Rasch-Based Assessment of Collaborative Problem-Solving Skills in Sound Waves among Senior High School Students

Salsabila Khairun Nisa[1], Winny Liliawati[2*], Ridwan Efendi[2], Lari Andres Sanjaya[3]

[1]*Study of Magister of Physics Education, Universitas Pendidikan Indonesia, Bandung 40254, Indonesia.*
[2]*Department of Physics Education, Universitas Pendidikan Indonesia, Bandung 40254, Indonesia.*
[3]*Faculty of Educational Sciences and Technology, Universiti Teknologi Malaysia, Skudai, Johor Bahru, Johor, Malaysia.*

Email: winny@upi.edu*

Abstract

This study developed and validated a Rasch-based assessment instrument to measure collaborative problem-solving (CPS) skills in sound wave learning among 185 eleventh-grade senior high school students. The instrument consisted of 15 tasks embedded in authentic sound wave contexts and was developed based on the OECD/PISA 2015 CPS framework, which includes establishing shared understanding, taking appropriate action, and managing team organization. Each task was scored using a structured rubric covering collaborative information use and problem-solving outcomes, and responses were analyzed using a dichotomous Rasch model in Winsteps version 4.7.0 after collapsing scoring categories into a single performance indicator per item. The results indicated acceptable measurement quality, with person reliability of 0.68 and KR-20 of 0.72. Item fit statistics (0.89–1.15 infit MNSQ) showed that all items functioned adequately within the model. However, item difficulty estimates were highly clustered (−0.18 to 0.20 logits), resulting in very low item separation and item reliability values and indicating limited variation in item difficulty levels. Overall, the instrument showed acceptable person measurement properties and adequate item fit for classroom-based CPS assessment in physics education. However, the narrow range of item difficulty indicates the need for further development to improve measurement precision across different levels of CPS ability.

Keywords: collaborative problem-solving; Rasch model; sound waves; physics education; OECD/PISA 2015; educational assessment; senior high school students.

INTRODUCTION

Collaborative problem-solving (CPS) has become one of the essential competencies in 21st-century education because students are increasingly required to work collaboratively to solve complex problems in both academic and real-world contexts (Wei et al. 2025). In physics education, CPS plays an important role as students are frequently engaged in collaborative activities such as conducting experiments, analyzing data, discussing scientific concepts, and making collective decisions. These learning activities require not only cognitive problem-solving abilities but also communication, coordination, and teamwork skills (OECD 2017). Consequently, the assessment of CPS has become an important concern in physics education to ensure that students' collaborative competencies can be measured accurately and systematically (OECD 2017).

Measuring collaborative problem-solving (CPS) skills remains challenging because CPS involves not only cognitive abilities but also social interaction, communication, and teamwork competencies

that are difficult to assess using conventional assessments (OECD 2017). The PISA 2015 framework defines CPS as an individual's capacity to engage effectively in a process whereby two or more agents attempt to solve a problem by sharing understanding, knowledge, and efforts to achieve a common goal. PISA 2015 identifies three major CPS competencies: establishing and maintaining shared understanding, taking appropriate action to solve the problem, and establishing and maintaining team organization. These competencies demonstrate that CPS integrates cognitive, social, and communication skills as essential components of effective collaborative learning and problem-solving (OECD 2017).

Because collaborative problem-solving is a multidimensional construct involving cognitive, social, and communication competencies, the availability of valid and reliable assessment instruments is essential. Instrument validity is required to ensure that each item or indicator accurately measures the CPS construct, particularly in terms of establishing and maintaining shared understanding, taking appropriate action to solve problems, and establishing and maintaining team organization (Setyawarno et al. 2025). Without reliable instruments, assessment results may not accurately represent students' actual CPS abilities, thereby limiting the accuracy of interpretation and decision-making based on the results.

The Rasch model offers several advantages over factor-analysis-based approaches, including specific objectivity, parameter separation, and sufficiency, enabling more precise and independent estimation of item difficulty and respondent ability (Suryadi et al. 2020). In educational assessment, the Rasch model has become increasingly important because it provides objective and reliable measurement of students' abilities while simultaneously evaluating the quality of assessment items (Kim et al. 2023). However, limited studies have applied Rasch-based CPS assessments in classroom-based physics learning contexts.

In physics learning, particularly on the topic of sound waves, students commonly engage in collaborative experimental activities such as designing investigations, conducting measurements, interpreting experimental data, discussing observations, and drawing conclusions together. These activities require students to share ideas, coordinate roles, negotiate solutions, and make collective decisions, which directly reflect the three collaborative competencies defined in the OECD/PISA 2015 framework. Therefore, sound waves provide an appropriate and authentic context for assessing collaborative problem-solving skills in physics classrooms (Oktaviana & Putranta 2024).

Although CPS has been recognized as an essential twenty-first-century competency, most physics assessments in schools continue to prioritize individual cognitive achievement over students' collaborative competencies during problem-solving activities (Alanazi et al. 2024). Furthermore, there is still a shortage of standardized and empirically validated instruments capable of assessing CPS skills, particularly in physics learning contexts involving sound wave materials among senior high school students (Rojas et al. 2025; Badmus & Jita 2025). Previous studies have primarily focused on evaluating students' cognitive achievement, implementing collaborative learning strategies, or assessing general collaboration skills. However, relatively few studies have developed assessment instruments specifically designed to measure CPS competencies based on the OECD/PISA 2015 framework within physics learning contexts. In addition, limited research has examined the psychometric properties of CPS instruments using Rasch analysis in domain-specific topics such as sound waves. To the best of our knowledge, studies that integrate the OECD/PISA 2015 CPS framework with Rasch-based psychometric validation in senior high school sound wave learning remain scarce. Therefore, this study aims to develop and psychometrically validate a Rasch-based CPS assessment instrument integrated with the OECD/PISA 2015 framework for sound wave learning, providing a more objective and evidence-based approach to assessing students' collaborative problem-solving skills in physics education.

METHODS

Research Design

This study employed a quantitative descriptive design to examine the psychometric properties of a Collaborative Problem-Solving (CPS) assessment instrument developed for sound wave learning. The analysis was conducted using the Rasch measurement model to evaluate item characteristics, person ability, and overall instrument quality on a common interval logit scale (Farida et al., 2024; Mahtari et

al., 2024). The Rasch model was selected due to its capability to provide objective measurement of latent constructs and to simultaneously estimate person and item parameters within a unified measurement framework.

Participants

The participants consisted of 185 eleventh-grade students from senior high school in Bandung, Indonesia, who were enrolled in physics courses. Participants were selected using convenience sampling from intact classes that had completed instruction on sound wave topics. All participants voluntarily agreed to participate in the study. The sample represented heterogeneous academic abilities to ensure variability in CPS performance during measurement.

Instrument Development

The Collaborative Problem-Solving (CPS) assessment instrument was developed based on the OECD/PISA 2015 framework for collaborative problem solving (OECD, 2017). The instrument was designed to measure three core CPS competencies: (1) establishing and maintaining shared understanding, (2) taking appropriate action to solve problems, and (3) establishing and maintaining team organization.

The instrument consisted of 15 task-based items embedded in authentic physics contexts related to sound waves, including sound velocity, sound intensity, resonance, standing waves, sound interference, organ pipes, and the Doppler effect. Each task was designed to simulate collaborative inquiry situations requiring students to interpret experimental data, negotiate meaning, propose solutions, and make collective decisions. To ensure content validity, each item was aligned with CPS indicators (A1–D3) derived from the OECD framework. This ensured balanced representation of cognitive, social, and regulatory dimensions of collaborative problem solving.

Scoring Procedure

Each CPS task was assessed using two analytically scored components grounded in the OECD/PISA collaborative problem-solving framework. The first component measured collaborative information use, where a score of 0 indicated ineffective use of team information and a score of 1 indicated effective integration of information contributed by group members. The second component assessed problem-solving performance, with scores of 0 indicating incorrect solutions, 1 indicating partially correct solutions, and 2 indicating correct solutions accompanied by appropriate reasoning.

For Rasch measurement purposes, both components were transformed into a single dichotomous CPS performance indicator per item. The transformation followed a rule-based decision:

a) A score of **1 (successful CPS performance)** was assigned when the student demonstrated effective collaborative engagement and achieved at least a partially correct solution supported by appropriate reasoning.
b) A score of **0 (unsuccessful CPS performance)** was assigned when the student failed to demonstrate effective collaboration and/or produced an incorrect solution.

This dichotomous transformation was necessary to meet the requirements of the Rasch dichotomous model implemented in Winsteps version 4.7.0. As a result, each of the 15 CPS items generated one binary response per student, producing a total score range of 0–15.

Data Collection Procedure

Data were collected during structured physics learning sessions on sound wave topics. Students engaged in collaborative group-based activities designed to promote discussion, data interpretation, experimentation, and joint decision-making.

Although learning activities were conducted in groups, assessment was performed at the individual level. Each student's CPS performance was evaluated based on their recorded responses and contribution outcomes within the group tasks. All responses were scored by the researcher using a standardized rubric aligned with the OECD/PISA CPS framework. Scoring was conducted by a single trained rater to ensure consistency in evaluation. Therefore, inter-rater reliability analysis was not conducted in this study.

Data Analysis

Data were analyzed using Winsteps version 4.7.0 based on the dichotomous Rasch measurement model. The model was used to place person ability and item difficulty on a common logit scale, enabling invariant measurement of CPS performance across respondents and items.

The analysis included person reliability and separation index to evaluate the consistency and spread of student abilities, as well as item reliability and separation index to examine the extent of item difficulty differentiation. Model fit was assessed using infit and outfit mean square (MNSQ) statistics, with acceptable values ranging from 0.5 to 1.5.

In addition, Wright map analysis was used to visualize the alignment between person ability distribution and item difficulty levels. Person fit statistics were also examined to identify response patterns that deviated from model expectations. Overall, the Rasch model was applied to evaluate the psychometric quality of the CPS instrument and to provide linear measurement of students' collaborative problem-solving abilities in sound wave learning.

RESULTS AND DISCUSSION

Results

Reliability and Measurement Quality of the CPS Instrument Based on Rasch Analysis

The Rasch analysis results on the person aspect indicate that the Collaborative Problem-Solving (CPS) instrument demonstrated an acceptable level of measurement quality. As presented in Figure 1, the person reliability value was 0.68, indicating adequate consistency in distinguishing students' CPS abilities. The person separation index of 1.46 suggests that the instrument can classify students into approximately two levels of ability.

The mean person measure was 0.01 logits with a standard deviation of 1.09, indicating that students' abilities were centered near the average item difficulty. The infit and outfit mean square (MNSQ) values were both 1.00, indicating good fit of the data to the Rasch model. This suggests that students responded consistently to the instrument according to model expectations. Students' abilities were distributed from −2.64 to 2.64 logits, indicating considerable heterogeneity in CPS performance. Of the 185 participants, 183 were successfully estimated by the Rasch model, whereas two respondents were identified as extreme persons. The small proportion of extreme respondents (1.1%) indicates that most data were suitable for Rasch estimation.

TABLE 3.1 C:\Users\Asus\OneDrive\Desktop\Data H. ZOU513WS.TXT May 14 2026 17:57
INPUT: 185 Person 15 Item REPORTED: 185 Person 15 Item 2 CATS WINSTEPS 4.7.0.0

SUMMARY OF 183 MEASURED (NON-EXTREME) Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	22.5	15.0	.01	.60	1.00	.10	1.00	.09
SEM	.2	.0	.08	.01	.00	.03	.00	.03
P.SD	3.3	.0	1.09	.12	.03	.37	.05	.39
S.SD	3.3	.0	1.09	.12	.03	.37	.05	.39
MAX.	29.0	15.0	2.64	1.04	1.09	1.82	1.19	1.83
MIN.	16.0	15.0	-2.64	.52	.93	-1.50	.83	-1.50

REAL RMSE .61 TRUE SD .90 SEPARATION 1.46 Person RELIABILITY .68
MODEL RMSE .61 TRUE SD .90 SEPARATION 1.47 Person RELIABILITY .68
S.E. OF Person MEAN = .08

MINIMUM EXTREME SCORE: 2 Person 1.1%

FIGURE 1. Summary of 183 Measured (Non-Extreme) Person.

Reliability Indices and Measurement Implications

The Cronbach's alpha (KR-20) coefficient was 0.72, indicating acceptable internal consistency of the CPS instrument. This suggests that the interaction between persons and items was sufficiently stable for measuring collaborative problem-solving ability. At the item level, item reliability and item separation were both 0.00. According to Fisher (1992), item separation indicates how well the sample distributes items along the measured variable and reflects the extent to which distinct item difficulty

strata can be identified. Wright (1996) further explained that separation is determined by the ratio between the true spread of measures and their measurement error. In the present study, item difficulty estimates were highly clustered within a narrow range (−0.18 to 0.20 logits), indicating minimal variability among items. As a result, the true spread of item difficulties was insufficient to establish distinguishable difficulty strata, leading to very low item separation and reliability indices. This finding should not be interpreted as evidence of item misfit, as all items demonstrated acceptable fit statistics. Rather, it reflects the restricted variation in item difficulty resulting from the instrument design, which focused on collaborative problem-solving skills within a single sound-wave learning context. Future studies should incorporate tasks with broader levels of collaborative complexity and more diverse physics contexts to increase item difficulty variation and improve item separation.

Item Fit and Difficulty Analysis

The item difficulty estimates ranged from −0.18 logits (Q13) to 0.20 logits (Q6), indicating a narrow spread of item difficulty within the CPS instrument.

All items demonstrated acceptable fit to the Rasch model, with infit MNSQ values ranging from 0.89 to 1.15 and outfit MNSQ values ranging from 0.83 to 1.18. Since all values fell within the recommended range of 0.5–1.5, no items were identified as misfitting, and all 15 items were retained for further analysis. Q6 was the most difficult item (0.20 logits), followed by Q14 (0.15 logits), whereas Q13 represented the easiest item (−0.18 logits). This limited spread of item difficulty confirms low item separation, indicating minimal differentiation among item difficulty levels while maintaining strong construct coherence within CPS measurement.

TABLE 10.1 C:\Users\Asus\OneDrive\Desktop\Data H ZOU513WS.TXT May 14 2026 17:57
INPUT: 185 Person 15 Item REPORTED: 185 Person 15 Item 2 CATS WINSTEPS 4.7.0.0

Person: REAL SEP.: 1.50 REL.: .69 ... Item: REAL SEP.: .00 REL.: .00

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
13	283	185	-.18	.17	1.05	.73	1.18	1.81	A .39	.44	67.2	69.2	Q13
4	272	185	.12	.16	1.05	.85	1.15	1.56	B .39	.44	65.6	68.5	Q4
12	279	185	-.07	.16	1.15	2.27	1.14	1.43	C .33	.44	62.8	68.9	Q12
6	269	185	.20	.17	1.07	1.13	1.05	.60	D .39	.44	64.5	68.7	Q6
11	273	185	.09	.16	1.07	1.06	1.05	.60	E .40	.44	61.7	68.6	Q11
9	282	185	-.15	.16	1.03	.53	1.01	.16	F .42	.44	68.9	69.0	Q9
1	272	185	.12	.16	.99	-.10	1.01	.11	G .45	.44	68.9	68.5	Q1
2	276	185	.01	.16	1.01	.25	.99	-.06	H .43	.44	66.7	68.8	Q2
3	275	185	.04	.16	.99	-.07	.96	-.46	g .45	.44	71.6	68.7	Q3
5	282	185	-.15	.16	.99	-.15	.96	-.35	f .45	.44	71.0	69.0	Q5
7	275	185	.04	.16	.95	-.84	.96	-.41	e .48	.44	72.7	68.7	Q7
15	282	185	-.15	.16	.93	-1.04	.94	-.61	d .49	.44	69.9	69.0	Q15
10	279	185	-.07	.16	.92	-1.25	.88	-1.25	c .51	.44	71.6	68.9	Q10
8	277	185	-.01	.16	.91	-1.50	.87	-1.43	b .52	.44	74.9	68.8	Q8
14	271	185	.15	.16	.89	-1.88	.83	-1.90	a .53	.44	69.9	68.5	Q14
MEAN	276.5	185.0	.00	.16	1.00	.0	1.00	.0			68.5	68.8	
P.SD	4.4	.0	.12	.00	.07	1.1	.10	1.1			3.6	.2	

FIGURE 2. Item Measures and Fit Statistics.

Wright Map Analysis

The Wright map shows that students' abilities ranged widely (−2.64 to 2.64 logits), while item difficulties were clustered around the mean of the scale. This indicates that the instrument was moderately well-targeted to average student ability. However, several high-ability students exceeded the most difficult item level, suggesting a ceiling effect. Conversely, some low-ability students were located below the easiest items. The clustering of item difficulty confirms limited item spread, consistent with the low item separation index. This suggests that future CPS instrument development should include items with broader difficulty levels to improve measurement precision.

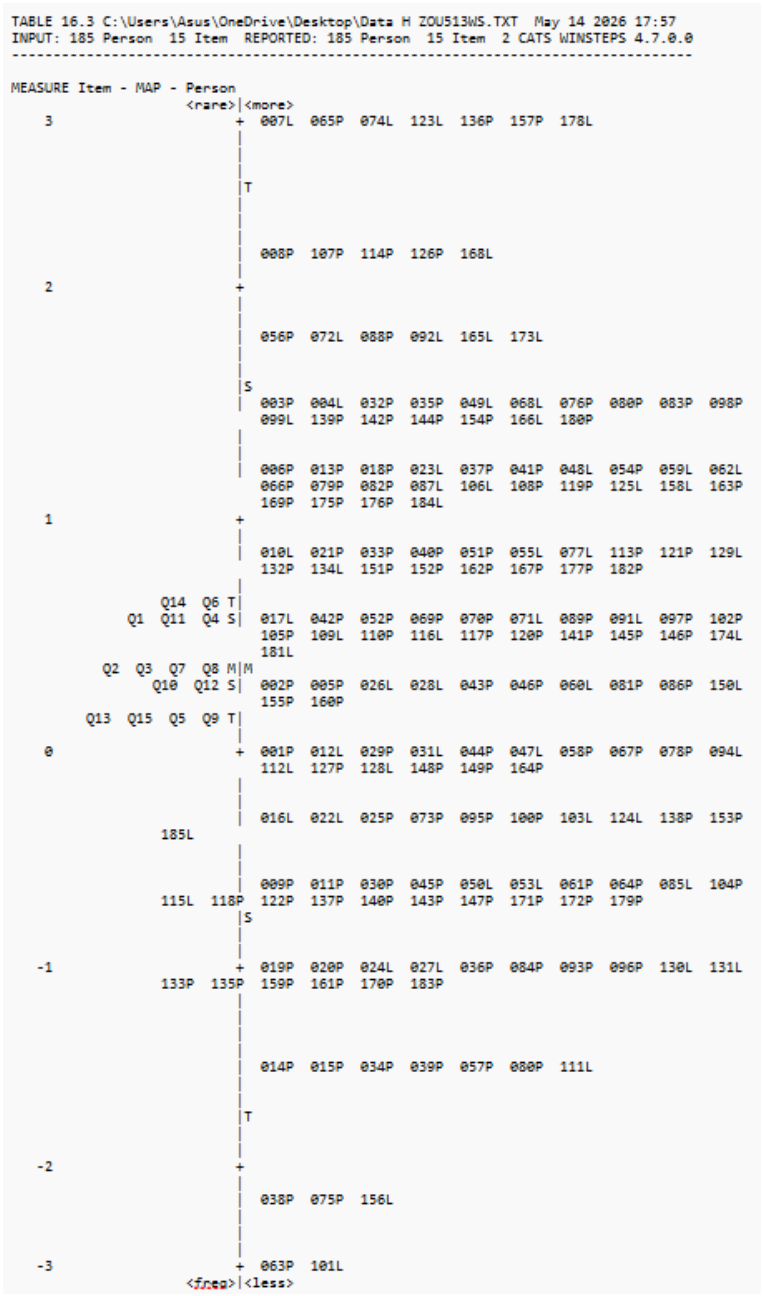

FIGURE 3. Distribution of Person Ability and Item Difficulty Based on Rasch Analysis.

Person Fit and CPS Ability Distribution

Person fit analysis was conducted using infit MNSQ, outfit MNSQ, and ZSTD statistics. Acceptable fit criteria were defined as 0.5–1.5 for MNSQ values.

TABLE 2. Person Misfit and Unexpected Response Patterns.

Category	Description
Misfitting persons	123, 136, 178, 36, 75, 8, 38, 173, 135 (all mild misfit, Outfit ≤ 1.19)
Unexpected high ability	007, 065, 074, 157
Unexpected low ability	156, 38, 75, 14, 111, 39

The results showed that most students demonstrated response patterns consistent with Rasch model expectations. Although a few respondents showed slightly elevated outfit values, all remained below 1.50. Therefore, no respondents were removed from the dataset.

The distribution of CPS abilities showed a mean of −0.04 logits with a standard deviation of 1.09, indicating slightly below-average performance but wide variation across students. High-ability students (1.5–2.6 logits) showed more consistent response patterns, while low-ability students (−1.5 to −2.6 logits) showed greater difficulty in completing CPS tasks.

Discussion

The results indicate that the developed Collaborative Problem-Solving (CPS) instrument demonstrated acceptable psychometric quality for assessing students' collaborative competencies in sound wave learning. Item functioning within the Rasch model showed that the instrument operated consistently in measuring the intended CPS construct aligned with the OECD/PISA 2015 framework. This supports the use of Rasch measurement as a robust approach in physics education assessment because it enables simultaneous evaluation of item quality and student ability on a shared measurement scale (Ying & Tiemann 2024; Farida et al. 2024).

At the item level, item reliability and item separation were both 0.00. According to Fisher (1992), item separation indicates how well the sample distributes items along the measured variable and reflects the extent to which distinct item difficulty strata can be identified. Wright (1996) further explained that separation is determined by the ratio between the true spread of measures and their measurement error. In the present study, item difficulty estimates were highly clustered within a narrow range (−0.18 to 0.20 logits), indicating minimal variability among items. As a result, the true spread of item difficulties was insufficient to establish distinguishable difficulty strata, leading to very low item separation and reliability indices. This finding should not be interpreted as evidence of item misfit, as all items demonstrated acceptable fit statistics. Rather, it reflects the restricted variation in item difficulty resulting from the instrument design, which focused on collaborative problem-solving skills within a single sound-wave learning context. Future studies should incorporate tasks with broader levels of collaborative complexity and more diverse physics contexts to increase item difficulty variation and improve item separation.

From a measurement perspective, the Wright map analysis suggests that the instrument was generally appropriate for the target group; however, student ability distribution was broader than item difficulty distribution. This indicates that the instrument is effective in capturing average CPS performance but has limited sensitivity in distinguishing students with higher levels of collaborative competence. In CPS assessment development, such a pattern typically suggests the need for incorporating more complex, open-ended, and higher-order collaborative tasks that require deeper negotiation, reasoning, and coordination among group members (OECD 2017).

The relatively narrow spread of item difficulty can be attributed to the deliberate design of tasks within a single instructional context, namely sound wave learning. While this design ensures contextual consistency and strong content validity, it may also limit the instrument's ability to represent the full continuum of CPS proficiency. Therefore, future development should consider expanding task heterogeneity, particularly by integrating multi-step problem-solving situations that require distributed reasoning and flexible role negotiation within groups (OECD 2017; Rojas et al. 2025).

The findings suggest that the developed instrument can be used to capture variations in students' collaborative problem-solving performance within sound-wave learning activities. However, the narrow spread of item difficulty indicates that the current instrument is more effective for measuring average levels of CPS performance than for differentiating higher and lower levels of collaborative competence. Future instrument development should therefore incorporate tasks with broader collaborative complexity to improve measurement precision across a wider range of student abilities.

The importance of assessing CPS through structured instruments such as the one developed in this study supports the view that CPS should be considered not only as a learning outcome but also as a process skill that requires explicit assessment within physics education (OECD 2017). Traditional physics assessments tend to emphasize individual conceptual understanding and procedural fluency, whereas CPS-based assessment extends evaluation to include communication, coordination, and shared problem representation. This is consistent with contemporary perspectives on scientific literacy that emphasize the integration of cognitive and social dimensions of scientific practice (OECD 2017).

Moreover, the use of sound wave contexts enhances the ecological validity of the assessment. Physics experiments provide authentic inquiry situations in which students interpret data, test

hypotheses, and make collective decisions. These conditions naturally require interdependence among group members, making them particularly suitable for observing CPS processes in action (Rojas et al. 2025; He et al. 2024).

Overall, this study demonstrates that integrating the OECD/PISA 2015 CPS framework with Rasch measurement provides a structured approach for assessing collaborative competencies in physics education. The findings indicate that the instrument demonstrated acceptable person measurement properties and adequate item functioning for classroom-based CPS assessment. However, the limited variation in item difficulty suggests that further refinement of task design is needed to improve the instrument's ability to measure a broader range of CPS performance. Therefore, future CPS-oriented assessment development should incorporate more diverse and complex collaborative tasks to enhance measurement precision and coverage of collaborative problem-solving competencies.

Future research is recommended to expand the item pool across different physics domains and to investigate CPS development longitudinally, enabling a more comprehensive understanding of how collaborative competencies evolve across contexts and over time.

CONCLUSION

This study developed and validated a Rasch-based assessment instrument for measuring collaborative problem-solving (CPS) skills in sound wave learning among senior high school students. The Rasch analysis indicated acceptable measurement quality, as reflected by adequate person reliability (0.68), acceptable internal consistency (KR-20 = 0.72), and item fit statistics within the recommended range. These findings suggest that the instrument functioned consistently in measuring the intended CPS construct.

The results showed substantial variation in students' CPS abilities, as evidenced by the broad distribution of person measures. However, item difficulty estimates were concentrated within a narrow logit range (−0.18 to 0.20), resulting in very low item separation and item reliability values. This indicates that the instrument was more effective for measuring average CPS performance than for distinguishing multiple levels of item difficulty or higher-order collaborative competence.

A limitation of this study is the relatively homogeneous difficulty level of the CPS tasks, which were developed within a single sound-wave learning context. Consequently, the instrument provided limited differentiation across item difficulty levels. Future research should expand the item pool by incorporating more complex collaborative tasks and a wider range of physics topics to improve item difficulty variation, measurement precision, and the generalizability of CPS assessment.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the supervisors, lecturers, and all parties who contributed to this research for their guidance, support, and valuable assistance throughout the study.

REFERENCE

Alanazi, A A, Osman, K & Halim, L 2024, 'Integrating digital assessment in physics education: enhancing higher-order thinking and problem-solving skills of students in technical colleges in the Kingdom of Saudi Arabia', in Digital assessment in higher education: navigating and researching challenges and opportunities, Springer Nature Singapore, Singapore, pp. 327–347, https://doi.org/10.1007/978-981-97-6136-4_15

Badmus, O T & Jita, L C 2025, 'Nature of science representations in South African Grade 10 physical sciences textbook on waves, sound and light', Social Sciences & Humanities Open, vol. 12, 101848, https://doi.org/10.1016/j.ssaho.2025.101848

Barker Scott, B A & Manning, M R 2024, 'Designing the collaborative organization: a framework for how collaborative work, relationships, and behaviors generate collaborative capacity', The Journal of Applied Behavioral Science, vol. 60, no. 1, pp. 149–193, https://doi.org/10.1177/00218863221106245

Caronni, A, Picardi, M, Scarano, S, Rota, V & Amadei, M 2026, 'Improving single-subject change assessment: deriving the minimal detectable change of questionnaires' ordinal scores from Rasch analysis measures', Disability and Rehabilitation, vol. 48, no. 7, pp. 2169–2186, https://doi.org/10.1080/09638288.2025.2547398

Farida, F, Aspat Alamsyah, Y, Anggoro, B S, Andari, T & Lusiana, R 2024, 'Rasch measurement validation of an assessment tool for measuring students' creative problem-solving through the use of ICT', Pixel-Bit. Revista de Medios y Educación, no. 71, pp. 83–106, https://doi.org/10.12795/pixelbit.107973

Fisher, WP Jr 1992, 'Reliability, separation, strata statistics', *Rasch Measurement Transactions*, vol. 6, no. 3, p. 238.

He, J, Ren, S & Zhang, D 2024, 'The relationship between personal-collaborative motivation profiles and students' performance in collaborative problem solving', Large-Scale Assessments in Education, vol. 12, no. 1, 34, https://doi.org/10.1186/s40536-024-00219-6

Kalyani, L K 2024, 'The role of technology in education: enhancing learning outcomes and 21st century skills', International Journal of Scientific Research in Modern Science and Technology, vol. 3, no. 4, pp. 5–10, https://doi.org/10.59828/ijsrmst.v3i4.199

Kim, H J, Lee, S J & Kam, K Y 2023, 'Reliability and validity of the School Function Assessment for children with disabilities in Korea: applying Rasch analysis', International Journal of Disability, Development and Education, vol. 70, no. 1, pp. 32–44, https://doi.org/10.1080/1034912X.2020.1870665

Mahtari, S, Syihabuddin, S, Setiawan, A & Wati, M 2024, 'Effectiveness of collaboration skills assessment instrument in physics practicum: a case study with the Rasch model', Dinasti International Journal of Education Management and Social Science, vol. 6, no. 2, pp. 1270–1278, https://doi.org/10.38035/dijemss.v6i2.3661

OECD 2017, PISA 2015 results (volume V): collaborative problem solving, OECD Publishing, Paris, https://doi.org/10.1787/9789264285521-en

Oktaviana, S & Putranta, H 2024, 'Identification of students' conceptual understanding of sound wave materials through the contextual teaching and learning (CTL) model', Research and Education, no. 10, pp. 124–145, https://doi.org/10.56177/red.10.2024.art.7

Rojas, M, Nussbaum, M & Moreno, C 2025, 'Validating a collaborative problem-solving assessment tool across educational stages', Computers & Education, vol. 227, 105228, https://doi.org/10.1016/j.compedu.2024.105228

Setyawarno, D, Maryati & Natadiwijaya, I F 2025, 'Promoting a valid question model for measuring computational thinking skills based on confirmatory factor analysis and Rasch model', Cogent Education, vol. 12, no. 1, 2505339, https://doi.org/10.1080/2331186X.2025.2505339

Suryadi, B, Hayat, B & Putra, M D K 2020, 'Evaluating psychometric properties of the Muslim daily religiosity assessment scale (MUDRAS) in Indonesian samples using the Rasch model', Mental Health, Religion & Culture, pp. 1–16, https://doi.org/10.1080/13674676.2020.1795822

Wallace, G H 2020, 'Improving Spanish classroom assessment via logistic regression: lessons from the Rasch model', Journal of Spanish Language Teaching, vol. 7, no. 1, pp. 51–63, https://doi.org/10.1080/23247797.2020.1771009

Wei, X, Wang, L, Lee, L K & Liu, R 2025, 'The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation: an experimental study', International Journal of Educational Technology in Higher Education, vol. 22, no. 1, 23, https://doi.org/10.1186/s41239-025-00526-0

Wright, BD 1996, 'Reliability and separation', *Rasch Measurement Transactions*, vol. 9, no. 4, p. 472.

Ying, Y & Tiemann, R 2024, 'Development of an assessment tool for collaborative problem-solving skills in chemistry', Disciplinary and Interdisciplinary Science Education Research, vol. 6, no. 1, 25, https://doi.org/10.1186/s43031-024-00116-6

Ying, Y & Tiemann, R 2026, 'Investigating the factors influencing collaborative problem-solving in chemistry education: a cross-national study of German and Chinese students', International Journal of Science and Mathematics Education, vol. 24, no. 2, 11, https://doi.org/10.1007/s10763-025-10629-9