

DAFTAR PUSTAKA

- Abdalla, M. E., & Gaffar, A. (2011). *Blueprints IN Health Profession Education Series Constructing A-Type Multiple Choice Questions (Mcqs): Step By Step Manual* (Issue January).
- Abdullah, H., Arsad, N., Hashim, F. H., Aziz, N. A., Amin, N., & Ali, S. H. (2012). Evaluation of Students' Achievement in the Final Exam Questions for Microelectronic (KKKL3054) using the Rasch Model. *Procedia - Social and Behavioral Sciences*, 60(c), 119–123. <https://doi.org/10.1016/j.sbspro.2012.09.356>
- Abdullah, N., & Lim, B. K. (2013). Parallel Circuit Conceptual Understanding Test (PCCUT). *Procedia - Social and Behavioral Sciences*, 90(InCULT 2012), 431–440. <https://doi.org/10.1016/j.sbspro.2013.07.112>
- Aeni, S., Mudzakir, A., & Hernani. (2013). Desain Pembelajaran Elektronika Menggunakan Konteks Keris sebagai Kearifan Lokal Indonesia untuk Meningkatkan Literasi Sains Siswa SMA. *Jurnal Riset Dan Praktik Pendidikan Kimia*, 1 No 1, 44–51.
- Agustine, P. C. (2018). *Pembangan Soal Matematika Model TIMSS di kelas IV Sekolah Dasar*. 3(1), 1–17.
- Ajayi, V. O. (2018). Course Title : Science and Society Course of Study : PhD Science Education Title : Scientific Literacy Author : Victor Oluwatosin Ajayi. *ResearchGate*, February. <https://doi.org/10.13140/RG.2.2.13345.92009>
- Alias, N., & Zainuddin, A. (2005). Innovation for better teaching and learning: Adopting the learning management system. *Malaysian Online Journal of Instructional Technology*, 2(2), 27–40.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Brooks/Cole publishing company.
- Almaiah, M. A., Jalil, M. @. M. A., & Man, M. (2016). Empirical investigation to

explore factors that achieve high quality of mobile learning system based on students' perspectives. *Engineering Science and Technology, an International Journal*, 19(3), 1314–1320. <https://doi.org/10.1016/j.jestch.2016.03.004>

Alshomrani, S. (2012). Evaluation of technical factors in distance learning with respect to open source LMS. *Asian Transactions on Computers*, 02(01), 11–17.

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.675.4267&rep=rep1&type=pdf%0Ahttp://www.asian-transactions.org/Journals/Vol02Issue01/ATC/ATC-40232015.pdf>

Andrich, D. (2010). Rasch models. In *International Encyclopedia of Education*. SAGE Publications Ltd. <https://doi.org/10.1016/B978-0-08-044894-7.00258-X>

Anelli, C. (2011). Scientific literacy: What is it, are we teaching it, and does it matter? *American Entomologist*, 57(4), 235–244. <https://doi.org/10.1093/ae/57.4.235>

Anonymous. (2017). *Faktor Analisis*.

Apriani, M. (2017). *Buku Pengayaan dan Penilaian Mozaik IPA untuk SMP Kelas VIII*. Yudhistira.

Azwar, S. (2012). *Dasar-Dasar Psikometri*. Pustaka Pelajar.

Azwar, S. (2013). *Reliabilitas dan Validitas*. Pustaka Belajar.

Barkaoui, K. (2019). Multifaceted Rasch Analysis for Test Evaluation. *Companion to Language Assessment*, November 2013, 1–46. <https://doi.org/10.1002/9781118411360.wbcla070>

Blyznyuk, T. (2019). Defining and Conceptualizing Geocultural Scientific Literacy. *Journal of Vasyl Stefanyk Precarpathian National University*, 6(1), 43–49. <https://doi.org/10.15330/jpnu.6.1.43-49>

Bond, T.G. and Fox, C. M. (2015). *Applying The Rasch Model, Fundamentals*

Measurement in the Human Sciences. 3rd edition. Routledge, Taylor and Francis Group.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model. In Applying the Rasch model: Fundamental Measurement in the Human Sciences* (3rd ed.). Routledge, Taylor and Francis Group.

Boztunç Öztürk, N., Şahin, M. G., & İlhan, M. (2019). An analysis of scoring via analytic rubric and general impression in peer assessment. *Turkish Journal of Education, October*, 258–275. <https://doi.org/10.19128/turje.609073>

Brown, G. T. L. (2009). The reliability of essay scores: The necessity of rubrics and moderation. *Tertiary Assessment and Higher Education Student Outcomes: Policy, Practice and Research, March*, 40–48.

Budiarti, I. S. (2021). Analysis On Students' Scientific Literacy of Newton's Law and Motion System in Living Things. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, 9(1), 36–51. <https://doi.org/10.24815/jpsi.v9i1.18470>

C. S. Wiseman. (2012). A Comparison of the Performance of Analytic vs . Holistic Scoring Rubrics to Assess L2 Writing. *Iran. J. Lang. Test*, 2, no, 59–92.

Cavus, N. (2015). Distance Learning and Learning Management Systems. *Procedia - Social and Behavioral Sciences*, 191, 872–877. <https://doi.org/10.1016/j.sbspro.2015.04.611>

Cavus, N., & Alhih, M. S. (2014). Learning Management Systems Use in Science Education. *Procedia - Social and Behavioral Sciences*, 143, 517–520. <https://doi.org/10.1016/j.sbspro.2014.07.429>

Christensen, R., & Knezek, G. (2017). Validating the Technology Proficiency Self-Assessment Questionnaire for 21st Century Learning (TPSA C-21). *Journal of Digital Learning in Teacher Education*, 33(1), 20–31. <https://doi.org/10.1080/21532974.2016.1242391>

Christopher D. Desjardins, & Okan Bulut. (2018). *Handbook of Educational*

Measurement and Psychometrics Using R (J. M. Chambers & D. T. Lang (eds.)). Taylor.

Churiyah, M., Sholikhan, S., Filianti, F., & Sakdiyyah, D. A. (2020). Indonesia Education Readiness Conducting Distance Learning in Covid-19 Pandemic Situation. *International Journal of Multicultural and Multireligious Understanding*, 7(6), 491. <https://doi.org/10.18415/ijmmu.v7i6.1833>

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-Based Early Maths Assessment. *Educational Psychology*, 28(4), 457–482. <https://doi.org/10.1080/01443410701777272>

Colton, D., & Covert, R. W. (2007). *Designing and Constructing Instruments for Social Research and Evaluation*. Jossey-Basse.

Conde, M. Á., García-Peñalvo, F. J., Rodríguez-Conde, M. J., Alier, M., Casany, M. J., & Piguillem, J. (2014). An evolving Learning Management System for new educational environments using 2.0 tools. *Interactive Learning Environments*, 22(2), 188–204. <https://doi.org/10.1080/10494820.2012.745433>

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart, and Winston, INC.

Dawati, F. M., Yamtinah, S., Rahardjo, S. B., Ashadi, & Indriyanti, N. Y. (2017). Uji Validitas Computerized Two-Tier Multiple Choice (CTTMC) Melalui Focus Group Discussion (FGD) Untuk Mendiagnosis Kesulitan Belajar Siswa. *Seminar Nasional Pendidikan Sains Universitas Sebelas Maret*, 21, 260–265.

Desiriah, E., & Setyarsih, W. (2021). *Tinjauan Literatur Pengembangan Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi (HOTS) Fisika di SMA*. 7.

Desnita, Delina, M., & Sri Rahayu, Y. (2017). *Gelombang Bunyi*.

Djaali, & Muljono, P. (2015). *Pengukuran dalam Bidang Pendidikan* (p. 65).

Intramedia.

- Djaali, & Muljono, P. (2012). *Pengukuran dalam Bidang Pendidikan*. Grasindo.
- Docktor, J. L., Strand, N. E., Mestre, J. P., & Ross, B. H. (2015). Conceptual problem solving in high school physics. *Physical Review Special Topics - Physics Education Research*, 11(2), 1–13. <https://doi.org/10.1103/PhysRevSTPER.11.020106>
- Duffy, N. (2016). *Literacy , Scientific Literacy and the Learning Needs of ESL Students* Author : Natalie A Duffy. November 2011. <https://doi.org/10.13140/RG.2.2.33252.58242>
- Elisa, Mustafa, & Syukri, M. (2018). *Konsep dan Aplikasi Gelombang dalam Fisika*. Bandar Publishing.
- Fakhriyah, F., Masfuah, S., Roysa, M., Rusilowati, A., & Rahayu, E. S. (2017). Student's science literacy in the aspect of content science? *Jurnal Pendidikan IPA Indonesia*, 6(1), 81–87. <https://doi.org/10.15294/jpii.v6i1.7245>
- Gable, Robert k., & Wolf, M. B. (1993). *Instrument Development in Effective Domain*. Kluwer Academic Publisher.
- Gen, M. (2015). The Effect of Scientific Studies on Students' Scientific Literacy and Attitude. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 34(1), 141–152. <https://doi.org/10.7822/omuefd.34.1.8>
- Glaze, A. L. (2018). Teaching and learning science in the 21st century: Challenging critical assumptions in post-secondary science. *Education Sciences*, 8(1), 1–8. <https://doi.org/10.3390/educsci8010012>
- Guilford, J. P., & Fruchter, B. (1956). *Fundamental Statistics in Psychology and Education*. McGraw-Hill Book.
- Gunawan, I., & Paluti, A. R. (2017). Taksonomi Bloom – Revisi Ranah Kognitif: Kerangka Landasan Untuk Pembelajaran, Pengajaran, Dan Penilaian. *E-Journal.Unipma*, 7(1), 1–8. <http://e-journal.unipma.ac.id/index.php/PE>

- Hadi, S., & Novaliyosi. (2019). TIMSS Indonesia (Trends in International Mathematics and Science Study). *Prosiding Seminar Nasional & Call For Papers Program Studi Magister Pendidikan Matematika Universitas Siliwangi*, 562–569.
- Heinemann, A. W., Michael Linacre, J., Wright, B. D., Hamilton, B. B., & Granger, C. (1994). Measurement characteristics of the Functional Independence Measure. *Topics in Stroke Rehabilitation*, 1(3), 1–15. <https://doi.org/10.1080/10749357.1994.11754030>
- Holbrook, J., & Rannikmae, M. (2009). The meaning of scientific literacy. *International Journal of Environmental and Science Education*, 4(3), 275–288.
- Hsiao, Y. Y., Shih, C. L., Yu, W. H., Hsieh, C. H., & Hsieh, C. L. (2015). Examining unidimensionality and improving reliability for the eight subscales of the SF-36 in opioid-dependent patients using Rasch analysis. *Quality of Life Research*, 24(2), 279–285. <https://doi.org/10.1007/s11136-014-0771-z>
- Hsueh, I. P., Wang, W. C., Sheu, C. F., & Hsieh, C. L. (2004). Rasch Analysis of Combining Two Indices to Assess Comprehensive ADL Function in Stroke Patients. *Stroke*, 35(3), 721–726. <https://doi.org/10.1161/01.STR.0000117569.34232.76>
- Huberty, J., Vener, J., Gao, Y., Matthews, J. L., Ransdell, L., & Elavsky, S. (2013). Developing an instrument to measure physical activity related self-worth in women: Rasch analysis of the Women's Physical Activity Self-Worth Inventory (WPASWI). *Psychology of Sport and Exercise*, 14(1), 111–121. <https://doi.org/10.1016/j.psychsport.2012.07.009>
- Huck, S. W. (2007). *Reading Statistics and Research*. Allyn & Bacon.
- Ibnu, M., Indriyani, B., Husnaini, & Guntra, Y. (2019). *Aplikasi Rasch Model: Pengembangan Instrumen Tes untuk Mengukur Miskonsepsi Mahasiswa pada Materi Mekanika*. 2(1), 205–210.

- Jafari Navimipour, N., & Zareie, B. (2015). A model for assessing the impact of e-learning systems on employees' satisfaction. *Computers in Human Behavior*, 53, 475–485. <https://doi.org/10.1016/j.chb.2015.07.026>
- John, D. S. (2021). *Strategies to Minimise Students' Weaknesses in Discussion Essay*. 7(March), 106–114.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Judge, D. S., & Murray, B. (2017). Student and Faculty Transition to a New Online Learning Management System. *Teaching and Learning in Nursing*, 12(4), 277–280. <https://doi.org/10.1016/j.teln.2017.06.010>
- Jufri, A. W., Setiadi, D., & Sripatmi. (2016). Scientific reasoning ability of prospective student teacher in the excellence program of mathematics and science teacher education in University of Mataram. *Jurnal Pendidikan IPA Indonesia*, 5(1), 69–74. <https://doi.org/10.15294/jpii.v5i1.5792>
- K. L. Gwet. (2012). *Handbook of Inter-rater reliability*.
- Kaba, Y., & Şengül, S. (2015). Developing the Rubric for Evaluating Problem Posing (REPP). *International Online Journal of Educational Sciences*, 8(1). <https://doi.org/10.15345/iojes.2016.01.002>
- Kaningan, M. (2019). *Mandiri IPA Fisika untuk SMP/MTS Kelas VIII*. Erlangga.
- Kerlinger, F. N. (2004). *Asas-Asas Penelitian Behavioral*. Gajah Mada University Press.
- Kurnia, F., . Z., & Fathurohman, A. (2014). Analisis Bahan Ajar Fisika Sma Kelas Xi Di Kecamatan Indralaya Utara Berdasarkan Kategori Literasi Sains. *Jurnal Inovasi Dan Pembelajaran Fisika*, 1(1), 43–47. <https://doi.org/10.36706/jipf.v1i1.1263>
- Kurniawan, Y., & Mulyani, R. (2019). A Development Design of Digital Story

Conceptual Change-Oriented in Physics Subject. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, 4(2), 110. <https://doi.org/10.26737/jipf.v4i2.1153>

Lailiyah, L., Supriyati, Y., & Komarudin, K. (2018). Analysis of Measures Items in Development of Instruments Self-Assessment (Rasch Modeling Application). *Jisae: Journal of Indonesian Student Assessment and Evaluation*, 4(1), 1–9. <https://doi.org/10.21009/jisae.041.01>

Lee, E. Y., & Jeon, Y. J. J. (2020). The difference of user satisfaction and net benefit of a mobile learning management system according to self-directed learning: An investigation of cyber university students in hospitality. *Sustainability (Switzerland)*, 12(7), 1–13. <https://doi.org/10.3390/su12072672>

Linacre, J. M. (2004). Rasch Model Estimation: Further Topics. *Journal of Applied Measurement*, 5(1), 95–110.

Linarce, J. M. (2012). *A user's guide to Winsteps Ministeps Rasch-model computer programs [version 3.74.0]*.

Litwin, M. S. (1995). *How to Measure survey Reliability and Validity*. Sage Publication.

Mardapi, D. (2012). *Teknik Penyusunan Instrumen Tes dan Nontes*. Mitra Cendikia Press.

Mardapi, D. (2012). *Pengukuran, Penilaian dan Evaluasi Pendidikan*. Nuha Medika.

Margono, G. (2005). *Pengembangan Instrumen Pengukur Afeksi terhadap Matematika*. Universitas Negeri Jakarta.

Margono, G. (2013). The Development of Instrument for Measuring Attitudes toward Statistics Using Semantic Differential Scale. *2nd International Seminar on Quality and Affordable Education, Isqae*, 241–250. <https://educ.utm.my/da/wp-content/uploads/2013/11/341.pdf>

Masruri, R. (2017). Penyusunan Instrumen Penilaian Pengetahuan. *Prosiding*

Seminar Nasional Profesionalisme, II(1), 49–63.

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument Development in the Affective Domain: School and Corporate Applications. In Instrument Development in the Affective Domain: School and Corporate Applications. 3.* <https://doi.org/10.1007/978-1-4614-7135-6>

Mincheva, K., & Planska-Simeonova, K. (2019). Scientific Research in the Field of Visual Competency. *EDULEARN19 Proceedings, I*(February 2020), 5110–5117. <https://doi.org/10.21125/edulearn.2019.1263>

Misbah, I. ., & Sumintono, B. (2014). *Instrument Development and Validation “Students’ Perception on Teachers Morality” in Indonesia using Rasch Model, presented in National Seminar on “Development of Valid Character Assessment at Faculty of Psychology of Universitas Muhammadiyah Surakarta. I.*

Moser, C. A., & Kalton, G. (1989). *Survey methods in social investigation.* Gower.

Muijs D. (2011). *Doing Quantitative Research in Education with SPSS.* SAGE Publications Ltd.

Muttaqin, M. Z., & Kusaeri, K. (2017). Pengembangan Instrumen Penilaian Tes Tertulis Bentuk Uraian Untuk Pembelajaran Pai Berbasis Masalah Materi Fiqh. *Jurnal Tatsqif, 15(1), 1–23.* <https://doi.org/10.20414/j-tatsqif.v15i1.1154>

Naga, Dali S. (2013). *Teori Skor pada Pengukuran Mental.* Nagarani Citrayasa.

Naga, Dali Santun. (2012). *Teori Skor pada Pengukuran Mental* (2nd ed.). PT. Nagarani Citrayasa.

Namaziandost, E. (2019). The Assessment of Oral Proficiency through Holistic and Analytic Techniques of Scoring: A Comparative Study. *Applied Linguistics Research Journal, May.* <https://doi.org/10.14744/alrj.2019.83792>

Ni'mah, F. (2019). Research trends of scientific literacy in Indonesia: Where are

we? *Jurnal Inovasi Pendidikan IPA*, 5(1), 23–30.
<https://doi.org/10.21831/jipi.v5i1.20862>

Nida, N., Patmawati, H., & Muhtadi, D. (2021). *Kerangka Penilaian Taksonomi Trends In International Mathematics And Science Study*. 3(1), 34–42.

Ningdyah, A. E. M., Greenwood, K. M., & Kidd, G. (2018). A Training-Model Scale's Validity and Reliability Coefficients: Expert Evaluation in Indonesian Professional Psychology Programs. *Makara Human Behavior Studies in Asia*, 22(1), 56. <https://doi.org/10.7454/hubs.asia.2190318>

Nja, C. O. (2019). Scientific literacy of undergraduate Science Education students in the University of Calabar Cross River State Nigeria. *Quest Journals: Journal of Research in Humanities and Social Science*, 7(5), 35–39. <https://www.researchgate.net/publication/333614087>

Nurjanah, & Marlianingsih, N. (2015). Analisis Butir Soal Pilihan Ganda Dari Aspek Kebahasaan. *Faktor Jurnal Ilmu Kependidikan*, 2(1), 69–78.

Nurul, mas'ud waqiah. (2020). Analysis of Interest and Scientific Literacy Skills of Senior High School in Learning Physics. *Jurnal Kependidikan Fisika*, 53(9), 1689–1699.

OECD. (2016). *Excellence And Equity In Education Volume: Vol. I*. <https://doi.org/10.1787/9789264266490-5-en>

OECD. (2017). Results from PISA 2015: Indonesia. In *OECD*. <https://www.oecd.org/pisa/PISA-2015-Indonesia.pdf>

OECD. (2019). PISA 2018 Results. Combined Executive Summaries. In *OECD* (Vol. 53, Issue 9). www.oecd.org/about/publishing/corrigenda.htm.

Patrícia, M., & Zvára, K. (2007). *Kybernetika Terms of use : Tolerance Automata*. 43(3), 315–326.

Perera, C. J., Sumintono, B., & Jiang, N. (2018). The Psychometric Validation Of The Principal Practices Questionnaire Based On Item Response Theory.

International Online Journal of Educational Leadership, 2(1), 21–38.
<https://doi.org/10.22452/iojel.vol2no1.3>

Pretz, C. R., Kean, J., Heinemann, A. W., Kozlowski, A. J., Bode, R. K., & Gebhardt, E. (2016). A Multidimensional Rasch Analysis of the Functional Independence Measure Based on the National Institute on Disability, Independent Living, and Rehabilitation Research Traumatic Brain Injury Model Systems National Database. *Journal of Neurotrauma*, 33(14), 1358–1362. <https://doi.org/10.1089/neu.2015.4138>

Priyadi, R., & Suryanti, K. (2017). Pengembangan Instrumen Tes Pemahaman Konsep Hukum Gravitasi Universal. *Jurnal Riset Pendidikan Fisika*, 2(2), 36–41.

Purwanto. (2015). *Metodologi Penelitian Kuantitatif*. Pustaka Belajar.

Putri, R. S., Purwanto, A., Pramono, R., Asbari, M., Wijayanti, L. M., & Hyun, C. C. (2020). Impact of the COVID-19 pandemic on online home learning: An explorative study of primary schools in Indonesia. *International Journal of Advanced Science and Technology*, 29(5), 4809–4818.

Quintal, V. A., & Murphy, J. (2012). LMS teaching versus Community Learning: A call for the latter. *Asia Pacific Journal of Marketing and Logistics*, 24(5), 826–841. <https://doi.org/10.1108/13555851211278529>

Rachmat, N. A., & Arfiandhani, P. (2019). “I Use Multiple-Choice Question in Most Assessment I Prepared”: Efl Teachers’ Voice on Summative Assessment. *ETERNAL (English, Teaching, Learning, and Research Journal)*, 5(1), 163. <https://doi.org/10.24252/eternal.v5i1.2019.a13>

Rasmitadila, Aliyyah, R. R., Rachmadtullah, R., Samsudin, A., Syaodih, E., Nurtanto, M., & Tambunan, A. R. S. (2020). The perceptions of primary school teachers of online learning during the covid-19 pandemic period: A case study in Indonesia. *Journal of Ethnic and Cultural Studies*, 7(2), 90–109. <https://doi.org/10.29333/ejecs/388>

- Retnawati, H. (2016). Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index). *Research and Evaluation in Education*, 2(2), 155–164. <https://doi.org/http://dx.doi.org/10.21831/reid.v2i2.11029> *Corresponding
- Rostikawati, D. A., & Permanasari, A. (2016). Rekonstruksi bahan ajar dengan konteks socio-scientific issues pada materi zat aditif makanan untuk meningkatkan literasi sains siswa. *Jurnal Inovasi Pendidikan IPA*, 2(2), 156–164. <https://doi.org/10.21831/jipi.v2i2.8814>
- Rusilowati, A., Kurniawati, L., Nugroho, S. E., & Widiyatmoko, A. (2016). Developing an instrument of scientific literacy asesment on the cycle theme. *International Journal of Environmental and Science Education*, 11(12), 5718–5727.
- Russell, waugh. (2009). *Appliation of Rasch Measurament In Education*. Nova Science Publishers, Inc.
- S. Surapranata. (2015). *Analisis, validitas, reliabilitas dan interpretasi hasil tes*.
- Sadler, T. D., & Zeidler, D. L. (2009). Scientific literacy, PISA, and socioscientific discourse: Assessment for progressive aims of science education. *Journal of Research in Science Teaching*, 46(8), 909–921. <https://doi.org/10.1002/tea.20327>
- Sahrani, R. (2019). Faktor-Faktor Karakteristik Kebijakan Menurut Remaja. *Jurnal Psikologi Sosial*, 17(1), 36–45. <https://doi.org/10.7454/jps.2019.6>
- Sally, V. K. (2016). *IPA Terpadu untuk SMP Kelas VIII*. Yudhistira.
- Samritin. (2016). DEVELOPING AN ASSESSMENT INSTRUMENT OF JUNIOR HIGH SCHOOL STUDENTS' HIGHER ORDER THINKING SKILLS IN MATHEMATICS. *Research and Evaluation in Education*, 2(1), 92–107.
- Santoso, S. (2012). *SPSS Statistik Multivariate*. Elex Media Komputindo Gramedia.

- Saputri, R., & Hariyadi, B. (2021). *Pengembangan Soal Higher Order Thinking Skills Berbasis Budaya Jambi*. 05(02), 1793–1806.
- Segarra, V. A., Hughes, N. M., Ackerman, K. M., Grider, M. H., Lyda, T., & Vigueira, P. A. (2018). Student performance on the Test of Scientific Literacy Skills (TOSLS) does not change with assignment of a low-stakes grade. *BMC Research Notes*, 11(1), 1–5. <https://doi.org/10.1186/s13104-018-3545-9>
- Septiani, D., Widiyawati, Y., & Nurwahidah, I. (2019). Pengembangan Instrumen Tes Literasi Sains Berbasis Pisa Pada Aspek Menjelaskan Fenomena Ilmiah Untuk Siswa Kelas Vii. *Science Education and Application Journal*, 1(2), 46. <https://doi.org/10.30736/seaj.v1i2.144>
- Shwartz, Y., Ben-Zv, R., & Hofstein, A. (2005). The importance of involving high-school chemistry teachers in the process of defining the operational meaning of ‘chemical literacy.’ *International Journal of Science Education*, 27(3), 323–344. <https://doi.org/10.1080/0950069042000266191>
- Sinnema, C., Meyer, F., & Aitken, G. (2017). Capturing the Complex, Situated, and Active Nature of Teaching Through Inquiry-Oriented Standards for Teaching. *Journal of Teacher Education*, 68(1), 9–27. <https://doi.org/10.1177/0022487116668017>
- Souza, A. C. de, Alexandre, N. M. C., & Guirardello, E. de B. (2017). Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saude : Revista Do Sistema Unico de Saude Do Brasil*, 26(3), 649–659. <https://doi.org/10.5123/S1679-49742017000300022>
- Sprinkle, S. D., Lurie, D., Insko, S. L., Atkinson, G., Jones, G. L., Logan, A. R., & Bissada, N. N. (2002). Criterion validity, severity cut scores, and test-retest reliability of the Beck Depression Inventory-II in a university counseling center sample. *Journal of Counseling Psychology*, 49(3), 381–385. <https://doi.org/10.1037/0022-0167.49.3.381>
- Srirahayu, R. R. Y., & Arty, I. S. (2018). Validitas dan reliabilitas instrumen

asesmen kinerja literasi sains pelajaran Fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(2), 168–181.
<https://doi.org/10.21831/pep.v22i2.20270>

Sudaryono, Rahardja, U., Aini, Q., Isma Graha, Y., & Lutfiani, N. (2019). Validity of Test Instruments. *Journal of Physics: Conference Series*, 1364(1).
<https://doi.org/10.1088/1742-6596/1364/1/012050>

Sudijono, A. (2012). *Pengantar Evaluasi Pendidikan*. PT. Grafindo Persada.

Sugiyono. (2015). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, kualitatif, dan R&D*. Alfabeta.

Sumintono, B., & Widhiarso, W. (2014). *Application of Rasch Model for Social Science Research* (B. Trim (ed.)). Tim Komunikasi Publishing House.

Sumintono, B., Widhiarso, W., & Mada, U. G. (2015). *Aplikasi Pemodelan Rasch pada Assesment Pendidikan* (Issue October). Trim Komunikata.

Sumintono, B., & Widiarso, W. (2014). *Application of Rasch Model in Scientific Research, revised edition*. Tim Komunikata Publishing House.

Supriyati, Yeti, & Dudung, A. (2017). *Penilaian Kelas*. KARIMA (Karya Ilmu Media Aulia).

Supriyati, Yetti, Zakiyah, Z., & Astra, I. M. (2021). *Essay questions on dynamic fluid physics material to measure intellection thinking ability of grade XI high school students* *Essay Questions on Dynamic Fluid Physics Material to Measure Intellection Thinking Ability of Grade XI High School Students*. 020016(March), 2–8.

Susongko, P., Kusuma, M., Arfiani, Y., Samsudin, A., & Aminudin, A. (2020). Develop and Analyze Instruments of Scientific Literacy Skills with Integrated Science (SLS-IS) Based on the 2015 PISA Standard via Rasch Model on Tegal-Students'. *Journal for the Education of Gifted Young Scientists*, December. <https://doi.org/10.17478/jegys.781583>

- Syawaludin, A., Supriyati, Y., & Rahayu, W. (2019). *RASCH Model Application for Validation of Measurement Instruments of Student Nationalism*. 5(2), 26–42. <https://doi.org/10.17501/24246700.2019.5204>
- Taherdoost, H. (2018). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electronic Journal*, January 2016. <https://doi.org/10.2139/ssrn.3205040>
- Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value in Health*, 7(SUPPL. 1), 22–26. <https://doi.org/10.1111/j.1524-4733.2004.7s106.x>
- Udompong, L., & Wongwanich, S. (2014). Diagnosis of the Scientific Literacy Characteristics of Primary Students. *Procedia - Social and Behavioral Sciences*, 116, 5091–5096. <https://doi.org/10.1016/j.sbspro.2014.01.1079>
- Ulfa, K., Khumaedi, M., Al, S. M. P., & Kedungwuni, F. (2018). The Developing of Performance Assessment to Calculate Scope and Volume of Cube and Block Competence in The Mathematic Learning of Junior High School. *Journal of Research and Educational Research Evaluation*, 7(1), 29–36. <https://doi.org/10.15294/jrer.v7i1.23045>
- Uysal, I. (2021). *Automated Essay Scoring Effect on Test Equating Errors in Mixed-format Test*. 8(2), 222–238.
- Van Helvoort, J. (2010). A scoring rubric for performance assessment of information literacy in Dutch higher education. *Journal of Information Literacy*, 4(1). <https://doi.org/10.11645/4.1.1256>
- Wahyuni, L. D., Gumela, G., & Maulana, H. (2021). Interrater Reliability: Comparison of essay's tests and scoring rubrics. *Journal of Physics: Conference Series*, 1933(1), 012081. <https://doi.org/10.1088/1742-6596/1933/1/012081>
- Whittaker, T. A., & Worthington, R. L. (2016). Item Response Theory in Scale

Development Research: A Critical Analysis. *Counseling Psychologist*, 44(2), 216–225. <https://doi.org/10.1177/0011000015626273>

William P. Fisher, J. (2007). Transactions of the Rasch Measurement SIG. *American Educational Research Association*, 21(1), 1087–1096.

Wong, W. S., & Bong, C. H. (2021). *Assessing Malaysian University English Test (MUET) Essay on Language and Semantic Features Using Intelligent Essay Grader (IEG)*. 29(2), 919–941.

Wu, H., Gao, X., & Shen, J. (2020). Principal leadership effects on student achievement: a multilevel analysis using Programme for International Student Assessment 2015 data. *Educational Studies*, 46(3), 316–336. <https://doi.org/10.1080/03055698.2019.1584853>

Yamtinah, S., Saputro, S., Mulyani, S., Ulfa, M., Lutviana, E., & Shidiq, A. S. (2019). Do students have enough scientific literacy? A computerized testlet instrument for measuring students' scientific literacy. *AIP Conference Proceedings*, 2194(December). <https://doi.org/10.1063/1.5139875>

Yasin, R. M., Yunus, F. A. N., Rus, R. C., Ahmad, A., & Rahim, M. B. (2015). Validity and Reliability Learning Transfer Item Using Rasch Measurement Model. *Procedia - Social and Behavioral Sciences*, 204(November 2014), 212–217. <https://doi.org/10.1016/j.sbspro.2015.08.143>

Zagorsek, H., Stough, S. J., & Jaklic, M. (2006). Analysis of the reliability of the leadership practices inventory in the item response theory framework. *International Journal of Selection and Assessment*, 14(2), 180–191. <https://doi.org/10.1111/j.1468-2389.2006.00343.x>