Development of Operational Level Practical Test Instrument by Using Competency-Based Assessment Model

Irwan*, Sarifuddin, Riki Wanda Putra, Hasri Devin
Politeknik Pelayaran Sumatera Barat, Padang Pariaman, Indonesia
*rikiii126@gmail.com

Abstract

The lack relevance of practical ability assessment with industry needs causes nautical technology graduates to not be fully prepared to work in the field. Therefore, it is necessary to develop a practical test instrument that is in accordance with Graduate Learning Outcomes. The purpose of this study is to produce an operational level practical test instrument with a valid, reliable, and effective competency-based assessment model. This type of research is development research by using the Wilson Model and the Oriondo and Antonio Model. Validity test was content validity. Reliability test was test-retest and inter-rater reliability. Effectiveness test used the N-Gain percentage. The results of the study prove that the developed instrument is valid with a validity value 3.8 and criterion A. The instrument is reliable because the t-count value is smaller than the t-table value, meaning that there is no difference between the average values of the three raters. The instrument is effective because the N-Gain percentage = 74.6%. The limitations of this study lie in the limited testing, only focusing on the operational level, and the variation of raters who only come from within the institution. The recommendation from this study is that the developed instrument should be tested in other maritime educational institutions. Then, it is necessary to develop similar instruments for the management level in the maritime sector.

Keywords: Operational Level; Practical Test Instrument; Competency-Based Assessment.

Abstrak

Kurang relevannya penilaian kemampuan praktik dengan kebutuhan industri menyebabkan lulusan teknologi nautika tidak sepenuhnya siap untuk bekerja di lapangan. Sehingga perlu dikembangkan instrumen uji praktik yang sesuai dengan Capaian Pembelajaran Lulusan (CPL). Tujuan penelitian ini yaitu untuk menghasilkan instrumen uji praktik tingkat operasional dengan model penilaian berbasis kompetensi yang valid, reliabel, dan efektif. Jenis penelitian ini adalah penelitian pengembangan dengan menggunakan Model Wilson dan Model Oriondo dan Antonio. Uji validitas menggunakan content validity. Uji reliabilitas dengan test-retest dan inter-rater reliability. Uji efektifitas menggunakan persentase N-Gain. Hasil penelitian membuktikan bahwa instrumen yang dikembangkan valid dengan hasil nilai validitas 3.8 dan kriteria A. Instrumen reliabel karena nilai t-hitung lebih kecil dari pada nilai t-tabel, artinya tidak terdapat perbedaan antara nilai rata-rata dari ketiga penilai. Instrumen efektif karena persentase N-Gain = 74.6%. Keterbatasan penelitian ini terletak pada pengujian yang terbatas, hanya berfokus pada level operasional, dan variasi penilai yang hanya berasal dari internal institusi saja. Rekomendasi dari penelitian ini yaitu agar instrumen yang dikembangkan diuji di berbagai institusi pendidikan maritime. Lalu, perlu dikembangkan instrumen serupa untuk level manajemen di bidang maritim.

Kata Kunci: Level Operasional; Instrumen Uji Praktik; Penilaian Berbasis Kompetensi

Introduction

One of the study programs at Merchant Marine Polytechnic of West Sumatera is the Technology Nautical Study Program. The Graduate Learning Outcome of Technology Nautical Study Program is to produce competent Human Resources in the maritime field, especially the ship engines competence (Asimakopoulos et al., 2023). The success of the Technology Nautical Study Program education is highly dependent on the ability of students to apply theory/concepts of the material into real practice (Bolstad et al., 2020). Therefore, practical learning contains of simulation activities has a vital role for students whether it holds in the workshop or in the laboratory or simulator (Hartikainen et al., 2022).

Lecturers of Technology Nautical Study Program play a very important role in producing practitioners who are capable for the ship engines in accordance with the demands of maritime industry. Lecturers must be able to design and implement learning practices in accordance with the expected procedures and competencies (Sanders et al., 2024). In addition, lecturers must be able to evaluate students' abilities according to established performance indicators as the results from the link and match process with the maritime sector (Al-Bahlani & Ecke, 2023). Evaluation of students' practical abilities requires assessment instruments that are able to measure each performance indicator from each competency (Kárpáti & Paál, 2022).

One of effective ways to evaluate students' practical abilities is through performance assessment with a competency-based assessment model (Kritkharuehart et al., 2024). Competency-based assessment is a model that emphasizes the evaluation of students' real skills and practical abilities (Sellberg et al., 2022). For students of the Technology Nautical Study Program, competency-based assessment must be adjusted to the standards set by the international regulations of the IMO (International Maritime Organization) (Mashartanto et al., 2024). One of the IMO regulations is STCW (Standards Training, Certification, and Watchkeeping) (Ghosh et al., 2024). STCW for students of Technology Nautical Study Program contains specific requirements for training and testing on three bases, namely able to use laboratory equipments and able to operate a simulator from the engine room (Mori & Manuel, 2024).

However, based on the results of interview with officers on board, especially engine room officers, it was found that technology nautical graduates were not fully prepared to work in the field or needed additional time to adjust to industry requirements. Furthermore, based on the results of discussions on knowledge sharing activities of lecturers at higher education institutions under the Ministry of Transportation, students had not fully mastered the skills relevant to the real work in the industry. Students of Technology Nautical Study Program at Merchant Marine Polytechnic of West Sumatera admitted that there were no test instruments provided by lecturers for practical learning in the engine room.

Students were only told the basic competencies that they had to master at each meeting. This proves that the assessments given are not yet competency-based, which often do not include the practical abilities and specific skills needed in the maritime industry (Kebede et al., 2024). The above problems can be caused by several things. First, the lack of practical ability assessments relevance to industry needs (Wakid et al., 2024). The assessments carried out by lecturers are still mostly in the form of traditional assessments because they are easier to design and assess (Sandal, 2023). Another cause is the lack of consistent and objective evaluation standards in measuring students' competency. Assessments that are not competency-based tend to be subjective and unstructured, so that evaluation results can vary between lecturers (Moore et al., 2024).

From the above problems, it is necessary to develop a practical test instrument to measure the competency of students in Technology Nautical Study Program. The development of a competency-based practice test instrument at the operational level aims to ensure that students are able to manage, monitor, and control ship engine systems in accordance with the expected performance standards (Holmström & Stjärnhagen, 2024). This instrument is designed with a competency-based assessment model that focuses on cognitive aspects, practical skills and professional attitudes (Richter & Kjellgren, 2024). There are several previous studies related to the development of assessment instruments in learning.

First, research on the development of critical thinking measurement instruments for students (Rahmawati & Harun, 2019). Second, research on the development of instruments to measure students' academic potential (Septianingsih and Jerusalem, 2021). Third, research on the development of the framework for assessing students' scientific competence (Intasoi et al., 2020). Fourth, research on the development of language assessment as a teacher's personal competence (Sukenti et al., 2020). Fifth, research on the development of students' mathematics ability assessment instruments (Kusmaryono et al., 2018). Sixth, research on the development of student project-based assessment instruments (Ofianto et al., 2024).

The novelty of this research is in the type of instrument developed, namely an operational level practical test instrument with a competency-based assessment model. In addition, this instrument was developed for maritime students, especially students of the technology nautical study program. Thus, the objectives of this study are to, (1) produce the validity of operational level practical test instruments with a competency-based assessment model, (2) produce the reliability of operational level practical tests with a competency-based assessment model, and (3) produce the effectiveness of operational level practical tests with a competency-based assessment model.

Method

This type of research is Research and Development (R&D) by using the Wilson Model and the Oriondo and Antonio Model. The steps taken in this R&D can be seen in Figure 1. The population used in this study were all students of Technology Nautical study program, Batch VII, Merchant Marine Polytechnic of West Sumatera, that consists of 6 (six) classes. Each class consists of 20 people, so that the total population is 120 people. The sample in this study was selected by using the probability sampling technique, so that all students have the same opportunity to become research samples. From the 120 population, 60 samples were selected randomly; 30 people were used as samples for validity and reliability tests, while 30 people were used for effectiveness tests. This research was conducted in the Engine Room Simulator of Merchant Marine Polytechnic of West Sumatera, for 8 months, starting from April 2024 to November 2024. The research instruments are instrument validation sheets by validators, and operational level practice test instruments. Data collection techniques are carried out by processing the results of instrument validation by 4 validators, Inter-Rater reliability and Test-Retest method to test the reliability and Calculating N-Gain Percentage to test the effectiveness of the instrument. Data analysis techniques are carried out by conducting validity tests, reliability tests, and effectiveness tests.

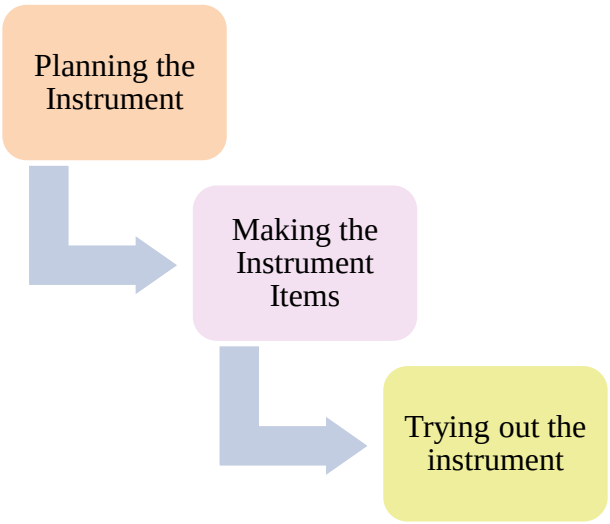

Figure 1. Wilson Model and Oriondo & Antonio Model

The Instrument Development Procedure Can Be Seen in Table 1.

Table 1. The Procedure of Practical Test Instrument Development

Development Stages	Activity
Planning The Instrument	Determining Graduate Learning Outcomes
	Determining the core competencies of STCW
	Determining the material tested in the practical test in the engine room for the operational level
Making the Instrument Items	Arranging an instrument grid by considering the core competencies and sub-competencies expected of students
	Arranging items for implementing competency standards with 2 assessment options (YES) or (NO)
	Designing practice scenarios
	Testing the validity of items by experts
Trying Out the Instrument	Conducting reliability tests on research samples
	Conducting the first trial in May 2024
	Conducting the second trial in August 2024
	Melakukan uji coba kedua pada bulan Agustus 2024
	Conducting instrument test analysis with the help of a calculation program from Microsoft Excel

Results and Discussion

1. Instrument Validity

The development of this practical test instrument used content validity and face validity indicators from Masuwai et al., (2024) which can be seen in table 2.

Table 2. Content Validity Indicators

No.	Indicator	Code
1	Reflection of Material Knowledge	CV1
2	Reflection of Technology Nautical Competence	CV2
3	Reflection of Ship Engineering Teaching	CV3
4	Suitability of material with indicator items	CV4
5	Suitability of language and sentence structure	CV5
6	Suitability of core competencies with Graduates Learning Outcomes	CV6

This validity test was conducted by 4 expert validators; 3 validators come from Technology Nautical experts, and 1 validator is language expert. The research validators can be seen in table 3.

Table 3. Instrument's Validators

No.	Aspect	Name	Position
1	Content/Material	Dr. Sarifuddin, M.Pd., M. Mar.E	Chief Engineer (ATT 1)
2	Content/Material	Dr. (Cand) Syamsyir, S.T., M.T., M. Mar. E	Chief Engineer (ATT 1)
3	Content/Material	Hasri Devin, S.T., M. Mar.E	Chief Engineer (ATT 1)
4	Language	Syafni Yelvi Siska, M. Pd	English Lecturer at Merchant Marine Polytechnic of West Sumatera

The results of the validity test from the four validators can be seen in table 4.

Table 4. Content Validity Score

No.	Validator	Evaluation Indicators					
		CV1	CV2	CV3	CV4	CV5	CV6
1	I	4	3	4	3	4	4
2	II	4	4	4	3	4	3
3	III	4	4	4	4	4	3
4	IV	4	4	4	3	4	4
Mean		4	3.75	4	3.25	4	3.5
Score		3.8					
Criteria		A					

Based on the assessment results of 4 validators above, the content validity of this practical test instrument are 3.8 with a value category of A. Thus, the operational level practical test instrument sheet with competency-based assessment model is valid and it can be used in engine room simulator practice learning.

2. Instrument Reliability

Instrument reliability is the consistency of the measuring instrument, whether the measuring instrument remains consistent if the measurement is repeated. Instrument reliability testing is usually carried out after the instrument is valid (Latifah et al., 2024). Reliability testing is carried out by using inter-rater reliability (Rohde et al., 2022). An instrument is said to be reliable if it produces the same assessment results even though the assessment is carried out repeatedly. Reliability testing is carried out using evaluation from the three raters (P1, P2, P3) during the test-retest. The results of the Reliability Test from P1, P2, and P3 can be seen in table 5.

Table 5. Results of Inter-Rater Reliability Test

Subject	P1 test	P1 re-test	P2 test	P2 re-test	P3 test	P3 re-test
1	80	82	82	82	82	80
2	79	82	82	80	80	80
3	82	82	82	82	82	82
4	80	80	80	82	79	80
5	80	80	80	80	80	80
6	80	82	82	82	82	80

7	82	80	80	82	79	82
8	82	82	82	82	82	80
9	80	80	80	80	80	80
10	82	82	82	80	80	82
11	79	82	82	80	80	82
12	80	80	80	82	82	80
13	82	82	82	82	82	82
14	80	82	79	80	80	80
15	79	82	82	80	79	80
16	82	82	82	80	80	82
17	80	80	80	82	79	80
18	82	80	80	82	82	80
19	80	80	80	80	80	80
20	82	82	79	82	82	82
21	79	82	82	80	80	80
22	80	82	82	80	80	80
23	82	80	80	82	82	82
24	79	80	82	80	80	86
25	82	82	79	82	82	82
26	80	82	82	80	80	80
27	79	82	82	80	82	82
28	80	82	82	80	80	80
29	82	80	80	80	82	82
30	80	82	82	80	80	80
Total	2416	2438	2431	2426	2420	2428
Mean	80.53	81.27	81.03	80.87	80.67	80.93
t-table	2.045		2.045		2.045	
t-count	-0.0952		0.017		0.027	

From the test-retest results of the three raters above, it can be stated that the t-count value is smaller than the t-table value (t count < t table), so it can be stated that there is no difference between the average values of the three raters. Thus, it can be concluded that the operational level practice test instrument with a competency-based assessment model is reliable and has met the requirements to be used as a practical test instrument in the engine room simulator.

3. Effectiveness Instrument

The effectiveness test aims to see the extent to which the operational level practice test instrument with a competency-based assessment model is effective to use. The effectiveness test of the instrument is carried out by calculating the N-Gain Score on the difference between the students' pre-test and post-test scores (Lai et al., 2023). The effectiveness test is carried out by using SPSS application. The calculation of the N-Gain Score using the SPSS application can be seen in table 6.

Table 6. Calculation of the N-Gain Score

No.	Pre	Post	Pre-Post	Ideal Score -Pre	N-Gain Score	N-Gain %
1	80	85	5	10	0.50	50
2	79	86	7	11	0.64	63.64
3	82	85	3	8	0.38	37.50
4	80	86	6	10	0.60	60

5	80	86	6	10	0.60	60
6	80	85	5	10	0.50	50
7	82	87	5	8	0.63	62.50
8	82	88	6	8	0.75	75
9	80	90	10	10	1	100
10	82	90	8	8	1	100
11	79	85	6	11	0.55	54.55
12	80	88	8	10	0.80	80
13	82	88	6	8	0.75	75
14	80	90	10	10	1	100
15	79	85	6	11	0.55	54.55
16	82	86	4	8	0.50	50
17	80	86	6	10	0.60	60
18	82	87	5	8	0.63	62.50
19	80	89	9	10	0.90	90
20	82	90	8	8	1	100
21	79	90	11	11	1	100
22	80	90	10	10	1	100
23	82	87	5	8	0.63	62.50
24	79	88	9	11	0.82	81.82
25	82	88	6	8	0.75	75
26	80	89	9	10	0.90	90
27	79	86	7	11	0.64	63.64
28	80	90	10	10	1	100
29	82	90	8	8	1	100
30	80	88	8	10	0.80	80

The descriptive statistical analysis of the N-Gain calculation from the above data can be seen in table 7.

Tabel 7. N-Gain Percentage

	N	Min.	Max.	Mean	SD
N-Gain Score	30	0.38	1	0.7461	0.1953
N-Gain Percentage	30	37.50	100	74.60	19.53

The Results Obtained Are Then Compared with The Intervals in Table 8 as follows.

Table 8. Effectiveness Conversion

Interval N Percentage	Conversion
85% < N ≤ 100%	Very Effective
72% < N ≤ 85%	Effective
58% < N ≤ 72%	Quite Effective
44% < N ≤ 58%	Not Effective
N ≤ 44%	Very Ineffective

Based on the results of SPSS calculation, it was found that the N-Gain percentage = 74.60% which is in the effective category. Thus, the operational level practice test instrument with a competency-based assessment model can be used as a practice test instrument on the engine room simulator for the next batch of technology nautical study program.

The results of validity test state that the operational level practice test instrument with a competency-based assessment model is valid for testing the operational level competency of technology nautical study program students in the engine room. Validity is the main test that must be carried out on a learning assessment material or instrument assessment (Yulia and Nasution, 2024). The validity test carried out in this study is content validity which is most often carried out by researchers in developing instrument or subject matter development research (Sabaruddin et al., 2022; Zain et al., 2022; Angraini et al., 2021). Furthermore, the results of reliability test for operational level practice test instrument with a competency-based assessment model using inter-rater reliability and test-retest are reliable. Inter-rater reliability is an evaluation carried out by more than one assessor (rater) which makes the assessment more objective and consistent (Kubicki et al., 2020).

In addition, inter-rater reliability increases the reliability of assessment results and minimizes individual bias, and makes decision-making more accurate, so that the developed instrument can be said to be reliable and can be used as a tool for evaluating student competency (Eltayar et al., 2022). Test-retest reliability is used to measure the consistency of test results or assessment instruments over time (Mehta and O'Connor, 2023). This is done by giving the same test or instrument to a group of individuals on two different occasions, then comparing the results to see if there is consistency. If the results of the two tests are very similar or show a high correlation, then the instrument is considered to have high reliability (Dragostinov and Mõttus, 2023). Then, the operational level practice test instrument with a competency-based assessment model is effective. It means that the measuring instrument can be used well in the context of learning testing, and the results obtained from its use can be trusted (Alt et al., 2023).

An effective instrument is able to accurately distinguish students with high, medium, and low levels of competence, thus allowing for proper assessment of different groups (Moore et al., 2024). The competency-based assessment model on the operational level practical test instrument is a novelty of this study. The development of instruments that have been carried out in previous studies is by using the Rasch Model (Yamtinah et al., 2022; Raof et al., 2021), using an assessment rubric (Andermo et al., 2023), using a practice scenario (Chanmas et al., 2023), nd using an assessment scale (Guvey Aktay and Mermi, 2022; Yildiz et al., 2022; Nada et al., 2022). A competency-based assessment model is important to be developed by lecturers for students of technology nautical study program because it focuses on measuring practical skills that are relevant to real work in the maritime industry (Abduh et al., 2022).

Conclusion

This study successfully developed a valid, reliable, and effective competency-based assessment (CBA) operational practice test instrument to measure the competency of students in technology nautical study program. This instrument provides a significant contribution in supporting the achievement of students' Competency Standards according to STCW and producing competent graduates in the maritime industry. The limitations of this study lie in the limited testing, only focusing on the operational level, and the assessors (raters) which only comes from within the institution. The recommendation from this study is that the developed instrument should be tested in other maritime educational institutions to ensure its reliability in a broader context. In addition, as a follow-up step, it is necessary to develop a similar instrument for the management level in the maritime field. This will ensure that competency assessment can cover all levels of hierarchy in seafarer education, from operational to managerial.

References

Abduh, M., Hasnur, J., & Siska, S. Y. (2022). The Effect Of Maritime English Vocabulary For Beginners Module On The Vocabulary Learning Outcomes. *Jurnal Pendidikan Vokasi*, *12*(2), 117-129.

Al-Bahlani, S. M., & Ecke, P. (2023). Assessment Competence And Practices Including Digital Assessment Literacy Of Postsecondary English Language Teachers In Oman. *Cogent Education*, *10*(2), 2239535.

Alt, D., Schneider, N. L., & Weishut, D. J. (2023). Competency-Based Learning And Formative Assessment Feedback As Precursors Of College Students' Soft Skills Acquisition. *Studies in Higher Education*, *48*(12), 1901-1917.

Andermo, S., Ehrlich, F. K., Larm, F. M., Bergström, L., Siljehag, A. P., & Broberger, E. (2023). Assessing Students Interprofessional Competence Using A Swedish Version Of The Interprofessional Collaborator Assessment Rubric. *Journal of Interprofessional Care*, *37*(4), 605-612.

Angraini, A. R. A., & Okfalisa. (2021). Measuring Information Security Policy Compliance: Content Validity Of Questionnaire. *Indonesian Journal of Electrical Engineering and Computer Science*, *22*(1), 469-475.

Asimakopoulos, I., Valencia, A. L. D., Lützen, M., & Rytter, N. G. M. (2024). Data-Driven Condition Monitoring Of Two-Stroke Marine Diesel Engine Piston Rings With Machine Learning. *Ships and Offshore Structures*, *19*(9), 1241-1253.

Bolstad, T., Wallin, P., Lundheim, L., Larsen, B. B., & Tybell, T. (2021). Emergent Premises In Student Experiences Of A First-Year Electrical Engineering Course. *European Journal of Engineering Education*, *46*(2), 302-317.

Chanmas, G., Taveekitworachai, P., Paliyawan, P., Thawonmas, R., Thawonmas, R., Nukoolkit, C., & Dajpratham, P. (2023). Driving Scenarios And Environmental Settings In Simulator-Based Driving Assessment Systems For Stroke: A Systematic Review. *Topics in Stroke Rehabilitation*, *30*(8), 872-880.

Dragostinov, Y., & Mõttus, R. (2023). Test-Retest Reliability and Construct Validity of the Brief Dark Triad Measurements. *Journal of Personality Assessment*, *105*(2), 143-148.

Eltayar, A. N., Aref, S. R., Khalifa, H. M., & Hammad, A. S. (2022). Do Entrustment Scales Make A Difference In The Inter-Rater Reliability Of The Workplace-Based Assessment?. *Medical Education Online*, *27*(1), 2053401.

Ghosh, S., Emad, G. R., & Ravi, A. (2024). Investigating The Characteristics Of Skills And Competency Frameworks Through A Systematic Literature Review: A Feasibility Study To Revise The Stcw Code For Seafarer Training. *Australian Journal of Maritime & Ocean Affairs*, 1-17.

Guvey, A. E., & Mermi, F. (2022). Speaking Skills Attitude Scale for Primary School Students: Validity and Reliability Study. In *Elementary School Forum (Mimbar Sekolah Dasar)*, 9(1), 43-57.

Hartikainen, S., Pylväs, L., & Nokelainen, P. (2022). Engineering Students' Perceptions Of Teaching: Teacher-Created Atmosphere And Teaching Procedures As Triggers Of Student Emotions. *European Journal of Engineering Education*, *47*(5), 814-832.

Holmström, O., & Stjärnhagen, O. (2024). Coordination Of Courses In University Programmes And Students' Experiences Of Their Studies: Student Perspectives On The Importance Of Course Coordination. *Quality in Higher Education*, *30*(2), 221-238.

Intasoi, S., Junpeng, P., Tang, K. N., Ketchatturat, J., Zhang, Y., & Wilson, M. (2020). Developing An Assessment Framework Of Multidimensional Scientific Competencies. *International Journal of Evaluation and Research in Education*, *9*(4), 963–970.

Kárpáti, A., & Paál, Z. (2022). Assessment Of Visual Sub-Competencies Through Visual Rubrics: Case Studies Based On The Common European Framework Of Reference of Visual Competencies (CEFR-VC). *Journal of Visual Literacy*, *41*(4), 224–246.

Kebede, A., Asgedom, A., & Asfaw, A. (2024). Is linking Technical And Vocational Training With Industry A Bridge To Employability? Lessons From A Literature Review For Ethiopia. *Cogent Education*, *11*(1), 2406721.

Kritkharuehart, S., Vimolsilp, U., & Photipussa, P. (2024). Development Of Competency-Based Assessment Model For KRU RAK THIN Scholarship'Recipients. *Cogent Education*, *11*(1), 2373230.

Kubicki, A., Brika, M., Coquisart, L., Basile, G., Laroche, D., & Mourey, F. (2020). The Frail'Bestest. An Adaptation Of The "Balance Evaluation System Test" For Frail Older Adults. Description, Internal Consistency And Inter-Rater Reliability. *Clinical Interventions in Aging*, *15*, 1249-1262.

Kusmaryono, I., Suyitno, H., Dwijanto, D., & Dwidayati, N. (2018). Developing Assessment Instrument as a Mathematical Power Measurement. *Journal of Education and Learning (EduLearn)*, *12*(3), 382-391.

Lai, Y. H., Wu, M. J., Chen, H. H., Lin, S. P., Wu, C. Y., Chin, C. S., Lin, C. H., Shiu, S. I., Lin, Y. C., Chen, H. C., Hou, S. C., & Chang, C. W. (2023). Impacts Of Huddle Intervention On The Patient Safety Culture Of Medical Team Members In Medical Ward: One-Group Pretest-Posttest Design. *Journal of Multidisciplinary Healthcare*, *16*, 3599-3607.

Latifah, M., Saripah, I., Suryana, D., & Sunarya, Y. (2024). Validity and Reliability of Self-Concept Instrument Using Rasch Model. *Jurnal Kajian Bimbingan Dan Konseling*, *9*(1), 26-35.

Mashartanto, A., Pranata, W., & Siska, S. Y. (2024). Development of Learning Media Maritime English Textbook for Ratings Forming. *JOLLT: Journal of Languages and Language Teaching*, *12*(3), 1238–1252.

Masuwai, A., Zulkifli, H., & Hamzah, M. I. (2024). Evaluation Of Content Validity And Face Validity Of Secondary School Islamic Education Teacher Self-Assessment Instrument. *Cogent Education*, *11*(1), 2308410.

Mehta, J., & Connor, O. A. (2023). Test Retest Variability In Stereoacuity Measurements. *Strabismus*, *31*(3), 188–196.

Moore, A. M., Fairhurst, P., Bennett, J. M., Harrison, C., Correia, C. F., & Durk, J. (2024). Assessment And Practical Science: Identifying Generalizable Characteristics Of Written Assessments That Reward And Incentivise Effective Practices In Practical Science Lessons. *International Journal of Science Education*, *46*(7), 643-669.

Mori, Y., & Manuel, M. (2024). An Exploratory Study Of Approaches To Onboard Training In Different Countries. *Journal of International Maritime Safety, Environmental Affairs, and Shipping*, *8*(3), 2386221.

Nada, Q., Herdiana, I., & Andriani, F. (2022). Testing The Validity And Reliability Of The Depression Anxiety Stress Scale (DASS)-21 Instrument For Individuals With Psychodermatology. *Psikohumaniora*, *7*(2), 153–168.

Ofianto, S. N., Ningsih, T. Z., Nababan, S. A., Asri, Z., Aman, & Setiawan, J. (2024). Developing An Online Project-Based Assessment Instrument For Assessing Students' Historical Thinking Skills. *International Journal of Evaluation and Research in Education*, *13*(6), 3993-4003.

Rahmawati, E., & Harun, H. (2019). Developing Instruments Of Teacher's Perception Of Critical Thinking In Elementary School. *Journal of Education and Learning (EduLearn)*, *13*(4), 559–566.

Raof, S. A., Musta'amal, A. H., Zamzuri, F. K., & Salleh, M. H. (2021). Validity and Reliability of Students Perceptions on OBE Approach in Malaysian VC Using Rasch Model. *Journal of Innovation in Educational and Cultural Research*, *2*(2), 44–50.

Richter, T., & Kjellgren, B. (2024). Engineers Of The Future: Student Perspectives On Integrating Global Competence In Their Education. *European Journal of Engineering Education*, *49*(3), 474–491.

Rohde, A., McCracken, M., Worrall, L., Farrell, A., O'Halloran, R., Godecke, E., David, M., & Doi, S. A. (2022). Inter-Rater Reliability, Intra-Rater Reliability And Internal Consistency Of The Brisbane Evidence-Based Language Test. *Disability and Rehabilitation*, *44*(4), 637–645.

Sabaruddin, S., Nur, M., Fadli, M., & Mazlan, M. (2022). Content Validity of KIP-K Scholarship Instruments at Higher Education Using Aiken's Coefficient Validity Scale. *AL-TANZIM: Jurnal Manajemen Pendidikan Islam*, *6*(3), 934–947.

Sandal, A. K. (2023). Vocational Teachers` Professional Development In Assessment For Learning. *Journal of Vocational Education and Training*, *75*(4), 654–676.

Sanders, J., Johnson, E., Mirabelli, J., Kunze, A., Vohra, S., & Jensen, K. (2024). Engineering Professor Perceptions Of Undergraduate Engineering Student Stress. *European Journal of Engineering Education*, 1–21.

Sellberg, C., Lundin, M., & Säljö, R. (2022). Assessment In The Zone Of Proximal Development: Simulator-Based Competence Tests And The Dynamic Evaluation Of Knowledge-In-Action. *Classroom Discourse*, *13*(4), 365–385.

Septianingsih, E., & Jerusalem, M. A. (2021). Developing Instrument Of Academic Potential Test Analogy Verbal Ability For Undergraduate Students. *Journal of Education and Learning (EduLearn)*, *15*(2), 234–241.

Sukenti, D., & Tambak, S. (2020). Developing Indonesian Language Learning Assessments: Strengthening the Personal Competence and Islamic Psychosocial of Teachers. *International Journal of Evaluation and research In education*, *9*(4), 1079-1087.

Wakid, M., Sofyan, H., Widowati, A., & Zaida, I. A. (2024). Learning-Oriented Assessment: A Systematic Literature Network Analysis. *Cogent Education*, *11*(1), 2366075.

Yamtinah, S., Dewi, M. C., Nurhayati, N. D., Susilowati, E., Saputro, S., Fakhrudin, I. A., Ramadhani, D. G., & Shidiq, A. S. (2022). Content Validity in Android-Based Augmented Reality Media for High School Science Students on Covalent Bonds Topic: Rasch Model Analysis. *Jurnal Pendidikan Sains Indonesia*, *10*(2), 240–249.

Yildiz, A., Butun, M., Turkdogan, A., & Kocak, E. (2022). Development of a Perception Scale of Private Lesson: A Validity and Reliability Study. *Mimbar Sekolah Dasar*, *9*(3), 517–534.

Yulia, P., & Nasution, E. Y. P. (2024). Geometry and Islamic Values: Validity of Teaching Materials Based on Modified Project-Based Learning Model. *Mosharafa: Jurnal Pendidikan Matematika*, *13*(1), 113–124.

Zain, N. H. M., Aziz, S. R. A., Moketar, N. A., Ishak, N. H., & Hendrayati, H. (2022). Content Validity Study: A Gamification Model To Drive Behavior Change In Defeating Coronavirus Disease 2019 Pandemic. *International Journal of Electrical and Computer Engineering*, *12*(5), 5375-5382.