PhET Simulations Performance to Measure Complex Problem-Solving Skills in Physics Subject Matter: Evidence of Validity and Reliability

Eko Juliyanto[1,2], Wiyanto Wiyanto[3,*], Sunyoto E. Nugroho[4], and Budi N. Mindyarto[4]

[1]Science Education Doctoral Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
[2]Department of Science Education, Faculty of Teacher Training and Education, Universitas Tidar, Magelang, Indonesia
[3]Science Education Doctoral Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
[4]Departement of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
Email: ekojuliyanto@untidar.ac.id (E.J.); wiyanto@mail.unnes.ac.id (W.W.); ekonuphysed@mail.unnes.ac.id (S.E.N.); budinaini@mail.unnes.ac.id (B.N.M.)
*Corresponding author

Manuscript received July 15, 2025; revised August 6, 2025; accepted September 17, 2025; published February 9, 2026

***Abstract*—This study aims to analyze the performance of the Physics-Complex Problem-Solving (Ph-CPS) skills test using Physics Education Technology (PhET) simulations to measure Complex Problem-Solving (CPS) skills using physics subject matter as the problem context. This study is instrumentation research that produces a CPS skill measurement instrument that will be used in physics classes. The results indicated that the Ph-CPS test meets the criteria for complex problems and minimal complex system criteria. Construct validity and reliability analysis using the Rasch model showed that only Ph-CPS 1, Ph-CPS 2, and Ph-CPS 4 items were valid and reliable. The results of the concurrent validity analysis indicated that Ph-CPS items 1, 2, and 4 align with the PISA 2012 CPS test. Ph-CPS item 3 is considered too easy because it has similarities with Angry Birds games, so it is classified as a "familiar problem". The results of the concurrent validity analysis show that Ph-CPS items 1, Ph-CPS 2, and Ph-CPS 4 have concurrent validity against the 2012 PISA test. Regression analysis between physics concept mastery and Ph-CPS skills showed no significant influence of physics concepts on Ph-CPS skills. It indicates that the Ph-CPS test measures domain-general skills.**

***Keywords*—complex problem-solving, physics education, PhET simulations, domain-general skill, instrumentation research**

I. Introduction

The complexity of problems in society is rapidly increasing due to advancements in technology and information [1]. The global community is facing problems that are dynamic and previously unknown. The COVID-19 pandemic is a recent example of a complex issue encountered by the global community [2]. The COVID-19 virus variant mutates very quickly [3], so the COVID-19 pandemic problem is classified as a dynamic problem. It has made finding solutions to the COVID-19 pandemic highly complex, as the influencing variables change. Vaccines developed for a particular COVID-19 variant just months earlier became less relevant in subsequent months due to the ongoing mutations of the virus. The global community will face complex problems such as the COVID-19 pandemic in the future. The issues of climate change and global warming are expected to become highly complex problems in the future [4], affecting various aspects of global society, such as agriculture, the economy, and public health.

The increasing complexity of societal problems requires the development of Complex Problem-Solving (CPS) skills. Therefore, CPS skills are the most relevant skills needed in all aspects of life [5]. However, CPS skills have not yet become a central focus in classroom instruction, particularly in K-12 education. Classroom problem-solving skills still focus on static problem solving, such as physics problem solving [6, 7] and mathematics problem solving [8, 9].

A consequence of implementing CPS-focused learning in the classroom is the need to incorporate subject matter content into the presentation of dynamic problems. Like teaching traditional static physics problem-solving skills, teaching CPS skills in physics classrooms must involve subject matter. This implies that developing CPS tests in physics education must incorporate physics content into the context of the problem. It is based on the perspective that subject matter is used to teach thinking skills [10].

The implication of teaching CPS skills involving physics subject matter is that the development of CPS tests must also include subject matter content. On the other hand, CPS skills are classified as domain-general skills [11, 12], so CPS skills should not require mastery of specific concepts. The solution is that the physics subject matter is used only as the problem context in the CPS test.

The CPS test requires a dynamic environmental situation in presenting problems [13]. PhET simulation is a learning application in physics subject matter that presents a dynamic environment [14]. Thus, PhET simulation can be used to present a complex problem.

This study aims to analyze the performance of the CPS test by involving subject matter as the context of the problem. The Physics Complex Problem-Solving (Ph-CPS) test was developed using PhET simulations to present dynamic environments for measuring Ph-CPS skills. This study examines whether the Ph-CPS test meets the criteria for complex problems. The validity and reliability of the Ph-CPS test were tested using the Rasch model. This study also analyzes whether the Ph-CPS test is classified as a domain-general or domain-specific skill.

This study aims to develop a CPS skills test instrument for physics classes by involving physics subject matter as the problem context. On the other hand, a CPS skills test must be categorized as a domain-general skills test. The next question is: when a CPS test involves physics subject matter as the problem context, can it still be classified as a domain-general skill? More specifically, this research question is formulated as follows:

RQ1: Does the physics-complex problem-solving skills test using PhET Simulations meet the complex problem criteria and minimal complex criteria?

RQ2: Does the physics-complex problem-solving skills test using PhET Simulations have good test validity and reliability?

RQ3: Is the physics-complex problem-solving skills test using PhET Simulations classified as a domain-general skill?

II. Literature Review

A. Physics-Complex Problem-Solving Skills

The ability to solve problems cannot be separated from the physics curriculum. Problem solving is essential in the physics curriculum at every level of education [15]. The attitudes and problem-solving approaches used by students in physics classes influence the development of their physics expertise and future problem-solving competence [16]. Thus, teaching physics problem-solving is essential in the classroom.

The increasing level of ambiguity as a characteristic of the Industrial Revolution 4.0 [17–19] has led to increasingly complex problems society faces. As a result, the ability to solve complex problems has become one of the most essential competencies required of future professionals [20]. Therefore, traditional physics problem-solving instruction alone is insufficient; it is also essential to incorporate the development of Physics-Complex Problem Solving (Ph-CPS) skills into the curriculum.

Some experts argue that CPS skills are domain-general skills [21], but many researchers have studied CPS skills using a domain-specific approach [22–25]. Developing CPS skills in the classroom through subject-based instruction inevitably requires the involvement of content material. Consequently, pursuing research that integrates domain-general problem-solving abilities (CPS) with investigations into domain-specific problem-solving, such as physics, is essential.

Physics problem-solving competence encompasses two key components: (1) conducting mathematical analyses of pertinent equations followed by qualitative interpretations of the resulting phenomena, and (2) examining physical phenomena and linking them to underlying physics theories [15]. On the other hand, the main characteristic of CPS is that it is a complex problem in a dynamic environment [16] and is unclear [26]. The characteristics of CPS include multiple highly interrelated variables that change over time (dynamic), underlying connections that are not transparent, and the requirement for participants to achieve several, sometimes contradictory, goals [27].

The CPS stage frameworks are depicted in the following. As Funke stated, the stages of CPS include exploration, knowledge acquisition, and knowledge application [27]. As OECD described the CPS frameworks used in the PISA 2012 test are exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting [28]. Those two CPS stage frameworks are related. The relevance of the CPS stage between the OECD [28] and Funke [27] frameworks is presented in Table 1.

Table 1. The relationship of CPS stages based on OECD [28] and Funke [27]

CPS Stages based on OECD [28]	**CPS Stages based on** Funke [27]
Exploring and understanding	Exploration
Represent and formulate	Acquisition of knowledge
Plan and execute	Application of Knowledge
Monitoring and reflecting	

An essential issue regarding current CPS research is how findings can be applied to formal educational environments. A debate exists regarding how the transferability of CPS to classroom content learning categorizes CPS as a specific-domain skill. Several studies position CPS as a domain-general skill applicable across contexts and regardless of conceptual mastery [29, 30]. This perspective encourages the development of microworld-based CPS measurement tools like MicroDYN and MicroFIN, which seek to assess CPS dimensions within a domain-general skill framework, limiting the emphasis on specific content knowledge [31, 32]. Conversely, complex problems in real-world and classroom environments inevitably involve subject matter, indicating that CPS performance is naturally dependent on conceptual mastery. Abstract and non-contextualized CPS tests yield biased measurements, overlooking students' ability to apply domain-specific knowledge in real-world contexts [33, 34].

This study integrates CPS skills into physics learning to appropriately position CPS within the current discussion in the CPS domain. This study maintains that integrating CPS skills into physics learning is classified as domain-general skills by restricting physics content to the problem context. Solving physics-complex problems facilitates knowledge acquisition and helps students learn physics content effectively. Integrating CPS skills into physics education should consider CPS a multidimensional competency encompassing both domain-general and domain-specific aspects [35, 36].

The transferability of CPS to educational environments, specifically in physics learning, relies on effectively integrating domain-general CPS with domain-specific physics content. As exemplified by PhET simulations, interactive learning environments possess the potential to connect these two domains effectively. PhET simulations offer dynamic environments that necessitate the application of domain-general skills essential for complex problem solving, including hypothesis construction, variable manipulation, and analysis of cause-and-effect relationships [37–39].

B. The Concept of Minimal Complexity

Many experts state that CPS test results can predict an individual's future success [27] because CPS correlates with a person's intelligence [40]. Tests to measure CPS are usually computer-based [28, 40].

One of the fundamental questions in complex problem research is whether the test used to measure CPS ability meets the criteria for complex problems. A problem is complex if it has multiple goals, involves many interdependent and connected variables with non-transparent relationships, and the conditions of the problem system's environment change over time [41]. Criteria for problem complexity are less operational for assessing whether a given task falls into the category of a complex problem or not [41].

At the beginning of the CPS study, researchers competed to present a high complexity of problems by involving more and more variables [33]. This perspective then changes with the emergence of the "minimal complex systems" [42, 43], establishing the minimal standard for a system to be considered a complex problem. This "minimal complex system" concept does not guarantee that a system involving many variables is classified as a complex problem.

A system is categorized as a problem with high complexity if it consists of at least two exogenous variables (input) and two endogenous variables (output) [43]. Each exogenous variable must not only have a "main effect" on one of the endogenous variables but also exhibit a "multiple effect" on both endogenous variables, and both endogenous variables must show "mutual dependence" on both exogenous variables. In a complex system, one of the endogenous variables must be "eigendynamic", and the other endogenous variable must have a "side effect" on the endogenous variable that possesses the "eigendynamic" property. *Eigendynamic* is the effect of an endogenous variable on itself [30]. A diagram categorizing complex systems with two exogenous variables and two endogenous variables (a 2×2 complex system) is presented in Fig. 1.

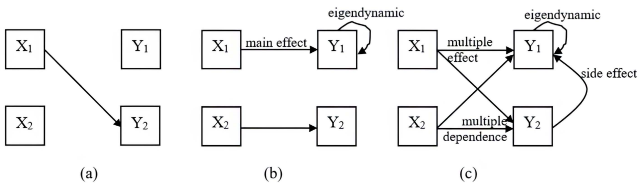

Fig. 1. System complexity categorization: (a) low complexity (b) medium complexity (c) high complexity (source: [NO_PRINTED_FORM] [43]).

C. PhET Simulations as Ph-CPS Measurement Test

Several tests can be used to measure CPS skills, such as MultiFlux [44], MicroDYN [45], Genetics Lab [46], and MicroFIN [47]. MultiFlux, MicroDYN, and Genetics Lab use the Linear Structural Equation (LSE) system framework, while MicroFIN uses the Finite State Automata (FSA) framework [48]. In CPS tests that use the LSE framework, research subjects are presented with a dynamic computer simulation containing exogenous variables (X) and endogenous variables (Y) [49]. The relationship between exogenous variables and endogenous variables involved in a microwolds is formulated using a system of linear equations. In FSA framework tests, subjects are also presented with a dynamic computer simulation similar to the LSE framework. The difference is that the output in the FSA simulation system is determined by predetermined inputs and transition functions [50]. This output produces a state that can change from one state to another depending on the input and the transaction function. More simply, the difference between the LSE and FSA frameworks is that the LSE framework presents a quantitative relationship between input and output variables. In contrast, the FSA framework presents a qualitative relationship between input and output variables [31].

In addition to MultiFlux, MicroDYN, Genetics Lab, and MicroFIN, various studies have employed existing dynamic computer applications to create CPS assessments. For instance, a CPS study by [22] utilized a virtual laboratory application to design a CPS test in chemistry. The test was developed using a virtual chemistry lab tool called ChemLab-Builder, which enabled researchers to simulate dynamic scenarios within a laboratory setting.

PhET Simulations is a virtual laboratory in the field of science that presents the dynamic situation of the science laboratory environment [51]. PhET Simulations is classified as a microworld that can be used to present a problem-solving test [52]. PhET Simulations is not explicitly used to measure CPS capability; however, it can be approached using the LSE framework because PhET presents simulations that quantitatively display the relationships between exogenous and endogenous variables.

The interactive and dynamic PhET simulation is well-suited for assessing CPS because it captures not only the results of problem-solving tasks but also the underlying cognitive processes. By utilizing dynamic and interactive simulations, such as PhET, researchers can distinguish between successful and unsuccessful problem solvers based on exploratory behavior. Recent CPS research has utilized log-file data to analyze strategies, reasoning patterns, and decision-making processes during CPS [1]. To enhance data validity, various methods of observing students during CPS can be employed, including sequence mining [53], behavioral profile clustering [54], and predictive modeling with machine learning [55].

Artificial Intelligence (AI) has the potential to enhance CPS measurement. AI and machine learning models have been successfully applied to predict problem-solving performance and provide real-time diagnostic feedback based on acquired data logs [56]. AI can be leveraged to track strategies implemented during the CPS process [57, 58].

PhET simulations are virtual laboratories that can be utilized to develop higher-order thinking skills (HOTS) in science learning domains, particularly physics. PhET simulations have the potential to be used in measuring and teaching CPS skills because students experience interactive exploration, hypothesis testing, and immediate feedback within a dynamic simulation [14, 59]. The dynamic nature of PhET simulations increases student engagement in solving higher-order problems. PhET simulations, as virtual laboratories, support inquiry-based learning, the foundation of CPS skills [60].

The application of PhET has been shown to improve students' domain-general skills, such as analysis, synthesis, and scientific communication, which are fundamental skills required for CPS [61, 62]. By providing a safe, flexible, and inquiry-based environment, PhET fosters domain-general skills, such as understanding causal relationships and the scientific principles underlying specific phenomena [38, 63]. In addition to its role in developing domain-general problem-solving skills, research shows that PhET simulations can also foster domain-specific skills in science learning [37, 64]. Research also indicates that using PhET in science learning can teach both domain-specific knowledge and problem-solving skills within the field context [39, 65]. Thus, PhET serves not only as a means of developing domain-specific knowledge but also domain-general skills.

III. Materials and Methods

A. Research Design

This study is instrumentation research [66, 67]. This study attempts to produce a CPS skill test that will be used in physics classes. This study began by designing the Ph-CPS skills test utilizing the dynamic environment presented by PhET Simulations. The simulations in PhET were identified using minimal complex system criteria [43]. Four of the 61 physics simulations in PhET meet the minimal complex system criteria and can be used to present real-world problems. Those four simulations are bending light, gravity and orbits, projectile motion and geometric optics. Contextual open-ended questions were prepared for each simulation to present complex physics problems.

A 5-point Likert scale questionnaire was created to assess the suitability of the developed test model with the minimal complex system criteria [43]. Fifteen experts in the field of physics education were involved in filling out the questionnaire. The suitability of the developed test model with the minimal complex system criteria constitutes a content validity analysis. The Ph-CPS skills test that meets content validity was trialled on students from the science education and physics education programs at six universities in Indonesia.

In the trial, the respondents also completed the PISA 2012 problem-solving test and a physics concept mastery test related to bending light, gravity and orbits, projectile motion, and geometric optics. The PISA 2012 problem-solving test is one of the standardized CPS tests [68, 69]. The PISA 2012 problem-solving test was used to analyze the concurrent validity of the developed Ph-CPS skill test, while the physics concept mastery test was used to analyze whether the developed Ph-CPS skill test belonged to domain-general or domain-specific skills.

B. Samples and Data Collection

Fifteen lecturers with doctoral degrees in physics education from eight universities in Indonesia working in the Department of Physics Education participated in the content validity questionnaire. They were familiar with and experienced using PhET simulations in their lectures during the COVID-19 pandemic. They possessed sufficient background knowledge to be involved as expert judges in this study due to their educational background and work experience. The questionnaire aims to analyze whether the developed Ph-CPS skill test meets a complex system's minimum criteria. The questionnaire consisted of seven items developed based on the indicators of the minimal complex system criteria [43]. The questionnaire used a 5-point Likert scale.

Two hundred forty-eight undergraduates in science education and physics education from six universities in Indonesia served as the test trial sample in this study. They had experience using PhET simulations in practical learning during and after the COVID-19 pandemic. They possessed sufficient basic physics knowledge to understand the context of the physics problems presented in Ph-CPS. The sample was selected randomly, regardless of their year of study. The Ph-CPS skills test [43] was piloted alongside the PISA 2012 problem-solving test and the physics concept mastery test. The scoring of each item of the Ph-CPS skill tests and the PISA 2012 problem-solving test was carried out with a minimum score of 0 and a maximum score of 3, while the physics concept mastery test was carried out with a minimum score of 0 and a maximum score of 5.

The scoring rubric for the Ph-CPS and PISA 2012 tests uses the following categories: 1) a score of 3 if the respondent successfully solves the problem by carrying out exploration, knowledge acquisition, and knowledge application; 2) a score of 2 if the respondent carries out the exploration and knowledge acquisition phases; 3) a score of 1 if the respondent carries out the exploration phase. The scoring rubric for the physics concept mastery test uses the following categories:1) a score of 5 if the respondent determines the known variables, determines the sought variables, writes the required physics formulas, applies the formulas, and draws conclusions; 2) a score of 4 if the respondent only applies 4 of the five skills above; 3) a score of 3 if the respondent only applies 3 of the five skills above; 2) a score of 2 if the respondent only applies 2 of the five skills above; 1) a score of 1 if the respondent only applies 1 of the five skills above. The scoring of the Ph-CPS and PISA 2012 tests is based on computer activity logs, while the scoring for the physics concept mastery test is based on the paper-based test results.

C. Data Analysis

RQ1. Expert assessment is used to assess the suitability of the Ph-CPS test with complex problem criteria and minimal complex system criteria. The conformity analysis between the Ph-CPS skills test using PhET Simulations and the criteria for complex problems based on the minimal complex system criteria was conducted through content validity analysis using V-Aiken [70] and calculated using Excel. The Ph-CPS skill test using PhET Simulation meets the content validity if the V-value exceeds the V-table. The Ph-CPS skills test that meets content validity implies that the Ph-CPS test developed using PhET Simulations satisfies the minimal complex system criteria.

RQ2. Construct validity and reliability were analyzed using the Rasch model [71–73], assisted by Winstep software. Construct validity analysis was conducted by evaluating the fit of each item and person to the developed test model and the test dimensionality. Empirical validity analysis was conducted using item validity tests and classical theory, which tested the correlation of each item to the total score. Concurrent validity was analyzed using Pearson's product-moment correlation value between the test model developed and the CPS PISA test value. Reliability analysis was conducted by analyzing the value of the Person Separation Index.

RQ3. A linear regression test between the results of the developed test model and physics concept mastery was conducted to analyze the extent of the influence of related physics concept mastery on respondents' success in completing the Ph-CPS skills test using PhET simulations. The influence of physics concept mastery on respondents' success in completing the Ph-CPS skills test using PhET simulations indicates that the test is classified as a domain-general skill.

IV. RESULT AND DISCUSSION

A. The Ph-CPS Test Utilizing the PhET Simulation Meets Complex Problem and Minimal Complex System Criteria

The Ph-CPS skill test was developed by utilizing the PhET

Simulation application. PhET Simulation is used to present dynamic problem situations using the LSE framework. Several material subjects are available in PhET Simulation, including: Physics, Math and Statistics, Chemistry, Earth and Space, and Biology. The development of the Ph-CPS skills test started with identifying simulations that use physics subject matter and the minimal complex system 2×2 matrix (two inputs and two outputs). The identification results found that at least four simulations meet both criteria. They were bending light (Ph-CPS 1), gravity and orbits (Ph-CPS 2), projectile motion (Ph-CPS 3), and geometric (Ph-CPS 4) [43].

B. The Item of Ph-CPS

1) Bending light (Ph-CPS 1)

Bending light simulation is a dynamic system that presents natural phenomena related to Snell's law where two mediums with different refractive indices pass light sources. The complex problems in the system are presented in Fig. 2.

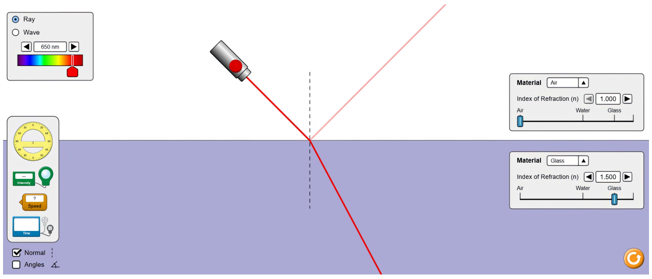

Fig. 2. Bending light simulation [74].

Task: The fibre optic cable working system utilizes the critical angle principle of internal total reflection on the refraction phenomenon. Total internal reflection occurs when a light ray travels from a denser medium to a less dense medium, and the light ray from the first medium is reflected 100 per cent by the second medium, with no light transmitted (refracted) into the second medium. The largest angle of incidence that results in total internal reflection is called the critical angle. The greater the critical angle, the more efficient the fibre optic cable is because less energy is lost. Show how to generate the greatest critical angle in the PhET simulation!

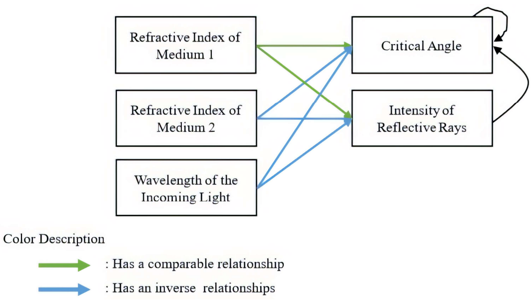

Fig. 3. System complexity of bending light simulation.

The Item Ph-CPS 1 presents complex problems related to the efficiency of fibre optic cables involving three exogenous variables and two endogenous variables (3×2 matrix). The exogenous variable in this system consists of the angle of incidence (initial), the refractive index of the first medium, the refractive index of the second medium and the wavelength of the incoming beam, while the endogenous variable in this system consists of the critical angle (refractive angle) and the intensity of the reflected beam. In Ph-CPS item 1, respondents were asked to produce internal total reflection with the greatest critical angle on the bending light phenomenon. The magnitude of the critical angle in total internal reflection determines the efficiency of electromagnetic wave transmission in optical fibre cables. Interaction between the respondent and the system is carried out by modifying and controlling the exogenous variables to achieve the intended goal. The system complexity matrix of the complex problem is presented in Fig. 3.

2) Gravity and orbits (Ph-CPS 2)

Simulation of gravity and orbits is a dynamic system that presents the phenomenon of the Earth's revolution. This simulation uses Kepler's Law to present planetary motion. The complex problems in the system are presented in Fig. 4.

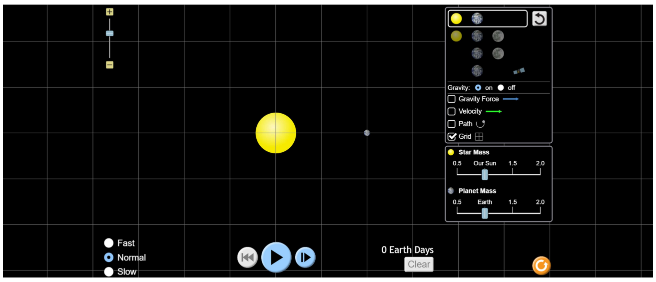

Fig. 4. Gravity and orbits simulations [75].

Task: The Earth revolves around the Sun in a revolving motion. Show in the PhET simulation how the Earth can continue to revolve ideally around the Sun at a distance of 1.5 times its original distance!

The Item Ph-CPS 2 presents complex problems related to planetary motion involving three exogenous variables and three endogenous variables (3×3 matrix). The exogenous variables involved in this system consist of the mass of the Earth, the mass of the Sun, and the distance between the Earth and the Sun, while the endogenous variables involved in this system consist of the shape of the Earth's revolution orbit, the Earth's revolution period, and the magnitude of the gravitational force. The system complexity matrix of the complex problem is presented in Fig. 5.

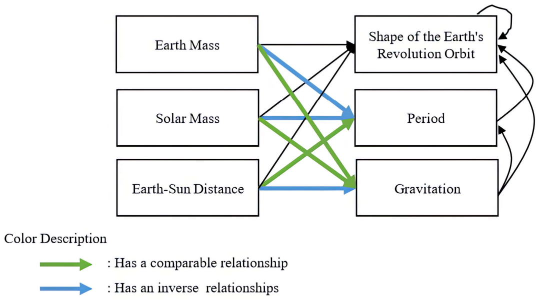

Fig. 5. System complexity of gravity and orbits simulation.

3) Projectile motion (Ph-CPS 3)

The projectile motion simulation is a dynamic system that presents the phenomenon of projectile movement. This simulation uses Newton's laws of motion to represent projectile motion. Respondents were asked to shoot targets using cannons like the Angry Birds game. The complex problems in the system are presented in Fig. 6.

Task: A cannon shot operates by utilizing parabolic motion. Shoot the available target with the highest level of accuracy! Earning three stars signifies the highest level of accuracy!

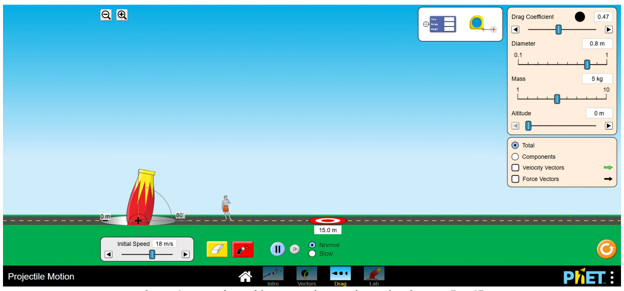

Fig. 6. Projectile motion simulations [76].

The item Ph-CPS 3 presents complex problems relating to the accuracy of cannon fire utilizing the concept of projectile motion. The system in this item involves three exogenous variables and two endogenous variables (3×2 matrix). Exogenous variables in this system consist of elevation angle, initial speed and initial height of the cannon. In contrast, endogenous variables in this system consist of the farthest horizontal distance and maximum height. In the Ph-CPS 3 item, respondents were asked to shoot a high-precision target marked with a three-star point. Increasing the elevation angle does not always improve the horizontal mileage. Increasing the elevation angle from 0 to 45 will increase the horizontal range, while increasing the angle from 45 to 90 will decrease the horizontal range. Interaction between the respondent and the system is done by modifying and controlling the exogenous variables to achieve the intended goal. The system complexity matrix of the complex problem is presented in Fig. 7.

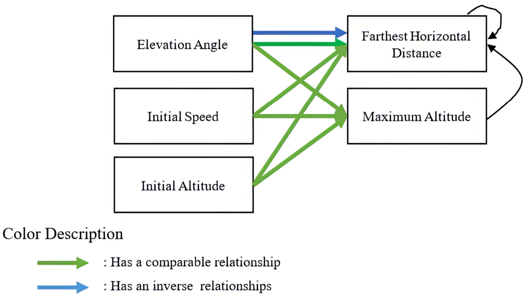

Fig. 7. System complexity of projectile motion simulation.

4) Geometric optics (Ph-CPS 4)

Geometric optics simulation is a dynamic system that presents natural phenomena related to the formation of images in lenses and curved mirrors using geometric optical approaches. Geometric optics provides practical solutions for designing optical instruments according to specific needs. The complex problems in the system are presented in Fig. 8.

Task: Produce a real image with a height 1.5 times the original height!

The item Ph-CPS 4 presents complex problems related to the formation of shadows on lenses that utilize the concept of geometric optics. The system in this item involves three exogenous variables and two endogenous variables (3×2 matrix). Exogenous variables in this system consist of object distance, lens refractive index and lens curvature radius, while endogenous variables in this system consist of shadow height and distance. In Ph-CPS item 4, respondents were asked to produce a shadow of an object with a certain height. Interaction between the respondent and the system is done by modifying and controlling the exogenous variables to achieve the intended goal. The system complexity matrix of the complex problem is presented in Fig. 9.

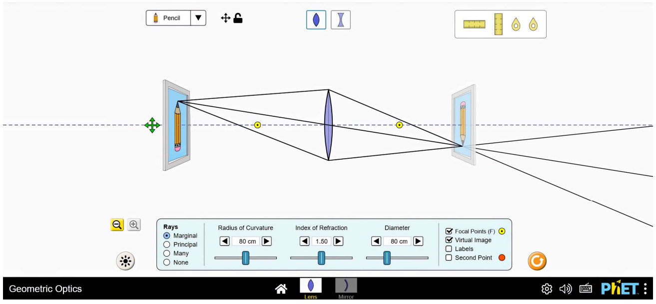

Fig. 8. Geometric optics simulations [77].

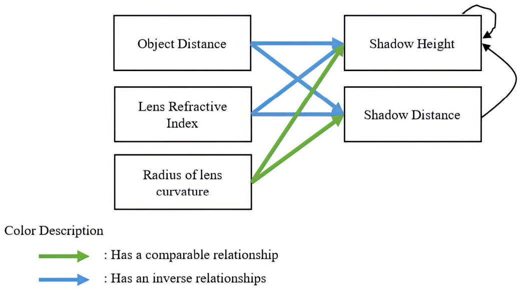

Fig. 9. System complexity of geometric optics simulation.

C. Expert Assessment of Ph-CPS

Table 2 presents the results of the expert assessment of the four Ph-CPS items. An expert assessment is conducted to determine whether the Ph-CPS item meets the criteria for a complex problem. Criteria for complex problems include: 1) classified as domain-general [11, 12]; 2) system problems involve many variables [27, 41]; 3) the variables involved are interconnected and interdependent [41]; 4) the problem is presented in a dynamic system [27]; 5) connections between variables involved in the problem are not transparently stated [27, 41]; 6) the problem has multiple objectives resulting in conflict of objectives [41]; and 7) the problem meets the criteria of a minimally complex system [43].

Table 2 shows that the four Ph-CPS items meet seven complex problem criteria. The four Ph-CPS items have a V_{value} greater than V_{table} (V_{table} = 0.67) for all criteria [70]. Thus, the results of the expert assessment show that the four Ph-CPS items meet the criteria as complex problems.

Table 2. V-Aiken value of expert validation results

Complex Problem Indicators	Ph-CPS 1	Ph-CPS 2	Ph-CPS 3	Ph-CPS 4
Classified as a domain-general skill (Problems can be solved without relying on mastery of the subject matter)	0.95*	0.88*	0.95*	0.97*
Problems with the system involve many variables	0.90*	0.87*	0.97*	0.97*
The variables involved in the problem are interconnected and interdependent with each other	0.87*	0.90*	0.98*	0.97*
The problem is presented in a dynamic system	0.88*	0.85*	0.97*	0.98*
The connection between the variables involved in the problem is not stated transparently	0.93*	0.87*	0.93*	0.98*
Problems have many goals resulting in conflicting goals	0.95*	0.88*	0.93*	1.00*
The problem meets the minimal complex system criteria	0.90*	0.92*	0.97*	1.00*
Average Score	**0.91***	**0.88***	**0.96***	**0.98***

Note: * The V_{count} is greater than V_{table} (V_{table} for 15 raters with a rating of 5 is **0.67**).

Computer-based tests that present dynamic simulations are used to measure CPS skills. The commonly used tests to measure CPS skills to date are MultiFlux [44], MicroDYN [45], Genetics Lab [46], and MicroFIN [47]. The four tests were developed in the domain-general skill test format. Teaching CPS in the classroom requires developing a CPS test using subject matter as the context of the problem. This Ph-CPS skill test was developed using physics material as the context of the problem, not as problem content. The context of the problem in Ph-CPS is presented in a dynamic system using PhET simulations.

The results of expert validation showed that the Ph-CPS test using PhET simulations met the criteria for complex problems. Although the Ph-CPS test uses physics subject matter as the context of the problem, experts agree that all question items on the Ph-CPS test can be solved without requiring mastery of physics concepts. Many studies state that prior knowledge contributes independently to CPS [29, 78]. However, prior knowledge is not always beneficial when it is only used superficially and does not align with deeper structural understanding [41]. The Einstellung effect proves that prior knowledge hinders problem-solving [79].

The experts also considered that the complex problems in the four Ph-CPS test items involved many variables. In the complex problem presented, each Ph-CPS test item consists of at least two endogenous and two exogenous variables (see Figs. 3, 5, 7, and 9). A system is categorized as a problem with high complexity if it consists of at least two exogenous variables and two endogenous variables [43].

Experts assess that the variables involved in the problem in each Ph-CPS test item developed are interconnected and interdependent. Connections and dependencies between variables can be explained through mathematical equations [80]. The connections and interdependencies among the variables involved in the Ph-CPS test are governed by the laws and principles of physics as applied in the problem context. The connection will be known after the respondent successfully explores the environmental system problems presented in the PhET simulation. The physical laws and principles used to present the connections and dependencies between variables in each Ph-CPS item are presented in Table 3.

Table 3. Laws and principles of physics on connections and dependencies between variables

Item	Context of the Issue	Laws or Principles of Physics
Ph-CPS 1	Bending lights	- Snell's Law of Refraction
Ph-CPS 2	Gravity and orbits	- Kepler's Law - Newton's Law of Gravitation
Ph-CPS 3	Projectile motion	- Newton's Law - Newton's Law of Gravitation
Ph-CPS 4	Geometric Optics	- Snellius's Law of Refraction

Experts assess the Ph-CPS test presents problems in a dynamic system using PhET simulations. PhET simulations are interactive simulations developed by the University of Colorado [81] that can present dynamic simulations. PhET simulations allow students to manipulate variables and observe real-time outcomes [37].

Experts assess that the connection between the variables involved in the Ph-CPS test problem is not stated transparently. The complex problems presented in the Ph-CPS test do not specify which variables are known and which are to be found. The Ph-CPS test only presents problem situations within the context of physics and asks respondents to achieve the desired goals without specifying the variables being asked. The non-transparency of a problem refers to the extent to which the target situation, the variables involved, the interaction and the dynamics cannot be ascertained [82].

Experts also consider that the Ph-CPS test problem has many objectives, resulting in a conflict of goals. Although the primary objective focuses on one of the endogenous variables, side effects from other endogenous variables make the Ph-CPS problem have many objectives. Side effects appear unwanted and unexpected due to interventions in complex systems [83]. Side effects can cause new problems or worsen existing ones [84] creating new urgent goals that must be addressed. Anticipating side effects is one of the keys to success in CPS [42], as side effects can create new goals that interfere with the primary objectives of the problem.

Experts assess that the Ph-CPS problems meet the criteria of a minimal complex system. In this criterion, the V_{values} for Ph-CPS 1, Ph-CPS 2, Ph-CPS 3 and Ph-CPS 4 were 0.90, 0.92, 0.97, and 1.0, respectively. These results indicate that the four Ph-CPS items meet the minimal complex system criteria with a high complexity category [43]. The minimal complex system criterion is the minimum limit of a problem categorized as a complex problem system [85].

Expert validation results confirm that the PhET simulation can help assess CPS skills in physics. These findings support the transferability of CPS to educational environments. Educators can employ the PhET simulation to assess and instruct CPS using physics materials. The Ph-CPS applies to instructing physics across all levels of students' cognitive capacities. Students with insufficient physics concepts will utilize domain-general skills to finish the Ph-CPS while acquiring new physics knowledge. In contrast, students with advanced proficiency in physics concepts will utilize physics-specific knowledge to complete the Ph-CPS. General cognitive skills are more influential in complex problems for younger students [86], but older and more expert students rely more on domain-specific knowledge [32, 35]. It indicates that the transferability of Ph-CPS to other CPS contexts is challenging due to its dependence on the overlap of the KSAO components (knowledge, skills, abilities, and other characteristics) required for both CPS contexts [87].

PhET simulations can assess and instruct on CPS across all science fields, including chemistry and biology. PhET simulations offer an interactive platform for students to actively explore and practice sense-making, facilitating the construction and testing of hypotheses [39]. PhET simulations can demonstrate real-time alterations due to student interactions, enabling learners to comprehend the cause-and-effect dynamics of systems [88]. These two elements are fundamental components of CPS. Nonetheless, not all PhET simulations satisfy the complexity criteria. Educators must discern PhET simulations that satisfy the minimum complexity standards for their practical application in assessing and instructing CPS in the classroom. Moreover, developing dynamic microworld-based simulations that integrate subject matter material is essential to enhance the transferability of CPS to educational environments as a long-term solution.

Although only four physics simulations in PhET satisfy the minimal criteria for assessing Physics-CPS (Ph-CPS), this finding paves the way for utilizing other simulation platforms to assess and teach CPS. Previous studies indicate that ChemLab Builder effectively assesses CPS skills in chemistry by providing a virtual laboratory environment where students may systematically plan experiments, manipulate variables, and test hypotheses [22]. Moreover, additional studies validate that diverse simulation-based microworlds—such as MicroDYN and Genetics Lab—effectively assess CPS across multiple domains by posing challenges that necessitate knowledge acquisition and the implementation of problem-solving strategies [89, 90].

Consequently, identifying and utilizing readily available simulations (such as PhET or similar platforms) is a practical, immediate answer for integrating CPS into learning environments. This corresponds with the perspective that the transferability of CPS can be improved by the development of simulation-based learning environments that highlight system dynamics, unclear relationships between variables, and necessitate active exploration by learners [27, 29]. This solution addresses the constraints of restricted simulations and enhances teaching methods centred on 21st-century competencies.

D. Validity and Reliability of Ph-CPS Test

Four Ph-CPS skill test items were tested on 248 respondents. The analysis results using the Rasch model on the fit item criteria are presented in Table 4.

Table 4 describes that three items meet all three validity standards, i.e. Ph-CPS 1, Ph-CPS 2, and Ph-CPS 4. Ph-CPS item 3 only meets the PTMEASUR-ALCORR criteria. It indicates that only Ph-CPS 1, Ph-CPS 2, and Ph-CPS 3 items meet construct validity. Table 4 shows that Ph-CPS item 3 has an MNSQ outfit value greater than 1.5. It indicates that item Ph-CPS 3 is incompatible with the other three items [91]. MNSQ values outside the range between 0.5 to 1.5 can indicate problems with the item [92]. MNSQ outfit values exceeding 1.3 are considered an indication of item misfit [93]. Item Ph-CPS 3 has an MNSQ outfit value of 1.91, which is classified as high. It is because the Ph-CPS item 3 is classified as a problem item that is too easy when compared to the other three items based on the logit map presented in Fig. 10.

Table 4. Item Fit (item = 4)

Item	Item Measure	Outfit MNSQ	Outfit ZSTD	PTMEASUR-AL CORR
Ph-CPS 1	1.45	0.82	−1.96	0.61
Ph-CPS 2	0.91	0.88	−1.40	0.66
Ph-CPS 3	−2.74	1.91	3.38	0.42
Ph-CPS 4	0.39	0.98	−0.20	0.63
Validity Standard [94]		0.77< Outfit MNSQ <1.5	−2< Outfit ZSTD <2	0.4<PTMEASUR-ALCORR < 0.85

Note: Cronbach Alpha: 0.69; Item Reliability: 0.99.

Table 4 explains that the Ph-CPS item 3 has a ZSTD outfit value greater than 2. It shows that Ph-CPS item 3 is classified as an outlier item [95]. Item Ph-CPS 3 has a ZSTD outfit value of 3.38, which is classified as high. It indicates that the Ph-CPS item 3 deviates considerably from the other three items, so the Ph-CPS item 3 can be considered for deletion.

Table 4 shows that the four Ph-CPS skill test items meet the PTMEASUR-ALCORR criteria. It indicates that each item correlates with the item's total score. Although each item contributes to the measurement model, Ph-CPS 3 has the smallest contribution value to the measurement model when compared to the other three items, which is 0.42. The lowest PTMEASUR-AL CORR value for the Ph-CPS 3 item indicates that this item has the weakest differentiating power compared to the other three items [92, 95].

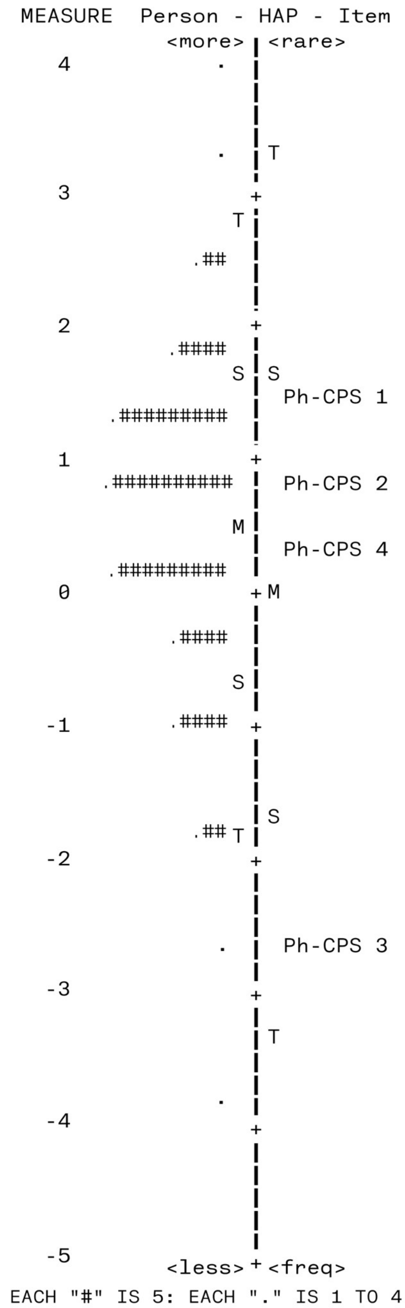

Fig. 10. Logit map.

The reliability test results on the four items showed that the Cronbach Alpha value was 0.69 and the item reliability value was 0.99. A low Cronbach's Alpha value indicates that the items in the developed test do not measure the same dimension, even though the internal consistency of the items, as shown by the item reliability, is high. It is usually caused by the presence of an item that deviates from the others because it is either too easy or too difficult. Table 1 shows that Ph-CPS item 3 is classified as an item that is too easy based on item measure value (−2.74 < −1.37) and MNSQ outfit value (**1.91** > 1.33). Thus, Ph-CPS 3 needs to be removed to obtain a reliable test. Table 4 also shows that Ph-CPS item 1 is classified as a complicated problem (item measure value = **1.45** > 1.37). However, Ph-CPS item 1 was retained because the validity criteria (Outfit MNSQ, Outfit ZSTD and PTMEASUR-ALCORR) were in the valid category.

The validity and reliability test results using the Rasch model on the four items showed that the Ph-CPS 3 item needed to be removed, and the other three items remained in use. Furthermore, re-validity and re-reliability calculations were performed on Ph-CPS 1, Ph-CPS 2 and Ph-CPS 4 items using the Rasch model. The results of the calculation of validity and re-reliability of Ph-CPS 1, Ph-CPS 2 and Ph-CPS 4 items are presented in Table 5.

Table 5. Item fit (item = 3)

Item	Item Measure	Outfit MNSQ	Outfit ZSTD	PTMEASUR-AL CORR
Ph-CPS 1	0.68	1.00	0.02	0.74
Ph-CPS 2	−0.01	0.92	−0.88	0.66
Ph-CPS 4	−0.67	1.13	1.51	0.68
Validity Standard [94]		0.77< Outfit MNSQ <1.5	−2< Outfit ZSTD <2	0.4<PTMEASUR-ALCORR < 0.85

Note: Cronbach Alpha: 0.77; Item Reliability: 0.96.

Table 5 illustrates that the three items meet the criteria of Outfit MNSQ, Outfit ZSTD and PTMEASUR-AL CORR. It indicates that the three items can be categorized as valid items. The results of the re-reliability test on the three items showed that the Cronbach Alpha value was 0.77 and the item reliability value was 0.96. A Cronbach Alpha value greater than 0.70 indicates that all three items measure the exact dimensions in the acceptable category [96]. The three items also demonstrated excellent internal consistency, as indicated by a high item reliability value (0.96). Thus, Ph-CPS 1, Ph-CPS 2, and Ph-CPS 4 items meet the construct validity and are reliable.

The concurrent validity test was conducted by correlating the Ph-CPS skill test items that met construct validity and reliability with the PISA 2012 CPS test. The correlation test results are presented in Table 6. Table 6 mentions that each item of Ph-CPS 1, Ph-CPS 2, and Ph-CPS 4 is significantly correlated to the PISA 2012 CPS test with consecutive correlation values of 0.436, 0.542, and 0.580. The Ph-CPS score also showed a significant correlation to the PISA 2012 CPS test, with a correlation value of 0.751. It depicts that the Ph-CPS skill test has concurrent validity with the PISA 2012 CPS test.

Table 6. Correlation test results between Ph-CPS skill test and PISA 2012 CPS test

Ph-CPS	Item	Total Ph-CPS	Total CPS PISA 2012
Ph-CPS 1	Pearson Correlation	0.650	0.436
	Sig.	0.000	0.000
Ph-CPS 2	Pearson Correlation	0.765	0.542
	Sig.	0.000	0.000
Ph-CPS 4	Pearson Correlation	0.666	0.580
	Sig.	0.000	0.000
Total Ph-CPS	Pearson Correlation	1.000	0.751
	Sig.	–	0.000
Total CPS PISA 2012	Pearson Correlation	0.751	1.000
	Sig.	0.000	–

Table 6 also demonstrates that Ph-CPS 1, Ph-CPS 2, and Ph-CPS 4 items meet the validity of classical theory items. It is indicated by each item of Ph-CPS 1, Ph-CPS 2, and Ph-CPS 4, which are significantly correlated to the total Ph-CPS score with consecutive correlation values of 0.650, 0.765, and 0.666. The correlation values of Ph-CPS 1, Ph-CPS 2 and Ph-CPS 4 belong to a strong correlation [97]. It can also be interpreted that Ph-CPS 1, Ph-CPS 2 and Ph-CPS 4 have strong validity.

Based on the results of item analysis using both classical test theory and the Rasch model, only three items were found to be valid and reliable: Ph-CPS 1, Ph-CPS 2, and Ph-CPS 4. Ph-CPS 3 was deemed invalid because the question was too easy, making it inconsistent with the other three items. By removing item Ph-CPS 3, the reliability of the Ph-CPS skills test increased [98]; thus, the Ph-CPS skills test was considered reliable.

Although the reliability value of the Ph-CPS increased after the deletion of Ph-CPS item 3, this deletion caused the reliability value of the remaining Ph-CPS to be undervalued [99]. The unstable Cronbach Alpha value due to the small number of items is a limitation of this study. However, the internal reliability value remains high, indicating that the Ph-CPS items consistently measure the exact dimensions.

The results of the validity and reliability test using the Rasch model show that the Ph-CPS test construct is valid and reliable if Ph-CPS item 3 is deleted. Ph-CPS 3 items that present problems in the context of projectile motion are classified as very straightforward compared to other Ph-CPS items. It is because problems related to projectile motion have similarities with Angry Birds games [100, 101], making these problems familiar beforehand. Respondents will find it easier to solve previously known problems. These previously known problems are more commonly referred to by the term “familiar problem.”

Individuals facing familiar problems, such as the Ph-CPS 3 item, will use strategies successfully applied before without going through an exploration phase [102]. They may skip the knowledge acquisition phase to immediately apply their strategies in the Angry Birds game when completing Ph-CPS 3 items.

One of the characteristics of a complex problem is that the connections between the variables involved in the Ph-CPS test problem are not stated transparently [27, 41]. Most respondents are familiar with the Angry Birds game, similar to the Ph-CPS 3 item. It makes them familiar with the connections between variables in Ph-CPS item 3. Thus, Ph-CPS item 3 is no longer a complex problem for respondents familiar with the Angry Birds game.

The Ph-CPS test has concurrent validity with the PISA 2012 test. The PISA 2012 test can be used to measure CPS skills [68, 103]. A new test construct with a significant correlation with a similar, already validated test means that the new test construct has concurrent validity [104]. Thus, the Ph-CPS test measures the same construct as what was measured on the PISA 2012 Test.

The result of the Rasch model analysis of the four Ph-CPS items showed that only three items (Ph-CPS 1, Ph-CPS 2, and Ph-CPS 4) met validity and reliability criteria, while Ph-CPS item 3 was invalid because it was categorized as a familiar problem. This finding aligns with previous studies that one of the main characteristics of complex problems is the opaque relationships between variables and the demand for active exploration by problem solvers [27, 48]. In solving Ph-CPS item 3, respondents showed familiarity with the context of projectile motion, similar to the game Angry Birds. It suggests that Ph-CPS item 3 only involves familiar strategies and no longer requires extensive exploration, thus reducing the authenticity of the CPS measurement [29].

Removing Ph-CPS item 3 improved the measurement's internal consistency, as reflected in increases in Cronbach's Alpha and item reliability. It confirms that the success of simulation-based CPS assessments is highly dependent on item quality. Items that are too easy or too familiar will reduce the item discrimination, while items that require the manipulation of dynamic variables with opaque causal relationships actually strengthen construct validity [89, 105].

Beyond the technical aspects of the test, these results have important pedagogical implications for the CPS's transferability to learning environments. The Ph-CPS test can be an authentic assessment that simultaneously measures domain-specific and domain-general learning outcomes relevant to 21st-century skills learning objectives [90, 106]. Teachers can integrate this test into inquiry-based or problem-based learning strategies so that students not only master physics material but also develop CPS skills such as variable control, cause-and-effect analysis, and decision-making in dynamic systems.

However, it is important to acknowledge the limitations of this test. Students' level of familiarity with the context of the questions can reduce the construct validity of the CPS, as was the case with Ph-CPS 3. Therefore, it is key for teachers to select simulations that meet the criteria of a minimum complex system in order to successfully integrate simulation-based CPS assessments into the classroom [27, 107]. Considering all the above factors, the Ph-CPS test can contribute to a more authentic competency-based assessment, such as the national assessment currently being developed by the Ministry of Primary and Secondary Education of the Republic of Indonesia.

This study is limited by the number of items tested, which consists of only four. This technical limitation stems from the application's exclusive use of PhET simulations, with the subject matter restricted to physics. In the PhET simulation, there are 66 physics simulations out of 120, and only four simulations meet the criteria for complex problems and minimal complex systems. In the short term, this study can be expanded by using PhET simulations to measure CPS skills not only limited to physics subject matter but also other STEM subject materials available in PhET, such as biology, chemistry, mathematics, and earth and space. In the long term, the key to successfully transferring CPS skills in educational environments is developing new applications through End-User Development (EUD) that enable teachers to create dynamic simulations based on the subject matter [108].

E. Domain of Ph-CPS Test

The regression test between the Ph-CPS test and mastery of physics concepts was carried out to determine whether the Ph-CPS test was helpful in domain-general or domain-specific skills. Ph-CPS is declared a domain-general skill if concept mastery does not contribute to the Ph-CPS test. The results of the regression test between the Ph-CPS test and mastery of physics concepts are presented in Table 7.

Table 7. Regression test results between Ph-CPS and physics concept mastery (N = 248)

R	Sig	R^2	% of variance explained	b	seb	Beta	Sig. F
0.004	0.475	0.000	0.000	42.457	0.069	0.004	0.949

Table 7 illustrates the correlation between mastery of physics concepts and Ph-CPS skills of 0.004 with a significance value of 0.475. A correlation significance greater than 0.05 indicates no significant relationship between mastery of physical concepts and Ph-CPS skills [109]. The R-squared value between mastery of physics concepts and Ph-CPS skills is 0.000. It explains that mastery of physics concepts does not contribute to Ph-CPS skills [110]. Table 7 shows that the Beta value is very small, 0.004. It shows that the effect of mastery of physics concepts on Ph-CPS skills is very small [111] so it can be concluded to have no effect.

The linear regression equation of the relationship model between Ph-CPS skills and mastery of physical concepts is written in Eq. (1). Table 7 shows that the value of sig. F is 0.949, which is greater than 0.05. It shows that the regression model presented in Equation 1 is insignificant in showing the relationship between Ph-CPS skills and mastery of physics concepts [112].

$$Y = 42.457 + 0.004X \tag{1}$$

where:

Y = Ph-CPS Skill

X = Mastery of physics concepts

The Ph-CPS test is a CPS test that uses physics subject matter as the context of the problem. The development of the Ph-CPS test is an effort to expand CPS research from the field of psychology into the field of education. Applying CPS skills research in education by involving specific subject matter must ensure that CPS skills in the context of the subject matter are classified as domain-general skills.

The regression test results showed that physics concept mastery does not affect Ph-CPS skills. It indicates that the Ph-CPS test construct, which measures CPS skills in physics subject matter, is classified as a domain-general skill. It aligns to develop the CPS test, namely measuring domain-general skills. [113]. CPS skills are classified as domain-general skills because they involve cognitive skills, such as reasoning, planning, and self-regulation, that are not tied to specific subject matter [114].

This study's results classify Ph-CPS as a domain-general skill, but it is important to avoid oversimplifying this conclusion. Debates in the CPS literature emphasize that domain-general and domain-specific aspects are not mutually exclusive but interact in complex ways [27, 90, 115]. Even when domain knowledge does not significantly predict performance, the physics context of Ph-CPS test can still activate embedded domain knowledge that supports reasoning and exploration [34]. In other words, using physics as a test context still involves domain representations, even if it is not the primary determinant of performance.

Transfer theory is also crucial to explaining these findings. CPS research shows that students employ general strategies across domains like hypothesis generation, variable control, and causal reasoning. However, these strategies are often anchored in domain representations available in the context [12, 116]. It suggests that Ph-CPS can simultaneously capture domain-general processes while remaining embedded in domain-specific contexts, thus enabling strategy transfer and new knowledge formation.

The absence of influence of mastery of physics concepts on Ph-CPS skills shows that Ph-CPS skills are not tied to physics material, so that Ph-CPS skills can be transferred to CPS with the context of other problems outside of physics subject matter. Transferring problem-solving strategies from one domain to another will increase a person's adaptability and flexibility in facing new challenges [117]. It explains why domain-general problem-solving skills, such as CPS, are needed in education in the 21st century [118].

However, a more comprehensive perspective suggests that the Ph-CPS is not entirely devoid of content; instead, it reflects domain-general cognitive processes functioning within a physics context, which implicitly supports reasoning. The dual nature of simulation-based CPS assessments renders them pertinent to education, evaluating transferable skills while maintaining alignment with the curriculum.

Teaching CPS skills with specific subject matter in the classroom, such as Ph-CPS, is required now, even though these skills belong to domain-general skills. Students will acquire physics knowledge that is used as the context of the problem after completing a Ph-CPS test as a construct of new knowledge. Complex problem solving stimulates the formation of new knowledge structures because students build internal models of the problem environment [27]. Students will gain knowledge while solving complex problems due to their active interaction with the problem [12].

The theory of embedded knowledge in CPS holds that complex problem solving stimulates the formation of new mental models by connecting domain-general strategies to domain-specific representations [29, 107]. This theory aligns with the perspective that CPS can be taught in classrooms using subject matter, such as physics. Thus, this study contributes to the domain-general versus domain-specific debate by demonstrating that, although the Ph-CPS tends to behave as a domain-general construct, it is still shaped by the surrounding physics context.

V. Conclusion

This study demonstrates that PhET simulations can be an effective test for assessing and teaching CPS skills in the context of physics, while also opening up opportunities for integration into other science disciplines. Ph-CPS exhibits domain-general characteristics that remain embedded in the material context, making it relevant to accommodate diverse levels of student conceptual mastery. These findings emphasize the importance of utilizing dynamic simulations based on minimally complex systems as a practical and sustainable strategy to strengthen authentic assessment and the development of 21st-century competencies. Thus, Ph-CPS contributes to the theoretical debate on the CPS domain and provides a practical foundation for transforming science assessment and learning toward a more innovative and relevant educational system for the challenges of the 21st century.

The results of this study confirm that the Ph-CPS test developed using the PhET simulation is valid and reliable after the overly familiar item (Ph-CPS 3) was removed. This finding emphasizes that item quality plays a central role in the success of simulation-based CPS assessments, where items requiring active exploration, variable control, and non-transparent cause-and-effect relationships strengthen construct validity. Thus, the Ph-CPS test is relevant as a measure of physics mastery and an authentic assessment that encompasses higher-order thinking skills, particularly CPS that is cross-domain. Practically, this test has the potential to be integrated into inquiry-based and problem-based learning so that students gain a more meaningful learning experience. Furthermore, the results of this study support the direction of competency-based assessment policies currently being developed nationally, while also opening up opportunities for further research to test the effectiveness of Ph-CPS in an interdisciplinary context, expand its application to other science fields such as chemistry and biology, and explore its integration with new technologies such as virtual reality and artificial intelligence to enrich the learning ecosystem and measurement of CPS skills in the 21st century.

This study shows that the Ph-CPS test functions primarily as a domain-general construct because mastery of physics concepts does not significantly influence test results. However, participant performance is still influenced by the physics context, so the Ph-CPS represents the interaction between domain-general strategies and domain-specific representations. These findings confirm the relevance of the PhET simulation as an authentic assessment that not only measures transferable cognitive skills, such as CPS, but also supports the formation of new knowledge in physics learning. Thus, the Ph-CPS can be a model for 21st-century competency assessment that integrates CPS skills into science education. These results contribute to an example of the transferability of CPS into educational environments, particularly physics learning.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Concept and design, EJ & WW; Reviewed the methodology research, WW, SEN, & BNM; Data acquisition, EJ & WW; Data analysis, EJ & BNM; statistical analysis EJ & BNM; Drafting manuscript, EJ & SEN; Writing, EJ; Supervision, WW, SEN, & BNM; Critical revision of manuscript, WW, SEN, & BNM; Providing technical support, SEN. All authors had approved the final version.

FUNDING

This research was funded by the Directorate of Research, Technology, and Community Service, Ministry of Education and Culture of the Republic of Indonesia based on Decree Number: 033/E5/PG.02.00/2022.

REFERENCES

[1] B. Eichmann, S. Greiff, J. Naumann, L. Brandhuber, and F. Goldhammer, "Exploring behavioural patterns during complex problem-solving," *J. Comput. Assist. Learn.*, vol. 36, no. 6, pp. 933–956, Dec. 2020. doi: 10.1111/jcal.12451

[2] C. Ansell, E. Sørensen, and J. Torfing, "The COVID-19 pandemic as a game changer for public administration and leadership? The need for robust governance responses to turbulent problems," *Public Manag. Rev.*, vol. 23, no. 7, pp. 949–960, 2021. doi: 10.1080/14719037.2020.1820272

[3] R. Wang, J. Chen, K. Gao, and G. W. Wei, "Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries," *Genomics*, vol. 113, no. 4, pp. 2158–2170, Jul. 2021. doi: 10.1016/j.ygeno.2021.05.006

[4] T. M. Letcher, "Global warming-a complex situation," in *Climate Change: Observed Impacts on Planet Earth*, Third Edition, T. M. Letcher, Ed., Amsterdam, The Netherlands: Elsevier, 2021. pp. 3–17. doi: 10.1016/B978-0-12-821575-3.00001-3

[5] C. Veríssimo, L. Pereira, A. Fernandes, and R. Martinho, "Complex problem solving as a source of competitive advantage," *J. Open Innov. Technol. Mark. Complex.*, vol. 10, no. 2, Jun. 2024. doi: 10.1016/j.joitmc.2024.100258

[6] J. Massolt and A. Borowski, "Perceived relevance of university physics problems by pre-service physics teachers: personal constructs," *Int. J. Sci. Educ.*, vol. 42, no. 2, pp. 167–189, Jan. 2020. doi: 10.1080/09500693.2019.1705424

[7] S. Lee, Y. il Choi, and S. W. Kim, "Roles of emotions induced by immediate feedback in a physics problem-solving activity," *Int. J. Sci. Educ.*, vol. 43, no. 10, pp. 1525–1553, 2021. doi: 10.1080/09500693.2021.1922778.

[8] N. Supriadi, W. Jamaluddin Z, and S. Suherman, "The role of learning anxiety and mathematical reasoning as predictor of promoting learning motivation: The mediating role of mathematical problem solving," *Think. Skills Creat.*, vol. 52, Jun. 2024. doi: 10.1016/j.tsc.2024.101497

[9] H. Chen, L. Disney, and L. Li, "Studying children's motives in mathematical problem-solving during transition from kindergarten to school: A conceptual PlayWorld approach," *Learn Cult. Soc. Interact.*, vol. 48, 100850, Oct. 2024. doi: 10.1016/j.lcsi.2024.100850

[10] R. J. Marzano, "What are the general skills of thinking and reasoning and how do you teach them?" *Clear. House A J. Educ. Strateg. Issues Ideas*, vol. 71, no. 5, pp. 268–273, May 1998. doi: 10.1080/00098659809602721

[11] G. Molnár, S. Greiff, S. Wüstenberg, and A. Fischer, "Empirical study of computer-based assessment of domain-general complex problem-solving skills," in *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning*, B. Csapó, Ed., Paris: OECD Publishing, ch. 8, 2017. doi: 10.1787/9789264273955-10-en

[12] S. Wüstenberg, S. Greiff, and J. Funke, "Complex problem solving—More than reasoning?" *Intelligence*, vol. 40, no. 1, pp. 1–14, Jan. 2012. doi: 10.1016/j.intell.2011.11.003

[13] S. Wüstenberg, M. Stadler, J. Hautamäki, and S. Greiff, "The role of strategy knowledge for the application of strategies in complex problem solving tasks," *Technol. Knowl. Learn.*, vol. 19, no. 1–2, pp. 127–146, 2014. doi: 10.1007/s10758-014-9222-8

[14] K. Perkins *et al.*, "PhET: Interactive simulations for teaching and learning physics," *Phys. Teach.*, vol. 44, no. 1, pp. 18–23, Jan. 2006. doi: 10.1119/1.2150754

[15] M. Niss, "What is physics problem-solving competency? The views of Arnold Sommerfeld and Enrico Fermi," *Sci. Educ. (Dordr)*, vol. 27, no. 3–4, pp. 357–369, May 2018. doi: 10.1007/s11191-018-9973-z

[16] A. Fischer, S. Greiff, and J. Funke, "The process of solving complex problems," *J. Probl. Solving*, vol. 4, pp. 19-42, 2012. doi: 10.7771/1932-6246.1118

[17] S. Leinweber, "Stage 3: Competency management," in *Strategic Human Resource Development: A Program in Eight Stages*, M. T. Meifert, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 144–178, 2010. doi: 10.1007/978-3-642-04401-4_7 (in German)

[18] F. Hecklau, M. Galeitzke, S. Flachs, and H. Kohl, "Holistic approach for human resource management in industry 4.0," *Procedia CIRP*, vol. 54, pp. 1–6, 2016. doi: 10.1016/j.procir.2016.05.102

[19] P. Fitsilis, P. Tsoutsa, and V. Gerogiannis, "Industry 4.0: Required personnel competences," *Industry 4.0*, vol. 3, no. 3, pp. 130-133, 2018.

[20] World Economic Forum. (Oct. 2020). The Future of Jobs Report 2020. [Online]. Available: https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf

[21] S. Greiff, "Assessment and theory in complex problem solving—A continuing contradiction?" *J. Educ. Develop. Psychol.*, vol. 2, no. 1, Apr. 2012. doi: 10.5539/jedp.v2n1p49

[22] R. Scherer, J. Meßinger-Koppelt, and R. Tiemann, "Developing a computer-based assessment of complex problem solving in chemistry," *Int. J. STEM Educ.*, vol. 1, no. 1, Dec. 2014. doi: 10.1186/2196-7822-1-2

[23] R. H. Stevens, A. C. Lopo, and P. Wang, "Artificial neural networks can distinguish novice and expert strategies during complex problem solving," *J. Am. Med. Inform. Assoc.*, vol. 3, no. 2, pp. 131–138, 1996. doi: 10.1136/jamia.1996.96236281

[24] W. K. Adams and C. E. Wieman, "Analyzing the many skills involved in solving complex physics problems," *Am J Phys*, vol. 83, no. 5, pp. 459–467, May 2015. doi: 10.1119/1.4913923

[25] K. VanLehn *et al.*, "Minimally invasive tutoring of complex physics problem solving," in *Intelligent Tutoring Systems*, S. A. Cerri, G. Gouardères, and F. Paraguaçu, Eds., Lecture Notes in Computer Science, Berlin, Heidelberg: Springer Berlin Heidelberg, vol. 2363, pp. 367–376, 2002. doi: 10.1007/3-540-47987-2_40

[26] B. Eichmann, F. Goldhammer, S. Greiff, L. Pucite, and J. Naumann, "The role of planning in complex problem solving," *Comput. Educ.*, vol. 128, pp. 1–12, Jan. 2019. doi: 10.1016/j.compedu.2018.08.004

[27] J. Funke, "Complex problem solving: A case for complex cognition?" *Cogn. Process.*, vol. 11, pp. 133–142, May 2010. doi: 10.1007/s10339-009-0345-0

[28] OECD, *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy,* Paris, France: OECD Publishing, 2013. doi: 10.1787/9789264190511-en

[29] H. M. Süß and A. Kretzschmar, "Impact of cognitive abilities and prior knowledge on complex problem solving performance—Empirical results and a plea for ecologically valid microworlds," *Front. Psychol.*, vol. 9, May 2018. doi: 10.3389/fpsyg.2018.00626

[30] W. Schoppek and A. Fischer, "Common process demands of two complex dynamic control tasks: Transfer is mediated by comprehensive strategies," *Front. Psychol.*, vol. 8, Dec. 2017. doi: 10.3389/fpsyg.2017.02145

[31] C. N. Herde, S. Wüstenberg, and S. Greiff, "Assessment of complex problem solving: What we know and what we don't know," *Appl. Meas. Educ.*, vol. 29, no. 4, Oct. 2016. doi: 10.1080/08957347.2016.1209208

[32] S. Greiff, S. Wüstenberg, G. Molnár, A. Fischer, J. Funke, and B. Csapó, "Complex problem solving in educational contexts—Something beyond g: Concept, assessment, measurement invariance, and construct validity," *J. Educ. Psychol.*, vol. 105, no. 2, pp. 364–379, May 2013. doi: 10.1037/a0031856

[33] D. Dörner and J. Funke, "Complex problem solving: What it is and what it is not," *Front. Psychol.*, vol. 8, Jul. 2017. doi: 10.3389/fpsyg.2017.01153

[34] J. Neubert, T. Lans, M. Mustafic, S. Greiff, and P. Ederer, "Complex problem-solving in a changing world: Bridging domain-specific and transversal competence demands in vocational education," in *Competence-based Vocational and Professional Education: Bridging the Worlds of Work and Education*, M. Mulder, Ed., Cham, Swiss: Springer International Publishing, 2017, pp. 953–969. doi: 10.1007/978-3-319-41713-4_44

[35] J. Schäfer, T. Reuter, J. Karbach, and M. Leuchter, "Domain-specific knowledge and domain-general abilities in children's science problem-solving," *Br. J. Educ. Psychol.*, vol. 94, no. 2, pp. 346–366, Jun. 2024. doi: 10.1111/bjep.12649

[36] J. E. Bartley *et al.*, "Meta-analytic evidence for a core problem solving network across multiple representational domains," *Neurosci. Biobehav. Rev.*, vol. 92, pp. 318–337, Sep. 2018. doi: 10.1016/j.neubiorev.2018.06.009

[37] C. E. Wieman, W. K. Adams, and K. K. Perkins, "PhET: Simulations that enhance learning," *Science*, vol. 322, no. 5902, pp. 682–683, Oct. 2008. doi: 10.1126/science.1161948

[38] N. D. Finkelstein *et al.*, "When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment," *Phys. Rev. ST Phys. Educ. Res.*, vol. 1, no. 1, 010103, Oct. 2005. doi: 10.1103/PhysRevSTPER.1.010103

[39] N. S. Podolefsky, K. K. Perkins, and W. K. Adams, "Factors promoting engaged exploration with computer simulations," *Phys. Rev. ST Phys. Educ. Res.*, vol. 6, no. 2, 020117, Oct. 2010. doi: 10.1103/PhysRevSTPER.6.020117

[40] P. Sonnleitner, U. Keller, R. Martin, and M. Brunner, "Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success," *Intelligence*, vol. 41, no. 5, pp. 289–305, Sep. 2013. doi: 10.1016/j.intell.2013.05.002

[41] J. Funke, "Complex problem solving," in *Encyclopedia of the Sciences of Learning*, N. M. Seel, Ed., Heidelberg, Germany: Springer, pp. 682–685, 2012. doi: 10.1007/978-1-4419-1428-6_685

[42] S. Greiff and J. Funke, "Measuring complex problem solving: The MicroDYN approach," in *The Transition to Computer-Based Assessment. New Approaches to Skills Assessment and Implications for Large-Scale Testing*, F. Scheuermann and J. Björnsson, Eds., Luxemburg: Office for Official Publications of the European Communities, pp. 157–163, 2009. doi: 10.11588/heidok.00015502

[43] J. Funke and S. Greiff, "Dynamic problem solving: Multiple-item testing based on minimally complex systems," in *Competence Assessment in Education: Research, Models and Instruments. Methodology of Educational Measurement and Assessment.*, D. Leutner, J. Fleischer, J. Grünkorn, and E. Klieme, Eds., Cham, Swiss: Springer, ch. 25, pp. 427–443, 2017. doi: 10.1007/978-3-319-50030-0_25

[44] S. Kröner, J. L. Plass, and D. Leutner, "Intelligence assessment with computer simulations," *Intelligence*, vol. 33, no. 4, pp. 347–368, Jul. 2005. doi: 10.1016/j.intell.2005.03.002

[45] S. Greiff, S. Wüstenberg, and J. Funke, "Dynamic problem solving: A new assessment perspective," *Appl. Psychol. Meas.*, vol. 36, no. 3, pp. 189–213, May 2012. doi: 10.1177/0146621612439620

[46] P. Sonnleitner *et al.*, "The genetics lab: Acceptance and psychometric characteristics of a computer-based microworld assessing complex problem solving," *Psychol. Test. Assess. Model.*, vol. 54, no. 1, pp. 54–72, 2012.

[47] J. C. Neubert, A. Kretzschmar, S. Wustenberg, and S. Greiff, "Extending the assessment of complex problem solving to finite state automata: Embracing heterogeneity," *Eur. J. Psychol. Assess.*, vol. 31, no. 3, pp. 181–194, Jul. 2015. doi: 10.1027/1015-5759/a000224

[48] S. Greiff, A. Fischer, S. Wüstenberg, P. Sonnleitner, M. Brunner, and R. Martin, "A multitrait-multimethod study of assessment instruments for complex problem solving," *Intelligence*, vol. 41, no. 5, pp. 579–596, Sep. 2013. doi: 10.1016/j.intell.2013.07.012

[49] J. Funke, "Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results," in *The Cognitive Psychology of Knowledge*, vol. 101, G. Strube and K. F. Wender, Eds., North-Holland: Elsevier B.V., ch. 14, pp. 313–330, 1993. doi: 10.1016/S0166-4115(08)62663-1

[50] A. Buchner and J. Funke, "Finite-state automata: Dynamic task environments in problem-solving research," *Q. J. Exp. Psychol.*, vol. 46, no. 1, pp. 83–118, 1993. doi: 10.1080/14640749308401068

[51] H. J. Banda and J. Nzabahimana, "The impact of Physics Education Technology (PhET) interactive simulation-based learning on motivation and academic achievement among Malawian physics students," *J. Sci. Educ. Technol.*, vol. 32, no. 1, pp. 127–141, Feb. 2023. doi: 10.1007/s10956-022-10010-3

[52] M. Renken, M. Pee, K. Otrel-Cass, I. Girault, and A. Chiocarriello, *Simulations as Scaffolds in Science Education*, Cham, Swiss: Springer, 2016. doi: 10.1007/978-3-319-24615-4

[53] Q. He, F. Borgonovi, and M. Paccagnella, "Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks," *Comput. Educ.*, vol. 166, 104170, Jun. 2021. doi: 10.1016/j.compedu.2021.104170

[54] X. Xu, S. Zhang, J. Guo, and T. Xin, "Biclustering of log data: Insights from a computer-based complex problem solving assessment," *J. Intell.*, vol. 12, no. 1, p. 10, Jan. 2024. doi: 10.3390/jintelligence12010010

[55] J. Shin, B. Wang, W. N. Pinto Junior, and M. J. Gierl, "An engagement-aware predictive model to evaluate problem-solving performance from the study of adult skills' (PIAAC 2012) process data," *Large Scale Assess. Educ.*, vol. 12, no. 1, 6, Mar. 2024. doi: 10.1186/s40536-024-00194-y

[56] J. Kim, Y. Ham, and S.-S. Lee, "Differences in student-AI interaction process on a drawing task: Focusing on students' attitude towards AI and the level of drawing skills," *Australas. J. Educ. Technol.*, vol. 40, no. 1, Feb. 2024. doi: 10.14742/ajet.8859

[57] C. Sun, V. Shute, A. E. B. Stewart, and S. K. D'Mello, "The relationship between collaborative problem-solving skills and group-to-individual learning transfer in a game-based learning environment," in *Proc. 15th Int. Learn. Anal. Knowl. Conf.*, Mar. 3-7, 2025, pp. 680–689. doi: 10.1145/3706468.3706560

[58] C. Sun, V. J. Shute, A. Stewart, J. Yonehiro, N. Duran, and S. D'Mello, "Towards a generalized competency model of collaborative problem solving," *Comput. Educ.*, vol. 143, 103672, Jan. 2020. doi: 10.1016/j.compedu.2019.103672

[59] W. K. Adams, "Student engagement and learning with PhET interactive simulations," *Il Nuovo Cimento*, vol. 33, no. 3, pp. 21–32, 2010.

[60] Z. C. Zacharia and G. Olympiou, "Physical versus virtual manipulative experimentation in physics learning," *Learn. Instr.*, vol. 21, no. 3, pp. 317–331, Jun. 2011. doi: 10.1016/j.learninstruc.2010.03.001

[61] T. W. Agustina, E. Paujiah, L. Lestari, A. Y. Nuryantini, Y. Sukmawardani, and R. L. Flara, "Development of PhET-assisted simulation e-worksheet using flipped laboratory approach for science communication skills," *JPBI*, vol. 10, no. 3, pp. 1086–1097, Nov. 2024. doi: 10.22219/jpbi.v10i3.36162

[62] R. Darwis and M. R. Hardiansyah, "Effect of PhET virtual laboratory implementation on students' higher order thinking skills," *J. Penelit. Pendidik. IPA*, vol. 9, no. 4, pp. 1922–1928, Apr. 2023. doi: 10.29303/jppipa.v9i4.1979

[63] T. Jong, M. C. Linn, and Z. C. Zacharia, "Physical and virtual laboratories in science and engineering education," *Science (1979)*, vol. 340, no. 6130, pp. 305–308, Apr. 2013. doi: 10.1126/science.1230579

[64] E. B. Moore, J. M. Chamberlain, R. Parson, and K. K. Perkins, "PhET interactive simulations: Transformative tools for teaching chemistry," *J. Chem. Educ.*, vol. 91, no. 8, pp. 1191–1197, Aug. 2014. doi: 10.1021/ed4005084.

[65] J. I. Smith and K. Tanner, "The problem of revealing how students think: Concept inventories and beyond," *CBE Life Sci. Educ.*, vol. 9, no. 1, pp. 1–5, Mar. 2010. doi: 10.1187/cbe.09-12-0094

[66] K. S. Kaiser, D. B. McGuire, T. J. Keay, and M. E. Haisfield-Wolfe, "Methodological challenges in conducting instrumentation research in non-communicative palliative care patients," *Appl. Nurs. Res.*, vol. 51, Feb. 2020. doi: 10.1016/j.apnr.2019.151199

[67] N. P.-M. Esomonu and O. J. Uzoma, "Development and standardization of instrument for measuring integration of university students into university academic culture," *Eur. J. Educ. Res.*, vol. 5, no. 4, pp. 201–212, Oct. 2016. doi: 10.12973/eu-jer.5.4.201

[68] H. Liu, Y. Liu, and M. Li, "Analysis of process data of PISA 2012 computer-based problem solving: Application of the modified Multilevel Mixture IRT model," *Front. Psychol.*, vol. 9, Aug. 2018. doi: 10.3389/fpsyg.2018.01372

[69] D. Ramalingam, R. Philpot, and B. McCrae, "The PISA 2012 assessment of problem solving," in *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning*, B. Csapó and J. Funke, Eds., Paris, France: OECD Publishing, ch. 5, pp. 75–91, 2017. doi: 10.1787/9789264273955-7-en

[70] L. R. Aiken, "Three coefficients for analyzing the reliability and validity of ratings," *Educ. Psychol. Meas.*, vol. 45, no. 1, pp. 131–142, Mar. 1985. doi: 10.1177/0013164485451012

[71] G. Rasch, *Probabilistic Models for Some Intelligence and Attainment Tests*, Chicago, Illinois, USA: University of Chicago Press, 1960.

[72] G. Rasch, "An individualistic approach to item analysis," in *Readings in Mathematical Social Science*, P. F. Lazarsfeld and N. W. Henry, Eds., Chicago, Illinois, USA: Science Research Associates, pp. 89–107, 1966.

[73] G. Rasch, "An item analysis which takes individual differences into account," *Br. J. Math. Stat. Psychol.*, vol. 19, no. 1, pp. 49–57, 1966. doi: 10.1111/j.2044-8317.1966.tb00354.x

[74] PhET Interactive Simulations. Bending Light. [Online]. Available: https://phet.colorado.edu/en/simulations/bending-light

[75] PhET Interactive Simulations. Gravity and Orbits. [Online]. Available: https://phet.colorado.edu/en/simulations/gravity-and-orbits

[76] PhET Interactive Simulations, Projectile Motion. [Online]. Available: https://phet.colorado.edu/en/simulations/projectile-motion

[77] PhET Interactive Simulations. Geometric Optics. [Online]. Available: https://phet.colorado.edu/en/simulations/geometric-optics

[78] J. J. Weise, S. Greiff, and J. R. Sparfeldt, "The moderating effect of prior knowledge on the relationship between intelligence and complex problem solving—Testing the Elshout-Raaheim hypothesis," *Intelligence*, vol. 83, 101502, Nov. 2020. doi: 10.1016/j.intell.2020.101502

[79] A. S. Luchins, "Mechanization in problem solving: The effect of Einstellung," *Psychol. Monogr.*, vol. 54, no. 6, pp. 1–95, 1942. doi: 10.1037/h0093502

[80] J. Funke, "Solving complex problems: Exploration and control of complex systems," in *Complex Problem Solving: Principles and Mechanisms*, 1st ed., R. J. Sternberg and P. A. Frensch, Eds., Mahwah, New Jersey, USA: Lawrence Erlbaum Associates, Inc., ch. 6, pp. 185–

222, 1991.
[81] J. Shen, J. Lei, H. Y. Chang, and B. Namdar, "Technology-enhanced, Modeling-Based Instruction (TMBI) in science education," in *Handbook of Research on Educational Communications and Technology: Fourth Edition*, J. Spector, M. Merrill, J. Elen, M. Bishop, Eds., New York City, New York, USA: Springer, pp. 529–540, 2014. doi: 10.1007/978-1-4614-3185-5_41
[82] U. Kipman, S. Bartholdy, M. Weiss, W. Aichhorn, and G. Schiepek, "Personality traits and complex problem solving: Personality disorders and their effects on complex problem-solving ability," *Front. Psychol.*, vol. 13, Aug. 2022. doi: 10.3389/fpsyg.2022.788402
[83] J. Funke, "Dynamic systems as tools for analysing human judgement," *Think. Reason.*, vol. 7, no. 1, pp. 69–89, 2001. doi: 10.1080/13546780042000046
[84] D. Dörner, *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations*, New York City, New York, USA: Metropolitan Books, 1996.
[85] S. Greiff and J. Funke, "Systematic research into complex problem-solving skills using minimally complex systems," in *Competence Modeling: Interim Assessment of the DFG Priority Program and Perspectives on the Research Approach*, E. Klieme, D. Leutner, and M. Kenk, Eds., Weinheim, Germany: Beltz, pp. 216-227, 2010. doi: 10.25656/01:3430 (in German)
[86] S. Greiff, S. Wüstenberg, T. Goetz, M.-P. Vainikainen, J. Hautamäki, and M. H. Bornstein, "A longitudinal study of higher-order thinking skills: Working memory and fluid reasoning in childhood enhance complex problem solving in adolescence," *Front. Psychol.*, vol. 6, Jul. 2015. doi: 10.3389/fpsyg.2015.01060.
[87] A. Fischer and J. C. Neubert, "The multiple faces of complex problems: A model of problem solving competency and its implications for training and assessment," *J. Dyn. Decis. Mak.*, vol. 1, pp. 1–13, 2015. doi: 10.11588/jddm.2015.1.23945
[88] S. B. McKagan *et al.*, "Developing and researching PhET simulations for teaching quantum mechanics," *Am. J. Phys.*, vol. 76, no. 4, pp. 406–417, Apr. 2008. doi: 10.1119/1.2885199
[89] G. Molnár, S. Greiff, and B. Csapó, "Inductive reasoning, domain specific and complex problem solving: Relations and development," *Think. Skills Creat.*, vol. 9, pp. 35–45, Aug. 2013. doi: 10.1016/j.tsc.2013.03.002
[90] S. Greiff and J. Funke, "Interactive problem solving: Exploring the potential of minimal complex systems," in *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning*, B. Csapó and J. Funke, Eds., Paris, France: OECD Publishing, ch. 6, pp. 93–105, 2017. doi: 10.1787/9789264273955-8-en
[91] M. Wilson, *Constructing Measures: An Item Response Modeling Approach*, 2nd Edition., New York City, New York, USA: Routledge, 2023.
[92] J. M. Linacre. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Meas. Trans.* [Online]. 16(2), p. 878. Available: https://www.rasch.org/rmt/rmt162f.htm
[93] J. J. Wang, T. A. Chen, T. Baranowski, and P. W. C. Lau, "Item response modeling: A psychometric assessment of the children's fruit, vegetable, water, and physical activity self-efficacy scales among Chinese children," *Int. J. Behav. Nutr. Phys. Activ.*, vol. 14, no. 1, Sep. 2017. doi: 10.1186/s12966-017-0584-x
[94] W. J. Boonee·, J. R. Staver, and M. S. Yale, *Rasch Analysis in the Human Sciences*, Dordrecht, Netherlands: Springer, 2014. doi: 10.1007/978-94-007-6857-4
[95] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Hciences*, New York City, New York, USA: Psychology Press, 2015. doi: 10.4324/9781410614575
[96] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *Int. J. Med. Educ.*, vol. 2, pp. 53–55, Jun. 2011. doi: 10.5116/ijme.4dfb.8dfd
[97] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., New York City, New York, USA: Routledge, 2013. doi: 10.4324/9780203771587
[98] A. A. Rezigalla, "Item analysis: Concept and application," in *Medical Education for the 21st Century*, M. S. Firstenberg and S. P. Stawicki, Eds., Rijeka, Croatia: IntechOpen, 2022. doi: 10.5772/intechopen.100138
[99] J. M. Cortina, "What is coefficient alpha? An examination of theory and applications," *J. Appl. Psychol.*, vol. 78, no. 1, pp. 98–104, Feb. 1993. doi: 10.1037/0021-9010.78.1.98
[100] C. Aldama and J. I. Pozo, "Do you want to learn physics? Please play Angry Birds (But with epistemic goals)," *J. Educ. Comput. Res.*, vol. 58, no. 1, pp. 3–28, Mar. 2020. doi: 10.1177/0735633118823160
[101] D. Robič and R. Repnik, "Computer game Angry Birds in teaching physics," in *Proc. 2015 38th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 25–29, 2015, pp. 680–684. doi: 10.1109/MIPRO.2015.7160358
[102] P. A. Frensch and J. Funke, *Complex Problem Solving: The European Perspective*, 1st ed., Mahwah, New Jersey, USA: Lawrence Erlbaum Associates, Inc., 1995.
[103] A. Graesser, B. C. Kuo, and C. H. Liao, "Complex problem solving in assessments of collaborative problem solving," *J. Intell.*, vol. 5, no. 2, 10, Mar. 2017. doi: 10.3390/jintelligence5020010
[104] R. J. Cohen, W. J. Schneider, and R. M. Tobin, *Psychological Testing and Assessment: An Introduction to Tests and Measurement*, 10th Ed., New York City, New York, USA: McGraw-Hill, 2022.
[105] S. Greiff, D. V. Holt, and J. Funke, "Perspectives on problem solving in educational assessment: Analytical, interactive, and collaborative problem solving," *J. Probl. Solving*, vol. 5, no. 2, pp. 71–91, Apr. 2013. doi: 10.7771/1932-6246.1153
[106] M.-J. Chen, H.-C. She, and P.-Y. Tsai, "The effects of online simulation-based collaborative problem-solving on students' problem-solving, communication and collaboration attitudes," *Educ. Inf. Technol. (Dordr)*, vol. 29, no. 14, pp. 19141–19162, Oct. 2024. doi: 10.1007/s10639-024-12609-y
[107] G. Molnár and S. Greiff, "Understanding transitions in complex problem-solving: Why we succeed and where we fail," *Think. Skills Creat.*, vol. 50, 101408, Dec. 2023. doi: 10.1016/j.tsc.2023.101408
[108] J. Pena, P. C. Shih, and M. B. Rosson, "Instructors as end-user developers: Technology usage opportunities in the inverted classroom," in *Handbook of Research on Applied Learning Theory and Design in Modern Education*, E. A. Railean, G. Walker, A. Elçi, and L. Jackson, Eds., Hershey, PA, USA: IGI Global Scientific Publishing, 2016, ch. 27, pp. 560–571. doi: 10.4018/978-1-4666-9634-1.ch027
[109] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 4th Edition., London, England: SAGE Publications Ltd, 2013.
[110] J. Frost, *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models*, State College, Pennsylvania, USA: Statistics by Jim Publishing, 2019.
[111] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 5th ed., Boston, Massachusetts, USA: Pearson Education, 2005.
[112] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, Sixth Edition., Hoboken, New Jersey, USA: John Wiley & Sons Inc, 2021.
[113] OECD, *PISA 2012 Results: Creative Problem Solving (Volume V)*, Paris, France: OECD Publishing, Apr. 2014. doi: 10.1787/9789264208070-en
[114] S. Greiff, S. Wüstenberg, D. V. Holt, F. Goldhammer, and J. Funke, "Computer-based assessment of complex problem solving: Concept, implementation, and application," *Educ. Technol. Res. Dev.*, vol. 61, no. 3, pp. 407–421, Jun. 2013. doi: 10.1007/s11423-013-9301-x
[115] M. Stadler, N. Becker, M. Gödker, D. Leutner, and S. Greiff, "Complex problem solving and intelligence: A meta-analysis," *Intelligence*, vol. 53, pp. 92–101, Nov. 2015. doi: 10.1016/j.intell.2015.09.005
[116] J. F. Beckmann and N. Goode, "The benefit of being naïve and knowing it: the unfavourable impact of perceived context familiarity on learning in complex problem solving tasks," *Instr. Sci.*, vol. 42, no. 2, pp. 271–290, Mar. 2014. doi: 10.1007/s11251-013-9280-7
[117] R. E. Mayer, *Thinking, Problem Solving, Cognition*, 2nd ed., New York City, New York, USA: W.H. Freeman, 1992.
[118] S. Greiff *et al.*, "Domain-general problem solving skills and education in the 21st century," *Educ. Res. Rev.*, vol. 13, pp. 74–83, 2014. doi: 10.1016/j.edurev.2014.10.002