

DEVELOPMENT OF BASIC TACTICAL SIGNAL INSTRUMENT: USING ITEM RESPONSE THEORY APPROACH WITH RASCH**Zidni Husnurofik¹, Anan Sutisna², Achmad Ridwan³,**

1,2,3 Doctoral Program in Educational Research and Evaluation, Universitas Negeri Jakarta (UNJ), Jakarta, Indonesia.

¹ huseinhusen44@gmail.com ² asutisna@unj.ac.id ³ achmadridwan.unj@gmail.com**ABSTRACT**

This study aims to develop a basic tactical progression test instrument using Item Response Theory with the Rasch Model approach. The test results to analyze the test instrument used to measure person abilities. Sampling using purposive sampling technique consisting of 300 people. The questions given in the form of multiple-choice questions as many as 15 items related to the basic tactical progression test. Data analysis techniques using quantitative descriptive analysis. Rasch Model is used to obtain fit items. The validity and reliability results were 0.87 and 0.83. All items fit the Rasch Model. The end of basic tactical progression test instrument has been adjusted for person with the ability of $-4 < \theta < 4$ which is indicated by the ability that has the maximum value for the information test or the minimum value for the standard error.

Keywords: *Item Response Theory, Rasch Model, basic tactical progression test***1. INTRODUCTION**

Close Quarter Battle (CQB) is one of the tactical signals faced by the police, especially in dealing with threats in closed areas (Ibrahim et al., 2024) such as buildings, vehicles, or narrow alleys. CQB operations are fast-paced, dynamic, and stressful, requiring high coordination between team members. Effective signaling ensures each team member can respond quickly without the need for verbal communication that risks detection (Chala & Kóczy, 2024; Han et al., 2024; Sangeetha et al., 2024; Wozniak & Zahabi, 2024). Mastery of signals by police personnel in CQB tasks requires intensive training and understanding of signals. Integrating signals in strategic operations, police teams can increase effectiveness, accelerate decision-making, and ensure member safety and overall mission success.

The principle of using signals can be applied in various fields, such as education, managerial, and industry, especially in the briefing stage or giving directions. The application of the field of education participants understand the role and effective use of signals. Police experience in the use of signals at CQB can be an inspiration for other fields to improve the effectiveness of briefing and task execution in various challenging situations. The use of CQB signals is carried out to measure the ability of police members in intensive training and understanding of signals as an assessment. Therefore, it requires an instrument as an assessment of police officers in improving the skills of using CQB signals.

Assessment is a systematic and continuous process or activity to collect information about the results of participants' abilities in the learning process (Kyndt & Baert, 2015; Maisarah et al., 2017; Rudibyani et al., 2020). Assessment is carried out to evaluate learning outcomes as a determinant of the direction of the next learning process (Putri et al., 2022; Supena et al., 2021) so that an assessment instrument is needed. Assessment instruments can be used to assess the results of feedback on learning (Ghomi & Redecker, 2019; Mertha & Mahfud, 2022; Soeharto et al., 2019). Assessment instruments have a function to determine the effectiveness of the participants' ability stage in understanding the material (Ramdiah et al., 2020; Sugiyanto et al., 2020) which includes aspects of the cognitive, affective, and psychomotor domains both in groups and individually. Assessment instruments are in the form of written, oral tests, or observation sheets (Bayram et al., 2021; Erfianti et al., 2019). The test instrument used as an assessment has valid measuring power. Instruments that do not have valid measuring power will not provide information regarding the ability of test takers (Hadi et al., 2022; Putra et al., 2021). There are several test instruments that have low question quality (Awada et al., 2021; Morawska et al., 2018; Trockel et al., 2018) so that the assessment becomes erroneous and has an impact on the ability of participants who are not measured.

Instruments used as assessments require item analysis (Dwijayani, 2019; Hamilton et al., 2021; Shaw et al., 2020; Supena et al., 2021) to determine the quality of questions through revision or discarding ineffective questions. Item analysis through Classical Test Theory and Item Response Theory. Classical test theory has weaknesses including measurement results will depend on the characteristics of the test used, item parameters depend on the ability of test takers and measurement errors can only be recognized by groups, not individuals (Mardapi, 2012). In contrast, the Item Response Theory (IRT) approach is a general framework for determining mathematical functions that describe the interaction of test takers with test items (Belzak, 2023; Kong & Wang, 2021; Min & Aryadoust, 2021). This suggests that low-ability test takers and high-ability test takers will have a greater chance of answering correctly (Retnawati, 2016; Zeigenfuse et al., 2020).

2. METHODOLOGY

2.1 Participants: Sampling using purposive sampling technique. The research subjects were 300 brimob members who were already active in battalion and detachment units throughout Indonesia random and were male.

2.2 Instruments: This research focuses on the development of a Basic Tactical policing test. There are 15 multiple choice questions with dichotomous scores on the test instrument. Test development is based on a grid with Understanding, applying, and analyzing basic knowledge of tactical signals in effective communication skills. Cognitive levels in the question matrix include C2 (understand), C3 (apply) and C4 (analyze).

2.3 Data Collection: The approach used is the Rasch model. Quantitative data analysis techniques conducted in this study to measure content validation and construct validation. Content validation for question items will be analyzed using the Aiken formula (Aiken, 1980). While construct validation is divided into prerequisite tests and Item Response Theory. Item characteristics are reviewed based on content validation, construct validity, test reliability, and item difficulty index.

2.4 Data Analysis: The analysis used R studio software. The results of R studio output are obtained from several item parameters in accordance with the Rasch model. The output in the form of z.outfit and z.infit as well as the correlation value of the item with the question as a whole shows the limit of items that are declared fit with the model. The reliability of the instrument relating to measurement error is indicated by the Cronbach's Alpha coefficient.

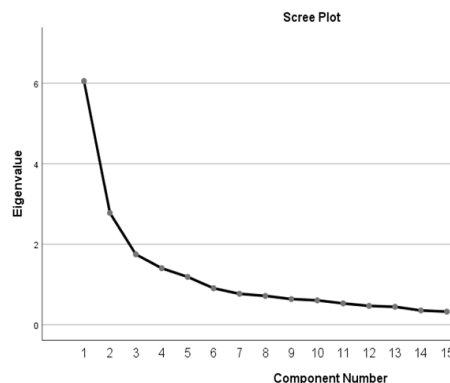
3. RESULTS

Data analysis using item response theory is carried out in testing the dimensions of empirical data (Dewanti et al., 2020). The results of unidimensional factor analysis with the KMO value to determine the feasibility of data with criteria > 0.5 with Barlett's significance value < 0.05 in Table 1 regarding the results of KMO analysis.

Table 1. KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,804
Bartlett's Test of Sphericity	Approx. Chi-Square	2177,869
	df	435
	Sig.	0,000

The number of factors can be determined by selecting factors that have an Eigen value greater than 1 so that 5 factors with Eigen value are formed (Dewanti et al., 2020). This can be seen in Figure 1 scree plot eigenvalue to determine the right number of components that are unidimensional by looking for arches.

**Figure 1. Scree Plot Eigen Value**

The results of data analysis in Table 2 show a Cronbach's Alpha reliability value of 0.83.

Table 2. Reliabilas of Instrumen Test

Cronbach's Alpha	N of Items
0,83	15

The data were analyzed with item response theory using the Rasch Model. The outfit value criteria can be seen in Table 3 in R studio. These results show the item validity information for all classified items fit the Rasch model.

Table 3. Criteria score outfit

Outfit Value	Description
>1.33	Not Fit Model
0.77-1.33	Fit Model
<0.77	Not Fit Model

The outfit value indicates that an ideal item must meet the requirements of the item fit test. All 15 items met the fit test by 100% as shown in Table 4.

Table 4. Item fit model

Item	outfit	Description
Item_01	1.027	Fit Model
Item_02	1.037	Fit Model
Item_03	1.011	Fit Model
Item_04	0.810	Fit Model
Item_05	0.898	Fit Model
Item_06	0.699	Fit Model
Item_07	0.862	Fit Model
Item_08	0.818	Fit Model
Item_09	0.822	Fit Model
Item_10	0.941	Fit Model
Item_11	1.146	Fit Model
Item_12	0.991	Fit Model
Item_13	0.970	Fit Model
Item_14	0.875	Fit Model
Item_15	1.080	Fit Model

The difficulty index for 15 items is classified as good with the categories of easy, medium and difficult questions. The average item characteristics are in the medium category. The item characteristics are shown with the item characteristic curve (ICC) in Figure 2.

Table 5. Question Item Difficulty Level

Item	Threshold	Criteria
Item_01	-1,05	Medium
Item_02	-0,64	Medium
Item_03	-1,08	Medium
Item_04	-1,97	Medium
Item_05	-1,03	Medium
Item_06	-2,34	Easy
Item_07	-0,95	Medium
Item_08	-0,66	Medium
Item_09	-1,19	Medium
Item_10	-0,14	Medium
Item_11	0,18	Medium
Item_12	1,01	Difficult
Item_13	0,18	Medium
Item_14	-1,08	Easy
Item_15	-0,97	Medium

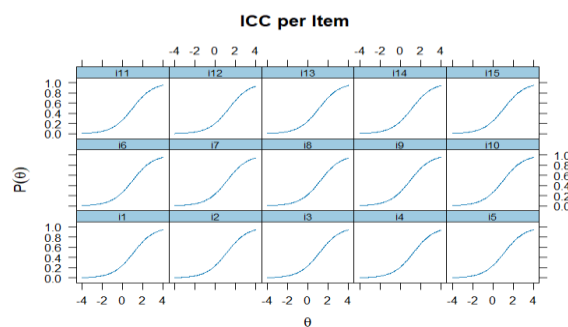


Figure 2. Characteristics ICC item

The smaller the measurement error, the better the instrument because the information function of the test will be better as shown in Figure 4 regarding Information Function and Standard Error Total Item.

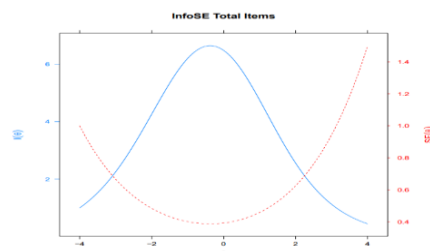


Figure 3. Informasi Function and Standar Error Total Items

The sum of the information functions of the items that make up the test can be expressed as the test information function (Hambleton et al., 1991). The results of the analysis using item parameters show that the developed test will provide maximum information of 6.6 with a standard error of 0.4.

4. DISCUSSION

Testing was conducted to determine content validity and construct validity. Content validation using Aiken met the criteria for valid items of 0.87 with a very good category. Content validity relates to a rational analysis of the domain to be measured to determine representation with the ability to be measured (Abidin & Retnawati, 2019). Construct validity uses exploratory factor analysis. Based on anti-image correlation, the Measures of Adequate Sampling (MSA) value is > 0.5 so that the requirements for factor analysis are met. The MSA value for the diagonal matrix shows that 15 items have an MSA value > 0.5 so that they qualify for factor analysis. Unidimensionality in the test instrument can be proven by the eigenvalues that form the factors in the previous factor analysis test. Test instruments can fulfill unidimensionality if there are factors that are more dominant than other factors. Based on empirical data, the eigenvalue on the scree plot shows that there are 5 components that have a value above 1 so that the test instrument is unidimensional. Test instrument development must meet reliability criteria. A reliable instrument will be able to measure a variable consistently even if it is tested repeatedly even at different times (Kriswanto et al., 2021) which has reliable measurement results credibility, trustworthiness, reliability, consistency, stability (Chakrabarty, 2013). Information function as a determinant of items that are suitable in the model so that items can be selected easily. The amount of item contribution to the information function value can reveal the ability of test takers (Safaruddin et al., 2012). The method of determining the index of test item parameters and the ability of test takers.

The test information function is correlated with the item information function. The value of the item parameter index as a result of estimation is inseparable from measurement error. The standard error of measurement has an inverse quadratic relationship with the information function (Hambleton et al., 1991). The Basic Tactical test instrument has been adjusted for police participants with abilities of $-4 < \theta < 4$ as indicated by abilities that have maximum values for test information or minimum values for standard error.

This study aims to develop a Basic Tactical policing test instrument. A good instrument has accurate results as the goal of assessment (Aldosari et al., 2023; Bajamal et al., 2023; Heydari et al., 2023). The validation process involves gathering evidence to demonstrate the scientific basis for interpreting scores as planned. Validity is the support for the interpretation of test scores against the intended use of the test based on evidence and theory (Mardapi, 2017). Construct validity refers to the extent to which test items can measure what is actually measured based on a particular concept or a specified concept definition (Sustekova et al., 2019). Item Response Theory is a quantitative analysis that is able to measure the ability of participants' characteristics in responding to a series of questions (items) of a test (Caycho-Rodríguez et al., 2023; Muranaka et al., 2023; Robitzsch & Lüdtke, 2023). The IRT model used is the Rasch Model.

The advantage of the Rasch Model is that two parameters are estimated including the test taker ability parameter and the item difficulty parameter (Bernshausen et al., 2022). The test taker ability parameter measures the ability level or characteristics of the test. Test takers with higher ability have higher logit values indicating that test takers answer correctly on more difficult items (Holzknecht et al., 2021; Sternler & Naples, 2021). The higher the ability parameter value, the higher the ability level of the test taker. The item difficulty parameter as a parameter to measure how difficult a test item is measured by a logit scale. Test takers with a higher probability of answering an item correctly have a higher test taker ability parameter than the item difficulty parameter. Therefore, the difficulty index for 15 items is classified as good with the categories of easy, medium, and difficult questions.

Easier items have lower difficulty values, while more difficult items have higher difficulty values. The Rasch Model can identify items tailored to the test's ability level as an accurate assessment of item quality and validity in instrument development (Feng et al., 2023; Meng et al., 2023). Analysis using the Rasch Model can measure and understand individual characteristics more accurately, identify items that need improvement, and improve overall test quality (Cao et al., 2023; Donhauser et al., 2023; Skyttberg et al., 2023). That the assumption of a test taker's probability of answering correctly on a given item depends only on the test taker's ability and item characteristics. The development of test instruments has been carried out a lot, including research conducted (Widyaningih et al., 2021) in developing physics questions with Rasch Model analysis through the QUEST program with the item fit test using INFIT MNSQ. Another research conducted by (Septiani & Paidi, 2021) in developing multiple choice test instruments, the goodness fit test uses Infit Mean Square, Outfit Mean Square, Infit t, and Outfit t which are analyzed by the Rasch model through the QUEST program. Other research conducted by (Harjo et al., 2019) in developing instruments using the Rasch model has similarities with this study in analyzing the quality of items.

5. CONCLUSION

Based on the test development results, the test is valid, reliable and has good item quality. Content validation was obtained valid with a value of 0.87 and 15 items on the quality characteristics of the items fit 100% with the Rasch Model. The item difficulty index is in the good category with the reliability of the test instrument of 0.83 which is in the reliable category. The KMO unidimensional prerequisite test results of 0.804 fulfill the instrument construct and item quality of the test instrument. The test instrument has been adjusted for the ability $-4 < \theta < 4$ which is indicated by the ability that has the maximum value for the information test or the minimum value for the standard error.

Acknowledgment

Thank you to the ranks of the battalion and detachment of the police brigade corps for contributing to this research.

Conflict of Interest

Researchers declare that there are no conflicts of interest regarding the publication of this paper

REFERENCES

- Abidin, M., & Retnawati, H. (2019). A diagnosis of difficulties in answering questions of circle material on junior high school students. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 23(2), 144–155. <https://doi.org/10.21831/pep.v23i2.16454>
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959. <https://doi.org/10.1177/001316448004000419>
- Aldosari, M., Heydarnejad, T., Hashemifardnia, A., & Abdalgane, M. (2023). The interplay among self-assessment, using reflection for assessment, classroom enjoyment, and immunity: into prospects of effective language learning. *Language Testing in Asia*, 13(1). <https://doi.org/10.1186/s40468-022-00213-1>
- Argo, A. R. B. (2020). PENGEMBANGAN ITEM BANK TES KOGNITIF KEPOLISIAN NEGARA REPUBLIK INDONESIA (POLRI). In *Universitas Islam Negeri Hidayatullah*. <https://doi.org/10.1016/j.jnc.2020.125798%0Ahttps://doi.org/10.1016/j.smr.2020.02.002%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/810049%0Ahttp://doi.wiley.com/10.1002/anie.197505391%0Ahttp://www.sciencedirect.com/science/article/pii/B9780857090409500205%0Ahttp://www.sciencedirect.com/science/article/pii/B9780857090409500205>
- Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, 38(1), 6–40. <https://doi.org/10.1177/0265532220927487>
- Awada, M., Becerik-Gerber, B., Hoque, S., O'Neill, Z., Pedrielli, G., Wen, J., & Wu, T. (2021). Ten questions concerning occupant health in buildings during normal operations and extreme events including the COVID-19 pandemic. *Building and Environment*, 188(November 2020), 107480. <https://doi.org/10.1016/j.buildenv.2020.107480>
- Bajamal, E., Anazi, S. Al, Esheaba, O., & Hantira, N. (2023). Psychometric properties of the arabic version of the positive/negative experiences of parents for school-aged students about online learning during COVID-19 pandemic assessment scale. *BMC Research Notes*, 16(1). <https://doi.org/10.1186/s13104-023-06299-x>
- Bayram, H. M., Ilgaz, F., Serel-Arslan, S., Demir, N., & Rakicioglu, N. (2021). The relationship between dysphagia, oral health, masticatory performance and activities of daily living in elderly individuals as assessed by the Eating Assessment Tool. *Progress in Nutrition*, 23(1), 1–8. <https://doi.org/10.23751/pn.v23i1.9101>
- Belzak, W. C. M. (2023). The Multidimensionality of Measurement Bias in High-Stakes Testing: Using Machine Learning to Evaluate Complex Sources of Differential Item Functioning. *Educational Measurement: Issues and Practice*, 42(1), 24–33. <https://doi.org/10.1111/emip.12486>
- Bernshausen, H., Fuhrmann, C., Harney, H.-L., Harney, K., & Müller, A. (2022). Form Invariance — An Alternative Answer to the Measurement Problem of Item Response Theory. *Mathematics and Statistics*, 10(3), 690–700. <https://doi.org/10.13189/ms.2022.100326>
- Brevik, T. B., Laake, P., Bjørkly, S., Leer-Salvesen, K., & Vatnar, S. K. B. (2024). An educational intervention study on mandatory reporting of intimate partner violence: changes in knowledge and attitudes among healthcare providers. *BMC Medical Education*, 24(1). <https://doi.org/10.1186/s12909-024-06120-8>
- Cao, C.-H., Liao, X.-L., Jiang, X.-Y., Li, X.-D., Chen, I.-H., & Lin, C.-Y. (2023). Psychometric evaluation of the depression, anxiety, and stress scale-21 (DASS-21) among Chinese primary and middle school teachers. *BMC Psychology*, 11(1). <https://doi.org/10.1186/s40359-023-01242-y>
- Caycho-Rodríguez, T., Carbajal-León, C., Vilca, L. W., Reyes-Bossio, M., Gallegos, M., Esteban, R. C., Noe-Grijalva, M., Gallegos, W. L. A., Delgado-Campusano, M., & Muñoz-del-Carpio-Toia, Á. (2023). Impact of COVID-19 on quality of life in Peruvian older adults: construct validity, reliability and invariance of the COV19—Impact on Quality of Life (COV19-QoL) measurement. *Psicologia: Reflexao e Critica*, 36(1). <https://doi.org/10.1186/s41155-023-00256-0>
- Chakrabarty, S. N. (2013). Best Split-Half and Maximum Reliability. *IOSR Journal of Research & Method in Education (IOSRJRME)*, 3(1), 01–08. <https://doi.org/10.9790/7388-0310108>

- Chala, T. D., & Kóczy, L. T. (2024). Intelligent Fuzzy Traffic Signal Control System for Complex Intersections Using Fuzzy Rule Base Reduction. *Symmetry*, 16(9). <https://doi.org/10.3390/sym16091177>
- Chan, S., Looi, C., & Sumintono, B. (2020). A Rasch model measurement analysis. *Journal of Computers in Education*, 8(2), 213–236.
- Che Lah, N. H., Tasir, Z., & Jumaat, N. F. (2023). Applying alternative method to evaluate online problem-solving skill inventory (OPSI) using Rasch model analysis. *Educational Studies*, 49(4), 644–666. <https://doi.org/10.1080/03055698.2021.1874310>
- Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa, I., & Nurhaswinda, N. (2021). Analysis of Rasch Model for the Validation of Chemistry National Exam Instruments. *Jurnal Pendidikan Sains Indonesia*, 9(3), 329–345. <https://doi.org/10.24815/jpsi.v9i3.19618>
- Debelak, R. (2019). An evaluation of overall goodness-of-fit tests for the rasch model. *Frontiers in Psychology*, 9(JAN). <https://doi.org/10.3389/fpsyg.2018.02710>
- Dewanti, S. S., Ayriza, Y., & Setiawati, F. A. (2020). The application of item response theory for development of a students' attitude scale toward mathematics. *New Educational Review*, 60, 108–123. <https://doi.org/10.15804/ner.2020.60.2.09>
- Donhauser, T., Apfelbacher, C., Kann, G., Masur, C., Kamudoni, P., Salek, S., Abels, C., & Gabes, M. (2023). Hyperhidrosis quality of life index (HidroQoL): further validation by applying classical test theory and item response theory using data from a phase III clinical trial. *Journal of Patient-Reported Outcomes*, 7(1). <https://doi.org/10.1186/s41687-023-00596-6>
- Dwijayani, N. M. (2019). Development of circle learning media to improve student learning outcomes. *Journal of Physics: Conference Series*, 1321(2). <https://doi.org/10.1088/1742-6596/1321/2/022099>
- Erfan, M., Mauliyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Tes Klasik Dan Model Rasch. *Indonesian Journal of Educational Research and Review*, 3(1), 11–19. <https://ejournal.undiksha.ac.id/index.php/IJERR/article/view/24080/pdf>
- Erfianti, L., Istiyono, E., & Kuswanto, H. (2019). Developing Lup Instrument Test to Measure Higher Order Thinking Skills (HOTS) Bloomian for Senior High School Students. *International Journal of Educational Research Review*, 4(3), 320–329. <https://doi.org/10.24331/ijere.573863>
- Falani, I., & Kumala, S. A. (2017). Kestabilan Estimasi Parameter Kemampuan Pada Model Logistik Item Response Theory Ditinjau Dari Panjang Tes. *SAP (Susunan Artikel Pendidikan)*, 2(2). <https://doi.org/10.30998/sap.v2i2.2028>
- Feng, Y., Ou-Yang, Z.-Y., Lu, J.-J., Yang, Y.-F., Zhang, Q., Zhong, M.-M., Chen, N.-X., Su, X.-L., Hu, J., Ye, Q., Zhao, J., Zhao, Y.-Q., Chen, Y., Tan, L., Liu, Q., Feng, Y.-Z., & Guo, Y. (2023). Cross-cultural adaptation and psychometric properties of the Mainland Chinese version of the manchester orofacial pain disability scale (MOPDS) among college students. *BMC Medical Research Methodology*, 23(1). <https://doi.org/10.1186/s12874-023-01976-8>
- Fisher, W. P., & Massengill, P. J. (2023). Explanatory Models, Unit Standards, and Personalized Learning in Educational Measurement. In *Explanatory Models, Unit Standards, and Personalized Learning in Educational Measurement*. Research Institute of Sweden (RISE). <https://doi.org/10.1007/978-981-19-3747-7>
- Gatlin, M. (2024). A One Health Approach to Public Safety: A Review of Police Canines in the United States. *International Journal of Environmental Research and Public Health*, 21(9). <https://doi.org/10.3390/ijerph21091235>
- Ghomi, M., & Redecker, C. (2019). Digital competence of educators (DigCompedu): Development and evaluation of a self-assessment instrument for teachers' digital competence. *CSEDU 2019 - Proceedings of the 11th International Conference on Computer Supported Education*, 1(Csedu), 541–548. <https://doi.org/10.5220/0007679005410548>
- Hadi, S., Haryanto, H., AM, M. A., Marlina, M., & Rahim, A. (2022). Developing Classroom Assessment Tool using Learning Management System-based Computerized Adaptive Test in Vocational High Schools. *Journal of Education Research and Evaluation*, 6(1), 143–155. <https://doi.org/10.23887/jere.v6i1.35630>
- Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamental of Item Response Theory*. Sage publication Inc.
- Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. In *Journal of Computers in Education* (Vol. 8, Issue 1). Springer Berlin Heidelberg. <https://doi.org/10.1007/s40692-020-00169-2>
- Han, Q.-X., Zhang, L., Wu, J.-Y., Liu, X.-H., Li, Y., Chen, T.-X., Yi, Y., & Yu, M.-Q. (2024). Early clinical efficacy study on the efficacy of a three-stage conservative Chinese medicine external treatment for acute lateral ankle ligament injuries. *China Journal of Orthopaedics and Traumatology*, 37(10), 997 – 1002. <https://doi.org/10.12200/j.issn.1003-0034.20240730>
- Harjo, B., Kartowagiran, B., & Mahmudi, A. (2019). Development of critical thinking skill instruments on mathematical learning high school. *International Journal of Instruction*, 12(4), 149 – 166. <https://doi.org/10.29333/iji.2019.12410a>
- Heydari, M., Mousazadeh, Y., Salmani, R., & Mehraeen, E. (2023). Assessment of virtual education during the covid-19 pandemic from the perspective of faculty members and students: a cross-sectional descriptive study in northwest Iran. *BMC Medical Education*, 23(1). <https://doi.org/10.1186/s12909-023-04378-y>
- Holeksa, J. (2024). “In Sweden you are worthless. In Denmark you get an identity again” – on being perceived and received as a person who uses drugs in different drug policy settings. *Harm Reduction Journal*, 21(1). <https://doi.org/10.1186/s12954-024-01035-5>
- Holzknicht, F., McCray, G., Eberharter, K., Kremmel, B., Zehentner, M., Spiby, R., & Dunlea, J. (2021). The effect of response order on candidate viewing behaviour and item difficulty in a multiple-choice listening test. *Language Testing*, 38(1), 41–61. <https://doi.org/10.1177/0265532220917316>
- Honeycutt, J. M., Frost, J. K., & Krawietz, C. E. (2019). Applying signal detection theory to conflict escalation as a consequence of victimization with physiological arousal covariates. *Journal of Aggression, Conflict and Peace Research*, 11(3), 200 – 212. <https://doi.org/10.1108/JACPR-10-2018-0386>
- Ibrahim, F., Feildboy, E., Huber, Y., Nagy, D., Hennig, J., & Herzberg, P. Y. (2024). Enhancing Close Quarters Battle Performance—The Effect of a Compact Training on the Tactical Performance and Stress Response of Non-Specialized and Special Forces. *Journal of Police and Criminal Psychology*. <https://doi.org/10.1007/s11896-024-09699-2>
- Keels, E., & Nichols, A. D. (2018). State repression and post-conflict peace failure. *Conflict, Security and Development*, 18(1), 17 – 37. <https://doi.org/10.1080/14678802.2017.1420313>
- Kim, W., Park, S., Woo, B. K., Kim, Y., Yoon, C., Lee, D., Kim, J., & Park, Y.-C. (2024). A Study on the Job Characteristics of Police Officers and the Status of Musculoskeletal Disorders. *Healthcare (Switzerland)*, 12(19). <https://doi.org/10.3390/healthcare12191983>
- Kong, S. C., & Wang, Y. Q. (2021). Item response analysis of computational thinking practices: Test characteristics and students' learning abilities in visual programming contexts. *Computers in Human Behavior*, 122(March), 106836. <https://doi.org/10.1016/j.chb.2021.106836>
- Kriswanto, Kartowagiran, B., & Rohaeti, E. (2021). A critical thinking assessment model integrated with science process skills on chemistry for senior high school. *European Journal of Educational Research*, 10(1), 285–298. <https://doi.org/10.12973/EU-JER.10.1.285>
- Kyndt, E., & Baert, H. (2015). PT NU SC. *Journal of Vocational Behavior*. <https://doi.org/10.1016/j.jvb.2015.07.002>
- Lin, C., & Zhao, J. (2024). Testing the determinants of job satisfaction among police administrative officers in Taiwan. *Policing*, 47(6), 880 – 896. <https://doi.org/10.1108/PIJPSM-11-2023-0145>
- Maisarah, I., Zaim, M., Syarif, H., & Bentri, A. (2017). Issues in designing instrument for affective assessment based on scientific approach: English teachers' reflection. *English Language Teaching and Research*, 1(1), 142–151.
- Mardapi, D. (2017). *Pengukuran Penilaian dan Evaluasi Pendidikan*. Nuha Medika.
- Mardapi, D. (2017). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Parama Publishing.
- Meng, R., Dong, L., Dzierzewski, J. M., Mastrotheodoros, S., Cao, M., Yu, B., Wang, J., Gong, B., Li, J., & Spruyt, K. (2023). The RU_SATED as a measure of sleep health: cross-cultural adaptation and validation in Chinese healthcare students. *BMC Psychology*, 11(1). <https://doi.org/10.1186/s40359-023-01203-5>
- Mertha, I. W., & Mahfud, M. (2022). History Learning Based on Wordwall Applications To Improve Student Learning Results Class X Ips in Ma As'Adiyah Ketapang. *International Journal of Educational Review, Law And Social Sciences (IJERLAS)*, 2(5), 507–612. <https://doi.org/10.54443/ijerlas.v2i5.369>
- Min, S., & Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Studies in Educational Evaluation*, 68(November 2020), 100963. <https://doi.org/10.1016/j.stueduc.2020.100963>
- Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., Bedini, A., Chai, F., Christensen, B., Dunbabin, M., Gao, J., Hagler, G. S. W., Jayaratne, R., Kumar, P., Lau, A. K. H., Louie, P. K. K., Mazaheri, M., Ning, Z., Motta, N., ... Williams, R. (2018). Applications of low-cost sensing technologies

- for air quality monitoring and exposure assessment: How far have they gone? *Environment International*, 116(February), 286–299. <https://doi.org/10.1016/j.envint.2018.04.018>
- Motley, R. O., Joe, S., McQueen, A., Clifton, M., & Carlton-Brown, D. (2023). Development, Construct Validity, and Measurement Invariance of the Modified Classes of Racism Frequency of Racial Experiences Measure (M-CRFRE) to Capture Direct and Indirect Exposure to Perceived Racism-Based Police Use of Force for Black Emerging Adults. *Cultural Diversity and Ethnic Minority Psychology*, 29(4), 575 – 589. <https://doi.org/10.1037/cdp0000525>
- Muranaka, S., Fujino, H., & Imura, O. (2023). Evaluating the psychometric properties of the fatigue severity scale using item response theory. *BMC Psychology*, 11(1). <https://doi.org/10.1186/s40359-023-01198-z>
- Putra, E. N., Syamsurizal, S., Lufri, L., & Zulyusri, Z. (2021). The Teacher-Made Multiple-Choice Tests Quality for Natural Science Subject. *Journal of Educational Research and Evaluation*, 5(3), 422–428.
- Putri, H., Susiani, D., Wandani, N. S., & Putri, F. A. (2022). Instrumen Penilaian Hasil Pembelajaran Kognitif pada Tes Uraian dan Tes Objektif. *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar*, 4(2), 139–148. <https://doi.org/10.36232/jurnalpendidikandasar.v4i2.2649>
- Rahim, A., & Haryanto, H. (2021). Implementation of Item Response Theory (IRT) Rasch Model in Quality Analysis of Final Exam Tests in Mathematics. *Journal of Educational Research and Evaluation*, 10(2), 57–65. <https://doi.org/10.15294/jere.v10i2.51802>
- Ramdiah, S., Abidinsyah, A., Royani, M., Husamah, H., & Fauzi, A. (2020). South Kalimantan local wisdom-based biology learning model. *European Journal of Educational Research*, 9(2), 639–653. <https://doi.org/10.12973/eu-jer.9.2.639>
- Retnawati, H. (2016). *Analisis Kuantitatif Instrument Penelitian*. Parama Publishing.
- Robitzsch, A., & Lüdtke, O. (2023). Comparing different trend estimation approaches in country means and standard deviations in international large-scale assessment studies. *Large-Scale Assessments in Education*, 11(1). <https://doi.org/10.1186/s40536-023-00176-6>
- Rudibyani, R. B., Perdana, R., & Elisanti, E. (2020). Development of problem-solving-based knowledge assessment instrument in electrochemistry. *International Journal of Instruction*, 13(4), 957–974. <https://doi.org/10.29333/iji.2020.13458a>
- Safaruddin, Anisa, & AF, M. S. (2012). Partial Credit Model (PCM) dalam Penskoran Politomi pada Teori Respon Butir. *Jurnal Matematika, Statistika, & Komputasi*, 9(1), 39–48. [https://journal.unhas.ac.id/index.php/jmsk/article/view/3397%0Ahttp://files/171/Safarudin et al. - 2012 - Partial Credit Model \(PCM\) dalam Penskoran Polit.pdf](https://journal.unhas.ac.id/index.php/jmsk/article/view/3397%0Ahttp://files/171/Safarudin%20et%20al.%20-%202012%20-%20Partial%20Credit%20Model%20(PCM)%20dalam%20Penskoran%20Polit.pdf)
- Sangeetha, R. G., Hemanth, C., Dipesh, R., Samridhi, K., Venetha, S., Abbas Alif, M., Arjun, S., & S, V. K. (2024). Density Based Real-time Smart Traffic Management System along with Emergency Vehicle Detection for Smart Cities. *International Journal of Intelligent Transportation Systems Research*, 22(2), 328 – 338. <https://doi.org/10.1007/s13177-024-00400-9>
- Septiani, V., & Paidi. (2021). Development of A Test Instrument to Measure Hots and Cognitive Knowledge in Biology. In O. B., W. L.L., & A. M. (Eds.), *Journal of Physics: Conference Series* (Vol. 1788, Issue 1). IOP Publishing Ltd. <https://doi.org/10.1088/1742-6596/1788/1/012028>
- Shaw, A., Liu, O. L., Gu, L., Kardonova, E., Chirikov, I., Li, G., Hu, S., Yu, N., Ma, L., Guo, F., Su, Q., Shi, J., Shi, H., & Loyalka, P. (2020). Thinking critically about critical thinking: validating the Russian HEIghten® critical thinking assessment. *Studies in Higher Education*, 45(9), 1933–1948. <https://doi.org/10.1080/03075079.2019.1672640>
- Simpson, R., & Nix, J. (2024). Police Shootings in Canada: An Empirical Analysis and Call for Data. *Crime and Delinquency*, 70(12), 3343 – 3368. <https://doi.org/10.1177/00111287231226182>
- Skyttberg, N., Kottorp, A., & Alenius, L. S. (2023). Sound psychometric properties of a short new screening tool for patient safety climate: applying a Rasch model analysis. *BMC Health Services Research*, 23(1). <https://doi.org/10.1186/s12913-023-09768-y>
- Soeharto, Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, 8(2), 247–266. <https://doi.org/10.15294/jpii.v8i2.18649>
- Stemler, S. E., & Naples, A. (2021). Rasch Measurement v. Item Response Theory: Knowing When to Cross the Line. *Practical Assessment, Research and Evaluation*, 26, 1–16. <https://doi.org/10.7275/v2gd-4441>
- Sugiyanto, S., Setiawan, A., Hamidah, I., & Ana, A. (2020). Integration of mobile learning and project-based learning in improving vocational school competence. *Journal of Technical Education and Training*, 12(2), 55–68. <https://doi.org/10.30880/jtet.2020.12.02.006>
- Sumintono, B. (2014). Model Rasch untuk Penelitian Sosial Kuantitatif. In *ITS Surabaya* (Issue November 201). <http://deceng3.wordpress.com>
- Supena, I., Darmuki, A., & Hariyadi, A. (2021). The Influence of Learning Model on Students' Learning Outcomes. *International Journal of Instruction*, 14(3), 873–892.
- Suraji, S., Sumaryanto, F.L. T., & Khumaedi, M. (2019). The Analysis of Instrument of The Ability to Acting and Thinking Creatively Based Rasch Model. *Journal of Educational Research and Evaluation*, 8(1), 48–56. <https://doi.org/10.15294/jere.v8i1.32123>
- Susdelina, S., Perdana, S. A., & Febrian, F. (2018). Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik Dan Rasch Model. *Jurnal Kiprah*, 6(1), 41–48. <https://doi.org/10.31629/kiprah.v6i1.574>
- Sustekova, E., Kubiátko, M., & Usak, M. (2019). Validation of critical thinking test on Slovak conditions. *Eurasia Journal of Mathematics, Science and Technology Education*, 15(12). <https://doi.org/10.29333/ejmste/112295>
- Syailendra, E. A. (2016). Inside papua: The police force as counterinsurgents in post-Reformasi Indonesia. *Indonesia*, 2016(102), 57 – 83. <https://doi.org/10.5728/indonesia.102.0057>
- Teremetskyi, V. I., Novichenko, A. V., Chudyk, N. O., Rudenko, M. M., Dubenko, O. M., Saporova, A. O., & Bodnar-Petrovska, O. B. (2024). INTERACTION BETWEEN POLICE AND SOCIETY AS BASIS FOR COMBATING CRIME: ANALYSIS OF UKRAINIAN AND FOREIGN EXPERIENCE. *Journal of Lifestyle and SDG'S Review*, 4(4). <https://doi.org/10.47172/2965-730X.SDGsReview.v4.n04.pe02927>
- Trockel, M., Bohman, B., Lesure, E., Hamidi, M. S., Welle, D., Roberts, L., & Shanafelt, T. (2018). A Brief Instrument to Assess Both Burnout and Professional Fulfillment in Physicians: Reliability and Validity, Including Correlation with Self-Reported Medical Errors, in a Sample of Resident and Practicing Physicians. *Academic Psychiatry*, 42(1), 11–24. <https://doi.org/10.1007/s40596-017-0849-3>
- Turner, J., Rogers, C., & Pepper, I. (2024). Police misconduct and social media: perceptions of aspiring future police officers. *Higher Education, Skills and Work-Based Learning*, 14(5), 1171 – 1186. <https://doi.org/10.1108/HESWBL-08-2023-0214>
- Uto, M., & Ueno, M. (2020). A generalized many-facet Rasch model and its Bayesian estimation using Hamiltonian Monte Carlo. *Behaviormetrika*, 47(2), 469–496. <https://doi.org/10.1007/s41237-020-00115-7>
- Westnes, E., & Hjortdahl, M. (2024). Firefighters and police search dog handlers' experiences working closely with paramedics in urban search and rescue incidents: a qualitative focus group study from Oslo. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 32(1). <https://doi.org/10.1186/s13049-024-01194-1>
- Widyaningsih, S. W., Yusuf, I., Prasetyo, Z. K., & Istiyono, E. (2021). The development of the hots test of physics based on modern test theory: Question modeling through e-learning of moodle lms. *International Journal of Instruction*, 14(4), 51 – 68. <https://doi.org/10.29333/iji.2021.1444a>
- Wozniak, D., & Zahabi, M. (2024). Cognitive workload classification of law enforcement officers using physiological responses. *Applied Ergonomics*, 119. <https://doi.org/10.1016/j.apergo.2024.104305>
- Yung, M., Du, B., Gruber, J., Hackney, A., & Yazdani, A. (2022). Fatigue measures and risk assessment tools for first responder fatigue risk management: A scoping review with considerations of the multidimensionality of fatigue. *Safety Science*, 154. <https://doi.org/10.1016/j.ssci.2022.105839>
- Zeigenfusse, M. D., Batchelder, W. H., & Steyvers, M. (2020). An item response theory model of matching test performance. *Journal of Mathematical Psychology*, 95, 102327. <https://doi.org/10.1016/j.jmp.2020.102327>