EUROPEAN JOURNAL OF OPEN, DISTANCE AND E-LEARNING

Human-Gen-AI Collaboration in Developing Situational Judgment Tests for Self-Regulated Problem Solving

RESEARCH

DAFNA AVIDOV
ORIT EZRA
GUY COHEN
ANAT COHEN
ALLA BRONSHTEIN

*Author affiliations can be found in the back matter of this article

ABSTRACT

Teachers play a crucial role in promoting self-regulated learning (SRL) and problem solving (PS) in classrooms, yet their knowledge in these domains remains understudied and challenging to evaluate. While traditional assessment tools like surveys exist, there is a need for diversification, particularly through situational judgment tests (SJT). This study leverages human-AI collaboration to develop an SJT tool for evaluating teachers' SRL-PS knowledge. Human interaction with generative artificial intelligence (Gen-AI) enables overcoming difficulties and mutually complementing each other. The purpose of this paper is to present an initial attempt to develop an SJT tool for evaluating teachers' SRL-PS knowledge using the assistance of ChatGPT and to examine how the researcher-ChatGPT interaction contributed to the development of this tool. Through systematic interaction between researchers and ChatGPT, we developed a comprehensive tool comprising 15 difficulty categories and 20 scenarios. The research demonstrates how scenarios created by researchers can be effectively complemented by ChatGPT-generated content. A synthesized map of difficulty categories and scenarios from both researchers and ChatGPT revealed patterns of over- and under-representation across categories, and identified opportunities for synthesized new difficulty categories that could facilitate new scenarios. The study advances SRL knowledge evaluation tools through SJT methodology while demonstrating the benefits of human-Gen-AI collaborative interactions.

CORRESPONDING AUTHOR:
Anat Cohen, Ph.D
School of Education, Tel Aviv University, Israel
anatco@tauex.tau.ac.il

KEYWORDS:
generative artificial intelligence; problem solving; situational judgment tests; self-regulated learning; human-machine interaction

TO CITE THIS ARTICLE:
Avidov, D., Ezra, O., Cohen, G., Cohen, A., & Bronshtein, A. (2025). Human-Gen-AI Collaboration in Developing Situational Judgment Tests for Self-Regulated Problem Solving. *European Journal of Open, Distance and E-Learning*, *27*(2): 5, 1–15.

1. INTRODUCTION

Self-regulated learning (SRL) is defined as a personal and active process in which learners acquire independent learning skills. During this process, they set goals for learning, choose learning strategies, and monitor and control their understanding, motivation, and behavior (Pintrich 2000; Zimmerman 1986). The promotion of SRL among learners is of great importance due to its connection to academic success, the development of effective learning habits, positive feelings towards learning, and high motivation for learning (Burns et al. 2018; Nota et al. 2004). It is also crucial in preparing learners for lifelong learning; therefore, it is important to cultivate these skills from the early stages of education (de Bruijn-Smolders et al. 2016). SRL encompasses cognitive, metacognitive, and motivational strategies (Dignath et al. 2008). Within the context of SRL, problem-solving (PS) strategies are essential cognitive skills for self-regulated learners (Assi & Cohen, 2023). To develop SRL-PS skills, student require appropriate support from their teachers in the classroom. Teachers who promote SRL-PS create learning environments that encourage students to be actively involved in their own learning process (Alvi & Gillies 2020; Dignath et al. 2008; Porter & Peters-Burton 2021) and provide opportunities for developing SRL-PS skills (Dignath-van Ewijk & Van der Werf 2012).

It is therefore crucial to provide teacher training that enhances educators' abilities in promoting SRL-PS among students (Cleary et al. 2022; Karlen et al. 2020; Porter & Peters-Burton 2021). However, there is a lack of clarity regarding what should be included in this SRL-PS knowledge component, and consequently, the method for evaluating this aspect of teachers' expertise remains ambiguous (Chen & Jang 2019). In the emerging field of SRL-PS competency evaluation, teachers are assessed using basic tools such as questionnaires, interviews, and observations (Dignath et al. 2008; Kistner et al. 2010). However, these methods are not sufficient, and there is a need for a more diverse range of evaluation approaches (Vosniadou et al. 2020). Situational Judgment Tests (SJTs) are used to assess decision-making abilities in various fields, such as medicine and military frameworks (Lievens et al. 2008; Patterson et al. 2016; Weekley & Jones 1999) and have been found to be reliable in these contexts. While studies have explored the use of SJTs in teacher evaluation (Aldrup et al. 2020; Chao et al. 2020; Klassen et al. 2020), their application in the context of SRL-PS remains unexplored.

Generative artificial intelligence (Gen-AI) is increasingly pervasive across all areas of life (Pegrum 2024), including assessment in education (Nasution 2023; Owan et al. 2023). As an emerging field, Gen-AI-assisted evaluation has shown promise in optimizing complex development processes for assessment tools, saving time and human resources while yielding favorable results (Xia et al. 2024). In the context of SJTs, Gen-AI has been used to test response accuracy and consistency (Borchert et al. 2023; Sareen 2023). However, to the best of our knowledge, no studies have explored the integration of Gen-AI in developing SJTs specifically for SRL-PS assessment. Moreover, research on effective human-Gen-AI interaction in educational settings, particularly in assessment, is still in its early stages (Drori & Te'eni 2024; Te'eni et al. 2023).

This study, based on an SRL-PS model implemented in a training program for K–12 in-service teachers, addresses calls to diversify assessment tools of teachers' SRL-PS knowledge beyond traditional methods. It presents an experience in developing a SJT tool using Gen-AI. It also examines the interaction between the researchers and Gen-AI in creating this evaluation tool. Thus, the potential benefits of collaboration between human and machine agents in the field of educational assessment are explored.

2. BACKGROUND

2.1 SELF-REGULATED LEARNING PROBLEM SOLVING (SRL-PS)

The development of self-regulated learning (SRL) among learners is known to be of great importance nowadays, due to its connection to academic success, effective learning habits, positive emotions and high motivation for learning (Burns et al. 2018; Li et al. 2023). SRL is also known to be of great importance in preparing learners for lifelong learning, and it is important to foster these skills from the early stages in education (de Bruijn-Smolders et al. 2016).

SRL is defined as an active process, during which the learner sets goals for learning, plans times, chooses learning strategies, monitors and controls understanding, motivation and behavior (Boekaerts 1999; Pintrich 2000; Zimmerman 2002). SRL is a circular process consisting of three stages: *The preparation stage,* in which students set goals and objectives and plan times; *the performance phase,* which involves monitoring and control processes carrying out the task; and *the evaluation phase,* in which reflection processes on the entire process take place (Boekaerts 1999; Jansen et al. 2017; Panadero 2017; Zimmerman 1998). SRL includes several key strategies. *Cognitive strategies* relate to handling information and are specific to various fields. *Metacognitive cognitive strategies* encompass planning, scheduling, monitoring, and evaluating learning. *Motivational strategies* involve learners' beliefs in their abilities and self-efficacy in light of failures and successes in previous tasks, as well as beliefs pertaining to values and interests in the task. These three types of strategies work together to support self-regulated learning. (Boekaerts 1999; Dignath et al. 2008; Pintrich 2000).

In the context of SRL, the problem-solving (PS) process was found to be an important cognitive skill in the 21st century by the OECD (OECD 2021). In the process of solving problems in learning activities, learners utilize a variety of cognitive and metacognitive strategies, such as time management, feedback, critical thinking, and the like, that help them overcome difficulties and solve problems. These PS strategies are cognitive skills that are required for self-regulated learners (Assi & Cohen, 2023).

2.2 TEACHER SRL-PS KNOWLEDGE EVALUATION

Teachers have a central role in promoting SRL in their classrooms. Teachers who promote SRL are increasing learners' responsibility and encouraging their participation in the learning process. They also help students develop cognitive, metacognitive, and motivational skills (Alvi & Gillies 2020; Dignath et al. 2008; Porter & Peters-Burton 2021). This is due to the fact that teachers largely control the learning environment through which they provide students with learning experiences that help students develop SRL skills and regulate their learning (Dignath-van Ewijk & Van der Werf 2012). In addition, teacher behavior can play a significant role in the development of students' SRL skills, especially in the early stages of SRL development when learners are primarily influenced by social supports (Porter & Peters-Burton 2021).

This fact requires a better understanding of which special abilities teachers should possess in the SRL domain, in order to develop SRL skills among their students (Karlen et al. 2020). Teachers often do not feel prepared or adequately supported to integrate SRL into their teaching (Cleary et al. 2022). Therefore, teachers need help to improve their abilities to better support students' SRL through professional development. Indeed, SRL is an important component of teachers' professional knowledge (Cleary et al. 2022; Karlen et al. 2020; Porter & Peters-Burton 2021). Despite this, teachers' required SRL knowledge as well as its evaluation remain unclear (Chen & Jang 2019). Furthermore, since very little is known on how teachers promote SRL in their classes, it is important to assess teachers SRL-PS knowledge (Dignath & Büttner 2018; Kistner et al. 2010).

There are a few basic ways to assess teachers' SRL-PS knowledge, such as using questionnaires or interviews to gather information about their SRL-promoting activities. These are usually self-reports that are collected before, after, or independently of a specific learning task (Dignath et al. 2008). Another common method of assessment is a systematic observation, to evaluate the teachers' teaching behavior during lessons (Kistner et al. 2010). Each of the tools has its advantages and limitations. There is a need for more tools to assess teachers' SRL-PS knowledge in a deeper and broader way (Vosniadou et al. 2020).

2.3 SJT EVALUATION TOOLS

Situational Judgment Tests (SJTs) are tools that can be used to characterize difficulties and scenarios across various situations and fields, serving both as a learning and evaluation tools (Chao et al., 2020). SJT tools have been used in military settings (Lievens et al. 2008), in retail organizations (Weekley & Jones 1999) and in the selection of medical candidates (Patterson et al. 2016) and aim to assess the decision-making ability of subjects. The subjects are presented with realistic or hypothetical scenarios related to their field of occupation and are required to choose or rate effective responses (Chao et al. 2020). Due to the lack of objective assessment tools designed to define the specific requirements for the teaching profession, it is difficult to empirically investigate what types of knowledge and skills are required for the teaching job in general and in SRL in particular (Aldrup et al. 2020; Gold & Holodynski 2015). Studies in recent years demonstrate the potential of SJT during admission procedures for elementary teacher training programs in middle school and in evaluating different practices in the teaching work (Aldrup et al. 2020; Chao et al. 2020; Klassen et al. 2020). The integration of SJT in evaluating SRL knowledge has not yet been explored.

Development of SJT tools is a long and complex process that follows three steps: The first step is job analysis, where scenarios of work situations are collected by experts in the field or from the literature. In the second step, experts or inexperienced workers create responses for each scenario in order to create a range of effective and ineffective responses, and in the last step, a coding key is developed for responses (Gold & Holodynski 2015; Lievens et al. 2008). In light of the stated above regarding the importance of promoting SRL-PS among learners and evaluating teachers' knowledge in SRL-PS using diverse tools, this study presents an attempt to develop an SJT tool for evaluating teachers' SRL-PS knowledge.

2.4 GEN-AI ASSISTED EVALUATION TOOLS

Artificial intelligence (AI) systems are taking more and more places in our lives in recent years. A significant breakthrough has occurred in the last few years with the developments of base models and large language models (LLM) as well as supporting conversational Gen-AI chatbots that provide the ability to chat with LLMs (Pegrum 2024). These models and chatbots allow the use of a variety of applications and have brought about changes in human-computer interaction. In Gen-AI, using prompts, content is created based on large sources of information that have been entered into the Gen-AI systems as part of its training. An example of such a system is ChatGPT, developed by OpenAI (European Parliamentary 2022).

The ubiquitous impact of Gen-AI throughout society has inevitably permeated the educational domain.

Gen-AI tools are now receiving a great deal of attention among those involved in education who deal with the implications of these tools on teaching, learning and assessment (Nasution 2023; Owan et al. 2023). The subject of Gen-AI assisted evaluation in education receives a lot of attention among researchers (Owan et al. 2023; Xia et al. 2024). A review of studies in the field of Gen-AI assisted evaluation in education spans different regions of the world and different disciplines, such as medicine, computer science, and teaching, the leading disciplines in Gen-AI assisted evaluation (Xia et al. 2024).

The development of evaluation tools with the help of Gen-AI can occur at different aspects of development, such as defining the objectives of the tool, identifying the key knowledge and skills to be assessed in the content area, creating a pool of potential test items, scoring the test items, and interpreting the test results (Owan et al. 2023). Previous studies that have used ChatGPT for developing standard multiple-choice tests indicate the development of a test which quality corresponds to experienced humans. Considering that building an evaluation tool is a complex and time-consuming process, development with the help of Gen-AI was found to be faster (Cheung et al. 2023; Ionescu & Enescu 2023; Lee et al. 2023; Nasution 2023).

Regarding SJT, studies in the medicine field used ChatGPT to respond to SJT tests (Borchert et al. 2023; Sareen 2023). In these studies, ChatGPT presented praiseworthy average accuracy response and consistent performance across domains. However, as best as we know, studies have not yet used ChatGPT to develop SJT tests. This paper presents a preliminary attempt to use ChatGPT to bridge the gap in academic research in the development of SJT to evaluate SRL-PS knowledge among teachers.

In light of the fact that the human-Gen-AI interaction is currently expanding, we must also recognize the risks that worry a wide variety of stakeholders (Fui-Hoon Nah et al. 2023). Risks in human-Gen-AI interaction manifest in over-reliance on Gen-AI that can lead to unsafe use and loss of social connection. Another risk is delegating essential decisions to Gen-AI, which may cause loss of skills and autonomy or feelings of disempowerment (Slattery et al. 2024). On the other hand, human-centered AI (HCAI) ensures avoidance of these risks. Delegating power to the various agents according to their capabilities and advantages is important (Drori & Te'eni 2024; Cohen et al. 2021; Te'eni et al. 2023). Thus, AI literacy has been coined to reflect an ability to collaborate with Gen-AI in an efficient and ethical manner with minimal risks (Fui-Hoon Nah et al. 2023). However, research on good practices for collaborating with Gen-AI in education, particularly in assessment, is still emerging (Drori & Te'eni 2024; Te'eni et al. 2023). In this study we examined the contribution of the collaboration between researchers and Gen-AI in the development of an SJT evaluation tool.

3. THE CURRENT RESEARCH

Given the importance of promoting SRL-PS and teachers' pivotal role therein, assessing teachers' SRL-PS knowledge post-professional development is crucial. This study, based on an SRL-PS model implemented in a training program for in-service teachers in Israel's K–12 system, addresses calls to diversify assessment tools of teachers' SRL-PS knowledge beyond traditional methods. Despite the educational benefits of SJT tools, a gap exists in SJT tools for evaluaing teachers' SRL-PS knowledge. Thus, this study developed a novel SJT tool for this purpose.

Considering the rise of consumer Gen-AI platforms and their potential in evaluation tool development, the complexities of creating SJT tools, and the nascent research on human-Gen-AI collaboration in educational assessment, this study aimed to develop an SJT tool for assessing teachers' SRL-PS knowledge using ChatGPT4. Additionally, the study sought to examine how the interaction between the researchers and ChatGPT contributed to the development of this tool. To this aim, the research questions were as follows:

1. What are the researchers' SJT scenarios for evaluating SRL-PS?
2. What are the ChatGPT's SJT scenarios for evaluating SRL-PS?
3. What is the contribution of the interaction between the researchers and ChatGPT to the development of an SJT tool for evaluating the SRL-PS knowledge of teachers?

4. RESEARCH METHODOLOGY

This study presents the development process of the SJT tool for evaluating teachers' SRL-PS knowledge created by the researchers and by ChatGPT4. It specifically examines the interaction between the researchers and Gen-AI and its contribution to the development of the tool. The development process was carried out in two successive stages. Initially, the tool was developed only by the researchers and then using ChatGPT4. Afterwards, an analysis comparing the developed outcomes by the researchers and ChatGPT4 was performed.

4.1 THE TEACHERS' TRAINING

The current study was conducted as part of a training program for in-service teachers held at the researches' university. The main goal was to provide the teachers with an educational program to promote SRL-PS. The

training instructed the teachers on a new SRL-PS model developed by the researchers. The SRL-PS model guides the learners along the PS prosses which is divided into four stages: problem identification; problem exploration; providing a solution; and solution evaluation. During these stages learners are instructed with a mix of SRL's cognitive, metacognitive, and behavioral strategies. The program also included a digital toolbox containing an online checklist of the PS stages enriched with supporting interactive videos (Assi & Cohen 2023; Cohen et al. 2022; Cohen & Cohen 2023). In the middle of the training, the teachers also started implementing the SRL-PS model in their classrooms.

4.2 PARTICIPANTS

The study involved ten in-service female teachers who participated in a training program. These teachers predominantly work in middle and high schools, with a minority teaching in elementary schools. They are teaching various subject areas, represent different sectors of society, and have teaching experience ranging from five to 21 years. The participants enrolled in the training program in response to an advertisement jointly issued by the Office of the Educational Chief Scientist and the university.

4.3 RESEARCH PROCEDURE

In this article, an attempt was made to harness Gen-AI for the enhancement of the SJT tool for evaluating teachers' SRL-PS knowledge while saving time and manpower resources. The research procedure was conducted in three sequential phases. The first two phases constituted the development processes of the researchers and Gen-AI of the SJT tool according to previously suggested procedures for developing SJT tools (Aldrup et al. 2020; Chao et al. 2020; Klassen et al. 2020). Thus, the *first phase* in the research procedure involved the development of the SJT tool by the research team based on teachers' experience. Subsequently, the *second phase* of the research procedure utilized ChatGPT4 for the SJT tool development. The first two phases in this development process facilitated the *third phase* in the research procedure, in which comparative analysis between the researcher-generated and Gen-AI-generated SJT products was enabled. Thus, we studied how human-machine interaction could potentially help improve the development of SJT assessment tools.

Phase one – the researchers' process

The SJT development process in the first phase was carried out by the researchers and took place as part of the teachers' training program aimed at promoting SRL-PS knowledge among teachers and their students. The process included six steps adjusted to this study based on previous procedural suggestions for developing SJT tools (Schubert et al. 2008):

- *Step 1—SRL-PS knowledge:* As is customary in SJT tools development, the first step involves defining the knowledge that should be evaluated by the tool. Accordingly, in this study, this step involved defining the SRL-PS knowledge that the teachers needed to have based on the SRL-PS model.
- *Step 2—SJT collection methods and categories design:* SJT tool development requires the design of diverse methods to collect information from experts regarding typical scenarios using wide spectrum of categories that relate to the knowledge defined in step one. Accordingly, in this study, this step describes building diverse methods to collect scenarios from the teachers who implemented the SRL-PS model in their classrooms. Specifically, the Scenarios Questionnaire was designed to collect scenarios from teachers based on their difficulties, representing a spectrum of difficulty categories relating to the SRL-PS knowledge, manifested through the SRL-PS model. Additionally, shared presentations and reflections methods were devised to enrich the collected scenarios pertaining to the SRL-PS model.
- *Step 3—teachers' training:* It is customary during SJT tool development to consult with experts in the field in order to cover the necessary breadth of knowledge and formulate the scenarios. Accordingly, in this study the teachers had to first be trained on the new SRL-PS model and implement it in their classrooms before they would become experts whose input could be used for drawing the scenarios. Therefore, this step involved training the teachers on the new SRL-PS model.
- *Step 4—scenario collection:* During SJT development, experts are required to come up with a wide scope of possible difficult scenarios to ideally cover the knowledge in the field. Accordingly, in this study this step was held during the training program after the teachers implemented the SRL-PS model in their classes. Difficulties regarding implementing the SRL-PS model in the classroom were collected using the methods built in step two. The researchers also added draft scenarios based on the SRL-PS model. Overall, 66 scenarios were identified during this step.
- *Step 5—scenario formulation:* SJT tool development recommends the formulation of scenarios according to tips and writing standards such as including adequate representation of names, gender, and social sectors (Aldrup et al. 2020; Chao et al. 2020; Gold & Holodynski 2015; Klassen et al. 2020). Accordingly, in this study this step involved the researchers formulating the pool of 66 scenarios.
- *Step 6—scenario selection:* A notable step in the development of SJT tools is the selection of the most relevant scenarios. In view of that, this study enacted this step in order to achieve optimal judgment to

select the top ten scenarios. Thus, blind inter-rater judgment was employed and followed by joint deliberations of two researchers. Each researcher defined separately the key difficulty and key SRL-PS knowledge in the original 66 scenarios. After that, the researchers held joint discussions and recursively selected scenarios to remove based on key difficulty and SRL-PS knowledge. That led to the selection of the ten most suitable scenarios for evaluating teachers' SRL-PS knowledge.

Phase two – the joint process: ChatGPT4 development

The researchers examined the potential points in the various steps of the SJT tool development process where Gen-AI could be integrated. It was decided that ChatGPT was relevant in all the steps except for step three. The researchers interacted with ChatGPT in the following way:

- *Step 1—SRL-PS knowledge:* The researchers examined ChatGPT's knowledge about SRL-PS by comparing the researcher's knowledge with ChatGPT's output. The researchers then uploaded the SRL-PS model as the main body of teachers' knowledge to be evaluated.
- *Step 2—SJT collection methods and categories:* The researchers asked ChatGPT to identify difficulty categories that could arise from the implementation of the SRL-PS model in a real classroom.
- *Step 4—scenario collection* and *step 5—scenario formulation:* The researchers asked ChatGPT to provide scenarios that reflect the difficulty categories and write them according to the SJT writing standards.
- *Step 6—scenario selection:* The researchers asked ChatGPT *to* select the top ten scenarios from the scenarios formulated in a previous step.

In order to interact with ChatGPT to perform these steps, gradual prompt strategies were employed in line with suggested guidelines for prompt engineering (Svendsen & Garvey 2023). Figure 1 shows the dialog with ChatGPT using the prompts. The first column shows the sequence number of the prompt, the second column the instruction given to ChatGPT, and the third column describes the purpose of the prompt.

Figure 2 illustrates the combined researchers and ChatGPT's processes for developing the SJT tool for evaluating teachers' SRL-PS knowledge. The six steps

01	*You are a researcher in the field of self-regulated learning and problem solving. you developed the model for solving problem, which is described in the attached table. can you please describe the model?*	Tell ChatGPT4 about SRL-PS model
02	*What about the time planning and management?*	
03	*In addition to all of that, we believe that the motivational and behavioral strategies of SRL should also be implemented in the model. Can you please list some motivational and behavioral strategies that are suitable?*	
04	*Now that you know our model, imagine that you are implementing it in real classrooms from 5th-8th grade. can you please describe the difficulties that may come up during the model implementation?*	Ask from ChatGPT4 to raise Difficulties in SRL-PS
05	*Thank you! now we want to share with you our aim.We want to develop a tool for evaluating teachers' knowledge of the model after they were trained on the model and implemented it in the class. We don't want to use the standard questionnaires and tests. We want to use the situational judgment tests (SJT). Are you familiar with the SJT method?*	Find out what ChatGPT 4 know about SJT
06	*Great! now we want you to write down SJT classroom scenarios that reflect the 8 difficulties you mentioned above, taking into consideration that teachers' answers to these scenarios point to teachers' knowledge of the model.*	Ask from ChatGPT 4 to raise SJT Scenarios
07	*Can you please give me four scenarios for each difficulty?*	
08	Can you please select the top ten scenarios?	

Figure 1 Dialog with ChatGPT using the gradual prompts strategies.

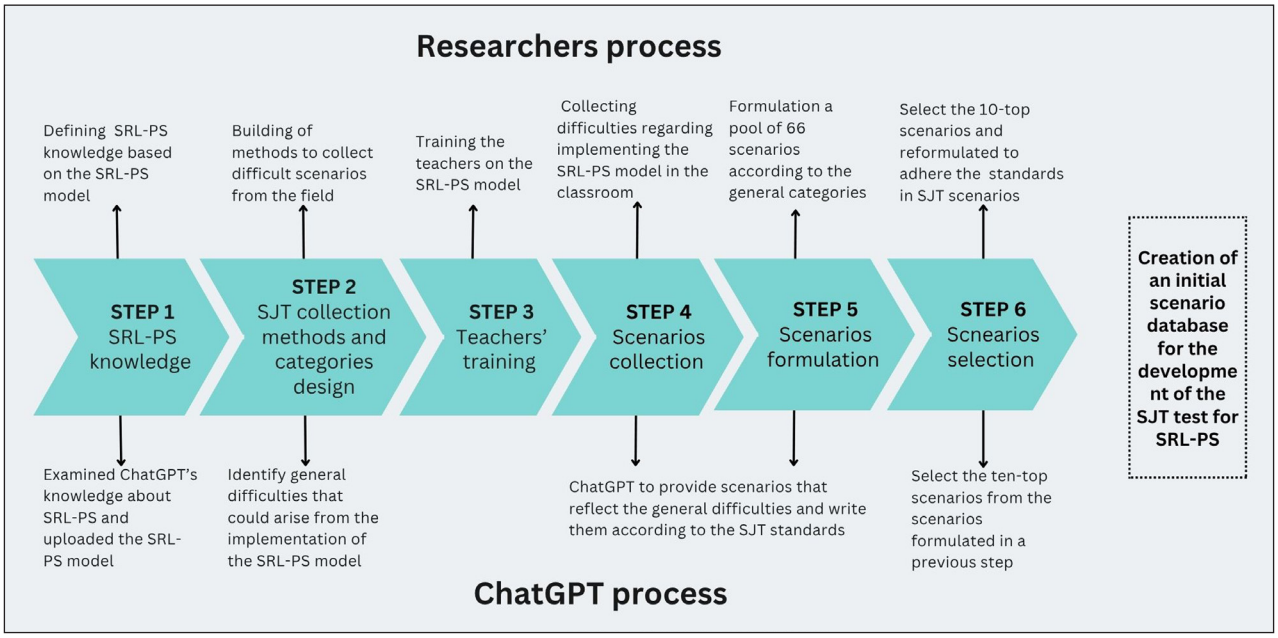

Figure 2 The development process of an SJT tool (by the researchers and ChatGPT) for evaluating teachers' SRL-PS knowledge (Phase one and two).

are presented in the middle. The researchers' role in the process is described in the upper part, and the ChatGPT's role is described in the lower part.

Phase three – comparative analysis

In order to assess the contribution of the interaction between the researchers and Gen-AI to the development of the SJT tool for evaluating teachers' SRL-PS knowledge, a comparison was conducted between the researchers' results with those generated by ChatGPT. This process enabled the charting of all 20 scenarios onto the map that combines the difficulties catgories from both the researchers' perspective and the ChatGPT's.

- *Step 1:* The difficulty categories produced by both the researchers and ChatGPT were compared to identify similarities and differences based on an initial overall impression.
- *Step 2:* The scenarios formulated by both the researchers and ChatGPT were compared to identify similarities and differences based on an initial overall impression.
- *Step 3*: The researchers' top ten selected scenarios were systematically assigned to ChatGPT's difficulty categories. Thus, each scenario articulated by the researchers was examined against all the difficulties categories proposed by ChatGPT. The analysis focused on two primary aspects: first, determining the nature of the challenge each scenario presented in relation to SRL-PS knowledge, and second, establishing whether a connection existed between the scenario of the researchers and the difficulties categories of ChatGPT.
- *Step 4:* ChatGPT's top ten selected scenarios were assigned to the researchers' difficulty categories in the same systematic manner.

4.4 RESEARCH TOOLS

After teachers started implementing the SRL-PS model in their classrooms their input regarding the new model was collected during a session dedicated for this purpose. This data was collected from the teachers in order to assist in developing the SJT tool scenarios, as teachers' raised difficulties constituted the basis for creating the researchers' developed scenarios. Additionally, data was collected from ChatGPT to provide its categories and scenarios.

Scenario Questionnaire: In the Scenario Questionnaire, the teachers were required to come up with authentic scenarios related to various difficulties applicable to the teacher or students, according to a literature-based set of predefined difficulty categories (Mishra & Koehler 2006). The categories covered difficulties that focused on teachers and students. For example, pedagogical difficulties focused on teachers, and social difficulties focused on students. These difficulty categories were defined to raise teachers' awareness of the various aspects of difficulties in implementing the SRL-PS model in their classroom.

Shared SWOT presentation: In a shared presentation the teachers reported about strengths, weaknesses, opportunities, and threats adhering to the SWOT method (Helms & Nixon 2010) to further their feedback about the model and its implementation.

Shared success story presentation: Drawing on the Learning from Success method (Rosenfeld 2016), we asked the teachers to identify one difficulty that they

had overcome and tell us how they faced this difficulty. We also asked teachers to comment on other teachers' responses.

Textual reflection: We asked the teachers to write an overall reflection about the training in the new SRL-PS model.

ChatGPT: We asked ChatGPT to provide difficulty categories and scenarios relating to the SRL-PS model.

4.5 DATA ANALYSIS

The data was qualitatively analyzed to enable the comparative examination between the products developed by the researchers and ChatGPT and made it possible to examine the potential contributions of human-machine collaboration in the development of the SJT evaluation tool. This comprehensive analysis was collaboratively performed by the five authors of this paper. Thus, discussions were first held with regard to scenarios comparison in general; then, the five researchers jointly assigned their top ten scenarios to ChatGPT's difficulty categories following a systematic assignment method. Subsequently, the five researchers followed a similar systematic method to assign ChatGPT's scenarios into the researchers' difficulty categories.

5. FINDINGS

5.1 RESEARCHERS' DIFFICULTIES AND SJT SCENARIOS FOR EVALUATING SRL-PS KNOWLEDGE (RQ1)

The SJT tool for evaluating SRL-PS knowledge, as formulated by the researchers, encompasses difficulty categories based on research literature, scenarios title, and the scenarios themselves. Some categories were focused on teachers' difficulties and some categories focused on students' difficulties. For example: the *pedagogical category*, which focused on teachers' difficulties, produced two scenarios: 'learning gaps between students,' and 'lack of knowledge about students' progress.' The *social category*, which focused on students' difficulties, produced one scenario: 'difficulty in group work.' Altogether, the scenarios built based on the data collected from the teachers, as well as the researchers own input, led to the formulation of scenarios in all difficulty categories. Figure 3 shows the difficulty categories (first column), the title of the scenario (second column), and the ten selected scenarios (third column). Each scenario is assigned an abbreviated code (for example: RS1 stands for research scenario 1).

5.2 CHATGPT'S DIFFICULTIES AND SJT SCENARIOS FOR EVALUATING SRL-PS KNOWLEDGE (RQ2)

The SJT tool for evaluating SRL-PS knowledge, as formulated by ChatGPT, includes difficulty categories and scenarios corresponding to these difficulties. The eight difficulty categories produced by ChatGPT are listed below. The explanation given by ChatGPT is written next to the category title. The titles of each difficulty category and their explanations are presented in the exact output from ChatGPT:

1. *Cognitive Development*: Students at this age have varying levels of cognitive development, which may affect their ability to grasp complex problem-solving processes.
2. *Motivational differences:* There can be significant variability in students' motivation levels, impacting their engagement with the model.
3. *Skill diversity:* Students may have diverse skill sets, making it challenging to tailor the model to all learning needs.
4. *Classroom management***:** Managing the classroom dynamics while implementing a structured model can be challenging, especially with larger class sizes.
5. *Time constraints*: The school curriculum may have tight schedules, making it difficult to allocate sufficient time for each stage of the model.
6. *Resource limitations*: Lack of adequate resources (materials, technology, etc.) can hinder the effective implementation of the model.
7. *Assessment alignment:* Aligning the model with standard assessment methods and ensuring it meets educational standards and objectives may be challenging.
8. *Teacher training*: Teachers may require additional training and support to effectively implement and adapt the model in their classrooms.

As seen, the difficulty categories raised by ChatGPT focused on teachers' and students' difficulties. Another insight is that the difficulty categories produced by ChatGPT referred to different aspects of SRL knowledge. More specifically, the difficulty categories produced by ChatGPT can be divided into two main class of categories: *Management* included teachers' management of classroom dynamics, time allocation, resources, assessment procedures, and professional knowledge. *Skill gaps* included the categories related to disparities among students in cognitive, motivational, and other skills. After producing the eight difficulty categories, ChatGPT formulated scenarios for these eight categories, and eventually selected the top ten scenarios. Thus, for each difficulty category, ChatGPT selected one or two scenarios. For example: the *cognitive development* difficulty category produced two scenarios that deal with simplifying a complex task for a fifth grade student and clarifying the constraints of the problem to the student. Figure 4 shows, in the first column, the eight difficulty categories, and in the second column the ten selected scenarios. Each scenario is assigned an

Researchers'SJT difficulties categories	Scenarios title	Researchers' SJT top ten Scenarios
		Difficulties Focused on teachers
Pedagogical difficulties	1. Learning gaps between students	RS1-Teacher Mia's class is a heterogeneous class. Mia begins to implement the method of PS in her history class. She discovers that there are gaps between the children in understanding the key concepts in history, which were learned using the new method. What should teacher Mia do?
	2. Lack of knowledge about students' progress	RS2-The teacher Yasmin applies the problem-solving method using the checklist and the videos. Yasmin does not know how each student progresses in the learning process. What should Jasmine do?
Emotional difficulties	3. Lack of teaching resources in the new method SRL-PS	RS3-Teacher Shula expressed her frustration at having to change all the plans she had been working on and now being expected to re-plan according to the problem-solving method. What should Shola do?
Difficulties in the school environment	4. Students' limited time	RS4-Teacher Erez teaches 9th grade two hours a week. The class started the learning method through problem solving to promote independent learners. Many classes are canceled because of school activities and exams. Erez shared the problem with the teacher. What should teacher Erez do?
		Difficulties Focused on students
Difficulties in the method	5. Inability to do homework's time-planning	RS5-The parents told teacher Hamudi that they really like the new method for solving problems and its importance, but they noted that the children still use them to plan and manage time. What should Hamudi do?
	6. Suggesting only one solution to the problem	RS6-Yael's teacher notices that Yael offers a nice solution to tasks using the problem-solving method, but she always deals with only one solution. What should Yael's teacher do?
	7. Undetailed enough solution	RS7-The teacher sees that Danny, who works according to the method of problem solving, does not specify the solution well in the step of proposing a solution. What should Danny's teacher do?
	8. Incomprehension of SRL-PS method transfer to other subjects	RS8-The teacher Misam taught in the citizenship class to work with the method of solving problems. Now the teacher wants to apply the method alone in the task she gave the students in the history lesson. The students said they did not understand how the method was related to the history assignment. What should they do?
Emotional difficulties	9. Difficulty to present to audience	RS9-Roy, a student in teacher Miri's class, appeared in front of the class and had difficulty overcoming his excitement to speak in front of an audience when required to present his solution in front of the class. What should Miri do?
Social difficulties	10. Difficulty in group-work	RS10-Ronan, a student in teacher Talia's class, was not satisfied with the members of his group when working together on the task using the problem-solving method. He was afraid he wouldn't get along with them. What should Talia do?

Figure 3 Researchers' difficulty categories and scenarios for evaluating SRL-PS knowledge.

abbreviated code (for example: CHS1 stands for the first ChatGPT scenario).

5.3 RESEARCHER-CHATGPT INTERACTION CONTRIBUTION TO SJT TOOL DEVELOPMENT (QR3)

Combining the findings from Figures 3 and 4 above, during the comparative analysis (phase three described in Research Methodology, above) revealed some preliminary insights concerning the researchers' interaction with ChatGPT. The researchers' difficulty categories and ChatGPT's difficulty categories complemented each other and presented similar and different ways to classify difficulties. Both the researchers and ChatGPT presented difficulty categories in the domain of SRL-PS that focused on students and teachers. However, some noticeable differences stood out: the researchers split difficulty categories into ones which explicitly focused on teachers' scenarios and ones which explicitly focused on students. ChatGPT did not explicitly suggest such sub-classification focusing on teachers and students' difficulty categories. However, the researchers noticed that the ChatGPT's difficulty categories could be divided into management and skill gap categories.

ChatGPT's difficulties categories in SRL	ChatGPT's SJT top ten scenarios
1. Cognitive Development	CHS1- A 5th grader is overwhelmed by the complexity of a problem. How would you simplify the task?
	CHS2- During Problem Identification, a student misunderstands the problem's constraints. How do you clarify this?
2. Motivational Differences	CHS3- Some students show little interest in collaborative problem-solving. What approach would you take?
3. Skill Diversity	CHS4- In a mixed-ability group, advanced students dominate the discussion. How do you ensure inclusivity?
4. Classroom Managemen	CHS5- A classroom activity gets too chaotic. How do you regain control?
	CHS6- You face difficulty in maintaining order during group work. How do you handle this?
5. Time Constraints	CHS7- You have less time than planned for a key stage. How do you adjust?
6. Resource Limitations	CHS8- Essential materials are unavailable for an activity. How do you improvise?
7. Assessment Alignment	CHS9- How do you ensure that your problem-solving activity aligns with the test formats students will face?
8.Teacher Training	CHS10- After training, you find some model concepts unclear. How do you seek clarification?

Figure 4 ChatGPT's difficulty categories and scenarios for evaluating SRL-PS.

Regarding the scenarios, occasionally, the researchers' scenarios and the ChatGPT's scenarios were quite similar (for example: lack of teaching materials for the teacher – RS3, CHS8). Alternatively, some scenarios formulated by ChatGPT expanded or presented another perspective of the difficulty (for example: evaluation and feedback – RS2, CHS9); yet in other instances, new scenarios were proposed by the ChatGPT (for example: Classroom Management – CHS5, CHS6). In addition, in certain cases the researchers' suggested scenarios which were absent from ChatGPT's pull of scenarios, primarily the scenarios reflecting lack of students' understanding of the SRL-PS model itself (RS5–RS8). This indicates that the researchers' and ChatGPT's scenarios complement each other.

This analysis led to the creation of the map, which assigns the seven difficulties categories created by the researchers (focused on teacher and focused on student) into the eight ChatGPT difficulty categories. Thus, the comparative analysis enabled the creation of a map consisting of the 20 scenarios created by the researchers and ChatCPT (Figure 5). The dark grey colored squares represent the researchers' scenarios, code named RS1, RS2, ... up to RS10, and the ChatGPT scenarios, codenamed CH1, CH2, ... up to CH10. The light grey squares represent difficulty categories where no scenarios were given, but this area has coverage by the parallel agent (researchers or ChatGPT). The white squares represent difficulty categories that lack any scenarios.

Thus, the map highlights several findings. First, difficulty categories were observed that lacked any scenarios (the white squares); these require further treatment. The second finding points to the difficulty categories consisting only of researchers' scenarios or ChatGPT scenarios (dark grey and light grey squares), which somewhat requires treatment. The third finding highlights the difficulty categories in which several scenarios are offered both by the researchers and ChatGPT (dark grey), suggesting further treatment might not be required. The final finding revealed by the map concerns the emergence of the new, refined integrated categories of the researchers and ChatGPT.

6. DISCUSSION AND CONCLUSIONS

Given the teachers' central role in promoting SRL-PS among learners, which is an issue of great importance nowadays (Burns et al. 2018; Li et al. 2023), a need arises to diversify the tools used to evaluate teachers'

Difficulties categories ChatGPT	Difficulties categories - researches															
		Focused on teachers								Focused on Students						
		01 Metod		02 Padgogical		03 Emotional		04 School Enviroment		05 Metod		06 Emotional		07 Scocial		
	01 Cognitive Development			RS1						RS6 RS7 RS8	CHS1 CHS2	RS9		RS10		
	02 Motivational Differences			RS1 RS2						RS6 RS7 RS8		RS9	CHS3		CHS3	
	03 Skill Diversity			RS1 RS2	CHS4					RS5 RS6 RS7		RS9		RS10		
	04 Classroom Management				CHS5 CHS6	RS3		RS4								
	05 Time Constraints		CHS7		CHS7	RS3		RS4		RS5 RS6 RS7	CHS7	RS9		RS10		
	06 Resource Limitations			RS2	CHS8	RS3										
	07 Assessment Alignment		CHS9	RS2	CHS9	RS3										
	08 Teacher Training		CHS10			RS3										

Figure 5 Researchers' and ChatGPT's scenarios according to difficulty categories.

SRL-PS knowledge (Vosniadou et al. 2020). Although SJT evaluation tools have been successfully used in various fields, including in education (Aldrup et al. 2020; Chao et al. 2020; Klassen et al. 2020), the method, as best as we know, has not been used to evaluate teachers' SRL knowledge. This study presents an initial development process of an SJT tool to evaluate teachers' SRL-PS knowledge following professional development. Meanwhile, the emergence of Gen-AI tools has also highlighted the potential of Gen-AI assisted evaluation (Owan et al. 2023; Xia et al. 2024). Indeed, the elaborate development process of SJT tools could draw support from Gen-AI tools such as ChatGPT (Borchert et al. 2023; Sareen 2023). Moreover, the interaction between humans and Gen-AI has sparked hope for new opportunities, such as expediting evaluation processes, alongside worries regarding potential risks, such as over-reliance (Slattery et al. 2024). Therefore, in order to prevent the risks and fulfill the opportunities, a human-centered AI (HCAI) approach has been promoted, where tasks are assigned according to the skills of each agent on the process (Fui-Hoon Nah et al. 2023). Despite that, in the Gen-AI flourishing discourse, the human-Gen-AI interaction has drawn relatively little attention thus far (Drori & Te'eni 2024; Te'eni et al. 2023; Cohen et al. 2021). This study aimed to develop an SJT tool for assessing teachers' SRL-PS knowledge using ChatGPT4 and examine the contribution of the interaction between the researchers and ChatGPT to the development of the tool.

The researchers employed Gen-AI in the development process by assigning tasks that seemed plausible to ChatGPT. The study examined the diverse perspectives reflected in the difficulties categories and scenarios for evaluating SRL-PS knowledge produced by both the researchers and ChatGPT, as well as the points of similarity between them. Our analysis led to the creation of a map that synthesized the difficulties categories and scenarios of both the researchers and Gen-AI. The research exemplifies an initial attempt at creating scenarios for the SJT tool that evaluate teachers' SRL-PS knowledge based on researcher-Gen-AI interaction. The comparative analysis of the researchers' and Gen-AIs' difficulty categories found that the researchers and Gen-AI present similarities as well as a different point of view. While the researchers' difficulty categories were based on Mishra and Koehler's (2006) model for teachers' knowledge and our SRL-PS model (Cohen et al. 2022), it seems that ChatGPT's proposed difficulty categories offer a perspective that differs from that of researchers. Additionally, the scenarios formulated by the researchers and ChatGPT complemented each other. The sources used by ChatGPT to establish the difficulty categories and formulate scenarios by GhatGPT remain unknown. Indeed, the process of creating categories and scenarios by GhatGPT is not transparent (Pegrum 2024). Therefore, the need for critical thinking and source validation when working with Gen-AI (Drori & Te'eni 2024; Schwartz & Te'eni 2024) is highlighted in this study. Altogether, this study supports the rising view that human-Gen-AI interaction constitutes a significant contribution (Fui-Hoon Nah et al. 2023) when both agents are utilized according to their relative advantages (Drori & Te'eni 2024; Te'eni et al. 2023; Cohen et al. 2021). Thus, the study exemplifies the interaction conducive to the

development of SJT tools, as the analysis suggested that each agent presents another perspective of the difficulty categories and provides alternate scenarios.

The formed map shows the researchers' and ChatGPT's scenarios along with the different difficulty categories. The map highlights the difficulty categories that need completion of scenarios and the difficulty categories that have a wide range of scenarios. Above all, the map reveals the potential to integrating difficulties categories of the researchers and ChatGPT. New synthesized difficulty categories could facilitate new scenarios. For example, when the school environment difficulty categories focused on teachers (researchers' difficulty categories) are integrated with the resource limitations or assessment alignment (Gen-AI's difficulties), new scenarios could be formulated. Overall, it is evident that the collaborative development process could enlarge the initial SJT scenarios pool. Therefore, this study illustrates the conducive role of the human-Gen-AI interaction to the development of evaluation tools, as was claimed in previous studies (Ionescu & Enescu 2023; Lee et al. 2023). Our case study exemplifies a good practice that demonstrates human-centered AI (Drori & Te'eni 2024; Cohen et al. 2021). The patterns of collaboration that we tried in this study mainly involved deciding where ChatGPT could be used as well as comparing the difficulty categories and scenarios which led to their synthesis. These practices adhere to previously discussed processes, namely the *specification* and *synthesis,* that are implicated in the human-Gen-AI collaborative interaction (Ezra et al. 2024).

6.1 CONTRIBUTIONS AND IMPLICATIONS

This paper outlines an initial attempt to develop an SJT tool, utilizing Gen-AI, to assess teachers' knowledge of SRL-PS following professional development. Additionally, the paper examines how the interactions between the researchers and ChatGPT contributed to the tool's development. Thus, the contributions of this study are twofold: First, it expands evaluation tools for teachers' SRL knowledge, which need diversification (Vosniadou et al. 2020) through the SJT methodology. Second, it highlights the beneficial elements of human-centered AI practices (Drori & Te'eni 2024; Te'eni et al. 2023), illustrating the advantages of collaborative interactions between humans and Gen-AI.

The map created serves as a foundation for the next step in developing the SJT tool for assessing teacher knowledge in SRL-PS. This map aids in formulating new and refined scenarios, as well as in selecting the final scenarios based on the integrated categories from researchers and ChatGPT. In terms of utilizing Gen-AI, this study establishes a basis for conducting further experiments on various Gen-AI platforms. As Gen-AI systems are rapidly evolving, they have the potential to spark significant advancements in SJT development, possibly widening the gap between researchers who choose to adopt these technologies and those who do not. Future research should investigate more human-centered AI practices to enhance Gen-AI-assisted evaluation.

ACKNOWLEDGEMENTS

This research is supported by the Chief Scientist office of the Israeli Ministry of Education and by the Science and the Technology Education Center (SATEC) at Tel-Aviv University.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Dafna Avidov
School of Education, Tel Aviv University, Israel

Orit Ezra orcid.org/0000-0001-6596-1345
School of Education, Tel Aviv University, Israel

Guy Cohen orcid.org/0000-0001-6264-781X
School of Education, Tel Aviv University, Israel

Anat Cohen, Ph.D orcid.org/0000-0002-8373-3511
School of Education, Tel Aviv University, Israel

Alla Bronshtein
Gray Faculty of Medical & Health Sciences, Tel Aviv University, Israel

REFERENCES

Aldrup, K., Carstensen, B., Köller, M. M., Klusmann, U., & **Frenzel, A. C.** (2020). Measuring teachers' social-emotional competence: Development and validation of a situational judgment test. *Frontiers in Psychology, 11*(May), 1–20. https://doi.org/10.3389/fpsyg.2020.00892

Alvi, E., & **Gillies, R. M.** (2020). Teachers and the teaching of self-regulated learning (SRL): The emergence of an integrative, ecological model of SRL-in-context. *Education Sciences, 10*(4), 98. https://doi.org/10.3390/educsci10040098

Assi, A., & **Cohen, A.** (2023, September). Video-based SRL program for teachers and students. In *Proceedings of the 15th International Conference on Education Technology and Computers* (pp. 175–182). https://doi.org/10.1145/3629296.3629323

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research, 31*(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

Borchert, R. J., Hickman, C. R., Pepys, J., & **Sadler, T. J.** (2023). Performance of ChatGPT on the situational judgement test—A professional dilemmas–based examination for doctors in the United Kingdom. *JMIR Medical Education, 9*(1), e48978. https://doi.org/10.2196/48978

Burns, E. C., Martin, A. J., & **Collie, R. J.** (2018). Adaptability, personal best (PB) goals setting, and gains in students' academic outcomes: A longitudinal examination from a social cognitive perspective. *Contemporary Educational Psychology*, *53*(February), 57–72. https://doi.org/10.1016/j.cedpsych.2018.02.001

Chao, T. Y., Sung, Y. T., & **Huang, J. L.** (2020). Construction of the situational judgment tests for teachers. *Asia-Pacific Journal of Teacher Education*, *48*(4), 355–374. https://doi.org/10.1080/1359866X.2019.1633621

Chen, Y. H., & **Jang, S. J.** (2019). Exploring the relationship between self-regulation and TPACK of Taiwanese secondary in-service teachers. *Journal of Educational Computing Research, 57*(4), 978–1002. https://doi.org/10.1177/0735633118769442

Cheung, B. H. H., Lau, G. K. K., Wong, G. T. C., Lee, E. Y. P., Kulkarni, D., Seow, C. S., ... & **Co, M. T. H.** (2023). ChatGPT versus human in generating medical graduate exam multiple choice questions—A multinational prospective study (Hong Kong SAR, Singapore, Ireland, and the United Kingdom). *PLoS One, 18*(8), e0290691. https://doi.org/10.1371/journal.pone.0290691

Cleary, T. J., Kitsantas, A., Peters-Burton, E., Lui, A., McLeod, K., Slemp, J., & **Zhang, X.** (2022). Professional development in self-regulated learning: Shifts and variations in teacher outcomes and approaches to implementation. *Teaching and Teacher Education, 111*, 103619. https://doi.org/10.1016/j.tate.2021.103619

Cohen, A., Ezra, O., Hershkovitz, A., Tzayada, O., Tabach, M., Levy, B., Segal, A., & **Gal, K.** (2021). Personalizing mathematical content in educational applets repository: human teacher versus machine-based considerations. *Educational Technology Research and Development*, *69*(3), 1505–1528. https://doi.org/10.1007/s11423-021-10002-x

Cohen, G., Assi, A., Cohen, A., Bronshtein, A., Glick, D., Gabbay, H., & **Ezra, O.** (2022). Video-Assisted Self-Regulated Learning (SRL) Training: COVID-19 Edition. In I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega Arranz & T. Farrell (Eds.), *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption. EC-TEL 2022. Lecture Notes in Computer Science* (pp. 59–73). Springer, Cham. https://doi.org/10.1007/978-3-031-16290-9_5

Cohen, G., & **Cohen, A.** (2023). Exploring Students' Problem-Solving Challenges in Self-regulated Learning Through Training Video Prompts. In O. Viberg, I. Jivet, P. Muñoz-Merino, M. Perifanou, & T. Papathoma (Eds.), *Responsive and Sustainable Educational Futures. EC-TEL 2023. Lecture Notes in Computer Science*, vol 14200. Springer, Cham. https://doi.org/10.1007/978-3-031-42682-7_38

De Bruijn-Smolders, M., Timmers, C. F., Gawke, J. C. L., Schoonman, W., & **Born, M. P.** (2016). Effective self-regulatory processes in higher education: Research findings and future directions. A systematic review. *Studies in Higher Education*, *41*(1), 139–158. https://doi.org/10.1080/03075079.2014.915302

Dignath, C., Buettner, G., & **Langfeldt, H. P.** (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review*, *3*(2), 101–129. https://doi.org/10.1016/j.edurev.2008.02.003

Dignath, C., & **Büttner, G.** (2018). Teachers' direct and indirect promotion of self-regulated learning in primary and secondary school mathematics classes–insights from video-based classroom observations and teacher interviews. *Metacognition and Learning*, *13*, 127–157. https://doi. org/10.1007/s11409-018-9181-x

Dignath-van Ewijk, C., & **Van der Werf, G.** (2012). What teachers think about self-regulated learning: Investigating teacher beliefs and teacher behavior of enhancing students' self-regulation. *Education Research International, 2012*(1), 741713. https://doi.org/10.1155/2012/741713

Drori, I., & **Te'eni, D.** (2024). Human-in-the-loop AI reviewing: feasibility, opportunities, and risks. *Journal of the Association for Information Systems*, *25*(1), 98–109. https://doi.org/10.17705/1jais.00867

European Parliamentary Research Service, Scientific Foresight Unit (February 2022). *AI Definition & Glossary Key Terms and Concepts*. https://www.europarl.europa.eu/cmsdata/244992/AI%20Glossary%20STOA%2014.02.2022.pdf

Ezra, O., Cohen, A., & **Shany, D.** (2024). Human experts and Generative AI interaction: From real-world and real-life language learning diary-research app to a diary-learning app. *Ubiquity Proceedings*, *4*(1), 19. https://doi.org/10.5334/uproc.141

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & **Chen, L.** (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of information technology case and application research*, *25*(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Gold, B., & **Holodynski, M.** (2015). Development and construct validation of a situational judgment test of strategic knowledge of classroom management in elementary schools development and construct validation of a situational judgment test of strategic knowledge of classroom management. *Educational Assessment*, *7197*(20), 226–248. https://doi.org/10.1080/10627197.2015.1062087

Helms, M. M., & **Nixon, J.** (2010). Exploring SWOT analysis–where are we now? A review of academic research from the last decade. *Journal of Strategy and Management, 3*(3), 215–251. https://doi.org/10.1108/17554251011064837

Ionescu, V. M., & **Enescu, M. C.** (2023, June). Using ChatGPT for generating and evaluating online tests. In *2023 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 1–6). IEEE. https://doi.org/10.1109/ECAI58194.2023.10193995

Jansen, R. S., van Leeuwen, A., Janssen, J., Kester, L., & **Kalz, M.** (2017). Validation of the self-regulated online learning questionnaire. *Journal of Computing in Higher Education, 29*(1), 6–27. https://doi.org/10.1007/s12528-016-9125-x

Karlen, Y., Hertel, S., & **Hirt, C. N.** (2020). Teachers' professional competences in self-regulated learning: An approach to integrate teachers' competences as self-regulated learners and as agents of self-regulated learning in a holistic manner. *Frontiers in Education, 5*(September), 1–20. https://doi.org/10.3389/feduc.2020.00159

Kistner, S., Rakoczy, K., & **Otto, B.** (2010). Promotion of self-regulated learning in classrooms: Investigating frequency, quality, and consequences for student performance. *Metacognition and Learning*, *5*(2), 157–171. https://doi.org/10.1007/s11409-010-9055-3

Klassen, R. M., Kim, L. E., Rushby, J. V., & **Bardach, L.** (2020). Can we improve how we screen applicants for initial teacher education? *Teaching and Teacher Education*, *87*, 102949. https://doi.org/10.1016/j.tate.2019.102949

Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & **Kim, H.** (2023). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies*, 1–33. https://doi.org/10.1007/s10639-023-12249-8

Li, L., Zhu, M. L., Shi, Y. Q., & **Yang, L. L.** (2023). Influencing factors of self-regulated learning of medical-related students in a traditional Chinese medical university: A cross-sectional study. *BMC Medical Education*, *23*(1), 87. https://doi.org/10.1186/s12909-023-04051-4

Lievens, F., Peeters, H., & **Schollaert, E.** (2008). Situational judgment tests: A review of recent research. *Personnel Review, 37*(4), 426–441. https://doi.org/10.1108/00483480810877598

Mishra, P., & **Koehler, M. J.** (2006). Technological pedagogical content knowledge: A framework for teacher knowledge related papers. *The Development of Technological Pedagogical Content Knowledge in Design Seminar*. https://doi.org/10.1177/016146810610800610

Nasution, N. E. A. (2023). Using artificial intelligence to create biology multiple choice questions for higher education. *Agricultural and Environmental Education, 2*(1), 4–8. https://doi.org/10.29333/agrenvedu/13071

Nota, L., Soresi, S., & **Zimmerman, B. J.** (2004). Self-regulation and academic achievement and resilience: A longitudinal study. *International journal of educational research*, *41*(3), 198–215. https://doi.org/10.1016/j.ijer.2005.07.001

OECD. (2021). *AI and the future of skills, volume 1: Capabilities and assessments.* OECD Publishing. https://doi.org/10.1787/5ee71f34-en

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & **Bassey, B. A.** (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(8), em2307. https://doi.org/10.29333/ejmste/13428

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*(Apr), 1–28. https://doi.org/10.3389/fpsyg.2017.00422

Patterson, F., Zibarras, L., & **Ashworth, V.** (2016). Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Medical Teacher, 38*(1), 3–17. https://doi.org/10.3109/0142159X.2015.1072619

Pegrum, M. (December 2024). *Generative AI*. https://markpegrum.com/tools-for-digital-learning/generative-ai/

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of self-regulation* (pp. 451–502). Elsevier. https://doi.org/10.1016/B978-012109890-2/50043-3

Porter, A. N., & **Peters-Burton, E. E.** (2021). Investigating teacher development of self-regulated learning skills in secondary science students. *Teaching and Teacher Education*, *105*, 103403. https://doi.org/10.1016/j.tate.2021.103403

Rosenfeld, J. M. (2016). *From exclusion to reciprocity: Learning from success.* Rowman & Littlefield.

Sareen, K. (2023). Assessing the ethical capabilities of Chat GPT in healthcare: A study on its proficiency in situational judgement test. *Innovations in Education and Teaching International*, 1–11. https://doi.org/10.1080/14703297.2023.2258114

Schubert, S., Ortwein, H., Dumitsch, A., Schwantes, U., Wilhelm, O., Kiessling, C., ... & **Kiessling, C.** (2008). A situational judgement test of professional behaviour: Development and validation. *Medical Teacher*, *30*(5), 528–533. https://doi.org/10.1080/01421590801952994

Schwartz, D., & **Te'eni, D.** (2024). AI for knowledge creation, curation, and consumption in context. *Journal of the Association for Information Systems*, *25*(1), 37–47. https://aisel.aisnet.org/jais/vol25/iss1/12/

Slattery, P., Saeri, A. K., Grundy, E. A., Graham, J., Noetel, M., Uuk, R., ... & **Thompson, N.** (2024). The AI risk repository: A comprehensive meta-review, database, and taxonomy of risks from artificial intelligence. arXiv preprint arXiv:2408.12622. https://airisk.mit.edu/. https://doi.org/10.70777/agi.v1i1.10881

Svendsen, A., & **Garvey, B.** (2023). *An outline for an interrogative/prompt library to help improve output quality from generative-AI Datasets.* (May 2023). https://doi.org/10.2139/ssrn.4495319

Te'eni, D., Yahav, I., Zagalsky, A., Schwartz, D., Silverman, G., Cohen, D., ... & **Lewinsky, D.** (2023). Reciprocal human-machine learning: A theory and an instantiation for the case of message classification. *Management Science*. https://doi.org/10.1287/mnsc.2022.03518

Vosniadou, S., Lawson, M. J., Wyra, M., Van Deur, P., Jeffries, D., & **Ngurah, D. I. G.** (2020). Pre-service teachers' beliefs about learning and teaching and about the self-regulation of learning: A conceptual change perspective. *International Journal of Educational Research, 99*, 101–495. https://doi.org/10.1016/j.ijer.2019.101495

Weekley, J. A., & **Jones, C.** (1999). Further studies of situational tests. *Personnel Psychology, 52*(3), 679–700. https://doi.org/10.1111/j.1744-6570.1999.tb00176.x

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & **Chiu, T. K. F.** (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education, 21,* 40. https://doi.org/10.1186/s41239-024-00468-z

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses?. *Contemporary educational psychology*, *11*(4), 307–313. https://doi.org/10.1016/0361-476X(86)90027-5

Zimmerman, B. J. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 1–19). Guilford.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

TO CITE THIS ARTICLE:
Avidov, D., Ezra, O., Cohen, G., Cohen, A., & Bronshtein, A. (2025). Human-Gen-AI Collaboration in Developing Situational Judgment Tests for Self-Regulated Problem Solving. *European Journal of Open, Distance and E-Learning, 27*(2): 5, 1–15.

Submitted: 25 December 2024 **Accepted:** 08 May 2025 **Published:** 19 June 2025