Education and Information Technologies
https://doi.org/10.1007/s10639-024-13037-8

Integrating 4C/ID model into computer- supported formative assessment system to improve the effectiveness of complex skills training for vocational education

Haoxin Xu[1] · Tianrun Deng[2] · Xianlong Xu[2] · Xiaoqing Gu[2] · Lingyun Huang[3] · Haoran Xie[4] · Minhong Wang[3]

Received: 26 April 2024 / Accepted: 2 September 2024

Abstract
In the 21st century, the urgent educational demand for cultivating complex skills in vocational training and learning is met with the effectiveness of the four-component instructional design model. Despite its success, research has identified a notable gap in the address of formative assessment, particularly within computer-supported frameworks. This deficiency impedes student self-awareness of skill mastery and limits effective monitoring of skill learning in the classroom by teachers. To address this gap, the study introduces an enhanced four-component instructional design model that seamlessly integrates formative assessment. Based on this model, an automated system for assessing complex skills was developed, with the aim of formative assessment and improving skill learning. A control experiment involving 54 industrial robot professional participants in vocational colleges has preliminarily verified the feasibility and effectiveness of computer-supported formative assessment. The findings reveal that this approach significantly enhances students' schema construction, knowledge, skill mastery, and transfer ability, thereby improving the overall effectiveness of complex skill learning. In addition, participants who underwent computer-supported formative assessment reported high levels of system satisfaction and usefulness, with no adverse impact on their learning attitudes, motivation, or cognitive load. This study contributes a robust theoretical framework and practical case study for computer-supported formative assessment in complex skill learning, providing empirical support for the advancement of computer-supported teaching. The integration of formative assessment within the four-component instructional design model offers a novel perspective, addressing a critical gap in the existing literature and laying the foundation for future developments in this educational domain.

Keywords Complex skill learning · 4C/ID model · Formative assessment · Computer-supported teaching

Extended author information available on the last page of the article

Published online: 23 September 2024

1 Introduction

With the age of artificial intelligence, the goals of vocational education are undergoing constant transformation. In response to the increasing demand for highly skilled professionals adept in mastering a diverse array of complex skills (Maddens et al., 2020), the focus is on cultivating the complex 21st-century skills, such as learning and innovation, digital literacy, as well as career and life skills (Van Laar et al., 2020). These complex skills require students to integrate knowledge, skills, attitudes, and coordinate various abilities to solve ill-structured problems in real-world scenarios (Alahmad et al., 2021), highlighting the need for an integrated approach in education to prepare students for the challenges of the modern workforce. Therefore, educational programs must be carefully organized to help learners master these increasingly complex skills (Hosseinzadeh et al., 2023).

Over the past ten years, the four-component instructional design model (4C/ID) has provided significant implications for the design of instructional methods for complex skills (Costa et al., 2022). It has been applied in various contexts, such as information problem-solving (Wopereis et al., 2015), engineering education (Martínez-Mediano & Losada, 2017; Mulders, 2022), teacher professional development (Frerejean et al., 2021) and medical education (Janesarvatan & Van Rosmalen, 2023), and has achieved good performance. However, existing research on 4C/ID model mostly focuses on instructional design and application, but neglects the potential educational value of formative assessment that can be further explored in the learning process (Bhagat & Spector, 2017).

Formative assessment is a dynamic response to the learning process, which can monitor learning in real time and make timely adjustments when the learning path deviates by collecting data on student behavior, performance, attitudes, and emotions when learning complex skills, which includes observation, measurement, evaluation and feedback (Maier et al., 2016). However, formative assessment in complex skill learning is a time-consuming and laborious task, and it is difficult to collect and process diverse and large amounts of learning data by human labor alone (Maier et al., 2016). Fortunately, computer-supported assessment techniques, such as formative assessment systems and a digital assessment tool, offer solutions to this problem (Spector et al., 2016). For example, the Snappet digital tool for primary school mathematics increased performance and motivation by providing detailed feedback and explanations (Faber & Visscher, 2018). Ackermans et al. (2021) with textual analytics rubrics, used to support formative assessment of complex skills, support feedback, reflection, and thus support the mental model development. Although existing research has achieved good results, it lacks structured theoretical guidance similar to the 4C/ID model to support complex skill learning, which also makes it difficult to apply widely in vocational education.

Previous research has incorporated the 4C/ID model into learning systems, providing summative assessments to gauge skill mastery (Mahantakhun et al., 2020; Mulders, 2022; Pontes et al., 2021). However, these computer-supported systems are predominantly utilized to support course delivery and still lack research on formative assessment. Furthermore, existing research has shown the positive impact of formative assessment on improving student learning outcomes, attitudes, and motivation (Ackermans et al., 2021; Hwang & Chang, 2011; Leenknecht et al., 2021). However, it is noteworthy that these methods may also increase students' external cognitive load, especially in complex skill learning contexts (Larmuseau et al., 2019; Marcellis et al., 2018). Therefore, a comprehensive assessment of the various aspects of the effects caused by interventions is crucial in research efforts.

In summary, the 4C/ID model offers a robust theoretical and practical framework to acquire complex skills, presenting a comprehensive approach that has garnered significant attention. However, its limited emphasis on formative assessment could potentially diminish the model's effectiveness in fostering complex skill acquisition. On the other hand, leveraging technology promises to enhance the implementation of formative assessments, suggesting a pathway to address this gap. Despite these advancements, a deeper exploration into the impact of technology on various student outcomes remains crucial. Understanding this impact is vital for more effectively integrating technology to bolster the effectiveness of formative assessment processes. To this end, a critical research question emerges: How can the integration of technology in the 4C/ID model be optimized to enhance formative assessment, thus improving the learning outcomes of complex skills? This study seeks to unravel the multifaceted effects of technological integration on formative assessment effectiveness and, ultimately, on student learning outcomes, paving the way for a more informed and effective application of the 4C/ID model.

This paper introduces an enhanced 4C/ID model that meticulously delineates the roles and timing of formative assessment mechanisms within the framework. Based on this model, we designed and developed an automated formative assessment system for the learning of complex skills. In this system, we automate the data collection, data analysis and assessment of formative assessment online, while the feedback is arranged in offline activities. The system not only streamlines the learning process and collects student data, but also provides timely and relevant feedback. The core objective of our research is to rigorously evaluate the impact of our proposed model and its associated automated system on a comprehensive range of educational outcomes. These outcomes include, but are not limited to, learning effectiveness, learner motivation, student attitudes towards learning, internal and external cognitive load management, and perceptions of the assessment system itself. Through a methodologically robust applied research approach, we seek to elucidate the synergistic effects of integrating theoretical frameworks with technological advancements in education. Our goal is to achieve a nuanced understanding of how these integrations can enhance educational practices and teaching strategies. The insights derived from this study are expected to offer substantial contributions to the field, providing actionable guidance for elevating the quality of future educational endeavors.

2 Literature review

2.1 The 4C/ID model for complex skills learning

Complex problem solving is an increasingly crucial skill of specialized problem-solving in the 21st century (Herde et al., 2016). Hence, helping learners master complex skills has always been one of the key issues in the field of education (Thima & Chaijaroen, 2021). The complex skills do not refer to a specific skill but rather the skills of integrating and transferring single skills, subject domain knowledge, and attitudes into problem-solving (Frerejean et al., 2021), a carefully designed instructional programs is needed their learning (Ackermans et al., 2021). The 4C/ID model is a comprehensive instructional design model that emphasizes the integration of knowledge, skills, and attitudes into real-life tasks. It is particularly effective for complex learning, offering learners a structured, yet flexible approach to mastering both the recurrent (routine) and non-recurrent (variable) aspects of tasks (Van Merriënboer, 2019). It has received widespread attention in effectively promoting complex skill learning (Costa et al., 2022). It reveals the process of learners' internal cognitive development during the learning process, and designs matching instructional programs for complex skill learning, as show in Fig. 1. Below is a brief introduction to the four components of the 4C/ID model (Van Merriënboer, 2019).

Learning tasks stand as the cornerstone of the 4C/ID model, meticulously crafted to mimic the complexities and challenges individuals encounter in the real world (Van Merriënboer, 2019). These tasks are not mere exercises but immersive experiences, authentically replicating real-world scenarios to ensure learners grasp how tasks unfold in actual settings. They engage learners in applying a balanced mix of skills, knowledge, and attitudes, reflecting the multifaceted nature of real-life tasks. The organization of these tasks follows a thoughtfully graduated complexity, starting from simpler tasks to more intricate ones (Van Merriënboer et al., 2002). The dashed square in Fig. 1 represents different levels of complexity. This progression allows learners to methodically build and expand on their existing knowledge and skill set. To support this journey, initial tasks are accompanied by substantial guidance (scaffolding), which is gradually scaled back as learners advance in competence, promoting a transition towards greater independence (Sweller et al., 2019). The size of the shaded area within the circle representing the learning task in the Fig. 1 indicates the level of support. A notable feature of these learning tasks is their high variability, which is instrumental in enabling learners to transfer and apply their acquired skills and insights across a broad spectrum of situations, thereby bolstering their adaptability and problem-solving capabilities. Different triangles within the circle representing the learning task represent different degrees of variation (Melo & Miranda, 2016).

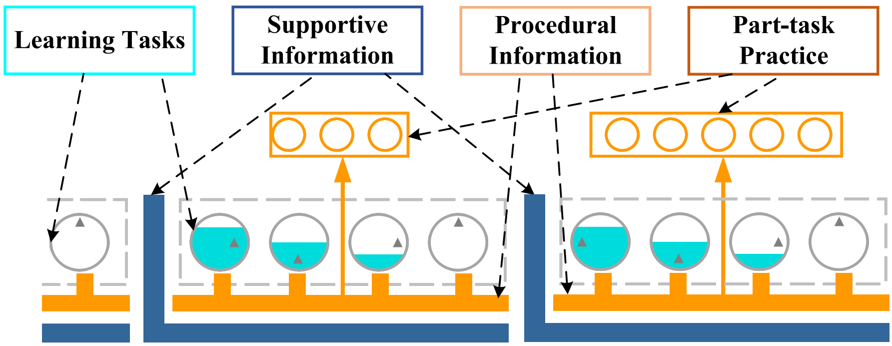


Fig. 1 Overview of 4C/ID model

Complementing the practical focus of learning tasks, **supportive information** provides a robust theoretical foundation, essential for understanding the non-recurrent (adaptive) aspects of tasks (Sarfo & Elen, 2006). This component provides comprehensive guidance on how to navigate and resolve problems within a domain, cultivating a strategic mindset. It also demystifies the organizational structure of the domain, aiding learners in efficiently organizing their knowledge (Musharyanti et al., 2021). Designed for each task class and perpetually accessible, this supportive information serves as a constant resource that learners can draw on to reinforce their understanding as needed.

Procedural information addresses the recurrent (routine) aspects of tasks, offering precise step-by-step instructions that facilitate the mastery of procedural skills (Van Merriënboer, 2019). Presented just-in-time, it aims to mitigate cognitive overload by providing relevant guidance precisely when it's most needed. As learners progress and their expertise deepens, this procedural guidance is progressively phased out, encouraging them to rely on their developing skill set and fostering a sense of autonomy (Corbalan et al., 2006).

Part-task practice is specifically designed to refine the proficiency in recurrent tasks to a level where execution becomes nearly automatic. By focusing on extensive repetition of select task aspects, learners can achieve a high degree of automaticity, enabling them to perform tasks with minimal conscious effort (Frerejean et al., 2019). This component is judiciously applied to task aspects where repetition is most beneficial, optimizing the learners' time and effort. Crucially, it is initiated only after these aspects have been introduced within the broader task context, ensuring learners understand how each element is integrated into the whole (Frerejean et al., 2023). This systematic approach ensures that the 4C/ID model not only equips learners with practical skills and knowledge, but also fosters a deep, strategic understanding and a high level of adaptability, preparing them to effectively navigate the complexities of real-world tasks.

In existing research, the 4C/ID model has been successfully applied in various fields such as medical education (Janesarvatan & Van Rosmalen, 2023; Maggio et al., 2015), teacher training, problem-solving (Wopereis et al., 2015), engineering education (Martínez-Mediano & Losada, 2017), and media literacy (Hosseinzadeh et al., 2023), highlighting its excellence and applicability in the learning of complex skills (Costa et al., 2022). However, these studies primarily focus on using the 4C/ID model to design courses supporting students' learning of complex skills, with insufficient attention to the assessment of these skills. Current research has summarized evaluations from perspectives like skill mastery, knowledge acquisition, skill transfer, and schema construction (Costa et al., 2022; Maddens et al., 2023). Yet, formative assessment are equally crucial for complex skill learning. For instance, Larmuseau et al. (2018) pointed out in a study involving 161 students that providing procedural information improved student learning outcomes, emphasizing the importance of data collection and evaluation to assist students in task and procedural information selection. Other researchers advocate for more in-depth measurements during teaching to offer timelier feedback to both teachers and students (Ndiaye et al., 2021). While Van Merriënboer (2019) underscore the significance of assessment design in 4C/ID-based course design, the current model lacks clarity regarding the location and timing of formative assessment, posing challenges to their implementation. In summary, these studies underscore the urgency of implementing formative assessment in the context of complex skill learning supported by the 4C/ID model.

2.2 Computer-supported formative assessment of complex skills learning

Classroom assessment includes both formative assessment, which is used to adapt instruction and help students improve, and summative assessment, which is used to grade students or otherwise demonstrate student achievement (Shepard, 2019). The role of formative assessment in education includes defining learning objectives, planning learning activities, classroom implementation, review of learning, and conclusion (Black, 2015). Informing the adaptation of processes and products to deepen learning and improve achievement, teachers are able to gather evidence about students' learning pathways and provide immediate feedback in the classroom (Andrade, 2019). Thus, this learning-orientated formative assessment empowers the opportunity to adjust and correct learning and promotes learning and achievement (Faber & Visscher, 2018). For example, Brookhart et al. (2010) documented, workshopped, and outlined the formative assessment methods, improvements, and reflections used by six teachers in their classrooms, and found that students with formative assessments scored higher on reading readiness than those who didn't use formative assessments, based on growth by all six teachers.

However, implementing formative assessment in the learning of complex skills is not an easy task, and the situation with formative assessment is so complex that they can only be understood in terms of the few theoretical perspectives needed to explore the different types of issues involved (Black & Wiliam, 2009). This leads to the fact that almost any formative assessment strategy can become teacher-centred rather than fully formative (Brookhart, 2013). The nature of formative assessment revolves around three issues of teacher and student practice: explicit learning objectives, evidence gathered during the learning process and instructional adjustments (Heritage, 2020). Conducting formative assessment can be challenging given the integrative and transferable nature of complex skills acquisition. Thanks to advances in technology, computer-supported assessment systems are expanding (Spector et al., 2016). For example, Wilson et al. (2021) developed an automatic evaluation system that can automatically assess users' writing and provide immediate feedback and evaluation. The automatic assessment draws on computer programs to assist teachers in monitoring students' learning and facilitates feedback. For training novice surgeons requiring precise operations, the Internet of Things and machine learning enable integrating sensors into objects and environments (Castillo-Segura et al., 2021). This collects abundant data, enabling more objective and automated assessments of complex skills like surgery. Moreover, the Pe(e)rfectly Skilled online method for higher education provided structured support for self-regulation, goal setting, feedback, and reflection during complex skills training (Rusman & Nadolski, 2023). However, these studies were conducted from a technical perspective and did not combine digital and computer technologies with the complex learning process supported by the 4C/ID model.

Several studies have been dedicated to developing learning systems based on the 4C/ID model (Larmuseau et al., 2018; Mahantakhun et al., 2020; Mulders, 2022; Qiu et al., 2012). For example, Qiu et al. (2012) explored the design of online learning systems guided by the 4C/ID model. However, this study did not consider formative assessment, and lacked practical validation. On the other hand, Mulders (2022) proposed a virtual reality-based learning system that shows potential. This system offers comprehensive learning support, including tasks, procedural information, supportive information, and part-task practice, to facilitate the effective implementation of 4C/ID model courses. Additionally, it provides summative assessments of students' skill mastery. Furthermore, Larmuseau et al. (2018) attempted to collect process-oriented learning behavior data to analyze the relationship between learner behavior and learning outcomes. However, there was a lack of sufficient formative assessment based on this data. In summary, these studies primarily focused on supporting course delivery, achieving success in summative assessments, but unfortunately, research on formative assessment remains limited.

Moreover, it is crucial to emphasize that existing research indicates positive impacts of computer-supported formative assessment on students' learning outcomes, attitudes, motivation and system perception (Ackermans et al., 2021; Leenknecht et al., 2021; Tapingkae et al., 2020). However, such technological interventions that alter the learning environment are prone to externally generated information that is irrelevant to learning, which may interfere with students' learning and increase their cognitive load (Larmuseau et al., 2019; Marcellis et al., 2018). Therefore, a comprehensive examination of these dimensions is necessary when implementing computer-supported formative assessment (Cidral et al., 2018). Current research on these dimensions is scattered across various literature, and there is a lack of a comprehensive exploration, especially in the context of complex skill learning guided by the 4C/ID model.

As outlined by the 4C/ID model, teachers are expected to assess students' abilities in understanding knowledge components, cognitive techniques, rule application in tasks, mental models, problem-solving sequences, and performance measurements (Van Merriënboer & Kirschner, 2017). However, in complex learning processes, it becomes challenging for teachers to observe the true mastery and transfer of skills by learners (Van Merriënboer & Dolmans, 2015). Additionally, faced with the complexity of composing skills, teachers find it difficult to visualize and effectively manage a diverse set of learning data, such as classroom demonstrations, task performances, and exam results (Webb et al., 2018). These data encompass the generation and regeneration aspects of complex skills, addressing knowledge, skills, and abilities. Therefore, there is an urgent need for a digitalized, visualized, and intelligent system for the automatic assessment of complex skills. Such a system can not only support complex skill learning guided by the 4C/ID model but also conduct comprehensive and accurate formative assessment. This will aid teachers in timely adjusting their teaching strategies and help students master complex skills.

In reviewing existing research, the 4C/ID model provides a solid theoretical foundation for complex skill learning, emphasizing the integrity and completeness of tasks. However, we also recognize that the lack of formative assessment weakens the model's effectiveness in practical applications, as formative assessment is a crucial means of facilitating ongoing improvement and deepening understanding for students. On the other hand, while the application of technology offers a more effective implementation pathway for formative assessment, a deeper understanding of technology's comprehensive impact on various aspects of students is still needed. Clarifying the effects of technological applications on students, will contribute to a more effective integration of technology, allowing it to play a more positive role in enhancing the effectiveness of formative assessment.

2.3 Research goal and questions

This study aims to explore the integration of formative assessment within the theoretical framework of the 4C/ID model to improve the efficacy of learning complex skills. We plan to conduct an in-depth investigation into the comprehensive impact of technological applications in the complex skill learning. Therefore, we have proposed four research questions (RQ) along with their respective four null hypotheses (H0) and four alternative hypotheses (H1). Additionally, for simplicity, we use the term "intervention" to refer to the implementation of computer-supported formative assessment when describing the research questions and hypotheses.

- **RQ1**: Does the intervention significantly improve students' post-test scores?
  - **H0**: The intervention does not significantly improve students' post-test scores.
  - **H1**: The intervention significantly improves students' post-test scores.
- **RQ2**: Does the intervention significantly improve students' learning attitudes and motivation?
  - **H0**: The intervention does not significantly improve students' learning attitudes and motivation.
  - **H1**: The intervention significantly improves students' learning attitudes and motivation.
- **RQ3**: Does the intervention significantly increase students' cognitive load?
  - **H0**: The intervention does not significantly increase students' cognitive load.
  - **H1**: The intervention significantly increases students' cognitive load.
- **RQ4**: Does the intervention significantly improve students' perceived ease of use, usefulness, and satisfaction with the learning system?
  - **H0**: The intervention does not significantly improve students' perceived ease of use, usefulness, and satisfaction with the learning system.
  - **H1**: The intervention significantly improves students' perceived ease of use, usefulness, and satisfaction with the learning system.

Through a thorough analysis of these research questions and hypotheses, we aim to provide valuable insights for both theoretical and practical domains, offering important recommendations for the future development of educational technology and teaching methodologies.

3 The present study

3.1 The enhanced 4C/ID model with integrated formative assessment

To address the first research question, we meticulously modified the 4C/ID model to provide clearer delineation of the positioning and timing of formative assessment within the learning process. In the enhanced 4C/ID model, we introduced two pivotal elements—formative assessment and formative feedback—represented by green and purple bars respectively, as illustrated in Fig. 2. Learning tasks are regarded as the foundation and cornerstone of the 4C/ID model (Van Merriënboer, 2019). We contend that formative assessment should be closely intertwined with learning tasks to ensure students receive effective guidance and support throughout the entire learning process. Specifically, we positioned formative assessment above learning tasks, emphasizing its parallel relationship with them. This design not only establishes a close association between formative assessment and learning tasks but also allows for it to accompany or integrate seamlessly into the learning tasks, providing students with immediate evaluation and guidance as they complete their assignments (Leenknecht et al., 2021).

In addition, we have highlighted feedback in formative assessment separately because in this study, the data collection, analysis and evaluation aspects of formative assessment were automated online, in conjunction with the learning tasks in the 4C/ID model. Feedback, on the other hand, is a direct intervention with students offline, distinguishing it from the four elements in the 4C/ID model. Therefore, we placed formative feedback to the right of learning tasks, underscoring its purpose of providing timely feedback and guidance. This intentional placement ensures that stu-

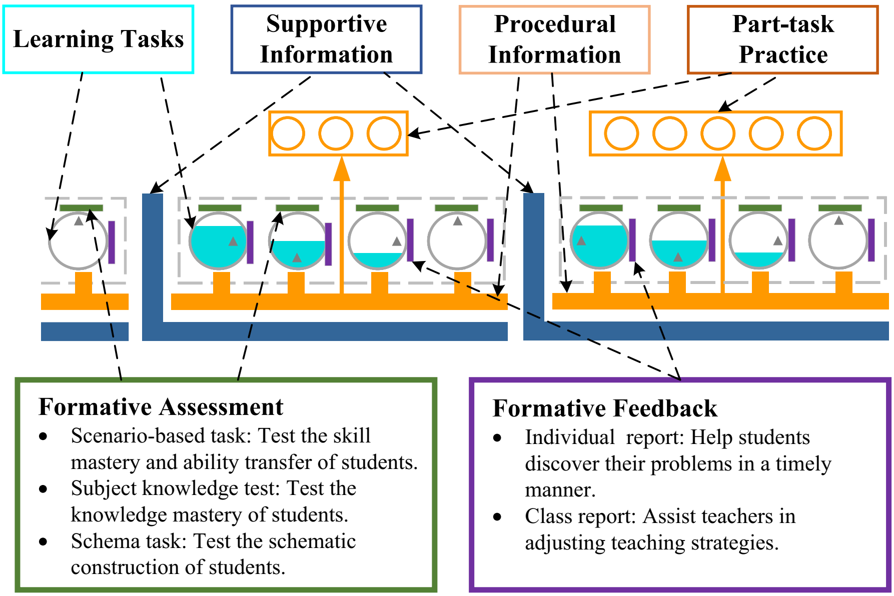


Fig. 2 Improvement of the 4C/ID model supporting formative assessment

dents can easily access and comprehend feedback information during the progression of learning tasks, facilitating better adjustment of learning strategies and continual improvement of their academic performance (Morris et al., 2021).

In conjunction with our modifications to the 4C/ID model, we further categorized formative assessments into three types of learning tasks: *scenario-based tasks*, *subject knowledge tests*, and *schema tasks*. Scenario-based tasks evaluate students' skill mastery and ability transfer. Subject knowledge tests help students construct declarative knowledge of critical rules of complex skills. Schema tasks are used to detect the schema construction of students. This categorization contributes to a comprehensive evaluation of students' skill mastery, ability transfer, and schema construction capabilities. For each type of formative assessment, we designed two forms of formative feedback to better support students' learning processes. Firstly, the individual feedback reports for students were developed, offering detailed information such as the presentation of correct answers, response time, question analysis, and teacher evaluations. This individual report aids students in quickly identifying areas of improvement and making timely adjustments during their learning journey. Secondly, we introduced class feedback reports for teachers, summarizing the overall performance of each formative assessment type. These feedback reports empower teachers to better understand the overall class proficiency, promptly identify common issues among students, and intervene with targeted teaching strategies.

By integrating this refined formative assessment mechanism into the 4C/ID model, our aim is to provide students with more targeted and comprehensive learning support, while offering teachers more effective means of class management and instructional adjustments. This integration contributes to aligning the 4C/ID model more closely with real-world teaching scenarios, thereby enhancing teaching quality and student learning outcomes.

3.2 Complex skill automatic assessment system

In order to achieve computer-supported formative assessment, we designed and developed a Complex Skill Automatic Assessment System (CSAAS) based on the enhanced 4C/ID model (see in Fig. 3). This system has two main characteristics. Firstly, it can support the development of courses designed based on the 4C/ID model, reducing the pressure of teachers' lesson preparation. Secondly, it can provide formative assessment and during the learning process, help teachers adjust teaching strategies in a timely manner, and help students master complex skills.

3.2.1 Support the implementation of courses designed based on the 4C/ID model

Firstly, teachers add complex skills that students need to learn through the skill management module. A complex skill will be divided into different sub skills, forming a skill hierarchy diagram (Frerejean et al., 2023), which will be linked to different learning tasks, teaching resources, and tests. Secondly, teachers can set a series of learning tasks with different complexity and variability through the task management module. These learning tasks are real-life tasks from a professional or daily perspective. They

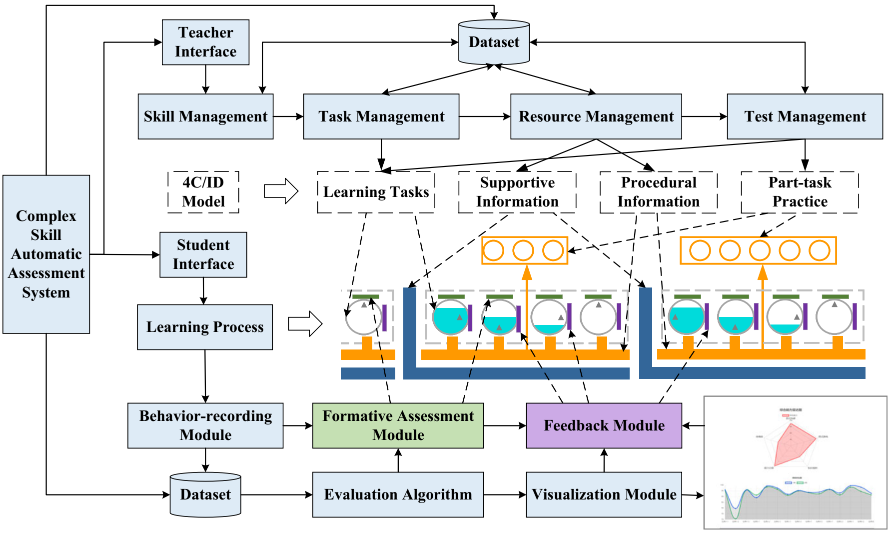


Fig. 3 The system architecture of CSAAS

are presented to students in order of ease, difficulty, high support, and low support per the 4C/ID model guidance (Van Merriënboer & Kirschner, 2017). For example, a task sequence starts with a worked example, where students complete the task following the example. Next is a completion task, where students need to complete the task on a solution with preliminary ideas. For example, when students complete the programming task, they are given a pre built programming environment, which needs to be built by the students in the conventional task. Finally, there is a conventional task where students find their own solutions. Thirdly, the resource management module is the hub of information provided by teachers. For example, teachers can provide supportive information (e.g., textbook chapters) to help students understand the rules critical for complex skills. They can also provide procedural information (e.g., videos) through this module to demonstrate the procedures of how to complete complex skills (Frerejean et al., 2019). Finally, the test management module is used to disseminate part-tasks practice, which promotes students' proficiency in knowledge rules through these practices (Van Merriënboer & Kirschner, 2017).

3.2.2 Provide formative assessment

Firstly, students attend classes offline and complete learning tasks online. During this process, the system provides support such as teaching videos, guidance documents, and exercises. During the process of students completing learning tasks, the behavior-recording module records students' learning behavior data to support subsequent formative assessments. Secondly, the formative assessment module mainly includes three types of tests: scenario-based tasks (Fig. 7), subject knowledge tests

(Fig. 8), schema tasks (Fig. 9). Finally, after each learning task is completed, the evaluation algorithm module will automatically rate the student's evaluation and visualize the results through the visualization module and feedback module. The feedback module is presented on the student and teacher interfaces, which includes both real-time individual and class reports. The system records the learning behavior of students through the behavior-recording module, and calculates information such as answering time (seconds), whether the answer is correct or not (0,1), score (0-100), accuracy (%), and review times through evaluation algorithms. These information are displayed to students through individual reports in the visualization module (Fig. 10). The provision of such information can encourage students to monitor and regulate their learning of complex skills (Marchisio et al., 2018). The partial content of the class reports are shown in Figs. 11 and 12, which includes comparison charts of different classes and overall data of all students on a certain test or question, such as average score, average time spent, and answer details. In summary, we offer three different types of feedback including knowledge of results, knowledge of correct response, and elaborated feedback (Mertens et al., 2022), and present them in a visualized form. Teachers can draw on the person and the class reports to diagnose student learning and provide cognitive feedback along with corrective feedback during the complex skill learning process.

The basic process of using CSAA for classroom teaching is shown in Fig. 4. Firstly, the teacher teaches course knowledge in the classroom and adjusts teaching based on feedback from CSAAS. Secondly, students learn knowledge, rules, and skills through listening or group collaboration in the classroom. Next, Students log in to the system and acquire complex skills by completing the tests published by the teacher. The CSAAS automatically records student behaviors, including the time allocated for each question and the frequency of reviewing and seeking other information. Finally, teachers and students view reports through the feedback module and make teaching and learning adjustments. Students go through the above four steps once for each learning task they complete, until all the learning tasks are completed.

4 Method

As mentioned in Section 3.1, we initially elucidated the seamless integration of formative assessment within the 4C/ID model framework. Subsequently, we explored the amalgamation of the enhanced 4C/ID model with computer-supported formative assessment, and introduce an automated assessment system tailored for the learning of complex skills. Next, we further validate the impact of computer-supported formative assessment on students' learning of complex skills across various aspects through the

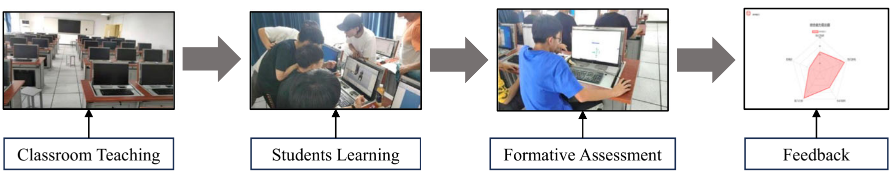


Fig. 4 The basic process of using CSAAS for classroom teaching

following experiment. This experiment was approved by the Institutional Research Ethics Board prior to its implementation, and participants signed the consent form before participating in the present study. During the experimental process, sensitive data collected was securely stored in the researchers' personal database and treated with strict confidentiality. All data will be completely destroyed within five years after the conclusion of the research.

4.1 Participants

The participants were 54 second-year college students who majored in industrial robotics in a vocational college in Eastern China. They were enrolled in the course entitled "Operation and Programming of Industrial Robots", and none of the participants had studied this course. They gave informed consent and agreed to participate in this study. Therefore, they fully understood the purpose and procedures of the study. Then, they were assigned to either the experimental group or the control group based on their existing class assignments, taught by the same teacher. This quasi-experimental design was chosen to maintain consistency in teaching content and style, minimizing their potential impact on the study's outcomes. Each group consisted of 27 participants. The experimental group used the system for formative assessment, while the control group only used the system to collect data. Participants of two groups were 19 years old on average, and belong to the same region. Prior to the experiment, we conducted a pretest on students from both groups regarding course-related conceptual knowledge. The pretest consisted of 10 multiple-choice items and had a duration of 30 minutes. The content of the pretest was selected by subject domain experts to ensure it covered key concepts relevant to the course. The result of independent sample t-test indicated that the difference in the pretest scores of the experiment ($M = 48.89$, $SD = 19.84$) and the control group ($M = 47.04$, $SD = 16.60$) did not reach a significant level ($t = 0.37$, $p = 0.71 > 0.05$). This result indicates that both groups had a comparable level of relevant knowledge before the learning activities.

4.2 Learning task and environment

The target complex skill of this study was trajectory planning for industrial robots, a core module in the industrial robotics training course offered in vocational colleges. Trajectory planning is a fundamental part of industrial robot applications for automation of production lines. It focuses on robot programming to generate accurate and repeatable robot trajectories, which is curial to manipulating and implementing robot motion. In this module, students are expected to acquire complex knowledge and skills to design appropriate robot motion paths, which are compliant with various production lines and implement the design in programming languages. Domain experts of industrial robots further divided this skill into four sub-skills: Creating virtual robotics workstations, Determine the coordination system, Six-point locating principle, and Robot programming. Under the guidance of the enhanced 4C/ID model, our researcher group and domain experts had also designed different learning tasks and formative assessments for these sub skills, as shown in Table 1.

Table 1 Outline of the formative assessment

Sub-skills	Description	Formative assessments
Creating virtual robotics workstations	Constructing a virtual industrial robot workstation in the Robot Studio software.	Subject knowledge tests 1-3 Scenario-based tasks 1-3 Schema task 1
Determine the coordination system	Determining the moving path of the industrial robot through the control of a teaching pendant.	Subject knowledge tests 4-6 Scenario-based tasks 4-6 Schema task 2
Six-point locating principle	Setting the coordinate axes of the tools used by the tool robot are set by determining six coordinate points.	Subject knowledge tests 7-10 Scenario-based tasks 7-10 Schema task 3
Robot programming	Compiling and debugging programs to achieve the automatic and accurate moving of the industrial robot.	Subject knowledge tests 11-14 Scenario-based tasks 11-14 Schema task 4

These tasks were designed to closely align with complete real-world scenarios, aiming to enable students to fully grasp the entire process of robot trajectory planning. The structure of the tasks demonstrated a gradual increase in complexity, allowing students to methodically develop and deepen their skills. Throughout all sub-skill learning, the primary task for students was to complete the robot's trajectory planning tasks. The difference was that in the first phase, the focus was mainly on the operation of creating a virtual robot workstation. Once these basic operations were mastered, students could complete the robot's basic trajectory planning tasks by calling preset programs, which were the most elementary entry-level tasks, thus set at the beginning of the course. As students moved into the second sub-skill phase, they needed to not only perform the task of creating a virtual robot workstation but also continue to complete the coordination system determination operation, and on this basis, use the default program to complete the robot's trajectory planning. The complexity of the tasks in this stage increased compared to the first sub-skill. The setting of the third sub-skill also followed this pattern, with its complexity level raised again. By the fourth sub-skill, students needed to further learn how to program and control the robot to precisely complete the trajectory planning based on the comprehensive operations from the first three sub-skills. At this stage, students needed to integrate all the knowledge and skills they had learned before, which not only increased the task's complexity but also aligned with the teaching principles of the 4C/ID model. To deepen the learning experience, for different tasks within each sub-skill of the same complexity, varying levels of support strategies were designed. Initially, students would complete the robot's trajectory planning tasks step by step by following specific cases. Then, necessary support materials, such as operation guides for creating virtual workstations and step-by-step instructions for determining the coordination system, were provided to assist students in completing the tasks. In the final stage of the sub-skill, students needed to complete tasks independently without any external support materials. This strategy of gradually withdrawing support not only helped students consolidate the skills they had learned but also encouraged them to develop problem-solving abilities and self-learning capabilities.

The formative assessments included fourteen scenario-based tasks, fourteen subject knowledge tests, and four schema tasks, their example is shown in Appendix A. They were implemented during the process of students completing learning tasks, and the results were presented to teachers and students through the feedback module of CSAAS. The scoring of subject knowledge tests was automatically completed by the system according to the answer to the question. The scoring of scenario-based tasks and schema tasks was conducted by the domain experts.

The learning environment of this study mainly included hardware and software environments. The hardware environment included: a multimedia classroom, equipped with electronic whiteboards, high-performance computers, and other devices. The software environment included: a browser for logging into the CSAAS for formative assessment, and the Robot Studio software for simulating the operation of industrial robots.

4.3 Measures and instruments

4.3.1 Post-test for academic performance

The post-test was designed to assess students' learning performance. It consisted of the following three tasks, each designed by multiple domain experts to ensure scientific rigor and precision, as show in Appendix C.

1) Post subject knowledge test. The duration of this task was 40 minutes. It was used to assess student subject knowledge by the end of the study. It consisted of fifteen multiple-choice questions, ten fill-in-the-blank questions, ten judgment questions, and five sorting questions. For brevity, only a portion of these questions is shown in Appendix C.1.

2) Post scenario-based task. The duration of this task was 60 minutes. The students were asked to complete a robot trajectory planning task from scratch and post a task compression package and program text. This task was to assess students' skills for trajectory planning for industry robots, which was a conventional task, as show in Appendix C.2.

3) Post schema task. The duration of this task was 20 minutes. As show in Appendix C.3, in the task of completing the robot trajectory planning from scratch constructed in the schema task, the students were asked to draw a mind map. The mind map was to describe the completion steps. The mind map included pointing out the industries that should be considered. This was to successfully complete the task parameters and key points on.

The full score of these three tasks was 100. The total duration of the post-test was two hours. Students completed these tasks individually in a classroom using the CSAAS system, under the supervision of a teacher.

4.3.2 Survey scales for various aspects

In order to delve into the comprehensive impact of computer-supported formative assessment on various aspects of students in complex skill learning, we employed six scales as detailed in Appendix D. In this study, the scales used are Likert-type five-point scales. These scales were adapted from existing instruments, specifically from the works on learning attitude, perceived ease of use, and perceived usefulness (Hwang et al., 2013), learning motivation (Wang & Chen, 2010), cognitive load (Leppink et al., 2014), and perceived satisfaction (Chu et al., 2010). The adaptation process was conducted collaboratively by an associate professor specializing in educational technology and an associate professor specializing in industrial machinery. Initially, each scale was evaluated for its suitability. Then, the items were adjusted to align with the study's context and objectives. Finally, multiple rounds of discussions and revisions were carried out to finalize the scales. Their specific description is as follows.

As shown in Table 6, the **learning attitude** scale had seven items full score was 35, revised based on Hwang et al. (2013) measure. The **learning motivation** scale had six items full score was 30, were revised based on Wang and Chen (2010) measure. The Cronbach's α coefficients for the two scales are 0.97 and 0.94, respectively.sss

As shown in Table 7, the **cognitive load** scale was used to measure the cognitive resources that students invest in learning. It consisted of 15 questions, further divided into three dimensions: intrinsic, extraneous, and germane, with six, five, and four items in each dimension, respectively, revised based on Leppink et al. (2014) measure. The Cronbach's α coefficients for these dimensions were 0.97, 0.98, and 0.98, indicating high reliability. The intrinsic cognitive load is related to students' own abilities and the difficulty of learning materials, while the extraneous cognitive load is generated by inefficient teaching procedures. The germane cognitive load is the cognitive resources that students invest in storing knowledge in long-term memory (Skulmowski & Xu, 2021).

As shown in Table 8, the **perceived ease of use** was the degree to which the students believed that the learning approach was easy to use. It had seven items full score was 35, revised based on Hwang et al. (2013) measure. The **perceived usefulness** was the degree to which the students believe that using the learning approach would enhance their learning performance. It had six items and full score is 30, were revised based on Hwang et al. (2013). The **perceived satisfaction** referred to students' degree of satisfaction with the learning approach. It had seven items and full score was 35, were revised based on Chu et al. (2010). The perception scales indicated high reliability, evidenced by the values of Cronbach's α: 0.96, 0.96, and 0.99, respectively.

The rationale behind selecting these six scales lies in their coverage of a wide range of learning-related variables, effectively capturing multidimensional feedback and experiences of students in both psychological and cognitive aspects during the learning process (Morris et al., 2021; Tapingkae et al., 2020). Furthermore, we utilized different independent tools and means to collect data from multiple sources, facilitating mutual validation. This approach contributes to ensuring the comprehensiveness and thoroughness of the analysis of experimental results. Overall, these efforts contribute to a holistic understanding of the role of computer-supported formative assessment in complex skill learning, enhancing the credibility and reliability of research outcomes.

4.4 Procedures

In this experiment, a total of 14 lessons were designed, each lasting 90 minutes, with two lessons scheduled per week. The experimental procedure is shown in Fig. 5. To ensure informed consent, all participants signed an informed consent form before the experiment, and the teacher provided an introduction to the course background and assessment requirements, ensuring their full understanding of the experiment's purpose and academic expectations. Guided by the 4C/ID model, the teaching of four sub-skills was organized in order of increasing complexity. Each sub-skill consisted of 3-4 learning tasks arranged in the sequence of example tasks, completion tasks, and conventional tasks. The first three lessons focused on the creating virtual robotics workstations, followed by three lessons on the determine the coordination systemm. The subsequent four lessons emphasized the six-point locating principle, while the final four lessons concentrated on the robot programming. In each lesson, students were required to complete formative assessments on the CSAAS, including a scenario-based task and a subject knowledge test. The first lesson of each sub-skill also included a schema task. Students in the experimental group received timely feedback from the system after completing formative assessments, and teachers made teaching interventions based on the evaluation results. The control group followed the same teaching process, but data obtained were not subject to formative evaluation, and no learning feedback was provided. One week after the end of the teaching, all students completed post-test and various survey scales. Teachers provided additional remedial

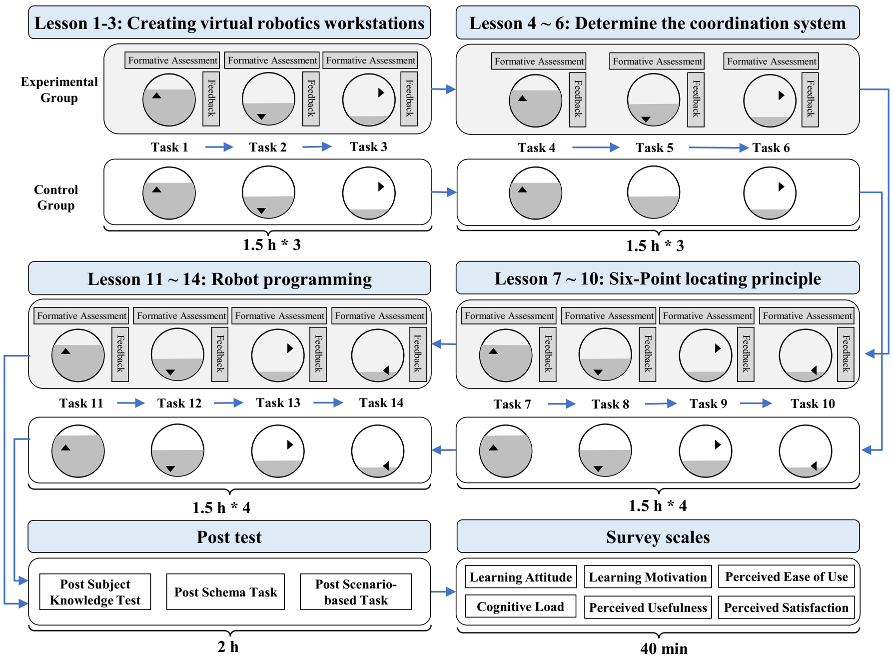


Fig. 5 The procedures of this experiment

lessons for the control group to ensure they could compensate for any learning losses caused by the experiment.

4.5 Data analysis

Firstly, to validate the feasibility and effectiveness of computer-supported formative assessment based on the enhanced 4C/ID model, descriptive statistical methods were employed to process and analyze formative assessment data. Additionally, a line graph was generated to visually illustrate the trends in the data.

Secondly, the purpose of this study is to explore the differences between the experimental group and the control group across multiple dimensions, including post-test scores, learning attitudes, learning motivation, cognitive load, and system perceptions. To achieve this, we employed the independent samples *t*-test, a method suitable for detecting whether there are significant differences in the means of two independent samples (different groups), which aligns well with our research questions. Then we verified the parameter hypothesis and the stability of the results of the independent sample *t*-test through the following four steps.

(1) Independence assumption verification. This study collected data on various continuous variables for both the experimental and control groups. These data are independent and continuous, meeting the independence assumption of the independent samples *t*-test.

(2) Normality assumption verification. This study used histograms and Q-Q plots to examine the normality of the distributions for all dimensions. The results indicated that the dimensions of learning attitude, extraneous cognitive load, perceived usefulness, ease of use, and satisfaction did not show clear characteristics of normal distribution. Therefore, we supplemented these dimensions with non-parametric tests (such as the Mann-Whitney U-test) to verify the reliability of the results, as detailed in the fourth step of the analysis. For the other dimensions, the results indicated that their data generally conformed to normal distribution characteristics: the histograms showed approximately symmetrical bell-shaped curves, and most data points were close to the line in the Q-Q plots. Given that the independent samples *t*-test has a certain tolerance for deviations from normality (Kyun & Hong, 2019; Weaver, 2011), we used the independent samples *t*-test for the analysis of these dimensions.

(3) Homogeneity of variance assumption verification. This study used Levene's test to verify the homogeneity of variance across all dimensions. The results of Levene's test showed that (Table 9 of Appendix E), except for germane cognitive load ($F = 4.30$, $p = 0.04$), post subject knowledge test ($F = 4.55$, $p = 0.04$) and post schema task ($F = 4.16$, $p = 0.05$), the variances of other dimensions were consistent ($p > 0.05$), satisfying the homogeneity of variances assumption of the independent samples *t*-test (Student's *t*-test). For the data on germane cognitive load, post subject knowledge test and post schema task, due to the heterogeneity of variances, we used Welch's *t*-test, a common method for handling data with unequal variances (Delacre et al., 2017).

(4) Result reliability verification. To ensure the reliability of the results, we conducted both non-parametric tests (Mann-Whitney U-test) and parametric tests (Independent samples *t*-test), with both yielding consistent results. The results of

Mann-Whitney U-test are shown in Table 10 of Appendix F. Therefore, we chose to report the results of the parametric tests, as they are more intuitive for most readers and are widely used in fields such as nursing, psychology, education, and public health.

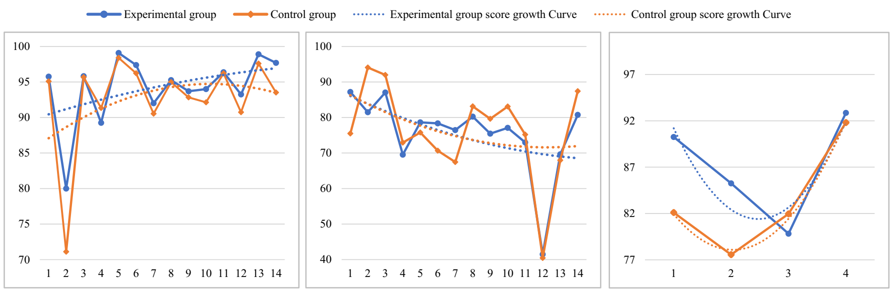
Thirdly, in the above independent samples *t*-test, a two-tailed test was used, and the statistical significance level was set at an alpha of 0.05. Regarding the handling of research hypotheses, we adopted the following strategies. For each research question, if the results of all dimensions in the research question are not significant, we accept the null hypothesis; if all dimensions are significant, we reject the null hypothesis and accept the alternative hypothesis; if some dimensions are significant, we discuss the significant dimensions separately.

Finally, we used Cohen's *d* to calculate the effect size with the support of Lenhard and Lenhard (2022). We also suggested that effects were negligible when the effect size was less than 0.2, small when the effect size was greater than or equal to 0.2 but less than 0.5, moderate when the effect size was greater than or equal to 0.5 but less than 0.8, and large when the effect size was greater than 0.8.

Through this series of analyses, we were able to explore variations in student performance across various aspects such as academic performance, learning attitude and motivation, cognitive load, and system perceptions, to ensure the comprehensiveness and reliability of the analysis.

5 Results

In this section, we first present the results of all formative assessments, as shown in Fig. 6. Next, we present the results for each research question and test the null and alternative hypotheses for each research question. Specifically, Table 2 presents the independent samples *t*-test results of the post-test scores for the experimental and control groups. Tables 3, 4 and 5 present the independent samples *t*-test results of the scores for learning attitude, learning motivation, intrinsic cognitive load, extraneous cognitive load, germane cognitive load, total cognitive load, perceived usefulness, perceived satisfaction, and perceived ease of use for the two groups.



a) Scenario-based tasks b) Subject knowledge tests c) Schema tasks

Fig. 6 Formative assessments results

Table 2 Results of the descriptive and t-test statistics of the post-test

Tests	Experimental group		Control group			
	M	SD	M	SD	t	d
Post subject knowledge test	61.85	19.34	52.37	14.84	2.02*	0.55
Post scenario-based task	94.19	4.48	91.3	4.05	2.49*	0.68
Post schema task	91.18	3.13	86.32	8.14	2.89**	0.79

Note: $*p < 0.05$. $**\ p < 0.01$. The sample size of both the experimental group and the control group is 27. M represents the mean, SD represents the standard deviation, t represents the t-test statistic, and d represents the effect size

Table 3 The results of independent sample t-tests for learning attitude and motivation

Dimensions	Experimental group		Control group			
	M	SD	M	SD	t	d
Learning attitude	30.81	4.76	29.85	4.56	0.45	0.21
Learning motivation	26.30	3.88	24.81	3.52	1.47	0.40

Note: $*p < 0.05$. $**\ p < 0.01$. The sample size of both the experimental group and the control group is 27. M represents the mean, SD represents the standard deviation, t represents the t-test statistic, and d represents the effect size

Table 4 The results of independent sample t-tests for cognitive load

Dimensions	Experimental group		Control group			
	M	SD	M	SD	t	d
Intrinsic cognitive load	17.93	7.48	18.19	6.10	−0.14	−0.04
Extraneous cognitive load	14.07	6.03	13.96	5.91	0.07	0.02
Germane cognitive load	12.89	5.11	12.11	4.06	0.62	0.17
Total cognitive load	44.89	17.60	44.26	15.62	0.14	0.04

Note: $*p < 0.05$. $**\ p < 0.01$. The sample size of both the experimental group and the control group is 27. M represents the mean, SD represents the standard deviation, t represents the t-test statistic, and d represents the effect size

Table 5 The results of independent sample t-tests for system preception

Dimensions	Experimental group		Control group			
	M	SD	M	SD	t	d
Perceived usefulness	26.63	4.25	24.48	4.12	2.13*	0.51
Perceived satisfaction	30.59	6.44	24.41	8.56	3.00**	0.82
Perceived ease of use	29.70	4.45	27.63	5.39	1.54	0.42

Note: $*p < 0.05$. $**\ p < 0.01$. The sample size of both the experimental group and the control group is 27. M represents the mean, SD represents the standard deviation, t represents the t-test statistic, and d represents the effect size

5.1 Results of formative assessments

The formative assessment results for both the experimental group and the control group are illustrated in Fig. 6. On the x-axis, it has the sequence number of the formative assessments, while the y-axis represents the mean score of each group for each assessment. Examining the solid lines representing the score curves, we observed an interleaving phenomenon, indicating that the scores of the two groups did not follow a discernible pattern. To analyze the evolving trends in each group's scores, quadratic polynomial growth curves (depicted as dotted lines) were fitted for both groups. Figure 6a reveals that, in the case of scenario-based tasks, the experimental group exhibited a gradual upward trend. In contrast, the control group initially improved but then declined, with the experimental group consistently scoring higher overall. Turning to Fig. 6b, we observed a slow decline in scores for both groups on subject knowledge tests. Although the experimental group started on par with the control group, it concluded with a lower overall score than the control group. In Fig. 6c, focusing on schema tasks, the experimental group consistently outperformed the control group. Both groups displayed a trend of initially decreasing and then rising scores. In summary, the results provide insights into the formative assessment outcomes. Specifically, the experimental group demonstrated varying trends across different assessment types, showcasing both advantages and challenges compared to the control group.

5.2 RQ1: Does the intervention significantly improve students' post-test scores?

For RQ1, the results of the independent sample t-test for post-test scores are shown in Table 2. In terms of post subject knowledge test, the average score of experimental group ($M = 61.85, SD = 19.34$) was higher than that of control group ($M = 52.37, SD = 14.84$). The t-test results indicated a significant difference between the two groups ($t = 2.02$, $p < 0.05$), with a moderate effect size ($d = 0.55$). This suggests that the experimental group outperformed the control group in subject knowledge.

In terms of post scenario-based task, the average score of experimental group ($M = 94.19, SD = 4.48$) was higher than that of control group ($M = 91.3, SD = 4.05$). The t-test results showed a significant difference between the two groups ($t = 2.49$, $p < 0.05$), with a moderate effect size ($d = 0.68$). This indicates that the experimental group performed significantly better on the scenario-based task compared to the control group.

In terms of post schema task, the average score of experimental group ($M = 91.18, SD = 3.13$) was higher than that of control group ($M = 86.32, SD = 8.14$). The t-test results revealed a significant difference between the two groups ($t = 2.89$, $p < 0.01$), with a moderate effect size ($d = 0.79$). This suggests that the experimental group performed significantly better on the schema task compared to the control group.

In summary, the results indicate that the experimental group scored significantly higher than the control group on all post-test measures. Therefore, the null hypothesis (H0) is rejected, supporting the alternative hypothesis (H1), which states that computer-supported formative assessment significantly improves students' post-test scores.

5.3 RQ2: Does the intervention significantly improve students' learning attitudes and motivation?

For RQ2, the results of the independent sample t-test for learning attitudes and motivation are shown in Table 3. In terms of learning attitude, the average score of experimental group ($M = 30.81$, $SD = 4.76$) was slightly higher than that of control group ($M = 29.85$, $SD = 4.56$). The t-test results indicated no statistically significant difference between the two groups ($t = 0.45$, $p > 0.05$), with a small effect size ($d = 0.21$). This suggests that computer-supported formative assessment did not significantly change students' learning attitudes.

In terms of learning motivation, the average score of experimental group ($M = 26.30$, $SD = 3.88$) was higher than that of control group ($M = 24.81$, $SD = 3.52$). The t-test results indicated no statistically significant difference between the two groups ($t = 1.47$, $p > 0.05$), with a moderate effect size ($d = 0.40$). This suggests that the improvement in learning motivation was not significant in the experimental group.

In summary, the results indicate that while the experimental group scored higher than the control group in learning attitudes and motivation, these differences were not statistically significant. Therefore, the null hypothesis (H0) is not rejected, and the alternative hypothesis (H1) is not supported, indicating that computer-supported formative assessment does not significantly improve students' learning attitudes and motivation.

5.4 RQ3: Does the intervention significantly increase students' cognitive load?

For RQ3, the results of the independent sample t-test for cognitive load are shown in Table 4. In terms of intrinsic cognitive load, the average score of experimental group ($M = 17.93$, $SD = 7.48$) was slightly lower than that of control group ($M = 18.19$, $SD = 6.10$). The t-test results indicated no statistically significant difference between the two groups ($t = -0.14$, $p > 0.05$), with a negligible effect size ($d = -0.04$). This suggests that the complexity of the task remained stable, regardless of the introduction of computer-supported assessments.

In terms of extraneous cognitive load, the average score of experimental group ($M = 14.07$, $SD = 6.03$) was slightly higher than that of control group ($M = 13.96$, $SD = 5.91$). The t-test results showed no statistically significant difference between the two groups ($t = 0.07$, $p > 0.05$), with a negligible effect size ($d = 0.02$). This suggests that the computer-supported instructional design did not add unnecessary extraneous cognitive load.

In terms of germane cognitive load, the average score of experimental group ($M = 12.89$, $SD = 5.11$) was slightly higher than that of control group ($M = 12.11$, $SD = 4.06$). The t-test results showed no statistically significant difference between the two groups ($t = 0.62$, $p > 0.05$), with a negligible effect size ($d = 0.17$). This suggests that using technology did not significantly enhance the deep processing of the material.

In terms of total cognitive load, the average score of experimental group ($M = 44.89$, $SD = 17.60$) was slightly higher than that of control group ($M = 44.26$, $SD = 15.62$). The t-test results indicated no statistically significant difference between the two groups ($t = 0.14$, $p > 0.05$), with a negligible effect size ($d = 0.04$).

In summary, the results indicate that there were no significant differences between the experimental and control groups across all dimensions of cognitive load. Therefore, the null hypothesis (H0) is not rejected, and the alternative hypothesis (H1) is not supported, indicating that computer-supported formative assessment does not significantly increase students' cognitive load.

5.5 RQ4: Does the intervention significantly improve students' perceived ease of use, usefulness, and satisfaction with the learning system?

For RQ4, the results of the independent sample t-test for system preception are shown in Table 5. In terms of perceived usefulness, the average score of experimental group ($M = 26.63$, $SD = 4.25$) was higher than that of the control group ($M = 24.48$, $SD = 4.12$). The t-test results indicated a significant difference between the two groups ($t = 2.13$, $p < 0.05$), with a moderate effect size ($d = 0.51$). This suggests that students using computer-supported formative assessment found the learning system more helpful for their learning performance.

In terms of perceived satisfaction, the average score of experimental group ($M = 30.59$, $SD = 6.44$) was higher than that of control group ($M = 24.41$, $SD = 8.56$). The t-test results showed a significant difference between the two groups ($t = 3.00$, $p < 0.01$), with a large effect size ($d = 0.82$). This suggests that using computer-supported formative assessment positively impacted students' satisfaction with the learning system.

In terms of perceived ease of use, the average score of experimental group ($M = 29.70$, $SD = 4.45$) was slightly higher than that of control group ($M = 27.63$, $SD = 5.39$). The t-test results indicated no statistically significant difference between the two groups ($t = 1.54$, $p > 0.05$), with a moderate effect size ($d = 0.42$). Although the experimental group scored slightly higher, this suggests that the introduction of computer-supported assessment did not significantly change students' perceptions of the system's ease of use.

In summary, the results indicate that the experimental group scored significantly higher than the control group in perceived usefulness and satisfaction, but not in perceived ease of use. Therefore, the null hypothesis (H0) is partially rejected, partially supporting the alternative hypothesis (H1), which states that computer-supported formative assessment significantly improves students' perceived usefulness and satisfaction with the learning system, but does not significantly improve perceived ease of use.

6 Discussion

6.1 The learning effect of complex skills in computer-supported formative assessment

Complex skill learning is a dynamic process, and the formative assessment of that needs to include different levels and aspects of skills. Based on the 4C/ID model, we designed three assessments including schema construction, subject knowledge, and mastery of complex skills as well as different types of feedback. And we used computer technology to incorporate these assessments and feedback into complex skills learning.

In order to verify the effectiveness of the computer-supported formative assessment on the mastery of complex skills, we assessed students' achievements during and after the complex skill learning. Firstly, the scores of both the experimental group and the control group on scenario-based tasks showed a slow upward trend in the first half. It indicated that as teaching progressed, students completed more and more tasks, and their proficiency in complex skills continued to improve. This is consistent with our general knowledge that as teaching progresses, students will continuously grasp the teaching content. Since the experimental group received feedback every time, the experimental group not only maintained an upward trend, but also overall surpassed the control group. The control group then showed a downward trend, which indirectly demonstrated the effectiveness of using CSAAS for formative assessment. This was consistent with previous research findings, which suggested that formative assessment could help students master complex skills (Ackermans et al., 2017).

Secondly, a noteworthy phenomenon was that the scores of both groups on subject knowledge tests showed a slow downward trend. One reason for this was that as the teaching process progresses, the amount of knowledge students needed to master also increases. As a result, students needed to devote more cognitive load to learning this increasing amount of material, which can lead to lower test scores (Kirschner, 2002; Larmuseau et al., 2019). On the other hand, in subject knowledge test 12, the scores of both groups showed a sharp decline, indicating that the test difficulty was too high and did not match the students' actual abilities. Therefore, after timely adjustment of the teaching content and assessment difficulty, the assessment was better able to accurately reflect the true level of the students (Maier et al., 2016).

Finally, in the schema task, both groups showed a trend of decreasing scores followed by increasing scores. This related to the teaching process, as the teacher provided guidance in the first schema task, while subsequent tests were completed independently by students, leading to an initial downward trend in grades. With continuous learning, self-reflection, and teaching feedback for students, grades also showed an upward trend over time (Kuklick et al., 2023; Roscoe & Craig, 2022; Tempelaar et al., 2013). Although there were fluctuations in the results of formative assessments, this study utilized CSAAS to provide formative assessment to both teachers and students. This helped teachers adjust their teaching methods and helped students engage in self-reflection. The ultimate goal is to improve students' mastery of complex skills and enhance their performance on post-test.

In investigating the effectiveness of the computer-supported formative assessment intervention in improving students' post-test scores, it was found that the post-test scores of the experimental group were significantly higher than those of the control group in terms of post scenario-based task, post subject knowledge test, and post schema task. Despite fluctuations in outcomes in formative assessments, the final learning achievements of students in the experimental group were significantly better than the control group, indicating the effectiveness of computer-supported formative assessment on complex skill learning. This aligns with the results of existing research. On one hand, Xu et al. (2023)'s study suggests that computer-based formative assessment can effectively improve the learning outcomes of complex skills. Through timely feedback and personalized guidance, students can master skills more quickly and correct mistakes. On the other hand, studies related to the 4C/ID model also indicate that by providing a structured learning environment and tasks, this model can effectively enhance students' mastery of complex skills (Hosseinzadeh et al., 2023; Kukharuk et al., 2023). Our research combines these two aspects by integrating computer-supported formative assessment into the theoretical framework of the 4C/ID model, further proving that this approach can significantly improve the learning outcomes of complex skills. The enhancement brought by this combination is comprehensive, including the mastery of knowledge and skills, schema construction, and the transfer of abilities, helping students solve complex problems in new and broader contexts. This also aligns with the emphasis on 21st-century skills (Van Laar et al., 2020).

In addition, like the results of formative assessment, both the experimental group ($M = 61.85$, $SD = 19.34$) and the control group ($M = 52.37$, $SD = 14.84$) scored lower on post subject knowledge test, while they scored relatively higher on post scenario-based task and post schema task. There are two possible reasons. First, the system did not provide sufficient support for the acquisition of subject knowledge essential for solving problems. Secondly, prior research found that students in vocational colleges prefer to learn hands-on operational skills rather than abstract knowledge rules like concepts and rules (Xu et al., 2019). Therefore, additional supportive information is necessary to facilitate the acquisition of knowledge rules and pay attention to students' mastery of subject knowledge in daily teaching (Musharyanti et al., 2021).

Overall, this addresses the first research question. The CSAAS in this study facilitated students' development of complex skills and assessed their learning process to generate feedback for teachers, who could then adapt their instruction accordingly. Despite some fluctuations in the formative assessments, the post-test results demonstrated that CSAAS could help improve the development of complex skills.

6.2 The impact of computer-supported formative assessment on students in various aspects

In this study, we conducted a comprehensive examination of the impact of computer-supported formative assessment on various aspects of students, including learning attitude, motivation, cognitive load, and system perception.

In investigating the effectiveness of the computer-supported formative assessment intervention in improving students' learning attitude and motivation, it was found that, contrary to previous studies, computer-supported formative assessment did not significantly improve students' learning attitude and motivation. This differs from the findings of Chu et al. (2019), who observed positive effects in both pre-test and post-test longitudinal analyses of experimental and control groups. However, our study conducted a cross-sectional analysis of post-test data for both groups, revealing that the addition of formative assessment did not directly influence students' learning attitude and motivation. Although there are some differences, the research results all point in one direction: computer-supported formative assessment do not diminish students' learning attitudes and motivation.

In investigating the concern of computer-supported formative assessment intervention on increasing students' cognitive load (Kuklick et al., 2023; Larmuseau et al., 2019), it was found that there were no significant differences in cognitive load dimensions between the two groups. This can be attributed to the attention to cognitive load within the 4C/ID model, including task sequencing from difficult to easy, providing timely and effective support procedures, and incorporating relevant knowledge designs (Van Merriënboer, 2019). Additionally, integrating formative assessment into learning tasks and offering timely feedback further reduced the likelihood of increased cognitive load for students. In a study of 39 students, Chu et al. (2019) pointed out that formative assessment in augmented reality environments did not increase cognitive load on students. Other studies using the 4C/ID model for teaching also demonstrated no significant differences in cognitive load among two groups of students (Xu et al., 2023). This finding alleviates concerns about increased cognitive load leading to a decline in learning performance, supporting the positive role of computer-supported formative assessment in complex skill learning processes.

In investigating the effectiveness of the computer-supported formative assessment intervention in improving students' system perception, it was found that the experimental group exhibited higher perceived usefulness and satisfaction with the system compared to the control group, consistent with the results of previous studies (Agustina & Purnawarman, 2020; Tapingkae et al., 2020). Most students in this study agreed or strongly agreed that the system helped them learn new knowledge, expressing a desire for other subjects to adopt the system. This may be attributed to the experimental group receiving more evaluations and feedback through the system, while the control group only used the system for data collection. This underscores the critical role of formative assessment in the learning process, particularly when combined with computer support. Regarding system perceived ease of use, it refers to how easy students perceive it to use the proposed system (Alshurideh et al., 2019). While there was no significant disparity in perception between the two student groups, the overall low scores suggest existing ease of use issues. Post-experiment feedback indicated concerns such as occasional delays in system response time and some aspects of interface design not being aesthetically pleasing. This feedback provides valuable suggestions for further iteration and system upgrades to enhance the user experience. In a meta-analysis study, Al-Fraihat et al. (2020) pointed out that although the ease of use of a system is an important feature, it does not have a significant impact on the actual use of learning systems. This might imply that while ease of use can enhance users' initial acceptance, it is not a decisive factor in long-term usage. In contrast, system satisfaction and usefulness have a more significant impact on the use of learning systems. This suggests that whether users feel the system meets their needs and provides value is a key factor in their continued use of the system.

In summary, this study emphasizes the positive role of computer-supported formative assessment in students' complex skill learning processes. It also offers practical guidance by providing suggestions to optimize system usability and further expand the impact on learning attitudes and motivation. The research holds practical significance in guiding the effective integration of formative assessment to enhance students' learning experiences and outcomes.

7 Conclusion

Since the beginning of the 21st century, the ability to solve complex problems has become increasingly important, emerging as an essential skill in daily life and career development. Cultivating students with complex skills to meet societal needs has emerged as an urgent priority. The 4C/ID model, as an instructional design framework, plays a pivotal role in facilitating complex skill learning. However, the 4C/ID model lacks explicit guidelines for formative assessment. Therefore, integrating computer-supported formative assessment with the 4C/ID model is necessary to better support complex skill learning. This study proposes an enhanced 4C/ID model and develops a system based on it to support complex skill learning, including formative assessment. Through a controlled experiment, this study presents the following conclusions.

First, the study results indicate that computer-supported formative assessment integrated with the 4C/ID model significantly enhance students' mastery of complex skills, demonstrating the high effectiveness of this method in teaching complex skills. Therefore, teachers can consider using the 4C/ID model in course design and utilizing computer-supported formative assessment tools to help students more effectively master complex skills. Schools and educational institutions should invest in developing and promoting these tools to enhance teaching quality.

Second, the study results show that students' learning attitudes and motivation did not significantly increase. This suggests that merely using computer-supported formative assessment tools is insufficient to boost students' learning motivation and attitudes. Therefore, it is necessary to explore additional motivational mechanisms to stimulate students' learning motivation and consider multiple factors for enhancing students' learning attitudes. For example, incorporating reward mechanisms, setting challenging tasks and goals in computer assessments, or providing more opportunities for autonomous learning can help boost students' learning motivation. Teachers can improve students' learning attitudes through more personalized teaching methods, increased teacher-student interaction, and peer assessment.

Additionally, the study results indicate that this method did not significantly increase students' cognitive load, due to the effective implementation of the 4C/ID model in mitigating cognitive load. This study meticulously designed a series of tasks for the target complex skill, presenting them in a sequence that progressed from simple to complex and from high support to low support, while providing both supportive and procedural information throughout the teaching process. This approach ensured that the cognitive load of students in both groups remained consistent. Therefore, this finding is significant for the field of computer-supported formative assessment. It demonstrates that, with the support of the 4C/ID model, computer-supported formative assessment tools can deliver effective feedback without increasing students' cognitive load. This offers robust support and valuable insights for the widespread application and further optimization of these tools in educational settings.

Finally, the study results indicate that students' perceived usefulness and satisfaction with the system significantly increased, showing that students received real learning help and support when using these tools. This demonstrates the practical utility of the 4C/ID model combined with computer-supported formative assessment. It helps promote the adoption of this combined method by more educational institutions and suggests that developers and educators should continue to optimize and improve these tools to meet students' needs. For example, the improvement in system ease of use was not significant, possibly because the design and operational complexity of the system did not meet students' expectations, indicating further enhancement of the user experience is needed. The development team needs to engage in more communication and testing with educators and students to understand their needs and usage habits, thereby improving system usability.

Overall, this study shows that computer-supported formative assessment integrated with the 4C/ID model have significant advantages in educational practice but also highlights areas needing further improvement. This provides valuable insights and guidance for the future development of educational technology and teaching practices.

8 Limitations and future directions

There are still some shortcomings in this study, we identified and discussed the following limitations, and proposed future improvements and research directions.

Firstly, the post-test was scheduled at the end of the term, and participants' performance might have been affected by the busyness and fatigue associated with the term's end. Despite taking some incentive measures, such as teacher supervision and providing gifts (pen and notebook), these measures might not have been sufficient to fully offset the impact of end-of-term stress. Future research should consider conducting tests during less busy periods to ensure participants are able to focus and perform at their best. This can be achieved by scheduling tests in the middle of the term or during less busy periods.

Secondly, the survey in this study included 47 items utilizing a five-point Likert scale. Lengthy surveys might cause participant fatigue, thus affecting the validity and reliability of the data. Future research could reduce the number of questions while ensuring data quality or employ factor analysis to select the most representative questions.

Thirdly, the study results might not be applicable to other types of complex skills or to different educational environments. Since participants were from the same vocational school and major, which might introduce bias. Future research should expand the sample range to include students from different types of schools and disciplines, and conduct cross-cultural or cross-regional comparative studies to verify the generalizability of the results. This will help determine the applicability of the findings within broader contexts.

Fourthly, this study focused primarily on the immediate effects of the intervention without evaluating its long-term impact on learning and the retention of complex skills. Future research should include long-term follow-up assessments to observe the sustained impact of the intervention on learning and skill retention. For example, follow-up tests could be conducted at intervals of three months, six months, and one year after the intervention to assess the longevity of its effects. This will help in understanding the long-term effectiveness of the intervention.

Finally, this study did not directly compare the proposed method with traditional teaching methods, making it difficult to assess the extent of improvement brought about by the CSAAS system. Future research should design controlled experiments to directly compare the proposed system with traditional teaching methods, in order to evaluate the effects of different approaches.

Appendix A. Formative assessments interface for students

Fig. 7 Scenario-based task

Fig. 8 Subject knowledge test

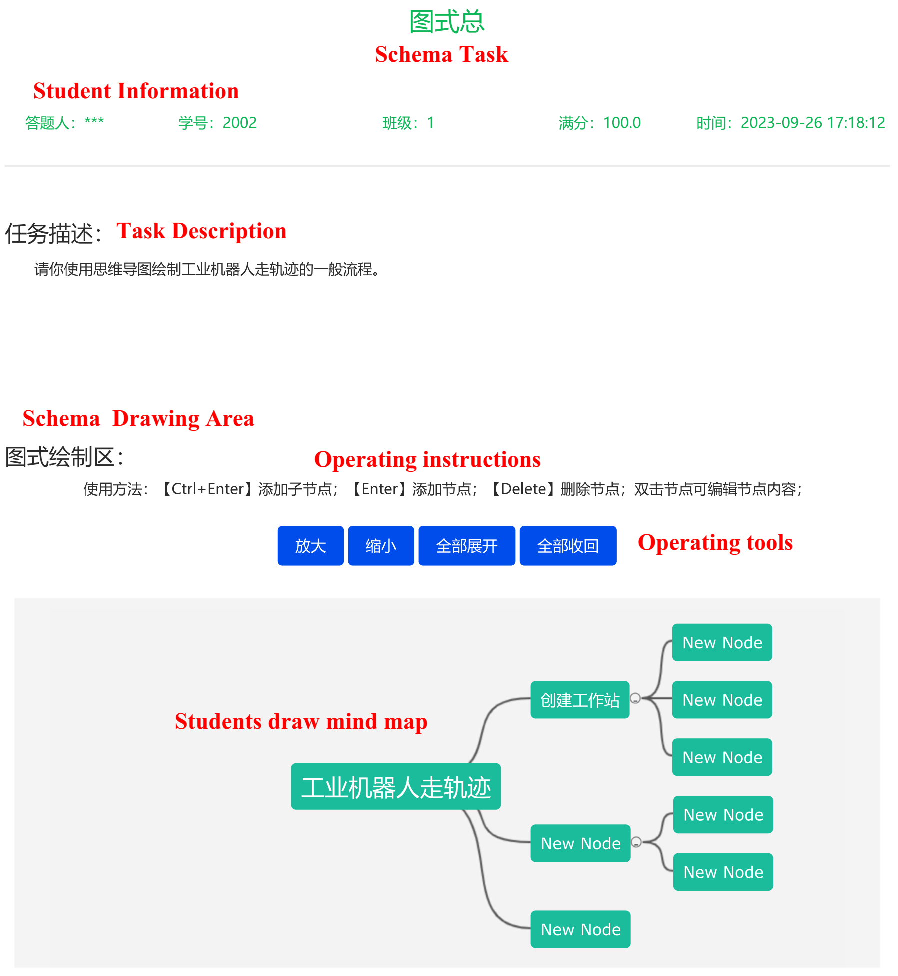


Fig. 9 Schema task

Appendix B. Interface of reports

Fig. 10 Interface of students' individual reports

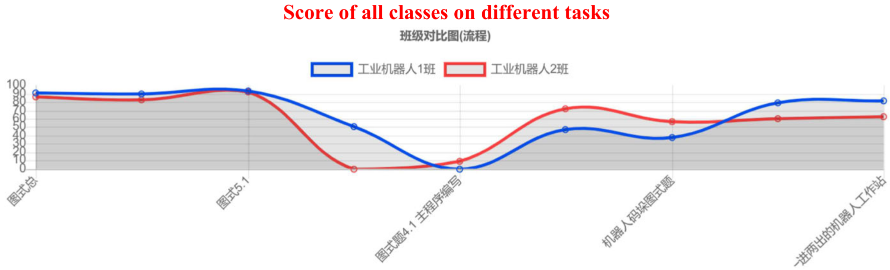


Fig. 11 Interface of class reports 1

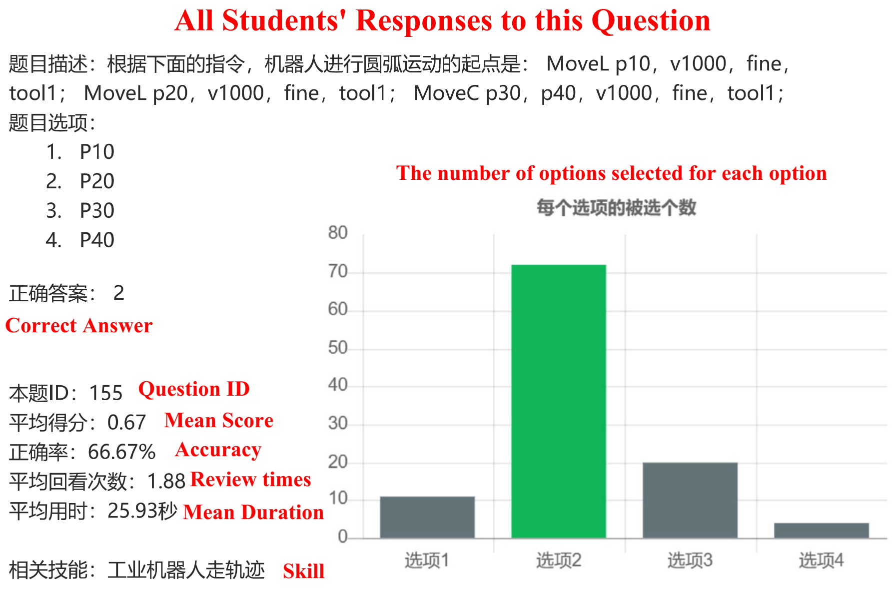


Fig. 12 Interface of class reports 2

Appendix C. Post subject knowledge test

C.1 Post subject knowledge test

Here are only part of the questions.
1. The (workpiece) refers to the object being processed in the mechanical machining process, while the tool denotes the instrument required for a robot to accomplish a specific task.
2. By default, when a single robot is in operation, the (world coordinate system) remains aligned with the base coordinate system.
3. The tool coordinate system is fixed at the end of the tool, and its coordinate origin is abbreviated as (TCP).
4. When creating tool coordinates using the six-point method in simulation software, it is advisable to switch to (B) mode when the reference point and fixed point are relatively close.
A. Normal B. Incremental C. Automatic D. Deceleration
5. When using the six-point method to create tool coordinates in simulation software, it is necessary to set (AB).
A. Center of gravity coordinates B. Tool mass C. TCP point D. Base coordinates
6. The recommended workflow for arranging peripheral devices outside the workstation is as follows: (C-D-A-B-E)
A. Rotate the external device model.
B. Directly move or use point-and-click to approximate the device’s position.
C. Import the required models.

D. Display the robot's workspace.

E. Use the "Set Position" function for fine-tuning the position.

7. In the incremental mode, the user increment in the teach pendant screen's bottom right corner can be set in size. (✓)

8. In the manual state of the robot, pressing the first gear of the enable button will stop the motors, putting the robot in a protective stop state. (×)

Note: Fill-in-the-blank questions: 1, 2, 3. Multiple-choice questions: 4, 5. Sorting question: 6. Judgment questions: 7, 8

C.2 Post scenario-based task

Task description: Please create a robotic trajectory workstation, name the workstation with your student ID, and then import necessary models such as the robot, tool, workpiece, peripheral devices, etc. Use the six-point method to determine the tool coordinates, name the tool coordinates as "tool" followed by the last two digits of your student ID, and save the corresponding TCP data. Finally, through point teaching and programming, make the robot follow the counterclockwise trajectory as shown in Fig. 13. Programming tasks include establishing initialization routines, trajectory walking routines, and returning home routines. Submit the task archive and program text upon completion.

Note: Ensure to perform programming tasks to establish initialization routines, trajectory walking routines, and returning home routines. Save all relevant data and submit the compressed task archive along with the program text.

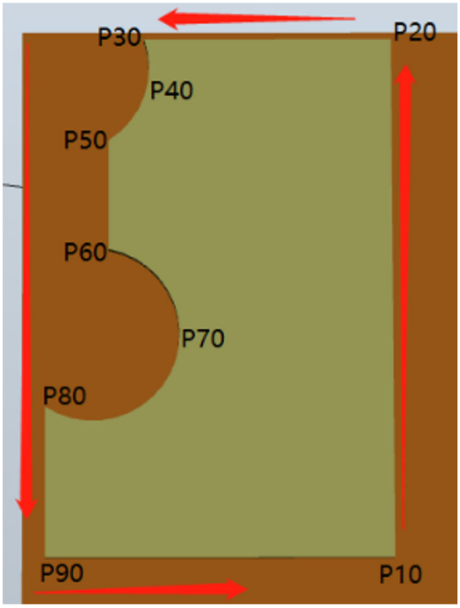


Fig. 13 The post-test for academic performance

C.3 Post schema task

Task description: Please draw a mind map illustrating the trajectory planning for operating industrial robots. Provide detailed descriptions for each step, including the purpose and significance of each step.

Note: There is an example of a student answering.

Appendix D. The survey scales for various aspects

Table 6 Learning motivation and learning attitude

Dimension	Items	Cronbach's α
Learning motivation	1. In this course, I prefer teaching content that is challenging because it allows me to learn new things.	0.94
	2. In this course, I prefer teaching content that arouses my curiosity, even if it is difficult.	
	3. If possible, I would choose courses where I can learn something, even if the grades are not high.	
	4. Getting good grades in this course is the most satisfying thing for me.	
	5. If possible, I would like to achieve the highest grades in this course.	
	6. Demonstrating excellent abilities in front of family, friends, teachers, or others is important to me.	
Learning attitude	7. I find learning this course meaningful and worthwhile.	0.97
	8. I believe acquiring knowledge related to this course is worthwhile.	
	9. I think it's worth studying this course well.	
	10. I believe learning and observing more about the content of this course is important.	
	11. I want to know more about the learning content of this course.	
	12. I will actively search for more content related to this course.	
	13. I think learning this course is important for everyone.	

Table 7 Cognitive load

Dimension	Items	Cronbach's α
Intrinsic	14. Learning the content of robots walking on trajectories is difficult for me.	0.97
	15. I feel anxious when learning the skill of robots walking on trajectories.	
	16. I often worry that I cannot complete tasks when learning the skill of robots walking on trajectories.	
	17. It took me a lot of time to master the method of robots walking on trajectories.	
	18. I have to invest a lot of effort to understand the content of the course during this period.	
Extraneous	19. I think the explanation of the content about robots walking on trajectories in the course is not very clear.	0.98
	20. I think the process explanation of robots walking on trajectories in the course is not very clear.	
	21. I think the teacher's time arrangement for course explanations, tasks, feedback, and Q&A sessions is not reasonable.	
	22. I always cannot get help in a timely manner when I encounter problems.	
	23. I think the standard explanation for completing tasks is not very clear.	
Germane	24. I have to put in a lot of effort to complete the tasks assigned in the classroom.	0.98
	25. I have to put in a lot of effort to master the complex operations of robots walking on trajectories in simulation software.	
	26. I have to put in a lot of effort to understand the process of robots walking on trajectories.	
	27. I need to take a lot of notes to understand the content of the course about robots walking on trajectories.	

Table 8 Perceptions of system

Dimension	Items	Cronbach's α
Perceived usefulness	28. I feel that using such a system enriches the content of learning activities.	0.96
	29. I feel that using this system is very helpful for me to learn new knowledge.	
	30. The learning mechanisms provided by such a system make my learning process smoother.	
	31. This system can help me obtain useful information when needed.	
	32. This system allows me to learn better.	
	33. In this learning activity, using this system is more effective than a general computer-assisted learning system.	
Perceived ease of use	34. Learning the operation of this system is not difficult for me.	0.96
	35. I only spent a short time to fully understand how to use this system.	
	36. The learning activities conducted using this system are easy to understand.	
	37. I quickly learned how to operate this system.	
	38. Using this system in this learning activity is not difficult for me.	
	39. I find all modules of this system easy to use.	
	40. Overall, the system used in this learning activity is easy to learn and use.	
Perceived satisfaction	41. Using this system for learning is more interesting than previous learning methods.	0.99
	42. Using this system for learning, I feel it can help me discover new problems.	
	43. Using this system for learning, I feel it can enable me to approach the content of learning in a new way.	
	44. I enjoy using this system for learning.	
	45. I hope other subjects can also use this system for learning.	
	46. I hope to have the opportunity to use this system for learning in the future.	
	47. I would recommend this system to other classmates.	

Appendix E. The results of Levene's test

Table 9 The results of Levene's test

Dimensions	*F*	*p*
Pre-test	0.06	0.81
Post subject knowledge test	4.55	0.04
Post scenario-based task	0.21	0.65
Post schema task	4.16	0.05
Learning attitude	0.14	0.71
Learning motivation	1.21	0.28
Intrinsic cognitive load	1.53	0.22
Extraneous cognitive load	0.16	0.69
Germane cognitive load	4.30	0.04
Total cognitive load	1.10	0.30
Perceived usefulness	0.01	0.92
Perceived satisfaction	1.87	0.18
Perceived ease of use	0.16	0.69

Appendix F. The results of Mann-Whitney U-test

Table 10 The results of Mann-Whitney U-test

Dimensions	*Z*	*p*
Pre-test	-0.70	0.49
Post subject knowledge test	-1.98	0.05
Post scenario-based task	-2.78	0.01
Post schema task	-2.84	0.01
Learning attitude	-0.71	0.48
Learning motivation	-1.58	0.12
Intrinsic cognitive load	-0.18	0.86
Extraneous cognitive load	-0.04	0.97
Germane cognitive load	-0.70	0.48
Total cognitive load	-0.12	0.90
Perceived usefulness	-2.19	0.03
Perceived satisfaction	-3.10	0.01
Perceived ease of use	-1.56	0.12

Acknowledgements The research will not have been possible without the cooperation of teachers and administrators from Shanghai Technical Institute of Electronics Information. We would particularly like to acknowledge our discussions with Dr. Wangqi Shen, who provided consultation in the preparation of this paper.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Haoxin Xu, Tianrun Deng, Xianlong Xu, Xiaoqing Gu, Lingyun Huang, Haoran Xie, and Minhong Wang. The first draft of the manuscript was written by Haoxin Xu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was funded by Key Project of Science and Technology Commission of Shanghai Municipality (17DZ2281800).

Data Availability Data will be made available on request.

Declarations

Statement Regarding Research Involving Human Participants and/or Animals The study was conducted with the approval of the East China Normal University Committee on Human Research Protection, and all subjects were adults. Prior to the start of the experiment, the subjects were informed of the purpose, method, process, and other information of the study, and written consent was obtained from all subjects.

Ethical Approval The questionnaire and methodology for this study was approved by the Human Research Ethics committee of the East China Normal University (Ethics approval number: HR692-2023).

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent to Publish The participant has consented to the submission of the case report to the journal.

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

Ackermans, K., Rusman, E., Brand-Gruwel, S., & Specht, M. (2017). A first step towards synthesizing rubrics and video for the formative assessment of complex skills. *Technology enhanced assessment* (pp. 1–10). Retrieved from https://doi.org/10.1007/978-3-319-57744-9_1

Ackermans, K., Rusman, E., Nadolski, R., Specht, M., & Brand-Gruwel, S. (2021). Video-enhanced or textual rubrics: Does the viewbrics' formative assessment methodology support the mastery of complex (21st century) skills? *Journal of Computer Assisted Learning, 37*(3), 810–824. https://doi.org/10.1111/jcal.12525

Agustina, M., & Purnawarman, P. (2020). Investigating learners' satisfaction utilizing google classroom as online formative feedback tool. *2020 6th International Conference on Education and Technology (ICET)* (pp. 26–31). Retrieved from https://doi.org/10.1109/ICET51153.2020.9276616

Alahmad, A., Stamenkovska, T., & Győri, J. (2021). Preparing pre-service teachers for 21st century skills education: A teacher education model. *GiLE Journal of Skills Development*, *1*(1), 67–86. https://doi.org/10.52398/gjsd.2021.v1.i1.pp67-86

Al-Fraihat, D., Joy, M., Masa'deh, R., & Sinclair, J. (2020). Evaluating e-learning systems success: An empirical study. *Computers in Human Behavior, 102*, 67–86. https://doi.org/10.1016/j.chb.2019.08.004

Alshurideh, M., Kurdi, B. A., & Salloum, S. A. (2019). Examining the main mobile learning system drivers' effects: A mix empirical examination of both the expectation-confirmation model (ecm) and the technology acceptance model (tam). *Advances in Intelligent Systems and Computing,* 406–417. https://doi.org/10.1007/978-3-030-31129-2_37

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education, 4*, 1–13. https://doi.org/10.3389/feduc.2019.00087

Bhagat, K. K., & Spector, J. M. (2017). Formative assessment in complex problemsolving domains: The emerging role of assessment technologies. *Journal of Educational Technology & Society*, *20*(4), 312–317. Retrieved from https://www.jstor.org/stable/26229226

Black, P. (2015). Formative assessment–an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice, 22*(1), 161–177. https://doi.org/10.1080/0969594X.2014.999643

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education)*, *21*, 5–31. https://doi.org/10.1007/s11092-008-9068-5

Brookhart, S. M. (2013). Develop a student-centered mind-set for formative assessment. *Voices from the Middle*, *21*(2), 21–25. https://doi.org/10.58680/vm201324462

Brookhart, S. M., Moss, C. M., & Long, B. A. (2010). Teacher inquiry into formative assessment practices in remedial reading classrooms. *Assessment in Education: Principles, Policy & Practice, 17*(1), 41–58. https://doi.org/10.1080/09695940903565545

Castillo-Segura, P., Fernández-Panadero, C., Alario-Hoyos, C., Muñoz-Merino, P. J., & Kloos, C. D. (2021). Objective and automated assessment of surgical technical skills with iot systems: A systematic literature review. *Artificial Intelligence in Medicine, 112*, 102007. https://doi.org/10.1016/j.artmed.2020.102007

Chu, H.-C., Chen, J.-M., Hwang, G.-J., & Chen, T.-W. (2019). Effects of formative assessment in an augmented reality approach to conducting ubiquitous learning activities for architecture courses. *Universal Access in the Information Society, 18*, 221–230. https://doi.org/10.1007/s10209-017-0588-y

Chu, H.-C., Hwang, G.-J., & Tsai, C.-C. (2010). A knowledge engineering approach to developing mindtools for context-aware ubiquitous learning. *Computers & Education, 54*(1), 289–297. https://doi.org/10.1016/j.compedu.2009.08.023

Cidral, W. A., Oliveira, T., Felice, M. D., & Aparicio, M. (2018). E-learning success determinants: Brazilian empirical study. *Computers & Education, 122*, 273–290. https://doi.org/10.1016/j.compedu.2017.12.001

Corbalan, G., Kester, L., & Van Merriënboer, J. J. (2006). Towards a personalized task selection model with shared instructional control. *Instructional Science, 34*(5), 399–422. https://doi.org/10.1007/s11251-005-5774-2

Costa, J. M., Miranda, G. L., & Melo, M. (2022). Four-component instructional design (4c/id) model: A meta-analysis on use and effect. *Learning Environments Research, 25*(2), 445–463. https://doi.org/10.1007/s10984-021-09373-y

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use welch's t-test instead of student's t-test. *International Review of Social Psychology, 30*(1), 92–101. https://doi.org/10.5334/irsp.82

Faber, J. M., & Visscher, A. J. (2018). The effects of a digital formative assessment tool on spelling achievement: Results of a randomized experiment. *Computers & Education, 122*, 1–8. https://doi.org/10.1016/j.compedu.2018.03.008

Frerejean, J., van Geel, M., Keuning, T., Dolmans, D., van Merriënboer, J. J., & Visscher, A. J. (2021). Ten steps to 4c/id: training differentiation skills in a professional development program for teachers. *Instructional Science, 49*(3), 395–418. https://doi.org/10.1007/s11251-021-09540-x

Frerejean, J., van Merriënboer, J. J., Kirschner, P. A., Roex, A., Aertgeerts, B., & Marcellis, M. (2019). Designing instruction for complex learning: 4c/id in higher education. *European Journal of Education, 54*(4), 513–524. https://doi.org/10.1111/ejed.12363

Frerejean, J., Van Merriënboer, J. J., Condron, C., Strauch, U., & Eppich, W. (2023). Critical design choices in healthcare simulation education: a 4c/id perspective on design that leads to transfer. *Advances in Simulation, 8*(1), 5. https://doi.org/10.1186/s41077-023-00242-7

Herde, C. N., Wüstenberg, S., & Greiff, S. (2016). Assessment of complex problem solving: What we know and what we don't know. *Applied Measurement in Education, 29*, 265–277. https://doi.org/10.1080/08957347.2016.1209208

Heritage, M. (2020). Getting the emphasis right: Formative assessment through professional learning. *Educational Assessment, 25*(4), 355–358. https://doi.org/10.1080/10627197.2020.1766959

Hosseinzadeh, A., Karami, M., Rezvanian, M. S., Rezvani, M. S., Bahmani, M. N. D., & Merriënboer, J. V. (2023). Developing media literacy as complex learning in secondary schools: the effect of 4c/id learning environments. *Interactive Learning Environments,* 1–16. https://doi.org/10.1080/10494820.2023.2244562

Hwang, G.-J., & Chang, H.-F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*(4), 1023–1031. https://doi.org/10.1016/j.compedu.2010.12.002

Hwang, G.-J., Yang, L.-H., & Wang, S.-Y. (2013). A concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education, 69*, 121–130. https://doi.org/10.1016/j.compedu.2013.07.008

Janesarvatan, F., & Van Rosmalen, P. (2023). Instructional design of virtual patients in dental education through a 4c/id lens: a narrative review. *Journal of Computers in Education,* 1–34. https://doi.org/10.1007/s40692-023-00268-w

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*(1), 1–10. https://doi.org/10.1016/S0959-4752(01)00014-7

Kukharuk, A., Goda, Y., & Suzuki, K. (2023). Designing an online pd program with 4c/id from scratch. *International Journal of Designs for Learning*, *14*(2), 72–86. https://doi.org/10.14434/ijdl.v14i2.34676

Kuklick, L., Greiff, S., & Lindner, M. A. (2023). Computer-based performance feedback: Effects of error message complexity on cognitive, metacognitive, and motivational outcomes. *Computers & Education, 200*, 104785. https://doi.org/10.1016/j.compedu.2023.104785

Kyun, K. T., & Hong, P. J. (2019). More about the basic assumptions of t-test: normality and sample size. *Korean Journal of Anesthesiology, 72*(4), 331–335. https://doi.org/10.4097/kja.d.18.00292

Larmuseau, C., Coucke, H., Kerkhove, P., Desmet, P., & Depaepe, F. (2019). Cognitive load during online complex problem-solving in a teacher training context. *Eden conference proceedings* (p. 466–474). Retrieved from https://www.ceeol.com/search/article-detail?id=847176

Larmuseau, C., Elen, J., & Depaepe, F. (2018). The influence of students' cognitive and motivational characteristics on students' use of a 4c/id-based online learning environment and their learning gain. *Proceedings of the 8th international conference on learning analytics and knowledge* (p. 171–180). Retrieved from https://doi.org/10.1145/3170358.3170363

Leenknecht, M., Wijnia, L., Köhlen, M., Fryer, L., Rikers, R., & Loyens, S. (2021). Formative assessment as practice: The role of students' motivation. *Assessment & Evaluation in Higher Education, 46*(2), 236–255. https://doi.org/10.1080/02602938.2020.1765228

Lenhard, W., & Lenhard, A. (2022). Computation of effect sizes. Retrieved from http://www.psychometrica.de/effect_size.html (2022, Oct)

Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C. P., & Van Merrienboer, J. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32–42. https://doi.org/10.1016/j.learninstruc.2013.12.001

Maddens, L., Depaepe, F., Raes, A., & Elen, J. (2020). The instructional design of a 4c/id-inspired learning environment for upper secondary school students' research skills. *International Journal of Designs for Learning*, *11*(3), 126–147. https://doi.org/10.14434/ijdl.v11i3.29012

Maddens, L., Depaepe, F., Raes, A., & Elen, J. (2023). Fostering students' motivation towards learning research skills: the role of autonomy, competence and relatedness support. *Instructional Science, 51*(1), 165–199. https://doi.org/10.1007/s11251-022-09606-4

Maggio, L. A., Ten Cate, O., Irby, D. M., & O'Brien, B. C. (2015). Designing evidence-based medicine training to optimize the transfer of skills from the classroom to clinical practice: applying the four component instructional design model. *Academic Medicine, 90*(11), 1457–1461. https://doi.org/10.1097/ACM.0000000000000769

Mahantakhun, C., Koraneekij, P., & Khlaisang, J. (2020). The effects of 4c/idbased adaptive procedural simulation on safety awareness in undergraduate students majoring in gems and jewelry. *Scholar: Human Sciences*, *12*(1), 296. Retrieved from http://www.assumptionjournal.au.edu/index.php/Scholar/article/view/3898

Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education, 95*(95), 85–98. https://doi.org/10.1016/j.compedu.2015.12.002

Marcellis, M., Barendsen, E., & van Merriënboer, J. (2018). Designing a blended course in android app development using 4c/id. *Proceedings of the 18th Koli calling international conference on computing education research* (p. 1–5). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3279720.3279739

Marchisio, M., Barana, A., Fioravera, M., Rabellino, S., & Conte, A. (2018). A model of formative automatic assessment and interactive feedback for stem. *2018 ieee 42nd annual computer software and applications conference (compsac)* (Vol. 1, p. 1016–1025). Retrieved from https://doi.org/10.1109/COMPSAC.2018.00178

Martínez-Mediano, C., & Losada, N. R. (2017). Internet-based performance support systems in engineering education. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 12*(2), 86–93. https://doi.org/10.1109/RITA.2017.2697778

Melo, M., & Miranda, G.L. (2016). The effects of the 4c/id model in the acquisition and transfer of learning: a meta-analysis. *RISTI (Revista Iberica de Sistemas e Tecnologias de Informacao)* (18), 114–131. Retrieved from https://link.gale.com/apps/doc/A464161712/AONE?u=anon~7cb43a6&sid=googleScholar&xid=5821580d

Mertens, U., Finn, B., & Lindner, M. A. (2022). Effects of computer-based feedback on lower-and higher-order learning outcomes: A network meta-analysis. *Journal of Educational Psychology, 114*(8), 1143–1772. https://doi.org/10.1037/edu0000764

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education, 9*(3), 1–26. https://doi.org/10.1002/rev3.3292

Mulders, M. (2022). Vocational training in virtual reality: A case study using the 4c/id model. *Multimodal Technologies and Interaction, 6*(7), 49. https://doi.org/10.3390/mti6070049

Musharyanti, L., Haryanti, F., & Claramita, M. (2021). Improving nursing students' medication safety knowledge and skills on using the 4c/id learning model. *Journal of Multidisciplinary Healthcare, 14*, 287–295. https://doi.org/10.2147/JMDH.S293917

Ndiaye, Y., Hérold, J. -F., & Chatoney, M. (2021). Applying the 4c/id-model to help students structure their knowledge system when learning the concept of force in technology. *Techne serien-Forskning i slöjdpedagogik och slöjdvetenskap*, *28*(2), 260–268. Retrieved from https://journals.oslomet.no/index.php/techneA/article/view/4319

Pontes, T., Miranda, G., & Santos, D. (2021). Virtual learning environments: What makes them effective. *Iceri2021 proceedings* (p. 407–417). IATED. Retrieved from https://doi.org/10.21125/iceri.2021.0159

Qiu, F., Liu, P., WangLiping, & Xie, Y. (2012). Exploring the architecture of a complex learning support platform based on the 4c/id model. *Research on Electrochemical Education* (4), 67–71. https://doi.org/10.13811/j.cnki.eer.2012.04.005

Roscoe, R. D., & Craig, S. D. (2022). A heuristic assessment framework for the design of self-regulated learning technologies. *Journal of Formative Design in Learning, 6*, 77–94. https://doi.org/10.1007/s41686-022-00070-4

Rusman, E., & Nadolski, R. (2023). Pe(e)rfectly skilled underpinnings of an online formative assessment method for (inter)active and practice-based complex skills training in higher education (he). *International Journal of Mobile and Blended Learning, 15*(2), 1–14. https://doi.org/10.4018/IJMBL.318646

Sarfo, F.K., & Elen, J. (2006). Technical expertise development in secondary technical schools: Effects of ictenhanced 4c/id learning environments. *Fourth ieee international workshop on technology for education in developing countries (tedc'06)* (pp. 62–65). Retrieved from https://doi.org/10.1109/TEDC.2006.25

Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *The ANNALS of the American Academy of Political and Social Science, 683*(1), 183–200. https://doi.org/10.1177/0002716219843818

Skulmowski, A., & Xu, K. M. (2021). Understanding cognitive load in digital and online learning: a new perspective on extraneous cognitive load. *Educational Psychology Review, 34*(1), 171–196. https://doi.org/10.1007/s10648-021-09624-7

Spector, J. M., Ifenthaler, D., Sampson, D., Yang, L. J., Mukama, E., Warusavitarana, A., . . . Gibson, D. C. (2016). Technology enhanced formative assessment for 21st century learning. *Journal of Educational Technology & Society*, *19*(2), 58–71. Retrieved 2023-10-04, from https://www.jstor.org/stable/jeductechsoci.19.3.58

Sweller, J., Van Merrienboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tapingkae, P., Panjaburee, P., Hwang, G.-J., & Srisawasdi, N. (2020). Effects of a formative assessment-based contextual gaming approach on students' digital citizenship behaviours, learning motivations, and perceptions. *Computers & Education, 159*, 103998. https://doi.org/10.1016/j.compedu.2020.103998

Tempelaar, D. T., Heck, A., Cuypers, H., van der Kooij, H., & van de Vrie, E. (2013). Formative assessment and learning analytics. Proceedings of the third international conference on learning analytics and knowledge (p. 205-209). Retrieved from https://doi.org/10.1145/2460296.2460337

Thima, S., & Chaijaroen, S. (2021). The framework for development of the constructivist learning environment model to enhance ill-structured problem solving in industrial automation system supporting by metacognition. *Innovative technologies and learning: 4th international conference, icitl 2021, virtual event, November 29–December 1, 2021, proceedings 4* (p. 511–520). Retrieved from https://doi.org/10.1007/978-3-030-91540-7_52

Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & de Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: A systematic literature review. *Sage Open, 10*(1), 2158244019900176. https://doi.org/10.1177/2158244019900176

Van Merriënboer, J. J. (2019). The four-component instructional design model. *Open Education Research*, *26*(3), 35–43. Retrieved from https://www.4cid.org/publications

Van Merriënboer, J. J., Clark, R. E., & De Croock, M. B. (2002). Blueprints for complex learning: The 4c/id-model. *Educational Technology Research and Development, 50*(2), 39–61. https://doi.org/10.1007/BF02504993

Van Merriënboer, J. J., & Dolmans, D. H. (2015). Research on instructional design in the health sciences: From taxonomies of learning to whole-task models. In J. Cleland & S. J. Durning (Eds.), *Researching medical education* (p. 193–206). John Wiley & Sons, Ltd. Retrieved from https://doi.org/10.1002/9781118838983.ch17

Van Merriënboer, J. J., & Kirschner, P. A. (2017). *Ten steps to complex learning: A systematic approach to four-component instructional design*. Retrieved from: Routledge. https://doi.org/10.4324/9780203096864

Wang, L.-C., & Chen, M.-P. (2010). The effects of game strategy and preferencematching on flow experience and programming performance in game-based learning. *Innovations in Education and Teaching International, 47*(1), 39–52. https://doi.org/10.1080/14703290903525838

Weaver, B. (2011). Silly or pointless things people do when analyzing data: 1. *Conducting a test of normality as a precursor to a t-test. manuskript eines vortrags auf der northern health research conference vom* (Vol. 10). Retrieved from https://www.researchgate.net/publication/299497976_Silly_or_Pointless_Things_People_Do_When_Analyzing_Data_1_Testing_for_Normality_as_a_Precursor_to_a_t-test

Webb, M. E., Prasse, D., Phillips, M., Kadijevich, D. M., Angeli, C., Strijker, A., & Laugesen, H. (2018). Challenges for it-enabled formative assessment of complex 21st century skills. *Technology, knowledge and learning, 23*(3), 441–456. https://doi.org/10.1007/s10758-018-9379-7

Wilson, J., Ahrendt, C., Fudge, E. A., Raiche, A., Beard, G., & MacArthur, C. (2021). Elementary teachers' perceptions of automated feedback and automated scoring: Transforming the teaching and learning of writing using automated writing evaluation. *Computers & Education, 168*, 104208. https://doi.org/10.1016/j.compedu.2021.104208

Wopereis, I., Frerejean, J., & Brand-Gruwel, S. (2015). Information problem solving instruction in higher education: A case study on instructional design. *Information literacy: Moving toward sustainability: Third European conference, ecil 2015, Tallinn, Estonia, October 19-22, 2015, revised selected papers 3* (p. 293–302). Retrieved from https://doi.org/10.1007/978-3-319-28197-1_30

Xu, X., Shen, W., Islam, A. A., & Zhou, Y. (2023). A whole learning process-oriented formative assessment framework to cultivate complex skills. *Humanities and Social Sciences Communications, 10*, 1–15. https://doi.org/10.1057/s41599-023-02200-0

Xu, X., Zhou, Z., Ji, Y., Wang, M., & Gu, X. (2019). Design and effectiveness of comprehensive learning for complex skills based on the 4c/id model. *China Educational Technology*, *10*(08), 124–131. Retrieved from https://kns.cnki.net/kcms/detail/11.3792.g4.20191008.1824.036.html

Authors and Affiliations

Haoxin Xu[1] · **Tianrun Deng**[2] · **Xianlong Xu**[2] · **Xiaoqing Gu**[2] · **Lingyun Huang**[3] · **Haoran Xie**[4] · **Minhong Wang**[3]

✉ Xianlong Xu
xlxu@eec.ecnu.edu.cn

Haoxin Xu
haoxin.xu@stu.ecnu.edu.cn

Tianrun Deng
51254108021@stu.ecnu.edu.cn

Xiaoqing Gu
xqgu@ses.ecnu.edu.cn

Lingyun Huang
lingyun.huang@hku.hk

Haoran Xie
hrxie2@gmail.com

Minhong Wang
magwang@hku.hk

1 Lab of Artificial Intelligence for Education, East China Normal University, Shanghai, China

2 Department of Education Information Technology, East China Normal University, Shanghai, China

3 Faculty of Education, The University of Hong Kong, Hong Kong, China

4 Department of Computing and Decision Sciences, Lingnan University, Hong Kong, China