



Assessing Physics Competence in Teaching Dynamics: A Survey of Upper Secondary Teachers and Students in Northern Mountainous Provinces of Vietnam

Cao Tien Khoa¹, Nguyen Van Hieu^{2*}

¹Thai Nguyen University of Education, Thai Nguyen, Vietnam

²Van Ban Centre for Vocational Education and Continuing Education, Lao Cai Province, Vietnam

ABSTRACT: This study describes current practices of assessing physics competence in the Dynamics chapter of Grade 10 Physics in order to inform the design of an assessment toolkit. Data were collected through two questionnaires from 34 teachers and 233 upper secondary students, mostly in Lao Cai province, and were analysed with descriptive statistics. Cronbach's α of the scales ranged from 0.913 to 0.983. Teachers used paper-based assessment tools ($M = 3.89$) more often than process-oriented tools ($M = 3.20$), and rubrics and peer assessment were used least. Teachers rated their understanding of the structure of physics competence as fairly good ($M = 3.63$), while 47.1% correctly identified all three competence components. Students reported lower frequencies than teachers for written tests, context-based tasks and experiments, and 49.1% of students named experimental inquiry as the task group assessed least often. Of the teachers, 82.4% considered a chapter-specific assessment toolkit necessary, provided that training is conducted on the chapter content itself and that record-keeping time is reduced. The findings suggest that the toolkit should include content-specific behavioural indicators, experimental tasks with rubrics, and criterion-referenced feedback templates.

Corresponding Author:

Nguyen Van Hieu

Published Online:

October 26, 2026

License:

This is an open access article under the CC BY 4.0 license:

<https://creativecommons.org/licenses/by/4.0/>

KEYWORDS:

competence-based assessment, physics competence, teacher assessment literacy, rubric, Dynamics

Cite the Article: Khoa, C.A., Hieu, N.V. (2026). *Assessing Physics Competence in Teaching Dynamics: A Survey of Upper Secondary Teachers and Students in Northern Mountainous Provinces of Vietnam*. *International Journal of Human Research and Social Science Studies*, 3(10), 959-969. <https://doi.org/10.55677/ijhrsss/10-2026-Vol03I10>

1. INTRODUCTION

The 2018 General Education Curriculum for Physics in Vietnam defines physics competence as comprising three components: physics cognition; inquiry into the natural world from a physics perspective; and application of acquired knowledge and skills (Ministry of Education and Training [MOET], 2018). Circular 22/2021/TT-BGDĐT regulates continuous and periodic assessment and combines descriptive comments with scores (MOET, 2021). These two documents call for assessment tools that collect evidence on each competence component, rather than focusing mainly on knowledge and problem-solving skills.

International reviews indicate that assessment for learning and guided feedback are associated with student learning outcomes (Black & Wiliam, 1998; Hattie & Timperley, 2007), and that the extent to which these principles are enacted depends on teachers' assessment literacy in specific contexts (Xu & Brown, 2016). In Vietnam, a survey of 438 teachers in four provinces and cities found that most teachers assessed students against the required learning outcomes, while comprehensive assessment tools, guidance materials and training remained limited, with differences between regions (Dương et al., 2024). In physics, the pressure of standardised examinations and the availability of equipment have been reported as factors affecting competence-based teaching (T. K. C. Nguyễn et al., 2026).

Existing surveys are usually conducted at the subject level or with only one group of respondents. Data comparing teachers and students for the same chapter are scarce, particularly in mountainous schools and ethnic minority boarding schools. The Dynamics chapter (Physics 10) contains many quantitative experiments (friction, elastic force, Newton's second law) and many

real-life situations (road safety), which makes it suitable for assessing all three competence components. This article reports the needs analysis phase of the project “Developing and using a toolkit for assessing upper secondary students’ physics competence: Application to teaching the Dynamics chapter, Physics 10” and addresses five research questions (RQs): (1) Which assessment forms do teachers use, how often, and how do teachers’ and students’ reports on these forms differ? (2) To what extent do teachers understand the structure of physics competence, as indicated by self-ratings and recognition items? (3) What difficulties do teachers face, and how do they view the limitations of current assessment tools? (4) How do students perceive assessment practices and rate their own abilities across the three competence components? (5) What needs and conditions do teachers report for adopting a new toolkit, and which assessment forms do students prefer?

2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1. *Physics competence and its assessment*

According to the 2018 Physics Curriculum, each competence component is described by specific indicators; for example, inquiry into the natural world includes posing questions, making predictions, planning, carrying out the plan, and writing and presenting reports (MOET, 2018). Assessing competence therefore requires three operations: identifying observable behavioural indicators; setting tasks in meaningful contexts; and describing performance levels for comparison. A test blueprint based on four cognitive levels (recognition, comprehension, application, high-level application) classifies item difficulty but does not directly show which competence component an item measures. Vietnamese studies have proposed procedures for designing PISA-based exercises that develop physics competence through context and data (T. H. H. Nguyễn et al., 2022). Regarding the inquiry component, school practical work often engages students in manipulating objects more than in reasoning with ideas and data (Abrahams & Millar, 2008), and teachers’ capacity to organise and assess experimental activities, including that of pre-service teachers, has also received attention (Lê & Võ, 2022).

2.2. *Process-oriented assessment, feedback and student involvement*

Wiliam (2011) describes assessment for learning through five strategies: clarifying learning intentions and success criteria; engineering activities that elicit evidence of learning; providing feedback that moves learners forward; activating students as learning resources for one another; and activating students as owners of their own learning. Hattie and Timperley (2007) distinguish feedback at the task, process, self-regulation and self levels; feedback that addresses “where am I going”, “how am I going” and “where to next” is associated with learning outcomes. Rubrics operationalise the strategy of clarifying criteria; the meta-analysis by Panadero et al. (2023) reported positive effects of rubrics on academic performance and, to a lesser extent, on self-regulated learning and self-efficacy. These arguments underpin the survey items on whether students know the criteria in advance, receive specific feedback, and take part in self- and peer assessment.

2.3. *Teacher assessment literacy*

Xu and Brown (2016) proposed a framework of teacher assessment literacy in practice comprising a knowledge base of assessment, an interpretive and guiding framework of conceptions, the conditions of school and policy contexts, and teachers’ learning and professional identity. Within this framework, theoretical knowledge is necessary but not sufficient; organisational conditions and external pressures may change how teachers carry out assessment. DeLuca et al. (2016) developed an instrument measuring teachers’ approaches to classroom assessment and showed that teachers may hold relevant knowledge while prioritising different approaches depending on the situation. A survey of Vietnamese primary teachers reported positive perceptions of competence-based assessment, while some teachers were not yet proficient or confident and some methods were not used regularly (Vũ et al., 2024).

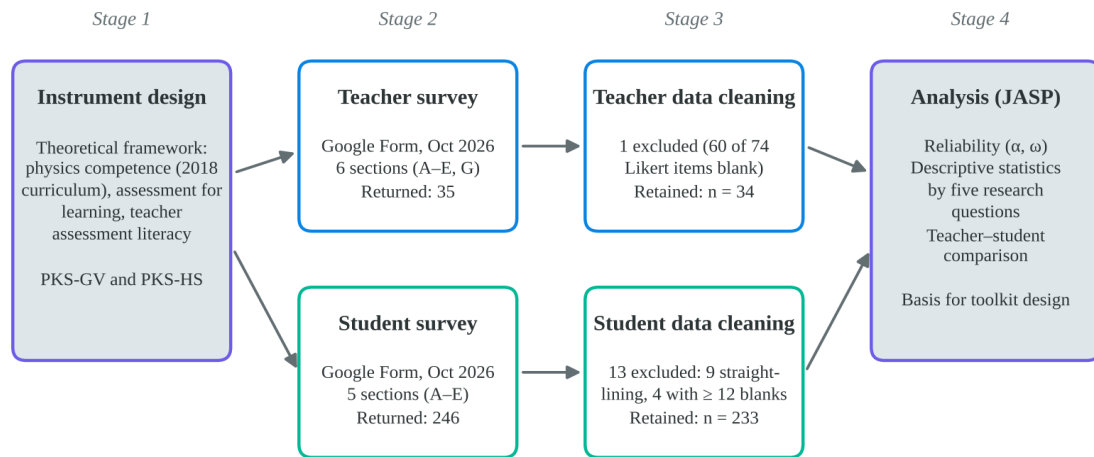
2.4. *Analytical framework of the study*

Building on these arguments, the study examines the assessment of physics competence along five aspects corresponding to the five RQs: assessment practice (Part B of both questionnaires); teachers’ understanding of the competence structure (Part C of the teacher questionnaire); difficulties and limitations in context (Parts D and E of the teacher questionnaire); students’ experiences and self-ratings (Parts C and D of the student questionnaire); and needs and conditions for adopting a new toolkit (Part G of the teacher questionnaire, Part E of the student questionnaire). Collecting data from two groups allows the same activity to be compared from the perspectives of those who organise assessment and those who are assessed.

3. METHODS

3.1. *Research design*

The study used a cross-sectional survey design with two questionnaires: one for teachers (PKS-GV) and one for students (PKS-HS). Data were collected online via Google Forms from 4 to 18 October 2026. The procedure comprised four stages, as shown in Figure 1.

**Figure 1. Research procedure**

3.2. Participants

Participants were recruited through convenience sampling. Of the 35 teacher and 246 student responses returned, responses not meeting quality criteria were excluded: one teacher response left 60 of 74 Likert items blank; nine student responses selected the same level for all 44 items; and four student responses left 12 or more items blank. The analytic sample comprised 34 teachers and 233 students (Tables 1 and 2).

Table 1. Characteristics of the teacher sample (n = 34)

Variable	Category	n (%)
Province	Lao Cai / Hanoi / Tuyen Quang / other provinces ¹	23 (69.7) / 4 (12.1) / 2 (6.1) / 4 (12.1)
School location	Mountainous, disadvantaged area / Rural / Urban, town	19 (57.6) / 5 (15.2) / 9 (27.3)
Teaching experience	5–10 years / 11–20 years / Over 20 years	3 (8.8) / 23 (67.6) / 8 (23.5)
Qualification	Bachelor / Master	23 (67.6) / 11 (32.4)
Class size	Under 35 / 35–40 / 41–45 / Over 45 students	2 (5.9) / 7 (20.6) / 23 (67.6) / 2 (5.9)
Training in competence-based assessment	None / Once / Several times / As a trainer	4 (11.8) / 4 (11.8) / 24 (70.6) / 2 (5.9)
Years teaching Dynamics (2018 curriculum)	1 year / 2 years / 3 years or more	3 (9.1) / 4 (12.1) / 26 (78.8)

¹ Hung Yen, Lai Chau, Ninh Binh and Phu Tho (one teacher each); one teacher did not report a province. Percentages are based on valid responses for each variable.

Table 2. Characteristics of the student sample (n = 233)

Variable	Category	n (%)
School	Van Ban No. 2 High School / Bac Ha Ethnic Boarding School / Lao Cai High School / Van Ban No. 1 High School / Van Ban Ethnic Boarding School / Ly Tu Tan High School (Hanoi)	71 / 60 / 45 / 26 / 23 / 2
Grade	10 / 11 / 12	1 (0.4) / 76 (33.6) / 149 (65.9)
Gender	Male / Female / Other, not stated	104 (44.6) / 124 (53.2) / 5 (2.1)
Physics achievement	Not achieved / Achieved / Good / Very good	7 (3.0) / 55 (23.6) / 139 (59.7) / 32 (13.7)
Orientation towards Physics	Taking Physics in exams / Compulsory subject only / Interested in Physics / Find Physics difficult	72 (30.9) / 79 (33.9) / 47 (20.2) / 35 (15.0)
Digital devices at home	None / Shared / Own device	14 (6.1) / 35 (15.2) / 181 (78.7)

School could not be identified for 6 students and grade for 7 students.

Most students were in Grades 11 and 12, so their self-ratings for the Dynamics chapter are retrospective. This characteristic is taken into account when interpreting the results in Section 4.5.

3.3. Instruments

The teacher questionnaire had six parts: A (general information, 10 items); B (assessment forms and frequency, including scale B1 with 14 items and questions B2 to B5); C (self-rated mastery scale C1 with 8 items and four recognition items C2); D (difficulty scale D1 with 26 items in five groups, D1a to D1e); E (scale E1 on limitations of current tools, with 14 items in four groups); and G (needs and conditions for adopting the toolkit, including scale G2 with 12 items). The student questionnaire had five parts: A (general information); B (scale B1 with 12 items on assessment forms experienced, questions B2 and B3); C (perception scale C1 with 15 items in three groups); D (self-rating scale D1 with 17 items in three groups corresponding to the three competence components, questions D2 and D3); and E (preferred assessment forms and suggestions). All scales used 5-point Likert responses. Table 3 presents the reliability coefficients computed in JASP.

Table 3. Reliability of the scales

Form	Scale	Items	Cronbach's α	McDonald's ω [95% CI]
Teacher	B1. Frequency of assessment forms used	14	0.951	0.953 [0.930, 0.976]
Teacher	C1. Mastery of the competence structure	8	0.948	0.949 [0.923, 0.975]
Teacher	D1. Difficulties in assessing physics competence	26	0.983	Not estimable ²
Teacher	E1. Limitations of current tools	14	0.948	0.949 [0.924, 0.974]
Teacher	G2. Needs for toolkit components	12	0.954	0.955 [0.933, 0.978]
Student	B1. Assessment forms experienced	12	0.915	0.916 [0.900, 0.932]
Student	C1. Perceptions of assessment	15	0.913	0.914 [0.898, 0.930]
Student	D1. Self-rated ability in Dynamics	17	0.956	0.956 [0.948, 0.964]

² The CFA model used to estimate ω did not converge with 34 teachers.

Alpha coefficients ranged from 0.913 to 0.983. According to Taber (2018), a high α does not imply unidimensionality and may reflect redundancy among items, particularly for scale D1 with 26 items. Student scale C1 contains both positively worded items (C1.1, C1.7 to C1.10) and items describing limitations; its α was computed without reverse coding. C1 results are therefore reported item by item, without an overall scale mean.

3.4. Data analysis

Data were cleaned, coded and analysed in JASP. Indicators included frequencies, percentages, means (M) and standard deviations (SD). Five-point means were interpreted in five intervals: 1.00 to 1.80; 1.81 to 2.60; 2.61 to 3.40; 3.41 to 4.20; and 4.21 to 5.00. Subscale scores were computed as item means when at least 75% of items were answered. Multiple-choice questions were coded 0/1 for each option; some responses exceeded the permitted number of choices (teachers: 10 responses for B5, 7 for G3 and 8 for G5; students: 59 for E1). These choices were retained, and percentages were computed on the number of respondents. Comparisons between teachers and students were based on means of item pairs with corresponding content and were interpreted descriptively, because the two groups reported on different aspects of the same activity (teachers: frequency of use; students: frequency experienced) and the samples were not matched by class. Open-ended responses were used to illustrate quantitative results, with respondent codes.

4. RESULTS

4.1. Assessment forms and frequency

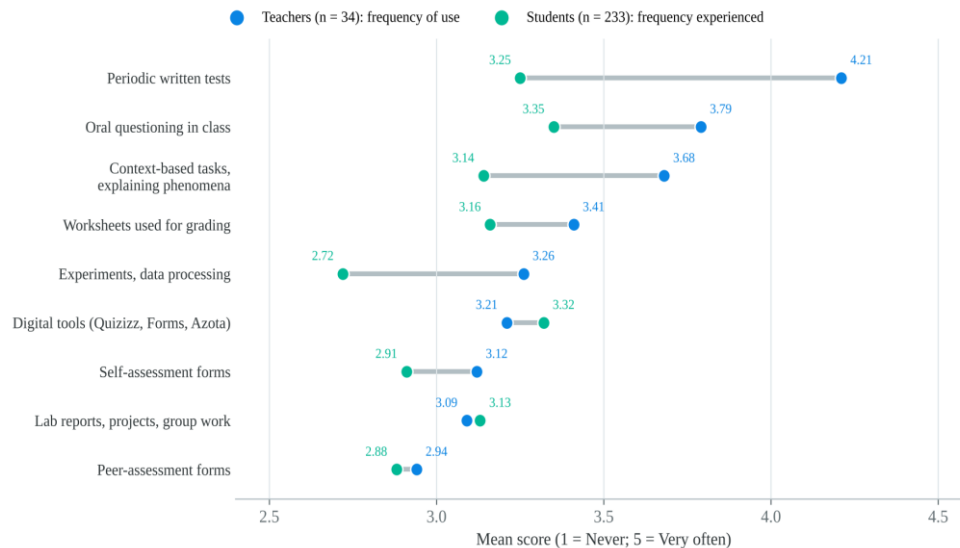
Table 4 shows that teachers used paper-based assessment tools ($M = 3.89$, $SD = 0.66$) more often than process-oriented tools ($M = 3.20$, $SD = 0.83$). The three most frequent forms were multiple-choice items, periodic tests based on a blueprint, and quantitative constructed-response problems (M from 4.09 to 4.21). Rubrics ($M = 2.88$) and peer-assessment forms ($M = 2.94$) had the lowest scores. Regarding assessment practice, 32.4% of teachers reported that at least 50% of continuous assessment scores were accompanied by comments describing competence indicators; 64.7% balanced process-oriented and summative assessment; and 35.3% reported that at least 30% of test items were set in real-life contexts or used experimental data. The most common sources of tools were textbooks, workbooks and teacher guides (64.7%) and the Internet or teacher groups on social media (52.9%); 35.3% of teachers wrote their own tools entirely, and 14.7% used toolkits from provincial or district education departments.

Table 4. Frequency of assessment forms used by teachers (n = 34)

Code	Assessment form	M	SD
Paper-based assessment tools: M = 3.89, SD = 0.66			
B1.1	Multiple-choice items	4.21	0.77
B1.3	Periodic tests based on a four-level blueprint	4.21	0.73
B1.2	Quantitative constructed-response problems	4.09	0.75
B1.4	Tasks set in real-life contexts	3.68	0.84
B1.5	Tasks on processing experimental data, reading and drawing graphs	3.26	0.79
Process-oriented assessment tools: M = 3.20, SD = 0.83			
B1.13	Oral questioning in class	3.79	0.95
B1.8	Worksheets used as a basis for comments or grading	3.41	0.96
B1.14	Digital assessment tools	3.21	0.85
B1.6	Observation checklists of student behaviour	3.18	1.11
B1.10	Learning portfolios	3.15	1.05
B1.11	Student self-assessment forms	3.12	0.98
B1.9	Lab reports, project products	3.09	0.84
B1.12	Peer-assessment forms	2.94	0.98
B1.7	Rubrics (criteria and performance levels)	2.88	0.95

Scale: 1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Often; 5 = Very often.

Figure 2 places teachers' means beside students' means for nine item pairs with corresponding content. The largest gaps appeared for periodic written tests (teachers 4.21, students 3.25), experiments and data processing (3.26 and 2.72), context-based tasks (3.68 and 3.14) and oral questioning (3.79 and 3.35). For process-oriented forms such as self-assessment, peer assessment, lab reports and projects, and digital tools, the two groups differed by no more than 0.21 and both were around the "sometimes" level. On the student B1 scale, "conducting experiments, measuring and recording data" had the lowest mean (M = 2.72).

**Figure 2. Teachers' and students' mean ratings of the frequency of corresponding assessment forms**

Note: Teacher items B1.3, B1.13, B1.4, B1.8, B1.5, B1.14, B1.11, B1.9 and B1.12 correspond to student items B1.3, B1.7, B1.4, B1.8, B1.5, B1.12, B1.10, B1.9 and B1.11, respectively. Between 230 and 233 students answered each item.

4.2. Teachers' understanding of the structure of physics competence

Teachers rated their mastery of competence-based assessment topics between 3.45 and 3.85 (overall M = 3.63, SD = 0.66), corresponding to basic to clear understanding (Table 5). The two highest-rated topics concerned regulations and test construction (Circular 22/2021; building test blueprints), while the two lowest concerned competence assessment directly (distinguishing competence assessment from knowledge and skill assessment; constructing performance levels in rubrics). On the recognition items,

47.1% of teachers correctly identified all three components of physics competence, and 41.2% selected “physics experimentation” as a separate component. On the two behaviour classification items, about one quarter of teachers confused the inquiry component with the application component (26.5% for each item). In total, 26.5% of teachers answered all four items correctly.

Table 5. Teachers’ understanding of the structure of physics competence (n = 34)

Code	Content	Result
C1. Self-rated mastery (1 = Not known; 5 = Mastered, able to guide colleagues)		
C1.7	Assessment regulations under Circular 22/2021	M = 3.85, SD = 0.78
C1.6	Building test blueprints from required learning outcomes	M = 3.76, SD = 0.87
C1.3	Specific indicators of each competence component	M = 3.68, SD = 0.73
C1.2	Three-component structure of physics competence	M = 3.65, SD = 0.65
C1.8	Checking reliability and validity of assessment tools	M = 3.65, SD = 0.77
C1.4	Turning competence indicators into measurable criteria	M = 3.50, SD = 0.71
C1.1	Concept of competence; distinction from knowledge and skill assessment	M = 3.48, SD = 0.80
C1.5	Constructing performance levels in rubrics	M = 3.45, SD = 0.87
C2. Recognition items: teachers answering correctly, n (%)		
C2.1	Identifying all three components of physics competence correctly	16 (47.1)
C2.2	Classifying the behaviour “proposing an experiment to test a prediction about friction” (key: inquiry into the natural world)	22 (64.7)
C2.3	Classifying the behaviour “explaining why a passenger lurches forward when a vehicle brakes suddenly” (key: application of knowledge and skills)	21 (61.8)
C2.4	Key difference between a rubric and a conventional marking scheme	30 (88.2)
	All four items correct	9 (26.5)

4.3. Teachers’ difficulties and limitations of current assessment tools

Teachers’ overall difficulty in assessing physics competence was at a moderate level ($M = 2.97$, $SD = 0.82$), and the five difficulty groups differed by no more than 0.19 (Table 6). The highest-rated items belonged to the organisational conditions and toolkit groups: the volume of records and paperwork ($M = 3.24$), lack of experimental equipment ($M = 3.21$), and the absence of a validated toolkit for the Dynamics chapter ($M = 3.12$). The item with the highest proportion of teachers rating it as “difficult” or above was “students and parents mainly care about scores” (47.1%). Among the 14 teachers who listed their three main difficulties, lack of experimental equipment (D1.15) was mentioned most often. One teacher wrote: “The curriculum needs adequate experiments to fully develop physics competence. But in reality, there is a lack of teaching equipment, and instructions for using the equipment are not specific” (GV13).

Table 6. Teachers’ difficulties in assessing physics competence (n = 34)

Difficulty group	M (SD)	Highest-rated item in the group: M (% rating difficult or above)
D1a. Teacher understanding and expertise (D1.1–D1.6)	2.87 (0.86)	D1.1 Unclear distinction between knowledge/skill and competence assessment; D1.6 Training too theoretical: 2.94
D1b. The toolkit itself (D1.7–D1.11)	3.02 (0.81)	D1.7 No validated toolkit for the chapter: 3.12 (38.2%)
D1c. Organisational conditions (D1.12–D1.17)	3.06 (0.87)	D1.16 Heavy records and paperwork: 3.24 (38.2%); D1.15 Lack of equipment: 3.21 (38.2%)
D1d. Processing and using results (D1.18–D1.22)	2.95 (0.98)	D1.19 Difficulty synthesising evidence into a competence judgement: 3.09

Difficulty group	M (SD)	Highest-rated item in the group: M (% rating difficult or above)
D1e. External context (D1.23–D1.26)	3.00 (0.95)	D1.24 Students and parents mainly care about scores: 3.12 (47.1%)
Overall (D1.1–D1.26)	2.97 (0.82)	Lowest-rated items: D1.2, D1.4 (2.79)

Scale: 1 = Not difficult; 2 = Slightly difficult; 3 = Moderately difficult; 4 = Difficult; 5 = Very difficult.

Teachers' views on the limitations of current assessment tools ranged from neutral to agreement ($M = 3.25$, $SD = 0.67$), and the three groups of statements had similar means (3.18 to 3.28) (Table 7). The statement with the highest agreement was "current tests mainly measure memorisation of formulas and problem-solving skills" (52.9%); 47.1% of teachers agreed that traditional tools are no longer sufficient for the requirements of the 2018 curriculum. One teacher's open response described this limitation in detail: "paper-based experiment" items with idealised data tables "only test data processing and cannot assess the ability to design a procedure, set up apparatus, or identify sources of error" (GV21).

Table 7. Teachers' views on the limitations of current assessment tools (n = 34)

Group, item	M (SD)	% agree
E1a. Range of competence measured: M = 3.28 (0.75)		
E1.1 Tests mainly measure memorisation of formulas and problem-solving skills	3.35 (1.01)	52.9
E1.4 Tools cannot assess attitudes, collaboration and responsibility in group work	3.32 (0.88)	47.1
E1.3 Paper-based tests cannot measure the inquiry component	3.09 (0.79)	35.3
E1b. Informational value of results: M = 3.18 (0.76)		
E1.9 Feedback is mainly a score or general comment, without specific guidance	3.35 (0.73)	44.1
E1.6 Scores do not show which competence component is weak	3.15 (0.89)	38.2
E1c. Backwash effects on teaching and learning: M = 3.26 (0.73)		
E1.11 Current assessment shifts teaching towards drilling problem-solving techniques	3.41 (0.93)	44.1
E1.10 Students obtain high scores by practising familiar problem types without understanding	3.35 (0.88)	47.1
E1.13 Students do not know the assessment criteria in advance	3.12 (0.81)	32.4
E1d. Overall conclusion		
E1.14 Traditional tools are no longer sufficient for the 2018 curriculum	3.38 (0.74)	47.1

Scale: 1 = Strongly disagree ... 5 = Strongly agree; % agree includes levels 4 and 5. Representative items are shown; group means are computed on all items in the group.

4.4. Students' perceptions of assessment

Regarding opportunities for experimentation, 31.7% of students reported no experiments in a typical month, 54.8% reported one or two, and 13.4% reported three or more. When tests were returned, 39.6% of students received only a score, 39.6% received a score with errors identified and guidance for correction, and 23.5% received comments on their application or practical skills ($n = 230$). Table 8 shows that most students viewed scores and feedback positively: 67.0% agreed that scores reflect their ability, and 72.1% agreed that teachers' comments help them know how to improve. Agreement was lower for knowing the marking criteria in advance (37.3%). In terms of preferences, 65.2% wanted specific strengths and weaknesses to be identified and 51.9% wanted to be assessed throughout the learning process. One student suggested: "I hope teachers announce the marking criteria clearly so that students can prepare better" (HS113).

Table 8. Students' perceptions of assessment in Physics (n = 233)

Item	M (SD)	% agree
C1a. Scores and actual ability		
C1.1 My current scores reflect my ability to understand and apply Physics	3.66 (0.88)	67.0
C1.2 I can score highly by memorising formulas and problem types without understanding	3.13 (1.06)	45.1
C1.3 I understand the lesson but score low for not reproducing textbook wording	3.03 (1.03)	38.2
C1.4 I struggle and lose marks with unfamiliar real-life situations	3.33 (0.95)	49.8
C1.5 Paper-based tests do not measure my experimental skills	3.22 (0.96)	40.8
C1.6 My group work is not assessed accurately and fairly	2.94 (1.00)	32.2
C1b. Information received after assessment		
C1.7 After each test, I know which skills I am weak in	3.68 (0.87)	67.8
C1.8 Teachers' comments help me know what to do to improve	3.80 (0.87)	72.1
C1.9 I know in advance the criteria teachers use for marking	3.10 (1.00)	37.3
C1.10 Self- and peer assessment help me understand lessons more deeply	3.58 (0.89)	62.2
C1c. Feelings and preferences		
C1.11 I feel stressed before Physics tests	3.28 (0.97)	45.1
C1.12 I prefer real-life problems to purely computational exercises	3.28 (0.99)	46.4
C1.13 I want to be assessed throughout the learning process	3.49 (0.93)	51.9
C1.14 I want specific strengths and weaknesses rather than only a score	3.67 (0.95)	65.2
C1.15 Dynamics has many applications, but tests do not reflect them	3.34 (0.93)	47.6

Scale: 1 = Strongly disagree; 3 = No opinion; 5 = Strongly agree; % agree includes levels 4 and 5.

4.5. Students' self-ratings across the three competence components

Students' self-ratings for the three task groups were similar: understanding and presenting knowledge ($M = 3.31$, $SD = 0.73$), inquiry through experiments ($M = 3.31$, $SD = 0.70$), and application to real life ($M = 3.37$, $SD = 0.73$), corresponding to "can do but still hesitant" (Table 9). Within the experimental group, procedural skills such as recording and tabulating data ($M = 3.50$) and reading measuring instruments ($M = 3.43$) were rated higher than skills related to experimental reasoning, such as proposing predictions ($M = 3.17$) and plotting graphs from data ($M = 3.22$). Self-ratings increased with achievement level (from 2.64 for "not achieved" to 3.70 for "very good") and were highest among students taking Physics in examinations (3.58). Students with no experiments in a month rated experimental inquiry lower than those with one or two experiments (3.09 vs. 3.41). Asked which task group was the most difficult, 46.9% of students chose application to real life; asked which group was assessed least often, 49.1% chose experimental inquiry (Table 10).

Table 9. Students' self-rated ability to perform tasks in the Dynamics chapter

Student group	n	Knowledge (D1.1–D1.6)	Experiments (D1.7–D1.13)	Application (D1.14–D1.17)	Overall
Whole sample	233	3.31 (0.73)	3.31 (0.70)	3.37 (0.73)	3.32 (0.67)
By Physics achievement					
Not achieved	7	2.50 (0.96)	2.71 (1.11)	2.71 (1.11)	2.64 (0.99)
Achieved	55	3.12 (0.63)	3.18 (0.62)	3.24 (0.67)	3.16 (0.60)
Good	139	3.33 (0.68)	3.31 (0.68)	3.38 (0.72)	3.33 (0.63)
Very good	32	3.73 (0.84)	3.68 (0.71)	3.68 (0.68)	3.70 (0.65)
By orientation towards Physics					
Find Physics difficult	35	3.00 (0.57)	3.01 (0.65)	3.02 (0.73)	3.00 (0.60)

Student group	n	Knowledge (D1.1–D1.6)	Experiments (D1.7–D1.13)	Application (D1.14–D1.17)	Overall
Compulsory subject only	79	3.18 (0.75)	3.24 (0.72)	3.26 (0.76)	3.23 (0.66)
Interested in Physics	47	3.33 (0.75)	3.27 (0.77)	3.43 (0.77)	3.33 (0.73)
Taking Physics in exams	72	3.58 (0.70)	3.57 (0.59)	3.62 (0.58)	3.58 (0.56)
By number of experiments per month					
None	73	3.12 (0.73)	3.09 (0.71)	3.23 (0.81)	3.13 (0.67)
1–2	126	3.42 (0.69)	3.41 (0.69)	3.43 (0.70)	3.42 (0.66)
3 or more ¹	31	3.26 (0.83)	3.40 (0.67)	3.39 (0.65)	3.35 (0.64)

Values are M (SD). Scale: 1 = Cannot do; 2 = Can do with step-by-step guidance; 3 = Can do but still hesitant; 4 = Can do independently; 5 = Can do well and guide peers. ¹ Combines 3–4 times (n = 24) and more than 4 times (n = 7).

4.6. Needs for a toolkit and students' preferences

Of the teachers, 82.4% considered a chapter-specific toolkit for assessing physics competence in Dynamics necessary or very necessary, and 82.3% were ready or very ready to use it (Table 10). Needs for toolkit components were at the “needed” level (M = 3.90, SD = 0.58); the three highest-rated components were sample tests with blueprints and specifications (M = 4.06), context-based tasks with marking guides (M = 4.03), and editable electronic versions (M = 4.00). The most frequently selected condition was training on the Dynamics chapter itself (82.4%), and the most frequently cited barrier was time for preparation and record-keeping (67.6%). One teacher suggested that each continuous assessment tool “should focus on only 3–4 core behavioural indicators that are easy to observe” (GV21). Among students, multiple-choice tests were the most preferred form (82.7%), followed by tasks explaining real-life phenomena (48.5%); experiments with reports were chosen by 32.9% of students.

Table 10. Teachers' needs and conditions for adopting the toolkit, and students' preferred assessment forms

Item	Result, n (%) or M
Teachers (n = 34)	
G1. Necessity of the toolkit	Slightly necessary 1 (2.9); Neutral 5 (14.7); Necessary 21 (61.8); Very necessary 7 (20.6)
G2. Desired components (M)	Highest: sample tests with blueprints 4.06; context-based tasks with marking guides 4.03; editable electronic versions 4.00. Lowest: portfolio templates 3.76; self- and peer-assessment forms 3.79
G3. Required conditions (up to 3)	Training on the Dynamics chapter 28 (82.4); additional equipment 18 (52.9); software for compiling results 15 (44.1); ready-made templates 12 (35.3); reduced paperwork 12 (35.3)
G4. Readiness to adopt	Slightly ready 2 (5.9); Undecided 4 (11.8); Ready 23 (67.6); Very ready 5 (14.7)
G5. Barriers (up to 2)	Preparation and record-keeping time 23 (67.6); lack of equipment 17 (50.0); curriculum pace and examination pressure 11 (32.4); limited confidence with rubrics and observation forms 10 (29.4)
G6. Willingness to give feedback or teach trial lessons (n = 33)	Yes 15 (45.5); Undecided 15 (45.5); No 3 (9.1)
Students	
D2. Most difficult task group (n = 226)	Knowledge 56 (24.8); Experiments 64 (28.3); Application 106 (46.9)
D3. Least assessed task group (n = 222)	Knowledge 40 (18.0); Experiments 109 (49.1); Application 73 (32.9)
E1. Preferred assessment forms (up to 3; n = 231)	Multiple-choice 191 (82.7); explaining real-life phenomena 112 (48.5); group projects 93 (40.3); observation of the learning process 93 (40.3); experiments and reports 76 (32.9); self- and peer assessment 54 (23.4); portfolios 43 (18.6)

5. DISCUSSION

The results indicate that assessment in the Dynamics chapter relies mainly on written work, while rubrics and peer assessment are used least. Compared with students' reports, the largest gaps were found for written tests, context-based tasks and experiments, whereas process-oriented forms had similar means in both groups. These gaps may arise because teachers and students reported on different aspects, because students did not recognise some activities as assessment, or because of social desirability in teachers' self-reports; the available data do not allow these explanations to be separated. With 39.6% of students receiving only a score and 37.3% knowing the marking criteria in advance, the strategies of clarifying criteria and providing feedback that moves learners forward (Wiliam, 2011; Hattie & Timperley, 2007) were not consistently enacted. This result is consistent with reported limitations of comprehensive assessment tools at the upper secondary level (Duong et al., 2024).

Teachers' self-rated understanding of the competence structure ($M = 3.63$) was higher than the understanding shown on the recognition items, where 47.1% identified the three components correctly and about one quarter confused the inquiry and application components. In the framework of Xu and Brown (2016), the knowledge base of assessment is the foundational layer of assessment literacy in practice; when this foundation is not secure, turning indicators into criteria and constructing performance levels in rubrics, the two topics with the lowest self-ratings, are likely to be hindered. The result is consistent with the gap between positive perceptions and proficiency reported among primary teachers (Vũ et al., 2024). However, four recognition items do not constitute a full measure of teachers' understanding, so this interpretation requires further verification.

Inquiry into the natural world through experiments is where data from both sides converge: students had few opportunities to conduct experiments, rated experiment-based assessment lowest, and named experimental inquiry as the least assessed group; teachers cited the lack of equipment among difficulties, conditions and barriers. In students' self-ratings, procedural skills were rated higher than proposing predictions and processing graphs, which is consistent with the observation by Abrahams and Millar (2008) that school practical work tends to focus on manipulation more than on reasoning with data. The comment by GV21 on "paper-based experiment" items with idealised data suggests one adjustment: using experimental data that contain measurement error in written tests when hands-on experiments cannot be organised. Although students considered application to real life the most difficult group, their self-ratings for this group were not lower than for the other two groups; this suggests that students may judge difficulty from their experience with tests rather than from their ability to perform the tasks.

Regarding feasibility, more than 80% of teachers considered the toolkit necessary and were ready to adopt it, but 45.5% were still undecided about taking part in trial teaching, and record-keeping time was the most frequently cited barrier. Class sizes of 41 to 45 students, reported by 67.6% of teachers, also limit the observation of individual students. Students' preference for multiple-choice tests (82.7%) may be linked to the structure of the national graduation examination, consistent with reports on the pressure of standardised examinations in physics teaching (T. K. C. Nguyễn et al., 2026). The toolkit could therefore combine multiple-choice items set in real-life contexts or based on experimental data (T. H. H. Nguyễn et al., 2022) with short rubrics and observation forms focusing on a few indicators, and provide editable and digital versions to reduce the time needed to compile results. Evidence on the effects of rubrics on performance and self-regulation (Panadero et al., 2023) supports including rubrics in the toolkit, together with activities that help students use the criteria in self-assessment.

6. CONCLUSION, LIMITATIONS AND RECOMMENDATIONS

In the surveyed sample, assessment of physics competence in the Dynamics chapter relied mainly on written work; process-oriented assessment, rubrics and peer assessment were used sometimes or less often. Teachers rated their understanding of the competence structure as fairly good, but the proportion correctly recognising the competence components did not correspond to these ratings. Teachers' difficulties were at a moderate level and concerned records, equipment and the lack of a validated toolkit. Students had limited opportunities to conduct experiments and to be assessed on experimental tasks, and they wanted to know the criteria and receive specific feedback. Teachers reported high levels of need and readiness for a new toolkit, provided that training is conducted on the chapter itself and that record-keeping time is reduced.

The study has several limitations: convenience sampling concentrated in Lao Cai province and a small number of teachers; most students were in Grades 11 and 12, so data on the Dynamics chapter are retrospective; the data are self-reported and teachers and students were not matched by class; and the analysis was descriptive, with the scale structure not yet tested by factor analysis.

Based on these results, the toolkit for the Dynamics chapter should prioritise: content-specific behavioural indicators with examples classified by the three competence components; experimental tasks with concise rubrics and alternatives when equipment is lacking; sample tests with blueprints organised by competence component; and criterion-referenced feedback templates with digital versions. Further studies should expand the sample, survey Grade 10 students immediately after the chapter, add classroom observation and analysis of test papers, test the structure of the scales, and conduct a teaching experiment with the toolkit.

REFERENCES

1. Abrahams, I., & Millar, R. (2008). Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30(14), 1945–1969. <https://doi.org/10.1080/09500690701749305>
2. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
3. DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248–266. <https://doi.org/10.1080/10627197.2016.1236677>
4. Dương, T. T. H., Phạm, Q. M., Vũ, T. A., & Bùi, T. N. (2024). Thực trạng đánh giá năng lực và phẩm chất học sinh trung học phổ thông theo Chương trình giáo dục phổ thông 2018 [Current status of assessing upper secondary students' competencies and qualities under the 2018 General Education Curriculum]. *Vietnam Journal of Education*, 24(Special Issue 13), 264–270. <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/3462>
5. Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. https://www.researchgate.net/profile/Nurhanis-Roslan-2/post/Im_searching_for_a_short_article_that_explains_the_value_of_feedback_feed_forward/attachment/59d64f6279197b80779a888a/AS%3A497433086234625%401495608697734/download/power-feedback.pdf
6. Lê, T. M. L., & Võ, N. T. A. (2022). Năng lực dạy học thí nghiệm của sinh viên sư phạm: Nghiên cứu thực trạng tại Trường Đại học Sư phạm Thành phố Hồ Chí Minh [Pre-service teachers' competence in teaching with experiments: A study at Ho Chi Minh City University of Education]. *Vietnam Journal of Education*, 22(20), 60–64. <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/download/574/335/390>
7. Ministry of Education and Training. (2018). *Chương trình giáo dục phổ thông môn Vật lý* [General education curriculum for Physics] (Issued with Circular No. 32/2018/TT-BGDĐT of 26 December 2018).
8. Ministry of Education and Training. (2021). *Thông tư số 22/2021/TT-BGDĐT quy định về đánh giá học sinh trung học cơ sở và học sinh trung học phổ thông* [Circular No. 22/2021/TT-BGDĐT on the assessment of lower and upper secondary students].
9. Nguyễn, T. H. H., Trần, V. N., & Nguyễn, T. N. (2022). Xây dựng bài tập phát triển năng lực vật lý cho học sinh trong dạy học chủ đề “Khí lý tưởng” (Vật lý 12) [Designing exercises to develop students' physics competence in teaching the topic “Ideal gas” (Physics 12)]. *Vietnam Journal of Education*, 22(14), 18–23. <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/download/476/251/291>
10. Nguyễn, T. K. C., Huỳnh, T. P. T., & Đào, X. H. (2026). Dạy học Vật lý theo định hướng phát triển năng lực trong bối cảnh chuẩn hóa thi cử: Đề xuất quy trình và kết quả thực nghiệm [Competence-oriented physics teaching in the context of standardised examinations: A proposed procedure and experimental results]. *Vietnam Journal of Education*, 26(Special Issue 5), 223–230. <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/5373>
11. Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analytic review. *Educational Psychology Review*, 35(4), Article 113. <https://doi.org/10.1007/s10648-023-09823-4>
12. Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
13. Vũ, P. L., Vũ, D. N., & Lê, T. H. (2024). Thực trạng nhận thức và thực hành đánh giá học sinh của giáo viên tiểu học theo tiếp cận năng lực [Primary teachers' awareness and practice of competence-based student assessment]. *Vietnam Journal of Education*, 24(2), 44–50. <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/download/1369/699/966>
14. Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://discovery.ucl.ac.uk/id/eprint/1507217/>
15. Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>