386/2024 approving the Profile and Professional Standards for Teachers in Pre-University Education, by career stage and level of education, Official Gazette of Romania, Part I, No. 1158 bis of 20 November 2024.
- Mishra, P., Koehler, M. J., Technological pedagogical content knowledge: A framework for teacher knowledge, *Teachers College Record*, 108(6), pp. 1017–1054, 2006. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Mollick, E. R., Mollick, L., Bach, N., Ciccarelli, L. J., Przystanski, B., Ravipinto, D., AI agents and education: Simulated practice at scale, *arXiv:2407.12796*, 2024. <https://doi.org/10.48550/arXiv.2407.12796>
- European Parliament, Council of the European Union, Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), *Official Journal of the European Union*, L series, 12 July 2024.
- Schön, D. A., *The reflective practitioner: How professionals think in action*, Basic Books, New York, 1983.
- Shulman, L. S., Knowledge and teaching: Foundations of the new reform, *Harvard Educational Review*, 57(1), pp. 1–22, 1987. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Stake, R. E., *The art of case study research*, Sage, Thousand Oaks, CA, 1995.
- Theelen, H., van den Beemt, A., den Brok, P., Classroom simulations in teacher education to support preservice teachers’ interpersonal competence: A systematic literature review, *Computers & Education*, 129, pp. 14–26, 2019. <https://doi.org/10.1016/j.compedu.2018.10.015>
- Tschannen-Moran, M., Woolfolk Hoy, A., Teacher efficacy: Capturing an elusive construct, *Teaching and Teacher Education*, 17(7), pp. 783–805, 2001. [https://doi.org/10.1016/S0742-051X\(01\)00036-1](https://doi.org/10.1016/S0742-051X(01)00036-1)
- Zawacki-Richter, O., Marín, V. I., Bond, M., Gouverneur, F., Systematic review of research on artificial intelligence applications in higher education – where are the educators?, *International Journal of Educational Technology in Higher Education*, 16, art. 39, 2019. <https://doi.org/10.1186/s41239-019-0171-0>
- Zeichner, K., Rethinking the connections between campus courses and field experiences in college- and university-based teacher education, *Journal of Teacher Education*, 61(1–2), pp. 89–99, 2010. <https://doi.org/10.1177/0022487109347671>
- Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., et al., Judging LLM-as-a-judge with MT-Bench and Chatbot Arena, *Advances in Neural Information Processing Systems*, 36 (Datasets and Benchmarks Track), pp. 46595–46623, 2023.
- Zheng, L., Jiang, F., Gu, X., Li, Y., Wang, G., Zhang, H., Teaching via LLM-enhanced simulations: Authenticity and barriers to suspension of disbelief, *The Internet and Higher Education*, 65, art. 100990, 2025. <https://doi.org/10.1016/j.iheduc.2024.100990>