

Building a diagnostic model for misconceptions in physics education using eye tracking technology

Jinmei Liu^{*,} Xuejuan Zhong, Gang Chen,[†] and Liqing Chen[‡]

School of Physics, East China Normal University, Shanghai 200241, People's Republic of China



(Received 27 August 2025; accepted 13 May 2026; published 22 July 2026)

Accurate diagnosis of student misconceptions is crucial for effective science instruction. Traditional diagnostic tools, however, often rely on unreliable self-reported confidence levels. This study developed and validated a two-dimensional diagnostic model that integrates objective eye tracking data with test responses to more robustly identify misconceptions in Newtonian mechanics. The study involved three phases: developing a four-tier diagnostic test, conducting an eye tracking experiment to analyze visual attention patterns (fixation duration, revisits, scan paths) among high school students, and constructing the final model by correlating eye movements with confidence judgments. Consistent with previous research, results showed that students with misconceptions had shorter fixation durations on correct options and displayed inefficient scan paths. Conversely, those with scientific conceptions showed focused attention. As a novel contribution of this study, a “reversed U-shaped” pattern was observed between confidence and revisits, indicating that revisits can serve as an objective measure of cognitive uncertainty. Based on these findings, we propose a two-dimensional “eye tracking-test” model. This model offers a more precise method for diagnosing student understanding, providing a foundation for targeted instructional interventions and fostering a dynamic “teach-learn-assess” feedback loop for personalized physics education.

DOI: [10.1103/ttrq-px6t](https://doi.org/10.1103/PhysRevPhysEducRes.22.020105)

I. INTRODUCTION

Constructivist learning theory posits that learning is not a passive reception of external knowledge but an active process where learners build new understanding upon their existing cognitive structures [1]. Before formal instruction, students develop intuitive theories about natural phenomena through everyday experiences and observations. These preinstructional ideas often diverge from scientifically accepted concepts and are commonly termed “misconceptions” [2]. In the context of this study, we use the terms misconceptions, preconceptions, and alternative ideas interchangeably to refer to such student cognitive structures that deviate from the scientific consensus, while acknowledging that subtle theoretical distinctions exist in the broader literature.

Misconceptions are notoriously robust, deeply ingrained in students’ cognitive frameworks, and can systematically impede science learning. The negative impact of

misconceptions is well documented. Specifically, they can obstruct the transfer and application of scientific knowledge, as preexisting incorrect mental models interfere with the assimilation of new, correct information [3]. They also hinder a deep understanding of scientific principles; a persistent misconception in one area can lead to erroneous judgments across an entire scientific domain, preventing the development of sound scientific reasoning [4]. Furthermore, the cognitive dissonance arising from the conflict between a student’s misconception and a scientific concept can lead to confusion, frustration, and even a rejection of scientific learning [5].

Given their detrimental effects, the early and accurate diagnosis of student misconceptions is of paramount importance for effective teaching. This aligns with a central tenet of pedagogical content knowledge (PCK), which emphasizes that effective teachers must possess a deep understanding of what makes specific topics easy or difficult for students, including their common preconceptions and misconceptions [6,7]. Teachers who are aware of their students’ conceptual difficulties are better equipped to design instruction that confronts these alternative ideas and facilitates meaningful conceptual change [8,9].

While knowing the most common incorrect responses and patterns of misunderstanding is a valuable part of PCK, these traditional diagnostic approaches alone often fail to fully diagnose the underlying “cognitive states” of

^{*}Contact author: jmliu2005@phy.ecnu.edu.cn

[†]Contact author: gchen@phy.ecnu.edu.cn

[‡]Contact author: lqchen@phy.ecnu.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

students [10]. Simply identifying what the incorrect answer is does not always reveal why the student holds that belief, nor the robustness or certainty with which they hold it [11]. This limitation is particularly critical when one needs to distinguish between a deeply held, confident misconception and a more fragile, uncertain error arising from a lack of knowledge or a simple guess [12].

To address the fact that traditional assessment methods often focus on final answers and fail to capture the underlying reasoning processes [13], multitier diagnostic tests have been developed, such as the two-tier, three-tier, and four-tier instruments [11,14,15]. Among these, the four-tier instrument offers a significant advantage [15]. By measuring confidence levels separately for both the answer and the reasoning, it allows for the detection of subtle cognitive states, such as when a student is confident in their factual answer but uncertain about the underlying scientific explanation—a distinction that simpler instruments fail to capture.

Monitoring these confidence levels is crucial for diagnosing the strength of students' beliefs [11,14,15]. It enables educators to distinguish between students who are merely guessing or responding with unstable knowledge (low confidence) and those who hold robust misconceptions (high confidence) [11,15]. This distinction is pedagogically critical because the remedial strategies required for these two groups differ substantially: the former needs instructional support to build foundational knowledge, whereas the latter requires conceptual change interventions to challenge and restructure entrenched alternative conceptions [8,12,15].

Yet, a critical limitation remains: the confidence tiers rely entirely on students' subjective self-assessments. Students may not be accurate reporters of their own cognitive states; they might be overconfident in an incorrect belief or lack confidence in a correctly memorized fact they do not truly understand. This subjectivity can compromise the reliability of the diagnosis [16]. In sum, traditional methods, even with confidence ratings, struggle to objectively capture the underlying cognitive states of students—their genuine certainty, uncertainty, or cognitive conflict. This fundamental reliance on self-report provides a compelling rationale for introducing objective measurement techniques.

To address the aforementioned limitation of subjectivity, recent advancements in technology offer new avenues for overcoming it. Eye tracking technology, in particular, provides a noninvasive, objective window into an individual's cognitive processes based on the eye-mind hypothesis, which posits a tight coupling between where a person is looking and what they are processing mentally [17,18]. Metrics such as fixation duration and saccades can reflect cognitive load and hesitation associated with uncertainty [19,20]. These capabilities make eye tracking a particularly promising tool for supplementing and validating the subjective confidence judgments central to multitier diagnostic instruments.

Therefore, eye tracking addresses the limitation of subjective reporting by providing objective, real-time physiological data. Based on the above analysis, this study aims to bridge the gap between the need for accurate misconception diagnosis and the limitations of current tools. Specifically, we integrate the four-tier diagnostic test with eye tracking technology, triangulating subjective confidence ratings with objective eye-movement data to validate and refine the diagnosis of student misconceptions and explore the visual attention patterns associated with different cognitive states.

We propose a novel “eye tracking-test” two-dimensional diagnostic model that integrates the rich qualitative data from a four-tier diagnostic test with objective, quantitative eye-movement data. By analyzing how students visually engage with problems—where they focus their attention, what information they revisit, and how their visual strategies differ based on their underlying knowledge—we seek to develop a more reliable and valid method for diagnosing misconceptions.

Ultimately, this research explores the potential of eye tracking to serve as an objective proxy for cognitive states like uncertainty and confidence, thereby enhancing the diagnostic power of established assessment formats. Its goal is to provide educators with a more precise tool to understand student thinking and to lay the groundwork for data-driven, personalized instructional strategies that can effectively address and remediate student misconceptions in physics.

II. THEORETICAL FRAMEWORK AND LITERATURE REVIEW

A. Misconceptions in Newtonian mechanics and diagnostic tools

In the domain of physics, specifically Newtonian mechanics, students often hold intuitive theories that contradict scientific principles. These misconceptions are not merely random errors but are based on consistent, albeit incorrect, explanatory frameworks derived from daily experience [21]. A wider array of misconceptions pervades students' understanding of Newtonian mechanics, often rooted in an “impetus theory” or a naive interpretation of motion [22]. These include the belief that a constant force is required to maintain constant velocity, that acceleration is always in the direction of motion, or that heavier objects fall faster in a vacuum [2,23]. Understanding the prevalence and nature of these misconceptions is crucial for designing effective instruction and assessment [24].

To effectively diagnose the aforementioned cognitive structures, researchers in physics education research (PER) have developed various assessment instruments. Standard multiple-choice concept inventories, such as the Force Concept Inventory (FCI) [25] and the Force and Motion Conceptual Evaluation [26], have been widely used to

identify students' conceptual difficulties. Specifically, the FCI includes a well-known question describing a collision between a small car and a large truck. This particular item has been extensively researched, providing rich data on common student misconceptions regarding Newton's third law [25]. Numerous studies employing the FCI have consistently reported that a significant percentage of students incorrectly believe that the truck exerts a greater force on the car, or vice versa, during the collision [27]. Common incorrect student responses often include rationales such as: the heavier truck has more "force," the car experiences a larger force because it sustains more damage, or that the forces are not equal because their effects (e.g., acceleration, damage) are different. These responses directly contradict Newton's third law, which states that the forces are equal in magnitude and opposite in direction. Analyzing these prevalent incorrect student responses is fundamental for understanding the depth and persistence of such misconceptions and for informing pedagogical strategies aimed at conceptual change [8]. However, traditional single-tier or two-tier tests have limitations: they cannot distinguish whether a correct answer stems from genuine understanding or a lucky guess, nor can they differentiate between a confident misconception and a lack of knowledge [28]. This limitation is addressed by the four-tier diagnostic test, which probes the answer, the confidence in the answer, the reasoning, and the confidence in the reasoning [15].

Building on this line of work, the present study categorizes students into distinct cognitive profiles: scientific conception (SC), misconception (MC), and lack of knowledge (LK) to classify students' conceptual understanding. In this framework, SC refers to correct answers and correct reasoning with high confidence, MC refers to incorrect answers and/or incorrect reasoning held with high confidence, and LK refers to responses characterized by low confidence, indicating insufficient or unstable knowledge rather than a robust alternative conception.

This categorization aligns with the well-established expert-novice framework in the literature [29]. The scientific knowledge (SK) group can be closely linked to "experts," possessing coherent and accurate conceptual understanding. Conversely, the MC and LK groups can be presented as two distinct subtypes within the broader "novice" category. This refinement in categorization is a significant finding that underscores the nuanced nature of student understanding.

Crucially, distinguishing between these knowledge states is not merely a taxonomical exercise but a pedagogical necessity. The instructional remediation required for a student lacking knowledge differs fundamentally from that required for a student with a misconception. Students with a "lack of knowledge" (low confidence) simply require direct instruction to build new knowledge structures. In contrast, students with misconceptions (high confidence in error) hold robust, resistant mental models that actively interfere with

learning. These students require specific "conceptual change" strategies, such as cognitive conflict, to dismantle their incorrect beliefs before scientific concepts can be established [8]. Therefore, a diagnostic framework that conflates these two groups into a single "incorrect" category fails to provide the specific guidance teachers need.

B. Eye tracking metrics as indicators of cognitive processes

As discussed in the Introduction, while the four-tier test theoretically categorizes these states, self-reported confidence can be subjective. Eye tracking technology offers objective physiological data that complement the subjective results of diagnostic tests. Based on the "eye-mind hypothesis" [18], visual attention is tightly coupled with cognitive processing. However, no single eye tracking metric can fully capture the complexity of student understanding. Therefore, this study combines the following four key metrics to triangulate the cognitive states of SK, MC, and LK.

1. **Fixation duration and count:** Total fixation duration and counts typically indicate the depth of cognitive processing. Existing literature suggests that longer fixations may reflect either deeper engagement with relevant information or greater difficulty, uncertainty, or conflict during processing, depending on the task context and the learner's strategy [18,19,30]. We hypothesize that students with LK will exhibit the longest fixation durations due to their struggle to process the problem. Conversely, students with SK or firmly held MC may show shorter durations, as their retrieval of (correct or incorrect) knowledge is more automatic and fluent.
2. **Revisits (transitions):** Transitions between areas of interest (AOIs) reflect the integration of information—that is, the extent to which a learner connects and cross-references distinct pieces of information (e.g., comparing the question stem with different answer options) during problem solving. High frequencies of revisits, particularly repeated shifts between the question stem and options, may reflect ongoing comparison, reevaluation, and uncertainty during multiple-choice science problem solving [20]. In contrast, confident students (both SK and MC) are expected to show fewer, more directed transitions.
3. **Scan paths:** While fixation duration and count measure how much processing occurs, scan paths reveal what is being processed—that is, the specific content and sequence of information that the learner attends to. Eye tracking research indicates that experts allocate attention more selectively to task-relevant information, whereas novices tend to show less efficient and more diffuse visual search patterns [31]. Thus, scan path analysis complements the quantitative metrics by providing qualitative insight into students' information processing strategies.

By interpreting these measures collectively, we can distinguish the “confident error” (misconception: low duration, directed path to incorrect option) from the “uncertain error” (lack of knowledge: high duration, chaotic path, high revisits). Specifically, this collective interpretation of eye tracking metrics provides a methodological basis for further clarifying novice learners—who all provide incorrect responses—into the distinct MC and LK subcategories: students providing confident error responses, characterized by short fixation durations and directed scan paths toward incorrect options, are classified as holding misconceptions; whereas students providing uncertain error responses, characterized by prolonged fixation durations, chaotic scan paths, and high revisit frequencies, are classified as having a lack of knowledge. This approach aligns with the diagnostic framework proposed by Hermita *et al.* [15], who demonstrated that confidence-based tiers can meaningfully differentiate among students’ cognitive states, and extends it by grounding the confidence dimension in objective behavioral evidence rather than self-report alone. This combined approach provides the rationale for using eye tracking to validate the four-tier test categories.

C. Eye tracking in physics education research (PER)

With the rapid development of eye tracking technology in educational research, the application of eye tracking in PER has become a prominent research topic in academia in recent years. Eye tracking has been increasingly used to reveal how students allocate visual attention during problem solving, going beyond whether their final answers are simply correct or incorrect. Existing studies can be organized into the following three strands based on their thematic focus.

The first strand of research focuses on differences in visual attention behaviors associated with problem-solving performance levels (e.g., experts vs novices). These studies aim to uncover distinctions in visual attention patterns between successful and unsuccessful problem solvers. For example, Madsen *et al.* investigated differences in visual attention between those who correctly and incorrectly answered physics problems, demonstrating that the two groups displayed distinctly different visual attention patterns [32]. This body of work reveals a fundamental consensus: there is a significant association between problem-solving accuracy and the distribution of visual attention. A question worth further exploring, however, is how students coordinate multiple sources of information during reasoning in more complex physics tasks—and this is precisely the focus of the second strand of research.

The second strand of research focuses on the visual behavior characteristics of students coordinating multi-source information during reasoning in more complex physical environments. These studies examine students’ gaze transitions and information integration processes in physics problems involving multiple concepts. For instance,

Ibrahim and Ding compared students’ gaze transitions in sequential and simultaneous synthesis problem solving, revealing how learners dealt with multiconcept physics problems [33]. It is evident that this second strand has a progressive relationship with the first—moving from simple correct-vs-incorrect comparisons to a deeper investigation of the process mechanisms of visual information coordination in complex tasks, thereby providing a more comprehensive understanding of visual cognition in physics problem solving.

The third strand of research focuses on student interactions with visual representations during conceptual change, and the linkage between visual attention and conceptual change processes. More recently, researchers have drawn on foundational eye tracking theories of visual information processing [34,35] to explore visual attention in physics-specific contexts. Studies have linked visual attention patterns to students’ problem solving in physics [36] and to conceptual understanding assessed via standardized instruments [37], while others have investigated the cognitive processes underlying scientific conceptual change using eye tracking [38]. For example, Han *et al.* employed eye tracking of visual attention in web-based assessment using the Force Concept Inventory to characterize general visual attention patterns [37]; She *et al.* unfolded the cognitive processes underlying computer-based scientific conceptual change with eye tracking, examining behavioral performance and sequential analysis of attention [38].

The above three strands of research have been instrumental in advancing the application of eye tracking in physics education, deepening our understanding of visual cognitive mechanisms during physics problem solving. However, a critical gap remains in the literature regarding the precise diagnosis of “confidence” in the misconceptions of novice learners—that is, the objective differentiation between LK learners and MC learners [15]. Although both types of learners provide incorrect responses, they differ fundamentally in the certainty with which they hold their erroneous beliefs. Specifically, the limitations of existing research are manifested in the following aspects:

1. **From the perspective of research design:** Existing PER studies have predominantly focused on the correctness of the solution process, while paying insufficient attention to the certainty of students’ beliefs. For example, Han *et al.* [37] employed eye tracking to characterize visual attention patterns in FCI-based assessment, yet their study did not incorporate a multitier confidence framework to distinguish between “confident misconceptions” and uncertain errors.
2. **From the perspective of diagnostic instrument integration:** Existing studies have largely relied on single-dimensional analytical frameworks (e.g., answer correctness), while rarely integrating four-tier diagnostic instruments that simultaneously account

for both response accuracy and confidence levels [15]. For instance, She *et al.* [38] investigated the cognitive processes underlying conceptual change, yet similarly did not explicitly integrate a four-tier diagnostic instrument with eye tracking technology to systematically classify learners into distinct cognitive profiles.

3. **From the perspective of cognitive classification granularity:** Existing studies have predominantly focused on a binary comparison between correct and incorrect responses, while rarely conducting fine-grained cognitive profiling within the population of incorrect responders—that is, distinguishing between MC learners who hold robust misconceptions (providing confident error responses) and LK learners who simply lack knowledge (providing “uncertain error” responses) [15,37,38].

In light of the above analysis, this study aims to investigate this topic from the novel angle of integrating eye tracking with a four-tier diagnostic framework. The objective is to use eye tracking technology to objectively characterize and differentiate the visual attention patterns associated with SC, MC, and LK categories—particularly to pinpoint the distinct cognitive signatures of students who hold misconceptions vs those who merely lack knowledge. This research employs eye tracking not merely to observe problem solving, but to validate the distinct cognitive profiles associated with the confidence levels of LK and MC learners in Newtonian mechanics concepts, thereby bridging the identified gap in the existing literature.

III. PURPOSE

The present study aimed to develop and validate a two-dimensional diagnostic model that integrates objective eye tracking data with subjective responses from a four-tier test to more accurately identify student misconceptions in physics. The study was designed to explore the differences in visual attention and cognitive processing among students with varying levels of conceptual understanding. By linking specific eye-movement patterns to different knowledge states (SC, MC, and LK), we seek to establish eye tracking as a viable, objective tool for enhancing cognitive diagnosis.

To achieve this goal, the study addresses the following research questions:

1. What are the differences in visual attention patterns (i.e., fixation duration, revisits, and scan paths) among students classified as having scientific conceptions, misconceptions, or a lack of knowledge when they solve physics problems?
2. How do students’ subjective confidence judgments, as captured by a four-tier test, correlate with their objective eye-movement behaviors, particularly the number of revisits to specific areas of interest?

3. Can a two-dimensional model combining four-tier test results and eye tracking data provide a more precise and reliable diagnosis of student misconceptions than a traditional test-based model alone?

IV. METHOD

A. Research design

This study employed a mixed-methods approach, combining quantitative analysis of test scores and eye tracking data with qualitative analysis of scan paths and post-task interviews. The research was conducted in three sequential phases:

Phase 1: A four-tier diagnostic test on Newtonian mechanics was developed based on established assessment instruments and administered to a large group of students to identify common misconceptions and to establish the instrument’s reliability and validity. Specifically, the answer-tier questions were derived from well-established concept inventories—primarily the FCI [25]—to which confidence tiers and reasoning tiers were added by the authors. The diagnostic structure builds on the three-tier framework proposed by Caleon and Subramaniam [14], which was subsequently extended to a four-tier format (answer, answer confidence, reasoning, and reasoning confidence) by Hermita *et al.* [15]. The present instrument adopts this four-tier architecture to enable simultaneous assessment of both conceptual understanding and metacognitive confidence.

Phase 2: Eye tracking experiment. A subset of students participated in an eye tracking study where they completed a selection of the diagnostic items while their eye movements were recorded. This phase focused on identifying characteristic visual behaviors associated with different conceptual understanding profiles.

Phase 3: Model construction. The data from the four-tier test and the eye tracking experiment were integrated to analyze the relationship between subjective confidence and objective visual behaviors, leading to the construction and validation of the eye tracking-test two-dimensional diagnostic model.

B. Participants

Participants were recruited from a high school in Shanghai, China. These grade 10 students, aged approximately 15 to 17 years, provided informed consent along with their guardians. For phase 1, 120 students participated in a paper-and-pencil four-tier test; after removing invalid responses, 89 valid questionnaires were used for analysis. For phase 2, 70 volunteers with normal or corrected-to-normal vision were recruited from three physics classes for an eye tracking experiment. They were selected based on preliminary screening to ensure a distribution of different

TABLE I. Content distribution and targeted misconceptions in the four-tier diagnostic test.

Concept area	Scientific concept	Targeted misconception
Newton's first law	Inertia is a property of an object	Inertia is a force [2,27]
Newton's second law	Force is the cause of a change in motion (acceleration)	Force is required to maintain motion (velocity) [22,23]
	Acceleration is proportional to net force	Velocity is proportional to net force [22,23]
Newton's third law	Action-reaction forces act on different objects	Action-reaction forces act on the same object (confusion with balanced forces) [25,27,28]
	Action-reaction forces are always equal in magnitude	The magnitude of action-reaction forces depends on the motion state or mass of the objects [25,27,28]

conceptual understanding levels. Ultimately, data from five participants were excluded as invalid due to calibration errors exceeding 1° or a gaze tracking ratio below 80% caused by excessive head movement.

C. Materials

A four-tier diagnostic test focusing on Newton's laws of motion was restructured from established sources rather than developed from scratch. The items were designed by integrating content from standard high school physics textbooks [39], the national physics curriculum standards [40], and established concept inventories like the FCI [25]. The reasoning-tier options were informed by documented incorrect student rationales reported in prior FCI studies [27] and physics education research on Newton's laws [28]. The confidence tiers—namely, tier 2 (confidence in the answer) and tier 4 (confidence in the reason)—were developed by the authors based on established multitier diagnostic frameworks, particularly those proposed by Caleon and Subramaniam [14] and Hermita *et al.* [15]. The test was designed to probe common misconceptions related to Newton's first, second, and third laws (see Table I). These targeted misconceptions have been extensively documented in the physics education research literature. The misconception that “inertia is a force” has been reported by Clement [2] and Halloun and Hestenes [27]. The misconception that “force is required to maintain motion (velocity)” and the related confusion between force and velocity are rooted in the intuitive impetus theory documented by McCloskey [22] and further confirmed by McDermott [23]. The misconception that “the magnitude of action-reaction forces depends on the motion state or mass of the objects” and the confusion between action-reaction pairs and balanced forces have been well documented by Hestenes *et al.* [25], Halloun and Hestenes [27], and Zhou *et al.* [28].

Each item in the test had four tiers: (i) an answer tier consisting of a multiple-choice question; (ii) an answer confidence tier using a five-point Likert scale; (iii) a reason tier consisting of a multiple-choice question about the reasoning behind the answer; and (iv) a reason confidence tier using a five-point Likert scale. For the eye tracking

experiment, one representative item was selected from the validated four-tier diagnostic test, specifically targeting the prevalent Newton's third law misconception that the larger or more active object exerts a greater force during an interaction.

From the initial item pool restructured from textbooks, curriculum standards, and established concept inventories such as the FCI, items were selected for inclusion in the final four-tier diagnostic test and for subsequent use in the eye tracking experiment based on two criteria: content validity and discrimination power. All items were aligned with the local high school physics curriculum standards and were informed by prior research identifying high-frequency misconceptions related to Newton's laws. Particular emphasis was placed on items that could effectively distinguish scientifically correct conceptions from common alternative conceptions.

For the eye tracking phase, we restructured a classic Newton's third law problem (the car-pushing-truck scenario, derived from the well-known FCI “collision” item [25]; the illustration used in the experimental material was also sourced from Ref. [25], as noted in Appendix A) into a digital format optimized for eye tracking. While standardized tests like the FCI are widely used, they often lack the specific spatial layout required for effective AOI analysis. Therefore, we restructured the original FCI item by modifying the spatial arrangement of answer options and reasoning choices to create clearly defined, nonoverlapping AOIs suitable for eye tracking analysis, while preserving the core conceptual content and the targeted misconception from the original item. This item requires students to integrate textual descriptions with diagrammatic representations and addresses a well documented and robust misconception—that the larger or more active object exerts a greater force during an interaction. Moreover, results from phase 1, namely the initial paper-and-pencil administration of the four-tier diagnostic test used for instrument validation and preliminary screening of students' conceptual understanding, indicated that this item had strong discriminatory capability in differentiating students with misconceptions from those holding scientific conceptions. This made it particularly suitable for process-level investigation in the subsequent phase 2 eye tracking experiment.

D. Apparatus

Eye movements were recorded using a Tobii Eye Tracker 4C, a remote system with a sampling rate of 90 Hz. The eye tracker was mounted at the bottom of a 14-inch laptop monitor with a screen resolution of 1920×1080 pixels. Tobii's Gaze Viewer software was used for calibration, data recording, and preliminary visualization. Participants were seated approximately 60–65 cm from the screen and were allowed free head movement within a reasonable range. A standard mouse was used for responses.

E. Procedure

Upon arrival at the laboratory, each participant was briefed on the experimental procedure and signed an informed consent form. They were then seated in front of the eye tracking setup. The procedure included:

1. Familiarization and calibration: Participants were given a brief 2–3 min tutorial on the task. This was followed by a nine-point calibration procedure to ensure the eye tracker could accurately record their gaze position. The calibration was repeated if the initial accuracy was low.
2. Task execution: Participants were presented with the four-tier diagnostic item on the screen. They were instructed to read the question and its options carefully and to think through their answer and reasoning. They recorded their final answers and confidence ratings on a separate sheet to avoid motor interference with gaze data. The entire session for one item lasted approximately 10–15 min.
3. Post-task interview: After the eye tracking session, a brief semistructured interview was conducted to probe the participant's understanding of the concepts in the test item, particularly their reasoning for distinguishing between balanced forces and action-reaction pairs.

Throughout the experiment, a researcher was present in the room but remained out of the participant's direct line of sight to minimize interference.

F. Data analysis

1. Classification of conceptual understanding

Our research is adapted from the classification of Hermita *et al.* [15]; participants were classified as experts (SC group) and novices (MC and LK groups) as follows:

Scientific conception (SC): Correct answer and correct reason, with high confidence (≥ 4 on a 5-point scale) in both.

Misconception (MC): Incorrect answer and/or incorrect reason, with high confidence (≥ 4) in the incorrect choice.

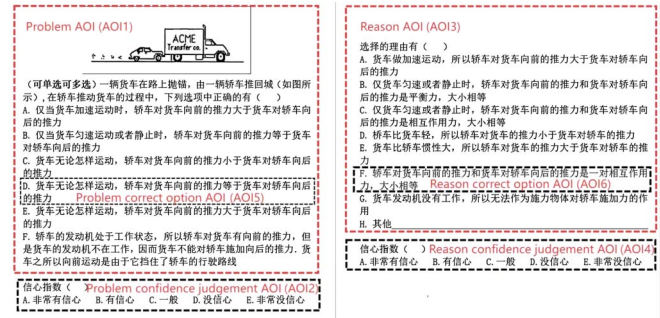


FIG. 1. The experimental material with six defined AOIs. AOI 1: question stem; AOI 2: answer confidence; AOI 3: reason stem and options; AOI 4: reason confidence; AOI 5: correct answer option; AOI 6: correct reason option.

Lack of knowledge (LK): Incorrect answer and/or incorrect reason, with low confidence (≤ 3) in their choices.

2. Eye tracking data processing

The screen displaying the test item was divided into six primary AOIs for analysis, as shown in Fig. 1. These were: AOI 1 (question stem), AOI 2 (answer confidence scale), AOI 3 (reason stem and options), AOI 4 (reason confidence scale), AOI 5 (correct answer option), and AOI 6 (correct reason option). The primary eye tracking metrics extracted for each AOI were:

Total fixation duration: The sum of the durations of all fixations within an AOI, reflecting the total amount of cognitive processing allocated to that area.

Number of revisits: The count of gaze entries into an AOI after having looked at another AOI, indicating comparative processing or reevaluation of information.

Scan path and heat map visualizations were also generated to provide a qualitative overview of attention distribution and processing strategies.

3. Statistical analysis

Statistical analyses were performed using SPSS. The reliability of the four-tier test was assessed using Cronbach's alpha. Prior to inferential analysis, the normality of the main eye tracking variables (e.g., total fixation duration and number of revisits across areas of interest) was examined using the Shapiro-Wilk test. The results indicated that these variables significantly deviated from a normal distribution (all $p < 0.05$). Therefore, the assumption of normality was not met, and nonparametric statistical methods were adopted. To compare the eye-movement data among the three conceptual understanding groups (SC, MC, and LK), the nonparametric Kruskal-Wallis H test was used, as the data were not normally distributed, likely due to the relatively moderate sample size inherent in eye tracking studies and the typical skewness of reaction time data. *Post hoc* pairwise comparisons were conducted using

the Mann-Whitney U test with Bonferroni correction. Effect sizes were calculated to determine the magnitude of the observed differences. A p value of less than 0.05 was considered statistically significant.

V. DATA ANALYSIS AND RESULTS

A. Instrument validation and misconception diagnosis

The four-tier diagnostic test developed in phase 1 demonstrated high reliability, with a Cronbach's alpha coefficient of 0.934. Content and construct validity were confirmed through expert review and exploratory factor analysis (KMO = 0.790, Bartlett's test $p < 0.001$). The analysis of the 89 valid responses revealed that while students had a relatively good grasp of Newton's first and second laws, a significant and persistent misconception existed regarding Newton's third law. Specifically, many students confused action-reaction pairs with balanced forces—for example, believing that the forces are equal only when objects move at constant velocity or remain stationary. This misconception has been extensively documented in the physics education research literature [25,27,28]. This confirmed the targeted misconception as a suitable focus for the eye tracking experiment.

From the 70 participants in the eye tracking study, the classification based on their responses to the target item was as follows: 28 participants (40%) were classified as having a SC, 24 participants (34%) held a MC, and 17 participants (24%) were classified as having a LK. A small number of participants (2%) were classified as making a careless error and were excluded from further group comparisons.

The classification results are summarized below. Table II provides representative examples of student responses and interview excerpts for each category to illustrate the classification criteria.

B. Differences in visual attention among conceptual groups

1. Quantitative eye-movement data

The Kruskal-Wallis H test revealed that there were no significant differences in the total fixation duration, fixation duration on AOIs 1–5, or any of the revisit measures across the three groups, except for fixation duration on the correct reason option (AOI 6). Specifically, among all fixation duration and revisit measures, only fixation duration on the correct reason option (AOI 6) differed significantly across the SC, MC, and LK groups. No significant group differences were found for total fixation time across the entire item, fixation duration on AOIs 1–5, or the revisit measures listed in Table III.

Post hoc tests showed that the MC group spent significantly less time fixating on the correct reason option (median = 2.2 s) compared to both the SC group (median = 4.3 s, $U = 189.5$, $p = 0.011$) and the LK group (median = 5.0 s, $U = 112.0$, $p = 0.025$). The LK group showed a numerically slightly higher median fixation duration than the SC group ($5.0 > 4.3$); however, the difference between the SC and LK groups was not statistically significant on this measure. The difference in confidence level between these two groups comes from the four-tier classification criteria used in this study: the SC group was defined by correct answer and correct reason with high confidence, whereas the LK group was defined by incorrect and/or irrelevant responses with low confidence (see Sec. IV F 1 and Table II). This suggests that similar fixation duration on the correct reason option (AOI 6); may reflect different cognitive states: SC students likely attended to AOI 6 because they recognized the scientific explanation, whereas LK students may have spent comparable time there because they were uncertain and searching for cues before making a decision.

TABLE II. Representative response patterns and interview excerpts for each conceptual group.

Group	Answer selection	Reasoning selection	Confidence level	Representative interview excerpt
SC	Correct (forces are equal)	Correct (Newton's third law)	High (certain)	"I looked for the option that mentioned Newton's third law; since they are an action-reaction pair, the magnitude of the forces must be equal regardless of their mass."
MC	Incorrect (larger object exerts more force)	Incorrect (mass/size determines force)	High (certain)	"I chose the option where the larger object exerts more force because it is heavier and moving faster, so it must hit harder than the small one."
LK	Incorrect/correct (by guessing)	Incorrect/irrelevant	Low (uncertain)	"I wasn't sure. I just guessed that maybe the faster one has more force, but I don't really remember the rule for this."

TABLE III. Comparison of median total fixation duration (in seconds) and revisits on key AOIs across conceptual groups. Values are median (P25, P75). Note: $p < 0.05$.

AOI	Scientific conception ($n = 28$)	Misconception ($n = 24$)	Lack of knowledge ($n = 17$)	H statistic	p value
Times					
Total item	142.9 (101.1, 238.7)	175.1 (131.9, 231.2)	198.0 (147.7, 247.2)	4.590	0.204
Question stem (AOI 1)	85.3 (59.8, 129.5)	108.7 (80.5, 149.4)	109.4 (80.4, 171.8)	5.374	0.146
In problem confidence level (AOI 2)	3.9 (2.1, 6.2)	3.0 (0.8, 5.4)	4.8 (1.7, 7.5)	2.125	0.547
In reason level (AOI 3)	43.2 (32.4, 74.7)	54.0 (31.7, 74.1)	59.4 (38.8, 79.2)	2.652	0.448
In reason confidence level (AOI 4)	6.7 (2.1, 11.3)	4.8 (3.0, 9.2)	4.6 (2.6, 7.0)	1.240	0.743
In problem correct option (AOI 5)	7.2 (4.0, 10.5)	7.5 (4.6, 12.8)	9.8 (2.7, 18.2)	3.076	0.380
Correct reason option (AOI 6)	4.3 (1.7, 5.7)	2.2 (0.8, 3.3) ^a	5.0 (2.8, 8.4)	10.683	0.014*
Revisits					
AOI 1	61.5 (26.5, 112.7)	75.5 (41.5, 169.2)	64.0(33.0, 109.5)	2.687	0.442
AOI 2	33.0 (19.2, 72.5)	40.5 (15.5, 53.7)	28.0 (15.5, 65.0)	2.874	0.411
AOI 3	58.5 (33.7, 89.0)	62.5 (28.2, 100.2)	71.0 (38.5, 120.0)	1.439	0.696
AOI 4	41.5 (14.2, 103.2)	44.0 (20.5, 64.5)	37.0 (16.5, 51.5)	0.981	0.806

^aIndicates a significant difference compared to both the scientific conception and lack of knowledge groups in *post hoc* tests.

2. Qualitative analysis of heat maps and scan paths

The visual data provided a richer interpretation of these quantitative findings. The heat maps in Fig. 2 show the distribution of visual attention for typical participants from each group.

Although the heat maps make some between-group contrasts visually salient, these patterns should be interpreted as descriptive differences in fixation density rather than as evidence of additional statistically significant differences across AOIs. Table III reports the inferential

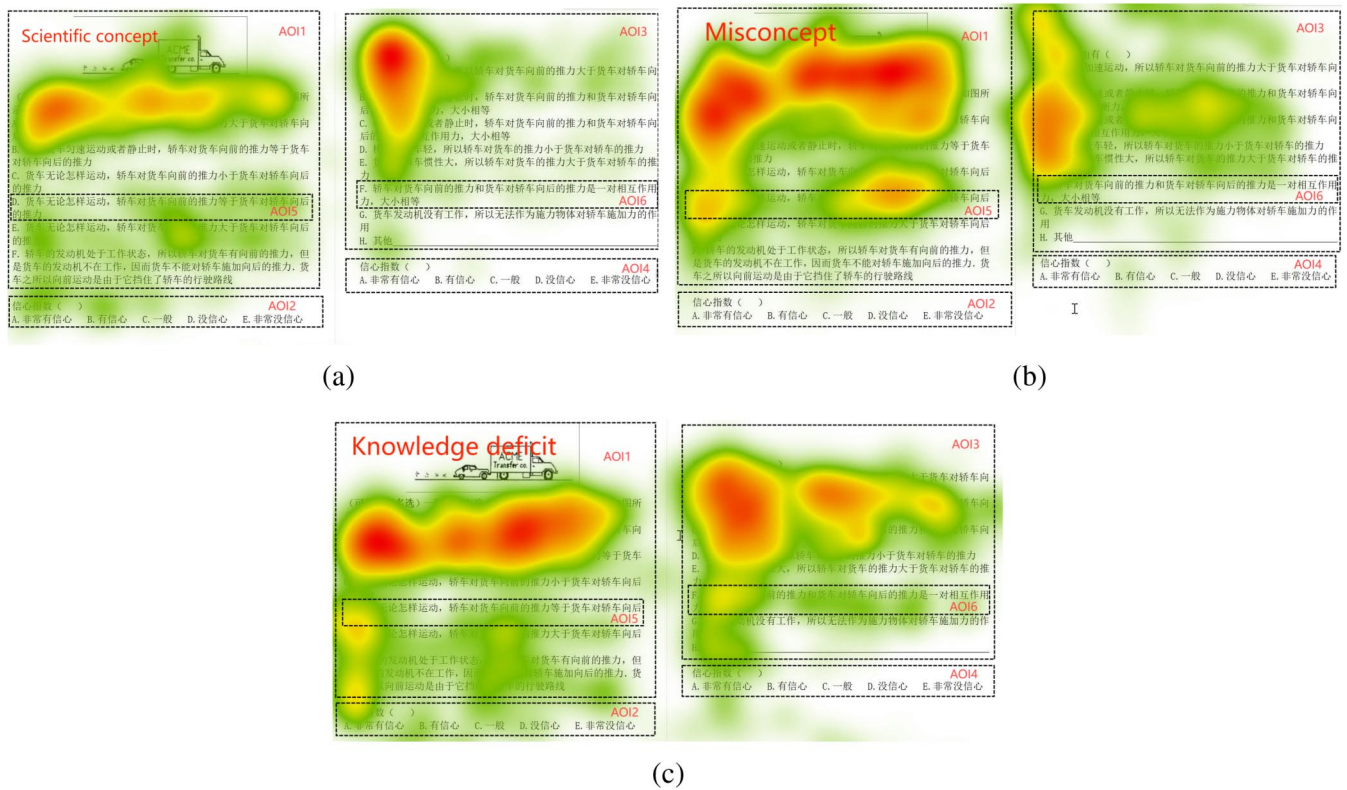


FIG. 2. Representative heat maps showing attention distribution. Heat map for the eye fixation of students with: (a) scientific conception—focused attention on key areas; (b) misconception—intense focus on incorrect regions; (c) lack of knowledge—diffuse attention.

statistical comparisons and shows that only fixation duration on the correct reason option (AOI 6) differed significantly among groups. By contrast, the heat maps illustrate how attention was spatially concentrated or dispersed within the item. In this sense, the visualizations enrich the interpretation of the quantitative results without contradicting them: students in the SC group showed concentrated attention to the correct scientific explanation (AOI 6), whereas students in the MC group showed relatively limited engagement with this region. The accompanying heat maps and scan paths should therefore be understood as illustrative qualitative evidence that clarifies the spatial distribution and sequence of attention, rather than as indicating additional statistically significant AOI-level differences beyond those reported in Table III.

Furthermore, the heat maps reveal the precision of attention. Unlike the MC group, whose attention was often drawn to misconception-consistent cues, such as statements referring to the larger or heavier object or to the motion state of the truck and car, the SC group's hotspots were tightly clustered around the scientifically correct reason option and the key textual information needed to judge the force relationship, indicating a precise extraction of relevant data rather than general viewing of the problem display.

The scan paths in Fig. 3 further illustrate the differences in information processing strategies.

For the SC group, the scan paths exhibit a distinct pattern of “strategic comparisons.” As shown in Fig. 3(a), there are repeated horizontal saccades (eye movements) directly connecting the mass of block A and the mass of block B (AOI 1 ↔ AOI 3). This “ping-pong” gaze behavior indicates that students were actively verifying the control variables, a hallmark of scientific reasoning.

In contrast, the LK group [Fig. 3(c)] displays a pattern characterized by random search behavior. The scan paths show fewer direct transitions between corresponding variables. Instead, their gaze frequently wanders to irrelevant regions (such as the table legs or empty background). This lack of a coherent search strategy aligns with their inability to formulate a valid solution plan. The MC participant's scan path [Fig. 3(b)] is highly chaotic, with frequent, rapid transitions back and forth between the question stem and the options, indicative of significant cognitive conflict.

To validate the search patterns observed in the scan paths, the frequencies of gaze transitions were compared among groups. The Kruskal-Wallis test revealed significant differences in the number of comparison transitions (AOI 1 ↔ AOI 3) among the three groups ($p < 0.05$). *Post hoc* tests indicated that the SC group made significantly more comparison transitions than the LK group. This quantitative finding supports the qualitative observation that SC

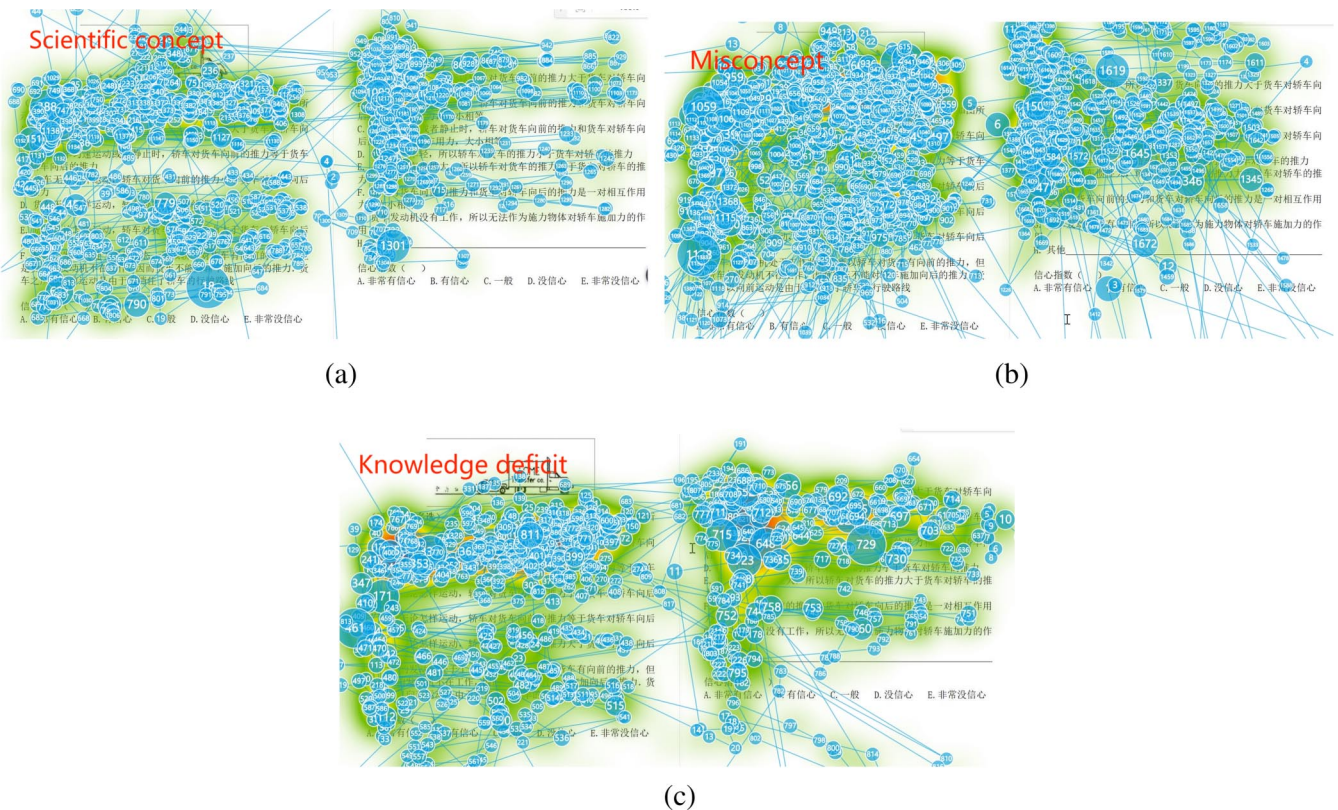


FIG. 3. Representative scan paths. The eye-movement patterns of students with: (a) scientific conception—a systematic, linear path; (b) misconception—a more chaotic pattern; (c) lack of knowledge—a disorganized path with few logical transitions.

TABLE IV. Eye-movement data analysis results of the confidence judgment process. Note: $*p < 0.05$.

	1 ($N = 18$)	2 ($N = 17$)	3 ($N = 47$)	4 ($N = 33$)	5 ($N = 25$)	H	p
Fixation time (second)	3.4 (2.2, 7.1)	5.2 (3.2, 11.8)	4.2 (2.1, 6.6)	4.2 (1.7, 10.2)	4.3 (2.1, 5.5)	3.165	0.531
Gaze (times)	192.0 (125.5, 410.7)	409.0 (145.5, 893.5)	236.0 (85.0, 382.0)	342.0 (45.0, 643.5)	240.0 (61.5, 483.5)	2.763	0.598
Revisits (times)	22.5 (11.0, 42.2)	45.0(26.5, 87.5) ^a	44.0(25.0, 73.0) ^a	34.0 (11.0, 67.0)	24.0 (11.0, 55.5)	10.646	0.031*
Percentage (%)	4.1 (2.4, 9.8)	6.1 (3.8, 14.5)	5.2 (1.8, 9.4)	3.4 (2.2, 17.1)	6.6 (4.6, 11.2)	3.155	0.532

^aIndicates a difference from “1”.

students actively engaged in checking and comparing control variables.

3. Interview results

Post-task semistructured interviews provided qualitative support for the eye tracking data. When asked about their reasoning, students in the MC group often expressed confusion regarding the interaction of forces, relying on intuitive but incorrect preconceptions. For example, one MC student stated: “I chose the option where the larger object exerts more force because it is heavier and moving faster, so it must hit harder than the small one.” This verbal report aligns with the heat map data showing their visual focus on mass-related distractors. In contrast, SC students demonstrated clear scientific reasoning, with one noting: “I looked for the option that mentioned Newton’s third law; since they are an action-reaction pair, the magnitude of the forces must be equal regardless of their mass.” This is consistent with their scan paths, which showed direct and sustained attention to the correct scientific explanation (AOI 6).

C. Eye-movement correlates of confidence judgment

To investigate the link between subjective confidence and objective eye movements, we analyzed the number of revisits to the confidence rating scales (AOI 2 and AOI 4) as a function of the participant’s self-reported confidence level on the five-point scale. The Kruskal-Wallis test revealed no significant differences in fixation time, gaze frequency, or percentage of time spent across the five confidence levels (all $p > 0.05$). Notably, there was a statistically significant difference in the number of revisits across the five confidence levels ($H(4) = 10.646$, $p = 0.031$), as shown in Table IV.

The median number of revisits for students with confidence ratings of 1, 2, 3, 4, and 5 was 22.5, 45.0, 44.0, 34.0, and 24.0, respectively. Pairwise comparisons showed that students with a confidence rating of 1 revisited the confidence stage significantly fewer times than those with ratings of 2 and 3 ($p < 0.05$).

A fascinating pattern emerged, as illustrated in Fig. 4. The relationship between confidence and the number of revisits followed a “reversed U-shape.” Participants who

were either highly confident (rating of 5) or highly unconfident (rating of 1) exhibited the fewest revisits. In contrast, participants with moderate levels of confidence (ratings of 2, 3, and 4) showed a significantly higher number of revisits. *Post hoc* tests revealed that participants with a confidence rating of 1 revisited significantly less than those with ratings of 2 and 3 ($p < 0.05$).

This result suggests that the number of revisits can serve as an objective behavioral marker for cognitive uncertainty. Those who are very certain (either correctly or incorrectly) make their confidence judgment decisively with little reevaluation. Those who are uncertain, however, engage in more cognitive effort, visually rechecking their answer and reasoning before committing to a confidence rating. This finding is crucial for the construction of the two-dimensional model, as it provides an objective data point that can validate or challenge a student’s subjective confidence report. While the present analysis focused on these confidence-related revisits, exploring whether similar nonlinear patterns exist for revisits to incorrect choices represents a valuable direction for future research.

VI. DISCUSSION

This study aimed to develop a more robust method for diagnosing student misconceptions than traditional four-tier diagnostic approaches that rely primarily on response

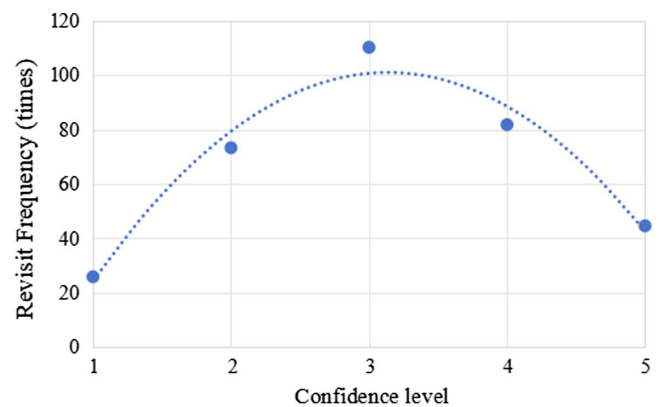


FIG. 4. The reversed U-shaped relationship between self-reported confidence level and the median number of revisits to the confidence AOIs. Error bars represent the interquartile range.

accuracy and self-reported confidence, by integrating objective eye tracking data with four-tier test responses. The results provide compelling evidence that specific certain aspects of eye movements not only differ significantly among students with varying levels of conceptual understanding but also correlate with their subjective feelings of confidence. These findings support the feasibility and value of the proposed eye tracking-test two-dimensional diagnostic model.

A. Distinct visual strategies and “attentional avoidance”

Our first research question examined whether visual attention patterns differed among students with SC, MC, and a LK. The results showed that between-group differences in eye movements were observed only for the correct reason, AOI 6, and this difference mainly distinguished the MC group from the other two groups. As shown in Table III, fixation times on the other AOIs (AOIs 1–5) did not differ significantly across groups.

Thus, the most important finding was that the MC group showed significantly shorter fixation durations on the correct reason (AOI 6). This suggests that the clearest group difference was not a general difference in attention across the entire item, but a specific difference in attention to the scientifically correct explanation. In other words, group differences became most apparent when students processed and evaluated the correct reason. Specifically, the MC group’s significantly reduced fixation on the correct scientific explanation (AOI 6), compared to both the SC and LK groups, indicates a pattern we term attentional avoidance—a tendency to quickly disengage from information that contradicts one’s firmly held belief. This pattern of attentional avoidance observed in the MC group suggests that students with strongly held misconceptions do not simply lack the correct knowledge; they actively ignore or quickly dismiss information that contradicts their preexisting mental model. This aligns with theories of cognitive dissonance and conceptual change, which posit that learners may resist new information that conflicts with their established beliefs.

In contrast, the SC group demonstrated a pattern of selective visual attention. As indicated by the heat maps and confirmed by the AOI transition analysis, their gaze was primarily directed to task-relevant regions of the diagnostic item, especially the question stem (AOI 1), the reason stem and options (AOI 3), and the correct reason option (AOI 6). They showed more focused processing of the scientifically relevant explanation and less attention to misconception-consistent distractors, such as options emphasizing object size, mass, or motion state. These students were able to efficiently identify the information needed to evaluate Newton’s third law—a finding that is consistent with Han *et al.* [37], who also reported that students with stronger conceptual understanding showed more targeted visual attention on FCI items, focusing attention on

scientifically relevant content rather than on misconception-consistent distractors. However, whereas Han *et al.* [37] characterized general visual attention patterns without distinguishing between MC and LK learners, the present study extends this line of inquiry by further differentiating the two novice subgroups through the integration of a four-tier confidence framework.

The LK group, on the other hand, showed diffuse attention and superficial processing, consistent with individuals who lack the necessary prior knowledge to engage deeply with the material. Their gaze patterns suggest that they are overwhelmed and do not know where to focus their cognitive resources.

These findings suggest that these groups rely on different visual search strategies when responding to physics conceptual questions. Specifically, students with misconceptions tended to focus more on salient but irrelevant features, whereas students with scientific conceptions demonstrated more systematic information processing. These distinct visual signatures demonstrate that eye tracking can effectively differentiate between knowledge states that might appear similar based on test scores alone. For example, although both MC and LK students answered the question incorrectly, their eye movements revealed fundamentally different underlying cognitive reasons for their errors, as evidenced by the significant difference in fixation duration between these two novice groups (MC and LK) on AOI 6.

B. The reversed U-shaped pattern: An objective marker for uncertainty

Our second research question investigated eye-movement behavior during the confidence judgment process. We adopted revisit count as the primary eye-movement metric. In this context, we use the term “cognitive monitoring” to refer to the metacognitive process by which learners track and evaluate their own knowledge states and certainty—a construct well established in the metacognition literature [41,42]—and “cognitive uncertainty” to denote the subjective sense of not knowing or being unsure about one’s response, which is operationalized here through the revisit behavior captured by the eye tracker.

The reversed U-shaped relationship between self-reported confidence and the number of revisits to the confidence AOIs (Fig. 4; Sec. V C) has several potential implications for diagnosing whether a student’s misconception is deeply entrenched or malleable for students’ concept assessment. For example, a student might report high confidence on a four-tier test, leading to a classification of misconception. However, if their eye movements show a high number of revisits, it could indicate that their confidence is actually brittle and they are experiencing underlying uncertainty. This inference is directly supported by our post-task interview data. For instance, one MC student who reported high confidence (rating of 4) but exhibited an elevated number of revisits to the

confidence scale admitted during the interview: “I chose the option where the larger object exerts more force because it is heavier and moving faster, so it must hit harder than the small one” (Table II), yet when probed further, added hesitantly that they “weren’t completely sure about the reasoning part.” This discrepancy between the student’s self-reported confidence and their hedging verbal response mirrors the elevated revisit behavior captured by the eye tracker, confirming that high revisits can reveal underlying uncertainty that self-report alone fails to detect. Conversely, another MC student who also reported high confidence but showed very few revisits stated unequivocally during the interview: “The truck is bigger, so obviously it pushes harder—that’s just common sense.” The absence of any verbal hedging in this case is consistent with the low revisit count, together suggesting a robust, deeply held misconception. Thus, the triangulation of eye tracking revisit data with interview responses validates the claim that revisits can objectively distinguish between brittle and entrenched confidence. This objective data can help teachers distinguish between deeply entrenched or potentially malleable, tentative one, allowing for more tailored interventions.

Conversely, individuals at the extremes of the confidence spectrum—both highly confident and highly unconfident—show fewer revisits. For the highly confident, their decision is quick and automatic, driven by a strong (though potentially incorrect) cognitive schema that requires little verification. Interview evidence supports this: SC students with high confidence showed decisive verbal reasoning (e.g., “I looked for the option that mentioned Newton’s third law; since they are an action-reaction pair, the magnitude of the forces must be equal regardless of their mass”), reflecting the same certainty captured by their low revisit counts. For the highly unconfident, the low number of revisits may reflect a form of cognitive disengagement or learned helplessness; they quickly conclude they do not know the answer and thus see no value in reexamining the information. This too was echoed in the interviews: several LK students with very low confidence (rating of 1) offered minimal verbal elaboration, with responses such as “I really don’t know this” or “I just picked something,” indicating rapid disengagement rather than deliberation—a pattern fully consistent with their low revisit counts. This pattern of low revisit count accompanying high confidence is consistent with the findings of Madsen *et al.* [32], who reported that students who answered physics problems incorrectly with high certainty showed more rapid, directed gaze patterns with fewer reinspections of key information regions compared to uncertain solvers. The present study extends this finding by linking revisit frequency specifically to confidence levels on a five-point scale, revealing a non-linear (reversed U-shaped) rather than simply a monotonic relationship between certainty and visual reevaluation behavior.

In sum, the convergence of the quantitative eye tracking data (the reversed U-shaped revisit pattern) with the qualitative interview evidence provides robust triangulated support for using revisits as an objective behavioral marker of cognitive uncertainty. This triangulation—a key methodological strength of the present study—demonstrates that the eye-movement patterns are not statistical artifacts but reflect genuine underlying cognitive and metacognitive processes that students themselves can articulate when prompted.

C. The eye tracking-test two-dimensional diagnostic model

Addressing the third research question, our findings strongly support the construction of a two-dimensional diagnostic model. Traditional one-dimensional diagnosis relies solely on test outcomes (correct/incorrect) and subjective confidence judgments [14]. Our model adds a second, objective dimension based on eye-movement behavior. By combining these, we can create a more nuanced and accurate diagnostic profile. Compared with the previous classification of experts (SC group) and novices (MC and LK groups), we made new findings among novices by using confidence level as the new grouping criterion. Specifically, Caleon and Subramaniam [14] established a foundational three-tier framework for diagnosing student conceptual states based on response accuracy and confidence; the present study not only replicates this core logic but also generalizes and expands their framework by integrating a fourth tier (reasoning confidence) and objective eye tracking metrics, enabling finer-grained differentiation within the novice category that was not achievable in their original classification. For example:

- (a) Students with robust misconceptions: These students provide incorrect answers, express high confidence in their responses, show low revisit frequency, and tend to avoid correct options. This profile characterizes students with deeply held misconceptions that are unlikely to yield to simple reteaching.
- (b) Students with emerging conflict: These students provide incorrect answers but also express high confidence while displaying high revisit frequency and chaotic scan paths. This profile suggests that although students hold a misconception, their visual behavior reveals underlying cognitive conflict—indicating a potentially malleable belief that may respond to guided inquiry.
- (c) Students with knowledge deficiency: These students provide incorrect answers and express low confidence, with diffuse attention patterns. This profile includes students with a simple knowledge deficit rather than a structured alternative conception, and they require direct foundational instruction rather than conceptual change strategies.

- (d) Students with insecure scientific knowledge: These students provide correct answers but express low confidence and display high revisit frequency. This profile highlights a group often overlooked in traditional assessments: students whose correct response may rest on fragile understanding. Their high revisit rate objectively signals that despite answering correctly, they lack the certainty of the SC group, warranting positive reinforcement rather than remediation.

This multifaceted diagnosis moves beyond simple labels and provides actionable insights into the student's cognitive state, enabling teachers to tailor their instructional strategies with much greater precision. The present study substantially extends the diagnostic framework established by Caleon and Subramaniam [14], who developed a three-tier test to distinguish misconceptions from lack of knowledge using subjective confidence ratings. While their work advanced the detection of cognitive states, it relied entirely on self-reported confidence, which is prone to bias. Our model goes further by integrating objective eye tracking data to validate and refine subjective judgments, which is a novel contribution not addressed in Ref. [14]. To our knowledge, the attentional avoidance pattern of students with robust misconceptions (significantly reduced fixation on the correct reasoning) and the reversed U-shaped relationship between confidence and revisits as an objective marker of cognitive uncertainty are novel findings that have not been previously documented in the literature. These results provide a more objective, process-oriented foundation for diagnosing student misconceptions beyond what was achievable using multitier tests alone.

VII. CONCLUSION AND IMPLICATIONS

A. Conclusions

This study integrated eye tracking technology with a four-tier diagnostic test to investigate how students with different knowledge states—SC, MC, and a LK—process physics conceptual items about Newton's third law. The principal conclusions, organized along three progressive instructional assessment layers, are as follows:

First, eye tracking data revealed distinct visual attention patterns that differentiate expertlike (SC group) from novicelike knowledge (MC and LK groups) states. Among all areas of interest examined (AOIs 1–6), significant between-group differences emerged exclusively on AOI 6, the correct scientific explanation. The SC group allocated substantial and focused fixation time to AOI 6, reflecting selective, expertlike information processing. In contrast, both the MC and LK groups—despite arriving at incorrect answers for fundamentally different cognitive reasons—exhibited significantly shorter fixation durations on AOI 6. Crucially, the MC group demonstrated what we have termed attentional avoidance, a tendency to

disengage rapidly from belief-inconsistent information, whereas the LK group displayed diffuse and superficial scanning indicative of an absence of relevant prior knowledge. The significant difference in fixation duration on AOI 6 between the MC and LK groups further confirms that identical behavioral outcomes (i.e., incorrect responses) can originate from qualitatively distinct cognitive processes—a distinction that traditional instructional assessment based solely on response accuracy cannot capture.

Second, the reversed U-shaped relationship between self-reported confidence and revisit count provides an objective behavioral marker for distinguishing between entrenched and potentially malleable misconceptions. Among MC students who reported high subjective confidence, those with low revisit counts exhibited verbal responses in post-task interviews that were unequivocal and unwavering, consistent with deeply entrenched misconceptions. In contrast, MC students who also reported high confidence but displayed elevated revisit counts revealed verbal hedging and underlying uncertainty during interviews, suggesting that their misconceptions, though superficially firm, harbor emerging cognitive conflict and are therefore potentially more amenable to conceptual change. This reversed U-shaped pattern—whereby both highly confident and highly unconfident students show fewer revisits, while moderately confident students show the most—offers educators an objective, real-time indicator of the epistemic status of a student's belief that self-report measures alone fail to detect.

Third, the convergence of these findings establishes the necessity and feasibility of the proposed eye tracking-test two-dimensional instructional assessment model. Traditional four-tier instruments classify students into categorical knowledge states but cannot distinguish within those categories—for instance, between a robust misconception and an emerging conflict, or between genuine scientific understanding and insecure correct knowledge. In the present sample, MC and LK students constituted approximately 34% and 24% of participants, respectively, meaning that over half of the students were classified as holding incorrect conceptions. Without the second, objective dimension provided by eye tracking metrics (fixation duration and revisit count), these students would receive undifferentiated instructional assessment labels and, consequently, undifferentiated instructional treatment. The two-dimensional model addresses this limitation by cross-referencing test-based classifications with eye-movement behavioral signatures, yielding four instructional assessment distinct profiles: robust misconception, emerging conflict, lack of knowledge, and insecure scientific knowledge. Each profile carries specific and actionable pedagogical implications.

B. Implications for instruction

The instructional assessment resolution afforded by the two-dimensional model enables educators to move beyond one-size-fits-all remediation and toward targeted, profile-specific instructional strategies:

For students with robust misconceptions (incorrect response, high confidence, low revisits, and attentional avoidance of AOI 6): These students' deeply entrenched alternative conceptions are unlikely to yield to simple reteaching or representation of the correct explanation, as their gaze behavior indicates they actively disengage from such information. Instructional approaches grounded in cognitive conflict—such as discrepant-event demonstrations, Socratic dialogue or bridging analogies are recommended to first destabilize the existing belief before introducing the scientific conception.

For students with emerging conflict (incorrect response, high confidence, high revisits, chaotic scan paths): The elevated revisit behavior signals that these students are already experiencing internal cognitive tension between their misconception and the correct explanation. This represents a critical “teachable moment.” Instructors should leverage this nascent uncertainty through guided inquiry, structured argumentation, or peer discussion that helps students explicitly confront and resolve the conflict, rather than allowing them to default back to their prior belief.

For students with a lack of knowledge (incorrect response, low confidence, low revisits, diffuse attention): These students require foundational knowledge construction rather than conceptual change. Their diffuse gaze patterns and rapid disengagement suggest cognitive overload or learned helplessness. Scaffolded instruction—including worked examples, visual representations, and step-by-step reasoning guides—is necessary to build the prerequisite knowledge structures before higher-order evaluation of competing explanations can be expected.

For students with insecure scientific knowledge (correct response, low confidence, and high revisits): Although these students arrive at the correct answer, their elevated revisit counts and low confidence indicate fragile understanding. Without reinforcement, they risk regressing toward misconceptions. Positive reinforcement, metacognitive reflection activities (e.g., asking students to explain why their answer is correct), and opportunities for successful application of the concept in novel contexts can help consolidate their emerging scientific understanding.

C. Implications for research methodology

From a methodological standpoint, this study demonstrates that eye tracking technology can serve as a useful complementary tool to traditional instructional assessment in physics education research. The combination of

fixation-based metrics (capturing what students attend to) and revisit-based metrics (capturing how certain they feel about their judgments) provides a two-dimensional behavioral window into cognitive and metacognitive processes that are otherwise inaccessible through paper-and-pencil instruments alone. We encourage future researchers to extend this paradigm to other conceptual domains beyond Newtonian mechanics, to larger and more diverse samples, and to longitudinal designs that track how eye-movement signatures evolve as students undergo conceptual change interventions. Additionally, the development of automated, real-time classification algorithms based on the instructional assessment profiles identified here could pave the way for adaptive learning systems that adjust instructional content dynamically based on students' gaze behavior.

D. Limitations

Several limitations should be acknowledged. First, the sample size, while sufficient for the exploratory analyses conducted, limits the generalizability of the findings. Second, the study focused on a single physics concept (Newton's third law); whether the observed eye-movement patterns generalize to other conceptual domains remains an empirical question. Third, eye tracking data were collected in a controlled laboratory setting, and ecological validity in authentic classroom environments has yet to be established. Finally, the reversed U-shaped revisit pattern, while supported by both quantitative data and interview evidence, warrants replication with larger samples and more fine-grained confidence scales to establish its robustness and boundary conditions.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers for their valuable comments and suggestions. We are grateful to Teacher Lu Xiaoxu from East China Normal University for technical support with the eye tracking experiment. This work was supported in part by the 2023 Association of Higher Education of P.R. China (Grant No. 23XXK0202), the National Natural Science Foundation of China (Grants No. U23A2075 and No. 12274132), the Fundamental Research Funds for the Central Universities, and the Innovation Program of Shanghai Municipal Education Commission (Grant No. 202101070008E00099).

DATA AVAILABILITY

The eye tracking metrics, four-tier diagnostic test responses, and student conceptual classification data that support the findings of this study are available from the corresponding author upon reasonable request. These research data are cited as Ref. [43].

APPENDIX A: ENGLISH TRANSLATION OF EXPERIMENTAL INSTRUMENTS

The following is a translation of the specific diagnostic problem and scales used in the eye tracking experiment. The problem is designed to assess students' understanding of Newton's third law and to identify common misconceptions regarding force and motion.



(The picture is from Ref. [25])

Problem Scenario: A truck breaks down on the road and is pushed back to the city by a small car (as shown in the figure). During the process of the car pushing the truck, which of the following statements is correct?

Part 1: Initial Conception Choice Please select the statement that best describes the forces involved:

- Only when the truck is accelerating, the forward force of the car on the truck is greater than the backward force of the truck on the car.
- Only when the truck is moving at a constant speed or is stationary, the forward force of the car on the truck is equal to the backward force of the truck on the car.
- Regardless of how the truck moves, the forward force of the car on the truck is smaller than the backward force of the truck on the car.
- Regardless of how the truck moves, the forward force of the car on the truck is equal to the backward force of the truck on the car. (*Scientific Conception*)
- Regardless of how the truck moves, the forward force of the car on the truck is greater than the backward force of the truck on the car. (*Common Misconception*)
- The car's engine is working, so the car exerts a forward force on the truck; however, the truck's engine is not working, so the truck cannot exert a backward force on the car. The reason the truck moves forward is that it blocks the car's path.

Part 2: Confidence Rating (First Pass) How confident are you in your choice above?

- Very Confident
- Confident
- Neutral/Uncertain
- Unconfident
- Very Unconfident

Part 3: Reasoning Choice Please select the reason that supports your answer:

- The truck is accelerating, so the forward force of the car on the truck is greater than the backward force of the truck on the car.
- Only when the truck is moving at a constant speed or stationary, the forward force of the car on the truck and the backward force of the truck on the car are *balanced forces*, so they are equal in magnitude. (*Misconception: confusing interaction with balance*)
- Only when the truck is moving at a constant speed or stationary, the forward force of the car on the truck and the backward force of the truck on the car are *interaction forces*, so they are equal in magnitude.
- The car is lighter than the truck, so the force of the car on the truck is smaller than the force of the truck on the car.
- The truck has greater inertia than the car, so the force of the car on the truck is greater than the force of the truck on the car.
- The forward force of the car on the truck and the backward force of the truck on the car are a pair of *interaction forces* (action and reaction forces), so they are always equal in magnitude. (*Correct Reason/AOI 6*)
- The truck's engine is not working, so it cannot act as an agent to exert force on the car.
- Other.

Part 4: Confidence Rating (Second Pass) How confident are you in your reasoning?

- Very Confident
- Confident
- Neutral/Uncertain
- Unconfident
- Very Unconfident

APPENDIX B

Conceptual understanding	Problem confidence	Reason confidence	Fixation time (second)						Revisit (times)				Gaze (times)		Percentage (%)	
			AOI1	AOI2	AOI3	AOI4	AOI5	AOI6	AOI1	AOI2	AOI3	AOI4	AOI2	AOI4	Problem	Reason
Scientific concept	3	3	171.5	0.4	30.2	0.5	18.3	5.3	109	25	68	25	40	43	0.26	1.77
	4	4	82.6	2.1	44.0	10.2	5.8	14.4	21	9	56	45	186	782	2.44	18.87
	4	4	113.3	6.6	57.5	18.4	7.7	2.7	338	79	46	7	598	1650	5.54	24.19
	3	3	248.4	0.5	133.1	2.1	13.9	7.1	68	16	89	47	41	139	0.18	1.57
	4	4	107.5	3.8	77.6	5.9	8.1	4.5	63	32	73	18	342	448	3.41	7.05
	4	3	126.2	4.3	61.7	5.6	5.6	5.8	46	20	43	21	383	395	3.26	8.37
	3	3	229.6	6.7	42.5	41.9	14.8	5.6	78	40	52	74	601	3649	2.83	49.65
	5	5	142.9	5.5	116.0	15.1	15.9	10.4	65	38	61	165	499	737	3.73	11.53
	4	4	178.6	4.1	63.8	11.7	10.7	5.5	127	82	90	54	371	817	2.26	15.49
	4	5	103.3	31.1	90.8	7.8	11.2	21.3	60	982	119	122	2803	574	23.17	7.92
	5	5	34.6	6.9	66.5	4.3	0.6	6.8	23	176	927	13	617	81	16.55	6.13
	5	5	52.2	2.7	20.0	4.4	3.2	4.5	28	46	96	120	240	271	4.86	18.06
	5	5	50.1	2.9	39.6	7.5	1.0	4.2	21	9	64	57	263	645	5.51	15.98
	5	5	35.7	4.6	36.3	4.6	2.2	5.0	37	54	23	13	414	373	11.41	11.16
	4	4	76.2	4.5	59.7	2.0	2.2	4.5	166	152	46	39	404	133	5.56	3.21
	4	4	105.5	6.7	83.7	9.3	7.0	9.3	56	78	74	56	606	783	6.00	10.02
	4	4	130.6	4.7	92.5	48.3	8.0	2.1	114	56	89	396	295	121	3.49	34.31
	3	3	88.0	8.9	44.2	1.2	8.5	0.9	108	43	32	4	236	21	9.22	2.64
	5	4	60.6	6.2	39.1	10.4	6.1	2.8	28	23	17	30	469	401	9.32	20.99
	5	5	77.7	0.6	36.5	0.7	10.1	2.2	19	10	11	9	42	4	7.8	19.4
	4	4	60.7	0.8	30.1	0.7	7.4	0.3	114	7	39	11	20	13	12.3	2.20
	3	3	169.3	3.0	114.1	2.2	71.3	1.2	817	32	843	34	193	146	1.72	1.85
	4	4	60.6	1.6	31.2	10.3	0.3	0.4	102	27	12	128	122	29	2.50	24.89
	3	3	61.8	1.9	8.6	18.8	5.5	0.8	26	14	101	113	85	713	2.95	68.58
	5	5	36.4	2.4	17.3	2	4	3.5	13	19	18	8	154	34	6.11	10.18
	5	3	35.3	2.5	28.7	7.6	4	2.6	14	21	23	44	160	120	6.61	20.99
	3	3	59.6	6.2	38.2	23.4	4.4	0.8	141	115	74	131	351	206	9.47	37.92
	4	4	102.2	2.7	36.2	2.5	7.5	1.6	60	34	42	31	120	51	2.59	6.48
Knowledge deficit	3	2	109.5	5.6	61.9	12.1	9.8	5.1	79	84	100	170	507	1024	4.89	16.34
	2	1	85.4	7.7	59.2	2.7	8.4	5.7	97	22	90	19	690	232	8.23	4.35
	2	2	130.1	5.4	90.9	4.7	14.2	6.1	136	92	107	59	488	415	4.00	4.87
	1	1	167.2	4.5	24.4	19.7	26.4	2.6	34	18	17	14	402	818	2.60	44.76
	1	2	192.6	2.4	33.3	4.8	22.2	1.5	26	11	55	41	214	396	1.22	12.62
	3	2	85.6	4.8	65.4	3.3	5.2	7.1	122	70	71	45	433	146	5.32	4.77
	2	2	79.8	0.0	60.4	2.3	0.1	0.8	549	0	624	29	0	17	0.00	3.63
	2	2	113.8	16.6	69.9	81.6	13.7	17.4	92	86	64	58	982	2574	12.71	53.85
	1	1	63.5	1.8	39.7	4.2	4.4	4.4	47	29	28	8	84	297	2.79	9.63
	2	1	81.0	5.3	59.5	7.1	2.4	3.0	20	13	25	24	409	244	6.10	10.65
	1	1	204.3	9.6	38.8	6.1	34.7	3.5	232	60	199	37	764	109	4.47	13.61
	1	1	176.4	7.3	44.7	2.7	3.1	9.8	32	21	21	11	437	142	3.99	5.63
	1	1	176.7	2.5	88.6	2.5	13.1	15.9	97	43	133	43	159	133	1.40	2.70
	2	2	160.5	1.1	100.8	3.5	23.8	4.9	64	28	71	25	40	145	0.67	3.35
	1	2	79.6	1.6	30.8	7.1	2.0	1.8	44	24	61	60	52	166	1.92	18.70
	1	1	37.8	1.6	88.9	0.5	0.2	18.0	13	7	192	9	131	12	4.01	0.58
	1	1	98.9	10.3	38.8	7.0	14.1	5.1	54	42	49	45	572	170	9.45	15.27

(Table continued)

Conceptual understanding	Problem confidence	Reason confidence	Fixation time (second)						Revisit (times)				Gaze (times)		Percentage (%)	
			AOI1	AOI2	AOI3	AOI4	AOI5	AOI6	AOI1	AOI2	AOI3	AOI4	AOI2	AOI4	Problem	Reason
Misconcept	3	3	163.3	1.9	58.6	4.0	10.9	3.3	63	47	72	61	167	249	1.12	6.32
	3	3	109.6	2.1	75.2	5.9	6.8	4.4	77	15	67	31	190	381	1.89	7.28
	3	3	135.4	4.2	50.5	5.0	11.1	0.7	50	40	18	10	382	316	3.04	9.05
	4	3	251.1	4.2	51.3	4.7	7.2	1.6	61	47	29	63	381	237	1.66	8.33
	3	3	154.0	0.8	70.7	3.9	16.8	2.1	172	26	71	86	73	308	0.52	5.24
	5	5	89.2	5.5	23.9	5.0	9.9	2.4	33	28	19	65	498	224	5.84	17.17
	3	3	95.3	0.9	36.2	0.7	4.6	1.3	58	43	28	7	85	54	0.98	1.78
	3	3	120.6	3.1	30.5	20.6	8.7	3.4	110	44	58	22	283	882	2.54	40.35
	4	4	78.2	0.1	26.0	3.0	3.4	0.4	6	3	6	4	5	268	0.07	10.33
	4	4	219.9	0.4	107.8	1.0	14.9	11.8	49	8	34	38	39	33	0.20	0.90
	3	3	135.8	2.9	85.2	4.0	15.9	4.3	105	56	94	60	260	277	2.08	4.47
	3	3	310.9	8.8	104.4	9.3	41.2	16.4	161	213	217	156	794	657	2.76	8.16
	4	4	251.1	7.6	136.5	33.2	10.8	2.3	123	148	74	44	681	1309	2.93	19.56
	3	3	8.9	2.9	35.3	30.5	0.1	0.1	193	126	2487	2739	1	2739	24.34	46.38
	2	2	102.6	11.5	75.7	20.5	8.0	3.4	187	89	477	398	805	1635	10.10	21.27
	5	5	87.6	0.1	66.1	3.1	0.8	1.5	19	6	33	28	6	13	0.10	4.45
	5	5	60.2	5.0	13.8	0.8	5.1	0.3	35	42	18	12	342	8	7.73	5.27
	3	3	126.4	2.8	14.8	6.4	24.2	0.4	388	33	242	46	75	99	2.20	30.19
	4	4	125.7	0.8	50.1	1.0	5.7	2.8	74	11	43	20	21	7	0.64	1.89
	3	4	61.0	4.3	14.0	29.5	6.5	0.5	17	17	15	3	284	2574	6.55	67.81
	2	3	50.1	3.2	56.8	6.3	2.8	2.1	89	31	95	74	204	213	6.06	9.98
	3	3	98.7	5.8	66.7	2.3	1.8	3.4	399	41	102	22	188	102	5.51	3.33
	3	3	108	0.6	40.7	3.6	13.4	1.6	39	9	38	44	29	35	0.55	8.07
	3	3	74.8	10.5	59.3	9.0	6.2	2.7	221	73	134	45	864	309	12.34	13.15
Mistake	5	5	48.1	1.7	23.3	2.4	0.4	2.4	21	6	34	24	154	142	3.43	9.26

[1] J. D. Bransford, A. L. Brown, and R. R. Cocking, *How People Learn: Brain, Mind, Experience and School* (National Academy Press, Washington, DC, 2000).

[2] J. Clement, Students’ preconceptions in introductory mechanics, *Am. J. Phys.* **50**, 66 (1982).

[3] P. Tao and R. Gunstone, The process of conceptual change in force and motion during computer-supported physics instruction, *J. Res. Sci. Teach.* **36**, 859 (1999).

[4] M. T. H. Chi, A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman, Two approaches to the study of experts’ characteristics, in the *Cambridge handbook of expertise and expert performance*, edited by K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman (Cambridge University Press, Cambridge, 2006), pp. 21–30.

[5] L. C. McDermott and P. S. Shaffer, Research as a guide for curriculum development: an example from introductory electricity. Part I: investigation of student understanding, *Am. J. Phys.* **60**, 994 (1992).

[6] L. S. Shulman, Those who understand: knowledge growth in teaching, *Educ. Res.* **15**, 4 (1986).

[7] L. S. Shulman, Knowledge and teaching: Foundations of the new reform, *Harv. Educ. Rev.* **57**, 1 (1987).

[8] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, Accommodation of a scientific conception: toward a theory of conceptual change, *Sci. Educ.* **66**, 211 (1982).

[9] J. H. van Driel, N. Verloop, and W. de Vos, Developing preservice teachers’ pedagogical content knowledge, *J. Res. Sci. Teach.* **35**, 673 (1998).

[10] D. F. Treagust, Development and use of diagnostic tests to evaluate students misconceptions in science, *Int. J. Sci. Educ.* **10**, 159 (1988).

[11] S. Hasan, D. Bagayoko, and E. L. Kelley, Misconceptions and the certainty of response index (CRI), *Phys. Educ.* **34**, 294 (1999).

[12] A. B. Arons and K. Kurki-Suonio, Teaching introductory physics, *Eur. J. Phys.* **19**, 315 (1998).

[13] R. Driver and J. Easley, Pupils and paradigms: A review of literature related to concept development in adolescent science students, *Stud. Sci. Educ.* **5**, 61 (1978).

- [14] I. Caleon and R. Subramaniam, Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves, *Int. J. Sci. Educ.* **32**, 939 (2010).
- [15] N. Hermita, A. Suhandi, E. Syaodih *et al.*, Constructing and implementing a four tier test about static electricity to diagnose pre-service elementary school teachers misconceptions, *J. Phys. Conf. Ser.* **895**, 012167 (2017).
- [16] R. E. Nisbett and T. D. Wilson, Telling more than we can know: Verbal reports on mental processes, *Psychol. Rev.* **84**, 231 (1977).
- [17] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, 2nd ed. (Springer, New York, 2017).
- [18] K. Rayner, Eye movements and attention in reading, scene perception, and visual search, *Q. J. Exp. Psychol.* **62**, 1457 (2009).
- [19] K. Rayner, Eye movements and information processing: 20 years of research, *Psychol. Bull.* **124**, 372 (1998).
- [20] M. J. Tsai, H. T. Hou, M. L. Lai *et al.*, Visual attention for solving multiple-choice science problem: An eye-tracking analysis, *Comput. Educ.* **58**, 375 (2012).
- [21] T. Berg and W. Brouwer, Teacher awareness of student alternate conceptions about rotational motion and gravity, *J. Res. Sci. Teach.* **28**, 3 (1991).
- [22] M. McCloskey, Intuitive physics, *Sci. Am.* **248**, 122 (1983).
- [23] L. C. McDermott, Millikan Lecture 1990: What we teach and what is learned—Closing the gap, *Am. J. Phys.* **59**, 301 (1991).
- [24] E. F. Redish and J. R. Burciaga, *Teaching Physics: With the Physics Suite* (John Wiley & Sons, Hoboken, NJ, 2003), Vol. 1, p. 216.
- [25] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [26] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, *Am. J. Phys.* **66**, 338 (1998).
- [27] I. A. Halloun and D. Hestenes, The initial knowledge state of college physics students, *Am. J. Phys.* **53**, 1043 (1985).
- [28] S. Zhou, Y. Wan, and C. Zhang, Pre-service science teachers' PCK: Inconsistency of pre-service teachers' predictions and student learning difficulties in Newton's third law, *Eurasia J. Math. Sci. Technol. Educ.* **12**, 373 (2016).
- [29] M. T. Chi, P. J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, *Cogn. Sci.* **5**, 121 (1981).
- [30] K. Holmqvist, M. Nyström, R. Andersson *et al.*, *Eye Tracking: A Comprehensive Guide to Methods and Measures* (Oxford University Press, Oxford, 2011).
- [31] H. Jarodzka, K. Scheiter, P. Gerjets *et al.*, In the eyes of the beholder: How experts and novices interpret dynamic stimuli, *Learn. Instr.* **20**, 146 (2010).
- [32] A. M. Madsen, A. M. Larson, L. C. Loschky, and N. S. Rebello, Differences in visual attention between those who correctly and incorrectly answer physics problems, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010122 (2012).
- [33] B. Ibrahim and L. Ding, Sequential and simultaneous synthesis problem solving: A comparison of students' gaze transitions, *Phys. Rev. Phys. Educ. Res.* **17**, 010126 (2021).
- [34] M. A. Just and P. A. Carpenter, A theory of reading: From eye fixations to comprehension, *Psychol. Rev.* **87**, 329 (1980).
- [35] J. Hyönä, R. F. Lorch, Jr., and J. K. Kaakinen Jr, Individual differences in reading to summarize expository text: Evidence from eye fixation patterns, *J. Educ. Psychol.* **94**, 44 (2002).
- [36] Q. Wang, Y. Zhu, L. Wei, and H. Deng, Visual attention pattern of middle school students during problem-solving in physics, *Mind Brain Educ.* **16**, 99 (2022).
- [37] J. Han, L. Chen, Z. Fu, J. Fritchman, and L. Bao, Eye-tracking of visual attention in web-based assessment using the force concept inventory, *Eur. J. Phys.* **38**, 045702 (2017).
- [38] H. C. She, M. J. Chen, L. Y. Huang, and C. Y. Hsueh, Unfolding the cognitive process underlying computer-based scientific conceptual change with eye tracker: Behavioral performance and sequential analysis of attention, *Educ. Inf. Technol.* **30**, 20157 (2025).
- [39] *Physics (Compulsory Volume 1)* (People's Education Press, Beijing, 2019).
- [40] Ministry of Education of the People's Republic of China, *General Senior Secondary Physics Curriculum Standards (2017 Edition)* (People's Education Press, Beijing, 2025).
- [41] T. O. Nelson and L. Narens, Metamemory: A theoretical framework and new findings, *Psychol. Learn. Motiv.* **26**, 125 (1990).
- [42] J. H. Flavell, Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry, *Am. Psychol.* **34**, 906 (1979).
- [43] J. Liu, X. Zhong, G. Chen, and L. Chen, Data for: Building a diagnostic model for misconceptions in physics education using eye tracking technology (2026), [10.1103/ttrq-px6t](https://doi.org/10.1103/ttrq-px6t).