Jurnal Penelitian Pembelajaran Fisika
Vol. 16 Issue 3 – July 2025, p301-309
p-ISSN 2086-2407, e-ISSN 2549-886X
Available Online at http://journal2.upgris.ac.id/index.php/JP2F
DOI: 10.26877/jp2f.v16i3.2103

Development of an Assessment Instrument for Conceptual Understanding, Collaboration Skills, and Learning Motivation on Gas Kinetic Theory

Siti Rahmatia[1,2], Yusman Wiyatmo[1] and Edi Istiyono[1]
[1]Departement of Physics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

[2]Email: sitirahmatia.2024@student.uny.ac.id

Received: 24 June 2025. Accepted: 25 July 2025. Published: 31 July 2025.

Abstract. This study aims to develop an assessment instrument on the Kinetic Theory of Gases (Boyle's Law) to measure the conceptual understanding, collaboration skills, and learning motivation of high school students in the context of the Independent Curriculum. The instrument developed consists of 10 multiple-choice questions to measure conceptual understanding and 10 questionnaire statements for collaboration skills and learning motivation. The development process follows systematic stages, including planning, expert validation, and empirical trials. Content validity was analyzed using the Aiken V index, and it was found that all items were in the high validity category. Empirical validation was carried out using the Rasch model through Winsteps software, which showed that most items were in the statistical range of feasibility (Infit MNSQ: 0.80–1.15 for conceptual understanding, 0.85–1.13 for collaboration, and 0.82–1.12 for motivation). The reliability coefficient ranged from 0.74 to 0.85, indicating high internal consistency. The distribution of the level of difficulty of conceptual understanding questions includes easy (30%), moderate (40%), and complex (30%), thus reflecting balance. As many as 9 out of 10 conceptual understanding questions were declared valid and feasible, while all items in the collaboration and motivation questionnaire were maintained. These results indicate that the developed instrument is valid and reliable as a holistic assessment tool in physics learning, supports the formative assessment that aligns with 21st-century competencies, and strengthens collaborative character and lifelong learning motivation.

Keywords: assessment instrument, boyle's law, concept understanding, collaboration skills, learning motivation

1. Introduction

The advancement of education in the modern era necessitates a comprehensive evaluation of learning that is not solely confined to the cognitive domain but also encompasses social and affective skills. A critical dimension in science education, particularly physics, is achieving deep conceptual understanding, which forms the bedrock for developing students' scientific reasoning [1–3]. Consequently, valid and reliable assessment instruments are indispensable for measuring learning outcomes objectively and fairly.

The National Curriculum directs the learning approach toward fostering 21st-century competencies, emphasizing critical thinking, collaboration, communication, and creativity (the 4Cs) [4–6]. Therefore, assessment instruments must be capable of measuring not only conceptual understanding but also students' collaboration skills and learning motivation. These three facets are intricately linked and vital for supporting successful, meaningful learning.

One topic within high school physics that demands robust conceptual comprehension is the Kinetic Theory of Gases, specifically Boyle's Law. This material involves theoretical concepts, yet is deeply intertwined with experimentation and practical application. This makes it an ideal context for evaluation using instruments grounded in both theory and practice [7–9]. The Boyle's Law practicum offers an excellent opportunity to measure these three core aspects concurrently.

A high-quality assessment instrument requires development based on a structured scientific approach, such as the instrument development theory proposed by Istiyono, which involves planning, tryout, and validation phases [10]. The validation process includes content validity testing using the Aiken's V index and empirical testing via the Rasch Model, which is proven effective for quantitatively evaluating the quality of test items [11,12]. This rigorous methodology guarantees that the developed instrument is not only theoretically sound but also tailored to the students' characteristics.

The measurement of collaboration skills is essential in contemporary education, as collaboration stands as one of the primary competencies graduates must possess in the digital age [13–15]. Collaboration assessment cannot be achieved solely through written tests; it necessitates valid questionnaires capable of accurately representing teamwork ability, effective communication, and conflict resolution. Recent studies consistently show that students with stronger collaboration skills tend to demonstrate superior academic outcomes [16–18].

Learning motivation is similarly a major determinant of educational success. Based on the theory advanced by Sudjana and corroborated by contemporary studies, intrinsic motivation significantly increases student engagement in the learning process, thereby driving achievement [20]. Consequently, it is important for educators to be able to objectively measure students' motivation levels using valid and dependable instruments.

The integration of the three aspects measured in this study, conceptual understanding, collaboration skills, and learning motivation, reflects a holistic assessment approach to education. This aligns perfectly with the new paradigm of educational assessment, which underscores the development of students' whole competencies across cognitive, affective, and psychomotor domains [21–23]. Assessment tools based on the National Curriculum must be designed to accommodate this diversity of competencies.

This research aims to develop a Kinetic Theory of Gases (Boyle's Law) based assessment instrument that can measure these three pivotal aspects in high school physics learning. Through this approach of theory-based instrument development and empirical validation, the findings are expected to make a significant contribution to providing tools widely usable in physics education and research in Indonesia. Furthermore, this development seeks to support the reinforcement of collaborative character and a lifelong learning spirit.

Despite the widespread acknowledgment of the importance of measuring conceptual understanding, collaboration skills, and learning motivation, and the high suitability of Boyle's Law material for comprehensive evaluation, the literature review indicates a crucial instrumental gap. A single, fully integrated assessment instrument that has undergone strict Rasch Model validation and can simultaneously measure these three critical aspects within one specific high school topic is currently unavailable. The development of this instrument introduces methodological novelty by not merely compiling three separate measuring tools, but by designing them as a coherent, unified package to evaluate student performance within the context of the Boyle's Law practicum. Thus, this study explicitly fills the void for a holistic assessment instrument that aligns with the demands of 21st-Century Competencies and the National Curriculum, providing physics educators with a more accurate and relevant tool.

2. Methods

2.1. *Research Design*

This study was designed to develop an assessment package related to the Kinetic Theory of Gases, specifically Boyle's Law, with the goal of measuring three primary aspects: conceptual understanding, collaborative skills, and learning motivation among high school students. The development process for both the test and non-test instruments adhered to the guidelines proposed by Istiyono, which were subsequently modified into three core stages: planning, instrument tryout, and measurement implementation [10]. However, for the current study, the research only proceeded up to the instrument tryout phase. Although the construction of non-test instruments was integrated into the overall test instrument development scheme, necessary adjustments were made to account for the intrinsic characteristics of the non-test (non-cognitive) scales.

2.2. *Population and research sample*

The population for this study consisted of all students at SMA Negeri 9 Yogyakarta in the same grade level (the relevant class) who had completed the Kinetic Theory of Gases material, particularly the topic of Boyle's Law. From this population, a sample of 70 students was selected to participate in the instrument tryout. The sampling procedure employed a random sampling technique from two classes within the same grade level at SMA Negeri 9 Yogyakarta.

2.3. *Research Instruments*

The instruments developed comprised 10 multiple-choice questions to measure conceptual understanding, a 10 item questionnaire for collaborative skills, a 10 item questionnaire to assess learning motivation, and an expert validation sheet. The primary focus of the instrument development was the Boyle's Law topic within the Kinetic Theory of Gases material, which is taught in Grade XI of high school, referencing the national curriculum for the 2025 academic year.

2.4. *Data Analysis Techniques*

The data analysis in this research utilized the Aiken's V content validity index formula. The validity calculation, based on the approach developed by Aiken is detailed in Table 1.

Table 1. Aiken's v index categories [24].

Aiken's V Index	Validity
$0 < V \leq 0{,}4$	Low
$0{,}4 < V \leq 0{,}8$	Moderate
$0{,}8 < V \leq 1$	High

Data analysis in this study covered the testing of validity, reliability, and the item difficulty level of the developed instruments based on empirical data. The evaluation of item difficulty was grounded in the one-parameter logistic (1PL) Rasch Model, where the primary focus lies on the difficulty parameter of each test item, which represents the students' ability to complete the test.

3. Results and Discussion

3.1. *Instrument Design Stage*

This study constitutes a type of developmental research aimed at producing an assessment package to measure students' conceptual understanding, collaborative skills, and learning motivation in high school physics, specifically on the Kinetic Theory of Gases, with an emphasis on Boyle's Law. The compiled instruments include test items for conceptual understanding and questionnaires (non-test) for measuring collaboration skills and learning motivation.

The selection of core competencies to be measured referred to the National Curriculum, starting with an analysis of the Learning Outcomes (Capaian Pembelajaran/CP) and the Learning Objectives Flow (Alur Tujuan Pembelajaran/ATP). This analysis determined the relevant main material, which subsequently served as the basis for formulating the learning indicators. The material focus developed was the Kinetic Theory of Gases, with an emphasis on Boyle's Law for Grade XI students.

The assessment was designed as a guideline for constructing high-quality items that align with the learning objectives. Its construction took into account the learning outcomes, material scope, item format, and the measurement indicators for the three target aspects: conceptual understanding, collaborative skills, and learning motivation. A synthesis of the used indicators reveals that the aspects derived from Bloom's Taxonomy [25] are consistent with the learning outcomes in the national curriculum, as summarized in Table 2. The aspect of conceptual understanding refers to Bloom's Taxonomy, which includes: (a) knowledge, (b) comprehension, (c) application, (d) analysis, (e) synthesis, and (f) evaluation. However, in this research, the assessment was limited only to the analysis level of ability.

Indicators for collaborative skills were formulated based on the concepts developed by Greenstein, encompassing: (a) the ability to listen and convey opinions within the group, (b) demonstrating empathy and a compromising attitude, (c) prioritizing group interests, (d) acknowledging and utilizing the contributions of group members, (e) collectively creating new ideas, (f) active participation and responsibility for group tasks, (g) respecting the views of other members, and (h) the ability to resolve conflicts and participate actively in group discussions [14]. Furthermore, indicators from Read write think were also used as a reference, covering: (a) contribution, (b) time management, (c) problem-solving, (d) cooperation, and (e) problem-solving strategies [15].

For the learning motivation aspect, indicators were developed by referencing Sudjana's work, which includes eight indicators of student activity in learning: (1) involvement in learning tasks, (2) participation in problem-solving, (3) courage to ask questions when facing difficulties, (4) initiative to seek additional information, (5) activeness in group discussions, (6) the ability to evaluate one's own learning results, (7) practice in solving problems, and (8) application of acquired knowledge in problem-solving contexts [19]. Each indicator for the conceptual understanding aspect was designed to have two to three test items. The complete instrument is presented in Table 2.

Table 2. Indicator for the conceptual understanding.

No	Learning Outcome	Aspect	Item Number
1	Recalling	Knowledge	1,2,3
2	Understanding	Comprehension	4,5
3	Applying	Application	6,7,8
4	Analyzing	Analysis	9,10

Table 3. Indicator for the collaboration skills instrument.

No	Aspect	Indicator	Item Number
1	Cooperation	Works cooperatively with diverse group members to solve problems encountered within the group.	1,2
2	Responsibility	Takes the initiative to self-regulate within the group when performing collaborative tasks. Each group member is involved in carrying out the assigned tasks.	3,4
3	Communication	Able to discuss problems with peers until a consensus is reached. Pays full attention to what is being listened to.	5,6
4	Compromise	Efforts to reach an agreement on the problem being solved. Involvement of group members in conducting discussions to reach a joint decision..	7,8
5	Flexbility	Completes group tasks in a timely manner. Adapts to each group member to solve problems.	9,10

Table 4. Indicator for the learning motivation instrument.

No	Aspect	Indicator	Item Number
1	Active Involvement	Students participate in carrying out their learning tasks.	1
2	Perseverance	Students are willing to engage in problem-solving activities during learning.	2
3	Curiosity	Students are willing to ask peers or the teacher if they do not understand the material or encounter difficulties.	3,4
4	Autonomy	Students strive to find the necessary information to solve the problem they are facing.	5
5	Social Interaction	Students conduct group discussions according to the teacher's instructions.	6
6	Self Evaluation	Students are able to assess their own ability and the results they have achieved.	7
7	Problem Solving	Students practice solving exercises or problems.	8,9
8	Concept Application	Students use or apply what they have acquired in solving assigned tasks or problems.	10

Referring to the indicator presented in Tables 2, 3, and 4, each indicator was subsequently elaborated into specific test items and questionnaire statements that form the final assessment instrument. The instrument consists of 10 multiple-choice questions to measure the conceptual understanding aspect, 10 questionnaire statements to assess collaboration skills, and 10 questionnaire statements to evaluate students' learning motivation.

Scoring for the multiple-choice questions utilized a dichotomous scoring system, where correct answers were assigned a score of 1 and incorrect answers a score of 0. Meanwhile, the questionnaires were assessed using a 4 point Likert scale, which allows educators to obtain measurable scores from students corresponding to their level of response.

To ensure the suitability and quality of the developed instrument, validation was carried out by two validators: one expert and one practitioner. The expert validator was an academic, specifically a lecturer from the Physics Education Study Program at the Faculty of Mathematics and Natural Sciences (FMIPA) UNY, while the practitioner validator was a physics teacher at a school relevant to the instrument's intended context. Both validators provided ratings as well as constructive feedback on every item constructed. The results of this validation were analyzed using the Aiken's V validity index, with the summary of the analysis presented in Tables 5, 6, and 7.

Table 5. Recapitulation of aiken's v results for conceptual understanding.

Item Number	V Index Score	Category
2,4,7,8,9	1	High
1,3,5,6,10	0,92	High

Table 6. Recapitulation of aiken's v results for collaboration skills.

Item Number	V Index Score	Category
2,3,5,6	0,98	High
1,4,7,8,9,10	0,83	High

Table 7. Recapitulation of aiken's v results for learning motivation.

Item Number	V Index Score	Category
1,2,4,7,8	0,92	High
3,5,6,9,10	0,82	High

Based on the data presented in Tables 5, 6, and 7, the Aiken's V index scores for all items conceptual understanding questions (items 1 through 10) and the collaboration skills and learning motivation questionnaire statements (items 1 through 10)—fall into the High Validity category. This outcome confirms that the developed assessment instrument is deemed content valid by the validators.

Nevertheless, the validators provided several essential notes and suggestions for instrument refinement. These included necessary improvements such as selecting more precise diction, adjusting the layout format of the questions and response options, and modifying the structure of interrogative sentences (questions) into descriptive statements using phrases like "is defined as..." or "which is...". After the instruments underwent revision based on this feedback, they were declared suitable for use in the subsequent tryout phase.

3.2. *Instrument Tryout Phase*

This research involved Grade XI students at SMA Negeri 9 Yogyakarta who had studied the Kinetic Theory of Gases material, specifically the sub-topic of Boyle's Law under the national curriculum. The selection of tryout subjects was performed randomly using a cluster random sampling method, where classes were randomly chosen from the population of students who had completed the material. The total number of tryout participants was 70 students, consisting of two classes with 35 students in each class.

The data gathered from the instrument tryout were subsequently analyzed using the Winsteps software. This analysis yielded essential psychometric information regarding the instrument's validity and reliability, as well as the item difficulty level for the conceptual understanding instrument.

Table 8. Winsteps analysis results of the assessment instrument.

Indicator	Analysis Results		
	Conceptual Understanding	Collaboration	Motivation
Infit MNSQ Range	0.80 – 1.15	0.85 – 1.13	0.82 – 1.12
Item Reliability	0.74	0.85	0.80
No. of Items in –2.0 to +2.0 Range	10	10	10
No. of Fit Items	9	10	10
No. of Misfit Items	1	0	0

3.3. *Instrument Validity*

Based on the results presented in Table 8, instrument validity was evaluated by assessing the item fit to the Rasch Model using Winsteps software. The analysis included items for conceptual understanding, the collaboration skills questionnaire, and the learning motivation questionnaire. An item is considered to fit the model if its Infit Mean Square (MNSQ) value falls within the acceptable range of 0.77 to 1.30 [26]. The analysis of the conceptual understanding instrument showed that most items were within the range of 0.80–1.15, with the exception of one item (item number 3), which was classified as a misfit.

Further analysis of the collaboration skills questionnaire revealed that all items had Infit MNSQ values between 0.85 and 1.13, indicating that all items fit the model. Similar results were found for the learning motivation questionnaire, with Infit MNSQ values ranging from 0.82 to 1.12. Consequently, all items in both questionnaires met the criteria for fitting the Rasch Model.

Overall, nine out of the 10 items in the conceptual understanding test were deemed valid based on the Rasch Model, and all items in both the collaboration and motivation questionnaires satisfied the fit criteria. Therefore, the developed instrument is suitable for measuring the conceptual understanding, collaborative skills, and learning motivation of Grade XI students concerning the Kinetic Theory of Gases (Boyle's Law).

3.4. *Instrument Realibility*

The instrument reliability values, as shown in Table 8, were obtained from the Summary of sStatistics section of the Winsteps analysis. A higher reliability value signifies greater consistency of the instrument's items with the measurement model. According to the classification of reliability coefficients by Sukiman, the reliability ranges are categorized as: very low (0.00–0.19), low (0.20–0.39), moderate (0.40–0.59), high (0.60–0.79), and very high (0.80–1.00) [27].

The conceptual understanding instrument showed an item reliability of 0.74 and a person reliability (case estimates) of 0.72; both fall into the High category. In contrast, the collaboration skills instrument yielded an item reliability of 0.85 and a person reliability of 0.84, both classified as Very High. For the learning motivation instrument, the item reliability was 0.80 and the person reliability was 0.75, classifying them into the Very High and High categories, respectively.

These results collectively demonstrate that all three instrument types possess an adequate level of consistency, both in terms of the items themselves and the subjects being measured. Thus, the instrument is suitable for repeated use and can be trusted to produce stable results. Reliability is a critical aspect of instrument development as it reflects the consistency of the measurement under various conditions.

3.5. *Item Characteristics*

The difficulty level of each test item was determined based on the Winsteps analysis using the Rasch Model or the 1-Parameter Logistic (1PL) Model. The output in the "thresholds" section of the software provides the difficulty estimate for each item. According to [28], the ideal difficulty value lies within the range of −2 to +2 logits. Items with values below −2 are generally considered too easy, while values above +2.9 indicate that the item is excessively difficult.

Based on the analysis of the 10 items for conceptual understanding, the distribution of item difficulty showed that 30% of the items were classified as easy, 40% were in the moderate category, and 30% were classified as difficult. The detailed difficulty category for each item can be found in Table 9.

Table 9. Students' conceptual undersatnding level by item difficulty.

Category	Item Numbers	Count	Percentage (%)
Easy	8,10,3	3	30
Moderate	7,4,2,9	4	40
Difficult	5,1,6	3	30

The overall difficulty level of the conceptual items ranged from −0.17 to +1.90. Considering both item difficulty and validity criteria, 9 test items successfully met the criteria for fitting the Rasch Model. The validity testing results in this study strongly affirm the quality of the developed items. Overall, every item designed to measure conceptual understanding, collaboration skills, and learning motivation demonstrated a good fit to the Rasch Model. Furthermore, the instrument exhibited a high reliability index, ranging from 0.74 to 0.85.

This finding is consistent with prior studies that utilized Rasch analysis in physics instrument development. For example, a study by Rasyid reported the development of a conceptual physics test on waves with an impressive reliability (0.95) and a perfect proportion of fit items (100%) [29]. A similar consistency was reported by Riza Andriani in developing a fundamental physics concept test, where the Outfit MNSQ value was within the accepted range (0.7–1.33) and high reliability was observed (0.8–0.91) [30]. The congruence of these findings underscores that Rasch analysis is a robust method for validating item quality due to its superiority in separating the subject's ability variable from the item's difficulty level. The key distinction of this research is the expansion of focus beyond the cognitive domain to include the affective constructs of collaboration and motivation. This approach is aligned with the National Curriculum's demand for a holistic assessment of students' social and emotional characteristics.

Moreover, the content validity of the instrument was strengthened by obtaining high Aiken's V index scores, indicating a substantial coherence between the instrument's content and the intended competencies. This conclusion is supported by Daeng Naba, who emphasized the importance of synergy between expert validation (expert judgment) and Rasch analysis to produce sound and reliable physics assessments [31].

On the other hand, this research also identifies avenues for future development. Mi demonstrated that applying the Cognitive Diagnostic Assessment (CDA) approach in electrostatic diagnostic instruments can provide a more detailed and in-depth mapping of conceptual difficulties [32]. The CDA model can serve as a reference framework for refining this instrument (the Rara instrument), particularly for more specifically identifying and analyzing student misconceptions regarding Boyle's Law. Furthermore, the relevance of measuring non-cognitive aspects is supported by Panthumas, who successfully developed a 21st-Century skills scale with layered validation (EFA & CFA) and excellent reliability (α=0.92).

Therefore, the novel contribution of this research lies in its methodological and substantive integration: combining three essential constructs conceptual understanding, collaboration skills, and learning motivation—into a single assessment package, all rigorously validated using the Rasch Model. This Rasch-based integrative approach represents a forward step that is relatively uncommon in high school physics education measurement studies.

4. Conclusion

The development of the assessment instrument for the Kinetic Theory of Gases (Boyle's Law) to measure conceptual understanding, collaboration skills, and learning motivation among high school students was conducted through the stages of planning, validation, and instrument tryout. The instrument comprises 10 multiple-choice questions for conceptual understanding and 10-item questionnaires each for the collaboration and motivation aspects. Content validity using the Aiken's V index showed that all items fall into the High category, while the empirical test using the Rasch Model indicated that most test items and questionnaires were within the acceptable Infit MNSQ range, with the overall instrument reliability categorized as high to very high. The item difficulty level also exhibited a balanced distribution across easy, moderate, and difficult categories. Thus, the developed instrument is declared valid and reliable and is suitable for use as a holistic assessment tool in physics learning based on the National Curriculum.

Acknowledgments
The authors extend their gratitude to all those who provided assistance and support, particularly the physics teachers and the administration of SMAN 9 Yogyakarta.

References

[1] Syahrir S, Kusnadin K and Nurhayati N 2017 Analisis Kesulitan Pemahaman Konsep Dan Prinsip Materi Pokok Dimensi Tiga Siswa Kelas Xi Smk Keperawatan Yahya Bima *Prism. Sains J. Pengkaj. Ilmu dan Pembelajaran Mat. dan IPA IKIP Mataram* **1** 89

[2] Suswati L, Yuliati L and Fisika-Stkip Bima P 2015 Pengaruh Integrative Learning Terhadap Kemampuan Berpikir Kritis dan Penguasaan Konsep Fisika Siswa *J. Pendidik. Sains* **3** 49–57

[3] Prastiwi V D, Parno P and Wisodo H 2018 Identifikasi pemahaman konsep dan penalaran ilmiah siswa SMA pada materi fluida statis *Momentum Phys. Educ. J.* **2** 56–63

[4] Ikhlas N and Murniati A 2025 Efektifitas dan Peran Guru PAI Dalam Kurikulum Merdeka Belajar *Jurnal Intelek Insan Cendikia* **2** 180–93

[5] Arifah M N 2024 Penerapan Kurikulum Merdeka dalam Pembelajaran Matematika untuk Meningkatkan Kreativitas dan Kolaborasi Siswa dalam Menghadapi Tantangan Abad 21 *J. Pendidik. Fis. Indones.*

[6] Wahyudin D, Subkhan E, Malik A, Hakim M A, Sudiapermana E, LeliAlhapip M, Nur Rofika Ayu Shinta Amalia L S, Ali N B V and Krisna F N 2024 Kajian Akademik Kurikulum Merdeka *Kemendikbud* 1–143

[7] Cho E, Lee J and Paik S-H 2021 A Study of Pre-service Chemistry Teachers' Perceptions on NOS -Focusing on Boyle's Air Pump Experiment- *J. Korean Chem. Soc.* **65** 333–46

[8] Saepuzaman D, Utari S and Nugraha M G 2019 Development of basic physics experiment based on science process skills (SPS) to improve conceptual understanding of the preservice physics teachers on Boyle's law *Journal of Physics: Conference Series* **1280**

[9] Vertchenko L and Dickman A G 2021 Verifying Boyle's law in a teaching laboratory using strictly measurable quantities *Rev. Bras. Ensino Fis.* **34**

[10] Istiyono E 2020 Pengembangan instrumen penilaian dan analisis hasil belajar fisika dengan teori tes klasik dan modern *Yogyakarta UNY Press. L, I.(2019). Eval. Dalam Proses Pembelajaran. J. Manaj. Pendidik. Islam* **9** 478–92

[11] Ari P, Abd M and Bunyamin H 2024 Investigation of Scientific Reasoning Skills Survey Research on Static Fluid Topics Investigasi Penelitian Survei Keterampilan Penalaran Ilmiah pada Topik Fluida Statis *Kasuari : Physics Education Journal (KPEJ)* **7** 396–408

[12] Bungaran S, Rahmatan H, Ulfa A and Pada T 2025 Measurement of Psychometric Attributes Scientific Literacy Tests for Madrasah Tsanawiyah Students **11** 1180–6

[13] Musyaddad M, Sinaga F P and Oktavia S W 2024 Analisis Keterampilan Kolaborasi Peserta Didik Pada Mata Pelajaran Fisika Di Sman Titian Teras Kabupaten Muaro Jambi *Sci. Educ. J.* **1** 48–65

[14] Greenstein L M 2012 *Assessing 21st Century Skills: A Guide to Evaluating Mastery and Authentic Learning* (SAGE Publications)

[15] Read write think 2005 Collaborative Work Skills Rubric *Int. Read. Assoc.*

[16] Groeneveld W, Vennekens J and Aerts K 2020 Engaging Software Engineering Students in Grading

[17] Pulgar J, Ram D, Umanzor A and Candia C 2022 Long-term collaboration with strong friendship ties improves academic performance in remote and hybrid teaching modalities in high school physics 1–19

[18] Santoso S and Pramesti G 2025 Multivariate analysis on students' cognitive assessment, attitude, and skill evaluation in problem-based learning *Journals Math. Educ. Univ. Sebel. Maret* **8** 172–

84
[19] Sudjana N 1995 *Penilaian hasil proses belajar mengajar* (PT Remaja Rosdakarya)
[20] Araujo A A, Kalinowski M, Paixao M and Graziotin D 2025 Towards Emotionally Intelligent Software Engineers : Understanding Students ' Self-Perceptions After a Cooperative Learning Experience
[21] Hindriana A F, Abidin Z and Setiawati I 2024 Workshop of Assessment Development in Curriculum Merdeka Pelatihan Pengembangan Asesmen pada Kurikulum Merdeka **5**
[22] Minarti I B, Dewi L R and Kurniawati A 2023 Implementasi Asesmen Autentik Pembelajaran Biologi Pada Kurikulum Merdeka Di SMA Negeri Se-Kabupaten Batang **5**
[23] Rifai M, Mahmud N, Kasim A, Muzakkir and Makduani R 2024 The Merdeka Curriculum In Character Education 838–48
[24] Aiken L R 1985 Three coefficients for analyzing the reliability and validity of ratings. *Educ. Psychol. Meas.* **45** 131–42
[25] Bloom B S, Engelhart M D, Furst E J, Hill W H and Krathwohl D R 1956 *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain* (New York: David McKay Company)
[26] Adams R J and Khoo S-T 1996 Acer quest: The interactive test analysis system *Aust. Counc. Educ. Res.* 1–104
[27] Sukiman 2015 *Pengembangan media pembelajaran* (Pedagogia)
[28] Hambleton R K, Swaminathan H and Rogers H J 1991 *Fundamentals of Item Response Theory* (SAGE Publications Inc.)
[29] Rasyid F, Istiyono E and Gunawan C W 2025 Developing and analyzing items of a physics conceptual understanding test on wave topics for high school students using the Rasch Model *REID (Research Eval. Educ.* **11** 1–16
[30] Riza Andriani, Widya, Nurul Fadieny, Muttakin and Permana N D 2023 Development of Conceptual Understanding Student Tests to The Basic Physics Subject: a Rasch Model Analysis *Co-Catalyst J. Sci. Educ. Res. Theor.* **1** 43–54
[31] Daeng Naba S, Istiyono E, Kurniawan A and Adrianto N 2024 Development and Validation of Physics Multiple-Choice Tests on the Nature of Physics Using Rasch Modelling Analysis *J. Pendidik. MIPA* **25** 671–85
[32] Mi S, Ye J, Yan L and Bi H 2023 Development and validation of a conceptual survey instrument to evaluate senior high school students' understanding of electrostatics *Phys. Rev. Phys. Educ. Res.* **19** 10114