

Development of Competency Test Instruments for PPG Students in the Automotive Field: Analysis of Validity, Reliability, and Discrimination Power

Syarif Suhartadi*, Widiyanti Widiyanti

Universitas Negeri Malang, Jl. Semarang 5, Malang, 65145, Indonesia
¹ syarif.suhartadi.ft@um.ac.id*; ² Widiyanti.ft@um.ac.id
* corresponding author

ARTICLE INFO	ABSTRACT
<p>Article history Received Sept 21, 2025 Revised Dec 15, 2025 Accepted Dec 28, 2025</p> <p>Keywords Competency test instrument Teacher professional education (ppg) Automotive vocational education Item validity Test reliability</p>	<p>This study aimed to develop and validate a competency test instrument for students in the Teacher Professional Education Program (PPG) in the automotive field. The instrument was developed through a systematic research and development process, including the preparation of measurement specifications, item writing based on a competency blueprint, expert review, test assembly, trial implementation, and item analysis. The trial involved 73 PPG automotive students at Universitas Negeri Malang. Psychometric analyses were conducted to examine item difficulty level, discrimination power, validity, reliability, and distractor effectiveness. The results indicated that out of 35 initial items, 23 items met the required psychometric criteria and were retained after revision and selection. Most retained items demonstrated medium difficulty levels and acceptable discrimination power. The reassembled instrument achieved a Cronbach's Alpha coefficient of 0.719, indicating high internal consistency. These findings confirm that the developed instrument is valid and reliable for assessing the competencies of PPG automotive students. Furthermore, this study offers a replicable framework for developing competency-based test instruments in vocational and technical education contexts.</p>

I. Introduction

Vocational education plays a strategic role in preparing learners to acquire competencies aligned with labor market demands and technological development. In Indonesia, vocational education is expected to produce graduates who are globally competitive and capable of supporting national industrial growth. Achieving this objective requires teachers who not only master theoretical knowledge but also possess strong pedagogical and practical competencies relevant to industry needs (Hadi et al., 2025).

The Teacher Professional Education Program (PPG) is a national policy designed to ensure that prospective teachers demonstrate pedagogical, personal, social, and professional competencies before entering the profession (Subroto, 2019). In Automotive Engineering Education, the urgency of ensuring competency is even greater due to the rapid development of automotive technology and industry standards. Automotive teachers are required to integrate theoretical concepts with applied skills so that learning outcomes remain relevant to workplace demands (Daniswara & Nuryana, 2023; Pangestika, 2015). Consequently, the competencies of PPG automotive students must be measured using evaluation instruments that are accurate, objective, and psychometrically sound.

Evaluation instruments are a crucial component in ensuring the quality of PPG graduates (Hardiyanta & Wagiran, 2023; Samosir et al., 2024). High-quality

instruments provide valid and reliable data for academic decision-making, whereas poorly constructed instruments

may lead to misleading conclusions about student competence. Achmad et al. (2024) emphasized that validity, reliability, item difficulty level, and discrimination power are essential indicators of test quality. Instruments that fail to meet these criteria cannot function effectively as tools for evaluating learning outcomes, particularly in vocational education, where both theoretical understanding and applied competence are required.

Empirical evidence indicates that the quality of evaluation instruments in vocational education remains problematic. Klotz and Winther (2017) reported that approximately 42% of vocational examination items failed to meet validity or discrimination criteria, limiting their ability to distinguish between high- and low-ability learners. Similarly, Hadi (2012) and Priti et al. (2023) found that many assessment instruments used in PPG lacked adequate reliability testing, resulting in inconsistent evaluation outcomes. These findings highlight the persistent gap between assessment practices and established psychometric standards in vocational teacher education.

In the context of Automotive Engineering Education, assessment instruments must address complex competencies, including mastery of engine systems, vehicle electronics, and emerging automotive

technologies. Studies by Moktavia et al. (2024) and Witorsa et al. (2017) demonstrated that field-specific instruments improve the accuracy of competency measurement because they reflect authentic occupational demands. Therefore, competency tests for PPG automotive students must be carefully developed to ensure alignment with professional standards and instructional objectives.

From a psychometric perspective, test quality is determined by item validity, reliability, difficulty level, discrimination power, and distractor effectiveness. Botterbusch (1987) suggested that items should demonstrate a minimum correlation coefficient of 0.3 to be considered valid, while reliability coefficients above 0.7 indicate acceptable internal consistency (Diao & Hu, 2022; Perdana, 2021). Additionally, proportional item difficulty distributions and adequate discrimination power are necessary to capture variation in student ability (Daryono et al., 2023; Jatmoko et al., 2025).

Empirical research in vocational education supports this view. Warju et al. (2020) found that only 25% of items used in PPG electrical engineering met validity and reliability criteria, indicating that existing instruments have not yet optimally reflected student competencies. Chikita (2019) also highlighted the importance of distractors in multiple-choice items. Non-functioning distractors reduce the instrument's quality, even if the difficulty level and validity are acceptable. This shows that every aspect of an item, including distractors, must be carefully designed. With comprehensive analysis, instruments can be refined to become genuinely suitable for evaluation.

Based on theoretical studies and prior research, developing evaluation instruments that are valid, reliable, and aligned with the automotive industry's needs is essential for improving the quality of PPG. High-quality instruments will provide accurate information about students' competencies, which, in turn, can improve the quality of vocational education for teachers. This aligns with PPG's goal of producing teachers who are not only professional but also relevant to the needs of the labor market and industry. Therefore, this study is driven by the urgent need to develop and validate competency test instruments for PPG automotive students.

II. Method

This study employed a research-and-development approach with the primary objective of developing a valid and reliable test instrument to assess students' competencies in the Teacher Professional Education Program (PPG) in the automotive field. The research subjects were 73 PPG students from Universitas Negeri Malang, selected as trial respondents. The instrument development process followed the standard stages of measurement tool construction: a) drafting instrument

specifications based on competency indicators derived from teacher competency standards, b) writing multiple-choice items according to the blueprint, c) expert review to ensure content validity, d) assembling the items into a test package, and e) administering the trial to the research subjects.

Data from the trial were analyzed using item analysis techniques to assess the instrument's quality across several aspects. The difficulty level was calculated as the proportion of respondents who answered each item correctly. Discrimination power was calculated by comparing the proportion of correct answers between high- and low-ability groups. Item validity was assessed by correlating individual item scores with total scores using the product-moment correlation, with a minimum criterion of 0.3 (Cronbach, 1990). Instrument reliability was calculated using Cronbach's Alpha coefficient through SPSS version 30, where values above 0.7 were considered adequate (Botterbusch, 1987). Distractor effectiveness was also analyzed to ensure that each option served as an adequate distractor, thereby enhancing the instrument's measurement accuracy.

Decisions on whether an item was retained, revised, or discarded were based on the combined results of validity, reliability, difficulty level, discrimination power, and distractor function. Items with high validity, proportional difficulty, medium-to-high discrimination, and functional distractors were retained. Items with borderline quality were revised, while items that did not meet the criteria were discarded. Through this procedure, the final instrument was expected to have sound psychometric quality and be suitable for accurately and consistently measuring PPG automotive student competencies.

Despite extensive research on item analysis in vocational education, prior studies have primarily evaluated existing tests rather than systematically developing and validating new instruments tailored specifically for PPG automotive students. Most studies address isolated competencies or single subject areas, leaving a gap in comprehensive instrument development that integrates multiple competency domains within a unified psychometric framework.

Based on this gap, the present study aims to develop and validate a competency test instrument designed explicitly for PPG students in the automotive field through a systematic development process, from blueprint construction to item analysis. The novelty of this study lies in its comprehensive approach to instrument development and its focus on the unique competency profile of PPG automotive students in Indonesia.

III. Results and Discussion

A. Difficulty Level

If an item is answered correctly by 30% to 69.9% of respondents, it is considered to have a medium or ideal level of difficulty. If more than 70% of respondents answer it correctly, the item is categorized as easy, meaning that over 70% of students answered it correctly. Of the 35 items tested on 73 students, none were classified as easy; 30 were categorized as medium or ideal, and 5 as difficult. If an item is answered correctly by fewer than 30% of respondents (0–29.9%), it is categorized as difficult, indicating that only a small number of respondents answered it correctly.

B. Discrimination Power

If the difference between the high- and low-ability groups is less than 30%, the item is categorized as having low discrimination power. This means that both the high- and low-ability groups answered the item correctly, with the difference in proportions being less than 30%. If the proportion of correct answers in the high group is the same as in the low group, then the difference is zero (0). This indicates that the item cannot distinguish between high- and low-ability students, since both groups answered it correctly. If the difference ranges from 30% to 69.9%, the item is considered to have medium or ideal discrimination power. If the difference exceeds 70%, the item has high discrimination power, meaning it can distinguish up to 70% of respondents in the low group who were unable to answer correctly. Of the 35 items tested, 22 were classified as having ideal discrimination power, 12 had low discrimination power, and 1 had high discrimination power.

C. Item Validity

Of the 35 test items administered, 2 for Indicator 1 mastering and applying student-centered learning management in automotive education were categorized as valid. In comparison, nine items were invalid or had low validity. For Indicator 2, possessing a strong personality, noble character, wisdom, and authority, and serving as a role model in automotive education, three items were valid, and three items were invalid or had low validity. For Indicator 3, the teacher's ability to communicate and interact effectively and efficiently with students, fellow teachers, parents/guardians, and the surrounding community in automotive education, 1 item was valid, and

five items were invalid or had low validity. For Indicator 4, mastery of subject matter broadly and deeply to determine learning objectives and organize content in automotive education, four items were valid.

In comparison, eight items were invalid or had low validity, with correlation values below 0.3. The validity of each item was calculated by correlating each item's score with the total score obtained by each respondent. The item score reflects the measurement result for a particular indicator, while the total score reflects the measurement result for the construct being measured overall. Cronbach (1990) used a criterion of 0.3 to determine item validity. If the correlation is below 0.3, it means the item contributes less than 30% to measuring the construct, and thus the item is categorized as invalid.

D. Distractor Function

Each answer option (distractor) serves as a decoy if respondents select it. If no respondent chooses an option, it does not function as a distractor. In other words, it is a poor distractor and must be revised. Out of the 35 test items administered, nine items still contained distractors that were not selected (ignored by respondents), indicating that these distractors need revision. Meanwhile, the other 26 items already had effective distractors and therefore did not require revision. Each item consisted of five answer options, A, B, C, D, and E, referred to as distractors or decoys.

E. Item Selection (Retained with No Revised, Revised, or Discarded)

The decision on whether an item is used without revision, used with revision, or discarded is primarily based on its validity, difficulty level, and, subsequently, its discriminant power. If an item has a validity score of 0.30 or higher and an ideal, easy, and/or difficult difficulty level, then the item can be retained. The researcher also considered input from the distractor analysis of each item to determine whether revision was necessary. If an item's validity score was close to 0.30, it could still be considered for use, if modifications were made beforehand. Revisions were made to the item stem and answer options while ensuring balance and representation of indicators in the test blueprint. The revised items were then reassembled for further use. Based on considerations of validity, discriminant power, and difficulty level, this study ultimately retained 23 items, while 12 were discarded.

Table 1. Items Reassembled After Trial

NO	Competency Aspect	Items Before Trial		Items After Trial		
		No	Revision	Revised	Total Retained	Discarded
1	Mastering and applying student-centered learning management in automotive education	11	2	5	7	4
2	Strong personality, noble character, wisdom, authority, role model in automotive education	6	3	1	4	2
3	Communication and interaction ability with students, peers, parents, community in automotive education	6	1	3	4	2

NO	Competency Aspect	Items	No	Items After Trial	Discarded	
		Before Trial	Revision	Revised Total Retained		
4	Mastery of subject matter broadly and deeply to set objectives and organize content in automotive education	12	4	4	8	4
		35	10	13	23	12

F. Reliability

After 23 items were reassembled based on validity, difficulty level, and discrimination power, their reliability was tested using Cronbach's Alpha, calculated with SPSS version 30. The results showed a Cronbach's Alpha reliability coefficient of 0.719. This indicates that the final test trial was highly reliable.

G. Validity, Difficulty, and Discrimination in Item Selection

In reducing the number of items from 35 to 23, the researcher based the decision on: (1) validity level, (2) difficulty level, (3) discrimination power, and (4) recommendations derived from the distractor analysis of each item. Based on these four criteria, it was determined whether an item would be used without revision, used with revision, or discarded. This decision was consistent with established test construction guidelines. The validity of each item was calculated by correlating each item's score with the total score obtained by each respondent. Cronbach (1949) set a criterion of 0.3 for determining item validity. If the correlation is below 0.3, the item contributes less than 30% to the measurement of the construct being assessed. If an item has a validity score of

0.30 or higher, an ideal, easy, and/or difficult level of difficulty, and medium-to-high discrimination power, then the item is retained without revision. Of the 35 items tested, 9 were classified as valid, while 26 were invalid or had low validity (below 0.3).

The purpose of developing this test was to predict students' learning achievement in Automotive Engineering Education. Therefore, the difficulty level and discrimination power of each item were taken into consideration. In this regard, the 23 retained items were further examined for their difficulty level and discriminant power and then aligned with the target specifications of a test instrument, namely, a high level of discrimination between students predicted to succeed and those expected to struggle in their academic journey in Automotive Education. The target distribution was 30% difficult items, 50% medium items, and 20% easy items; and for discrimination power, 50% of items with ideal discrimination, 30% with high discrimination, and 20% with low discrimination. The comparison between these target proportions and the actual distribution of items meeting the required difficulty levels and discrimination power is presented in the following table.

Table 2. Targeted vs Actual Item Distribution

NO	ASPECT	Total Items	Target & Actual (Difficulty Level)			Target & Actual (Discrimination Power)		
			Difficult (30%)	Medium (50%)	Low (20%)	High (30%)	Medium (50%)	Low (20%)
1	Mastering and applying student-centered learning management in automotive education	11	3 (2)	5 (7)	3 (2)	3 (3)	5 (6)	3 (2)
2	Strong personality, noble character, wisdom, authority, role model in automotive education	6	2 (2)	3 (2)	1 (2)	2 (2)	3 (2)	1 (2)
3	Ability to communicate and interact effectively with students, peers, parents/guardians, and the community in automotive education	6	2 (1)	3 (4)	1 (1)	2 (2)	3 (3)	1 (1)
4	Mastery of subject matter broadly and deeply to set objectives and organize content in automotive education	12	3 (3)	6 (7)	3 (2)	3 (2)	6 (8)	3 (2)
		35	10 (8)	17 (19)	8 (8)	10 (9)	17 (19)	8 (7)

Based on the previous table, although the instrument already demonstrated high reliability (0.719), further improvements were still needed. In terms of difficulty level, almost all aspects, except for certain ones, required upgrading to a greater difficulty category. Similarly, regarding item discrimination, nearly none of the items across all aspects showed high discrimination. Therefore, the items needed to be revised to increase their difficulty level to achieve higher discrimination. The target that was completed was the proportion of items classified as easy.

For items in the medium (ideal) category, some aspects exceeded the target, while others were not yet met.

On the other hand, for items with ideal and low discrimination, almost all aspects met the target, except for certain elements that still showed low discrimination. This means that, from the perspective of discrimination power, the items still require revision to their difficulty levels to achieve high discrimination. To evaluate the 23 reassembled items, data on the distribution of items, along

with information on their discriminant power and difficulty level, were needed. Based on the data presented in the tables, the researcher made revisions only to specific items within each aspect measured by the test instrument under development.

From Table 3 below, two items were used without revision, five items were revised, and four items were discarded. From Table 4, three items were used without revision, 1 item was revised, and two items were discarded. From Table 5, 1 item was used without revision, three items were used after revision, and two items were discarded. From Table 6, four items were used without revision, four were revised, and four were discarded.

Table 3. Distribution of revised items in the aspect of mastering and applying student-centered learning management in automotive education, based on discrimination power and difficulty level.

Old Item No.	Discrimination (d)	Correct Response Proportion (p)	Conclusion	New Item No.
1	0,42 9 Fair	0,58 8 Medium	Used With Revision	1
2	0,16 7 Poor	0,16 2 Difficult	Discarded	-
3	0,45 2 Fair	0,48 5 Medium	Used With Revision	2
4	0,45 2 Fair	0,26 5 Difficult	Used With Revision	3
5	0,33 3 Fair	0,45 6 Medium	Used With Revision	4
6	0,50 0 Fair	0,45 6 Medium	Used With Revision	5
7	0,00 0 Poor	0,36 8 Medium	Discarded	-
8	0,23 8 Poor	0,23 5 Difficult	Discarded	-
9	- 0,02 Poor	0,52 9 Medium	Discarded	-
10	0,66 7 Fair	0,52 9 Medium	Used With Revision	6
11	0,38 1 Fair	0,82 4 Medium	Used With Revision	7

Table 4. Distribution of revised items in the aspect of strong personality, noble character, wisdom, and authority, as well as serving as a role model in automotive education, based on discrimination power and difficulty level.

Old Item No.	Discrimination (d)	Correct Response Proportion (p)	Conclusion	New Item No.
12	0,35 7 Fair	0,92 6 Medium	Used With Revision	8
13	0,35 7 Fair	0,63 2 Medium	Used With Revision	9
14	0,21 4 Poor	0,92 6 Medium	Discarded	-

Old Item No.	Discrimination (d)	Correct Response Proportion (p)	Conclusion	New Item No.
15	0,23 8 Poor	0,82 4 Medium	Discarded	-
16	0,50 0 Fair	0,80 9 Medium	Used With Revision	10
17	0,50 0 Fair	0,85 3 Medium	Used With Revision	11

Table 5. Distribution of revised items in the aspect of teachers' ability to communicate and interact effectively and efficiently with students, fellow teachers, parents/guardians, and the surrounding community in automotive education, based on discrimination power and difficulty level.

Old Item No.	Discrimination (d)	Correct Response Proportion (p)	Conclusion	New Item No.
18	0,35 7 Fair	0,79 4 Medium	Used With Revision	12
19	0,50 0 Fair	0,66 2 Medium	Used Without Revision	13
20	0,26 2 Poor	0,69 1 Medium	Discarded	-
21	0,31 0 Fair	0,89 7 Medium	Used With Revision	14
22	0,28 6 Poor	0,70 6 Medium	Discarded	-
23	0,40 5 Fair	0,38 2 Medium	Used With Revision	15

Table 6. Distribution of revised items in the aspect of mastering subject matter broadly and deeply to determine learning objectives and organize content in the field of automotive education, based on discrimination power and difficulty level.

Old Item No.	Discrimination (d)	Correct Response Proportion (p)	Conclusion	New Item No.
24	0,31 0 Fair	0,64 7 Medium	Used With Revision	16
25	0,09 5 Poor	0,04 4 Difficult	Discarded	-
26	0,59 5 Fair	0,35 3 Medium	Used Without Revision	17
27	0,50 0 Fair	0,41 2 Medium	Used With Revision	18
28	0,78 6 Very Good	0,63 2 Medium	Used Without Revision	19
29	0,16 7 Poor	0,94 1 Medium	Discarded	-
30	0,09 5 Poor	0,55 9 Medium	Discarded	-
31	0,35 7 Fair	0,88 2 Medium	Used With Revision	20

32	0,35 7	Fair	0,75 0	Mediu m	Used With Revision	21
33	0,61 9	Fair	0,55 9	Mediu m	Used Without Revision	22
34	0,02 4	Poor	0,19 1	Difficul t	Discarded	-
35	0,35 7	Fair	0,91 2	Mediu m	Used Without Revision	23

The results of this study show that of the 35 test items trialed, only 23 met the criteria for reuse after selection based on validity, reliability, difficulty level, discrimination power, and distractor analysis. This finding indicates that nearly one-third of the initial items did not meet proper psychometric standards. Such a condition is understandable, as instrument development often faces challenges in both item construction and mapping competency indicators. Jatmoko et al. (2025) emphasized that test development is a complex process, and not all items will be of high quality without undergoing trial and revision.

From the perspective of validity, only 9 out of 35 items achieved a correlation above 0.3, as recommended by Cronbach (1949). This means that most items were not sufficiently representative in measuring the intended construct. This issue aligns with findings by Sensus et al. (2022), who also reported low validity in many PPG electrical engineering student assessments. Low validity may result from poorly formulated indicators, ambiguous language, or mismatched contexts between the questions and students' abilities. Therefore, substantial revisions are necessary to improve item quality.

In terms of reliability, the reassembled instrument achieved a Cronbach's Alpha coefficient of 0.719, exceeding the 0.7 threshold considered adequate by Menchetti & Rusch (1988). This reliability level indicates strong internal consistency, meaning the instrument can produce relatively stable scores when measuring PPG automotive student competencies. Similar results were reported by Murwantini (2023) and Winther & Klotz (2013), showing that reliability improves after item selection and revision.

The analysis of difficulty levels revealed that most items fell into the medium category, a smaller proportion were difficult, and none were easy. This distribution generally aligns with recommended test construction principles (Chikita, 2019; Warju et al., 2020), which suggest that most items should be of medium difficulty to effectively capture variation in students' abilities. However, in vocational education, having more difficult items is also essential to distinguish students with higher mastery. Aayn & Listiadi (2022) and Monnier et al. (2016) stressed that vocational education assessments must distinguish students who are genuinely ready for industry from those who require further support. Thus, while

reliability is satisfactory, the addition of more difficult items remains necessary to strengthen the test's discriminatory function.

Regarding discrimination power, most items fell in the medium range, some were low, and very few were high. This outcome suggests that the instrument still requires enhancement to better distinguish between students of different competency levels. Antonietti et al. (2022) emphasized that discrimination is a key quality indicator, as items with low discrimination tend to yield homogeneous and less meaningful evaluations. Compared to the expected distribution (30% high, 50% medium, 20% low), the results here were not yet ideal, confirming the need for further revisions to increase item discrimination.

The analysis of distractors provided additional insights. Of the 35 items, 9 had non-functioning distractors, meaning respondents never selected those answer choices. This weakens the effectiveness of those items since good distractors should attract some respondents with lower mastery of the material. Cattaneo et al. (2022) pointed out that non-functioning distractors are a serious flaw, as they reduce the overall measurement power of an item. Therefore, revising distractor wording and ensuring alignment with the intended difficulty level is essential to optimizing item effectiveness.

When compared with previous studies, the present findings are consistent with the general pattern that not all drafted items can meet psychometric standards. Venatius et al. (2023) underscored that item analysis is mandatory to guarantee usable instruments. Similarly, Sensus et al. (2022) and Witarsa et al. (2017) also found a large proportion of invalid or weak items in vocational test instruments. This study, therefore, strengthens prior evidence and underscores the importance of trial and revision in developing valid and reliable instruments for vocational education.

In practice, this study demonstrates that the reassembled competency test for PPG automotive students can serve as a valuable tool for more accurately assessing student achievement. A reliable and valid instrument enables lecturers and program organizers to conduct more objective evaluations. Furthermore, this study highlights the importance of increasing item difficulty and discrimination in future revisions. Such improvements will not only enhance evaluation quality but also ensure vocational education fulfills its goal: producing competent automotive teachers capable of preparing vocational high school students for industry demands.

Nevertheless, this study has limitations. First, the sample size was limited to 73 students from a single university, thereby limiting the generalizability of the findings. Second, the instrument only assessed cognitive aspects through a written test, while automotive teacher competencies also include affective and psychomotor domains, which remain unmeasured. Future research

should therefore develop authentic or practice-based assessments to complement written tests, thereby providing a more comprehensive evaluation of PPG automotive students' readiness to become future vocational educators

IV. Conclusion

The results of this study show that, of the 35 items tested with 73 students, 23 were retained after reassembly. These consisted of: 7 items from the aspect of mastering and applying student-centered learning management in automotive education, four items from the element of possessing a strong personality, noble character, wisdom, and authority while serving as a role model in automotive education, four items from the aspect of teachers' ability to communicate and interact effectively and efficiently with students, fellow teachers, parents/guardians, and the surrounding community in automotive education, and eight items from the aspect of mastering subject matter broadly and deeply to determine learning objectives and organize content in automotive education. From these 23 items, a reliability coefficient of 0.719 was obtained, which is categorized as high.

References

- [1] Aayn, S. L., & Listiadi, A. (2022). The influence of school field introduction, teacher profession perception, and self-efficacy on readiness to become a teacher (Students of the Accounting Education Study Program at UNESA). *Journal of Teacher and Learning Studies*, 5(1), 132–140.
- [2] Achmad, I. A., Asmas, M. A., & Sukawati, Y. (2024). Evaluation of the Work Skills Education Program for Automotive Training in Improving Student Skills. *Continuing Learning Society Journal*, 2(1), 40–58.
- [3] Antonietti, C., Cattaneo, A., & Amenduni, F. (2022). Can teachers' digital competence influence technology acceptance in vocational education? *Computers in Human Behavior*, 132, 107266.
- [4] Ardhani, Y. (2020). Quality of year-end exam items for Basic Automotive Technology, Grade X, Light Vehicle Engineering at SMK Muhammadiyah Gamping, 2018/2019. *Journal of Vocational Automotive Education*, 3(1), 85–94.
- [5] Botterbusch, K. F. (1987). *Vocational Assessment and Evaluation Systems: A Comparison*. ERIC.
- [6] Cattaneo, A. A. P., Antonietti, C., & Rausedo, M. (2022). How digitalised are vocational teachers? Assessing digital competence in vocational education and looking at its underlying factors. *Computers & Education*, 176, 104358.
- [7] Chikita, W. (2019). Analysis of item quality in the odd semester final exam for PKSM subject, Grade XI Motorcycle Engineering. *Journal of Vocational Automotive Education*, 2(1), 23–32.
- [8] Cronbach, L. J. (1949). *Essentials of psychological testing*.
- [9] Daniswara, A. A. A., & Nuryana, I. K. D. (2023). Data preprocessing patterns in student assessment of the Teacher Professional Education Program. *Journal of Informatics and Computer Science (JINACS)*, 5(01), 97–100.
- [10] Daryono, R. W., Ramadhan, M. A., Kholifah, N., Isnantyo, F. D., & Nurtanto, M. (2023). An empirical study to evaluate the student competency of vocational education. *International Journal of Evaluation and Research in Education (IJERE)*, 12(2), 1079–1086.
- [11] Diao, J., & Hu, K. (2022). Preparing TVET teachers for sustainable development in the information age: Development and application of the TVET teachers' teaching competency scale. *Sustainability*, 14(18), 11361.
- [12] Hadi, S. (2012). Evaluation of the implementation of a competency-based curriculum in automotive course and training institutions in Yogyakarta. *Journal of Vocational Education*, 2(2), 267–283.
- [13] Hadi, S., Pd, M., Pidato, M. E., Jabatan, P., & Besar, G. (n.d.). *Vocationalism–liberalism: Implications for governance and leadership of vocational education in Indonesia*.
- [14] Hardiyanta, R. A. P., & Wagiran, W. (2023). Evaluation of competency tests in light vehicle engineering at vocational schools in Yogyakarta. *Journal of Vocational Automotive Education*, 5(2), 67–86.
- [15] Jatmoko, D., Basuki, B., Pambudi, A., Setiyono, K. A., & Nugroho, B. (2025). Analysis of basic automotive engineering items using cognitive diagnostic Anates Version 4.02. *Journal of Educational Science Research*, 4(1), 461–471.
- [16] Klotz, V. K., & Winther, E. (2017). Assessing Tomorrow's Potential: a competence measuring approach in vocational education and training. In *Competence Assessment in Education: Research, Models and Instruments* (pp. 221–243). Springer.
- [17] Menchetti, B. M., & Rusch, F. R. (1988). Reliability and validity of the vocational assessment and curriculum guide. *American Journal of Mental Retardation: AJMR*, 93(3), 283–289.
- [18] Muktavia, M., Isnandar, I., & Ichwanto, M. A. (2024). Evaluation of block system learning in the Department of Modeling Design and Building Information at SMK Islam 1 Blitar. *Proceedings of Life and Applied Sciences*, 4(1).
- [19] Monnier, M., Tschöpe, T., Srbeny, C., & Dietzen, A. (2016). Occupation-specific social competences in vocational education and training (VET): the example of a technology-based assessment. *Empirical Research in Vocational Education and Training*, 8(1), 10.
- [20] Murwantini, S. (2023). Optimization of assessment for vocational high schools under the Independent Curriculum. *Steam Engineering*, 4(2), 105–113.
- [21] Pangestika, R. R. (2015). *Teacher Professional Education (PPG): A strategy for developing teacher professionalism and improving the quality of Indonesian education*.
- [22] Perdana, N. S. (2021). Analysis of student readiness in facing minimum competency assesment. *MUKADIMAH: Jurnal Pendidikan, Sejarah, Dan Ilmu-Ilmu Sosial*, 5(1), 15–20.

- [23] Priti, P., Sofyan, H., Budiman, A., & Sriyanto, J. (2023). Evaluation of industrial class programs in the Automotive Engineering major at SMK Negeri 2 Yogyakarta. *Journal of Learning Innovation in Schools*, 4(2), 602–610.
- [24] Samosir, D. M., Munawar, W., & Noor, R. A. M. (2024). Development of cognitive and psychomotor evaluation instruments for replacement of transaxle, motor, and generator. *ATIKANOTO: Journal of Automotive Engineering Education*, 1(1).
- [25] Sensus, M., Arifin, K., & Munir, A. (2022). Item validity in the Minimum Competency Assessment for Grade X High School Ecology. *Wahana-Bio: Journal of Biology and Learning*, 14(1), 1–10.
- [26] Subroto, J. G. (2019). Improving education quality: Teacher Professional Education (PPG) as pre-service professional teacher preparation in Indonesia. Scientific Article. <http://download.garuda.kemdikbud.go.id/article.php>
- [27] Venatius, A. S., Musta'amal, A. H., & Ekwok, O. B. (2023). Developing Framework for Teaching Technical Skills to Metalwork Undergraduates in Vocational-Based Education in Nigeria. *Journal of Technical Education and Training*, 15(2), 11–20. <https://doi.org/10.30880/jtet.2023.15.02.002>
- [28] Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Quality analysis of HOTS-type items in brake system competency at vocational high schools. *Journal of Technology and Vocational Education*, 17(1), 95–104.
- [29] Winther, E., & Klotz, V. K. (2013). Measurement of vocational competences: an analysis of the structure and reliability of current assessment practices in economic domains. *Empirical Research in Vocational Education and Training*, 5(1), 2.
- [30] Witarsa, B. N., Munawar, W., & Berman, E. T. (2017). Development and analysis of test items for Basic Automotive Technology in vocational schools. *Journal of Mechanical Engineering Education*, 4(2), 146–155.