

DAFTAR PUSTAKA

- Abdullah Sani, R., & Prayitno, W. (2020). *Asesmen Kompetensi Minimum* (N. Nur M (ed.)). PT. Remaja Rosdakarya.
- Abidin, Y., Mulyati, T., & Yunansah, H. (2017). Pembelajaran Literasi: Strategi Meningkatkan Kemampuan Literasi Matematika. In *Bumi Aksara* (2 ed.). Bumi Aksara.
https://books.google.co.id/books?hl=en&lr=&id=M_UrEAAQBAJ&oi=fnd&pg=PA1&dq=info:_q1mv1O3480J:scholar.google.com&ots=_L8b6P4Q6l&sig=Kgrb-gdo35_liZrb9IFX6RtoVVo&redir_esc=y#v=onepage&q&f=false
- Ad'hiya, E., & Laksono, E. W. (2018). Development and validation of an integrated assessment instrument to assess students' analytical thinking skills in chemical literacy. *International Journal of Instruction*, 11(4), 241–256.
<https://doi.org/10.12973/iji.2018.11416a>
- Aditomo, A., Rahmawati, Felicia, N., Shihab, N., & Handayani, F. (2019). *Kajian Akademik dan Rekomendasi Reformasi Sistem Asesmen Nasional*. Badan Penelitian dan Pengembangan dan Perbukuan Kementerian Pendidikan dan Kebudayaan.
- 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>
- Ambiyar. (2011). Pengukuran & Tes Dalam Pendidikan. In *Padang: Universitas Negeri Padang*. UNP Press.
- Arends, R. I. (2013). Learning to teach. In *Handbook of Educational Psychology* (9 ed., Vol. 2). McGraw-Hill New York.
<https://doi.org/10.4324/9780203053874-29>
- Arikunto, S. (2016). Dasar–Dasar Evaluasi Pendidikan Evaluasi Pendidikan. In *PT. Bumi Aksara* (2 ed.). PT Bumi Aksara.
- Arikunto, S. (2017). *Pengembangan Instrumen Penelitian dan Penilaian Program*. Pustaka Pelajar.
- Asrijanty. (2020). Asesmen Kompetensi Minimum (AKM) dan implikasinya pada pembelajaran. In *Pusat Asesmen Dan Pembelajaran Badan Penelitian Dan Pengembangan Dan Perbukuan Kementerian Pendidikan Dan Kebudayaan* (hal. 3–38). Pusat Asesmen dan Pembelajaran.
- Ayanwale, M. A. (2021). Calibration of Polytomous Response Mathematics Achievement Test Using Generalized Partial Credit Model. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3824911>
- Azami, M. I., & Rahmawati, N. M. (2023). Designing an English Coursebook for Senior Executive Administrative Assistants for English Department Students

- Based on Indonesian National Work Competency Standards (SKKNI). *J-Shmic Journal of English for Academic*, 10(2), 108–122. [https://doi.org/10.25299/jshmic.2023.vol10\(2\).12944](https://doi.org/10.25299/jshmic.2023.vol10(2).12944)
- Bachman, H. J., Elliott, L., Duong, S., Betancur, L., Navarro, M. G., Votruba-Drzal, E., & Libertus, M. (2020). Triangulating Multi-Method Assessments of Parental Support for Early Math Skills. *Frontiers in Education*, 5. <https://doi.org/10.3389/feduc.2020.589514>
- Bahari, J. I. (2023). Evaluation of Cognitive, Psychomotor, and Affective Aspects in the Subject of Islamic Education. *Incare*, 4(2), 136–162. <https://doi.org/10.59689/incare.v4i2.702>
- Baker, F. B. (2001). *The basics of item response theory* (2 ed.). ERIC.
- Bare, Y., Ra'o, P. Y. S. B., & Putra, S. H. J. (2021). Pengembangan Media Teka-Teki Silang Biologi Berbasis Android Materi Sistem Gerak Untuk Meningkatkan Keterampilan Berpikir Kreatif Siswa. *Jurnal Pendidikan Mipa*, 11(2), 158–167. <https://doi.org/10.37630/jpm.v11i2.508>
- Barham, A. I., Ihmeideh, F., Al-Falasi, M., & Alabdallah, A. (2019). Assessment of first-grade students' literacy and numeracy levels and the influence of key factors. *International Journal of Learning, Teaching and Educational Research*, 18(12), 174–195. <https://doi.org/10.26803/ijlter.18.12.11>
- Bellini, D., Crescentini, A., Zanolla, G., Cubico, S., Favretto, G., Faccincani, L., Ardolino, P., & Giancesini, G. (2019). Mathematical Competence Scale (MCS) for primary school: The psychometric properties and the validation of an instrument to enhance the sustainability of talents development through the numeracy skills assessment. *Sustainability (Switzerland)*, 11(9). <https://doi.org/10.3390/su11092569>
- Borg, W. R., & Gall, M. D. (1984). Educational research: An introduction. *British Journal of Educational Studies*, 32(3).
- Bürkner, P.-. C., Schwabe, R., & Holling, H. (2018). Optimal Designs for the Generalized Partial Credit Model. *British Journal of Mathematical and Statistical Psychology*, 72(2), 271–293. <https://doi.org/10.1111/bmsp.12148>
- Candiasa, I. M. (2010). *Pengujian Instrumen Penelitian disertai Aplikasi Iteman dan Bigstep* (U. P. U. P. Ganesha (ed.)).
- Chairunnisa, C. (2016). Manajemen Pendidikan dalam Multi Perspektif. In *Raja Grafindo Persada*. PT Raja Grafindo Persada. <http://digilib.iainkendari.ac.id/1828/>
- Cheng, S., Liu, Q., Chen, E., Huang, Z., Huang, Z., Chen, Y., Ma, H., & Hu, G. (2019). DIRT: Deep learning enhanced item response theory for cognitive diagnosis. *International Conference on Information and Knowledge Management, Proceedings*, 2397–2400. <https://doi.org/10.1145/3357384.3358070>

- D.M. Andikayana, N. Dantes, & I.W. Kertih. (2021). Pengembangan Instrumen Asesmen Kompetensi Minimum (Akm) Literasi Membaca Level 2 Untuk Siswa Kelas 4 Sd. In *Jurnal Penelitian dan Evaluasi Pendidikan Indonesia* (Vol. 11, Nomor 2, hal. 81–92). Universitas Pendidikan Ganesha. <https://doi.org/10.23887/jpepi.v11i2.622>
- Degest. (2020). Asesmen Kompetensi Minimum (AKM). In N. Nur M (Ed.), *Guru Berbagi (Berbagi Informasi Pendidikan)* (Nomor September). PT Remaja Rosdakarya.
- DeMars, C. (2010). *Item response theory*. Oxford University Press.
- Department of Education and skills. (2011). Literacy and Numeracy for Learning and Life: The National Strategy to improve Literacy and Numeracy among children and young people 2011-2020. *Igarss 2014*, 1, 1–5.
- DeSimone, J. A., & James, L. R. (2015). An Item Analysis of the Conditional Reasoning Test of Aggression. *Journal of Applied Psychology*, 100(6), 1872–1886. <https://doi.org/10.1037/apl0000026>
- DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A Psychometric Toolbox for Testing Validity and Reliability. *Journal of Nursing Scholarship*, 39(2), 155–164. <https://doi.org/10.1111/j.1547-5069.2007.00161.x>
- Dewayani, S., Retnaningdyah, P., Antoro, B., Susanto, D., Ikhwanudin, T., Fianto, F., Muldian, W., Syukur, Y., & Setiakarnawijaya, Y. (2021). *Panduan Penguatan Literasi dan Numerasi di Sekolah Dasar*. Sekretariat Direktorat Jenderal PAUD, Pendidikan Dasar dan Menengah. http://repositori.kemdikbud.go.id/22599/1/Panduan_Penguatan_Literasi_dan_Numerasi_di_Sekolah_bf1426239f.pdf
- Djaali, H., & Muljono, P. (2008). Pengukuran dalam bidang pendidikan. *Jakarta: Grasindo*, 2(8), 55.
- Essen, C. B., Idaka, I. E., & Metibemu, M. A. (2017). Item Level Diagnostics and Model - Data Fit in Item Response Theory (IRT) Using BILOG - MG v3.0 and IRTPRO v3.0 Programmes. *Global Journal of Educational Research*, 16(2), 87. <https://doi.org/10.4314/gjedr.v16i2.2>
- Fadel, C. (2008). Century Skills: How Can You Prepare Students for the new. Global Economy? Global Lead Education. Retrieved February, 20, 1–22.
- Flanagan, D. P., & Harrison, P. L. (2005). Contemporary Intellectual Assessment. In *the Guilford Press* (Vol. 40).
- Gall, M. D., Borg, W. R., & Gall, J. P. (2003). *Educational research: An introduction* (7 ed.). Longman Publishing.
- Garvin, A. D., & Ebel, R. L. (1980). Essentials of Educational Measurement. *Educational Researcher*, 9(9), 21. <https://doi.org/10.2307/1175572>

- Griffin, P., McGaw, B., & Care, E. (2012). Assessment and teaching of 21st century skills. *Assessment and teaching of 21st century skills*, 9789400723, 1–345. <https://doi.org/10.1007/978-94-007-2324-5>
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2009). Multivariate Data Analysis. In *Book* (Vol. 87, Nomor 4).
- Hambleton, R. K., Swaminathan, H., & Jane Rogers, H. (1992). Fundamentals of Item Response Theory. In *Contemporary Sociology* (Vol. 21, Nomor 2). Sage. <https://doi.org/10.2307/2075521>
- Han, W., Susanto, D., Dewayani, S., & Pandora, P. (2017). *Materi Pendukung Literasi Numerasi*. Kementerian Pendidikan dan Kebudayaan.
- Hanemann, U. (2019). Examining the application of the lifelong learning principle to the literacy target in the fourth Sustainable Development Goal (SDG 4). *International Review of Education*, 65(2), 251–275. <https://doi.org/10.1007/s11159-019-09771-8>
- Hanif, M. (2024). Pengembangan Model Pembelajaran History Work Camp (HWC) Untuk Meningkatkan Kemampuan Berpikir Kreatif. *Jurnal Pendidikan Dan Pembelajaran Indonesia (Jppi)*, 4(4), 1564–1575. <https://doi.org/10.53299/jppi.v4i4.839>
- Harlen, W., Gipps, C., Broadfoot, P., & Nuttall, D. (1992). Assessment and the improvement of education. *The Curriculum Journal*, 3(3), 215–230. <https://doi.org/10.1080/0958517920030302>
- Hessen, D. J. (2020). Random Effects and Extended Generalized Partial Credit Models. *British Journal of Mathematical and Statistical Psychology*, 74(2), 232–256. <https://doi.org/10.1111/bmsp.12213>
- Hevia, F. J., Vergara-Lope, S., Velásquez-Durán, A., & Calderón, D. (2022). Estimation of the fundamental learning loss and learning poverty related to COVID-19 pandemic in Mexico. *International Journal of Educational Development*, 88. <https://doi.org/10.1016/j.ijedudev.2021.102515>
- Hidayati, K. (2018). Pengembangan Instrumen Penilaian Kompetensi Matematika Materi Statistika Siswa Smp Berdasarkan Kurikulum 2013. In *Program Studi Penelitian Dan Evaluasi Pendidikan Program Pascasarjana Universitas Negeri Yogyakarta* (hal. 1–235).
- UU Republik Indonesia nomor 20 tahun 2003, (2003).
- Indrakusuma, A. D. (2021). Evaluasi pendidikan. In *Jakarta: Rineka Cipta*. Rineka Cipta.
- Jabar, A. dan. (2009). Evaluasi Program Pendidikan Pedoman. In *Jakarta: Bumi Aksara* (2 ed.). Bumi Aksara.
- Jacobs, E. A., Walker, C. M., Miller, T., Fletcher, K. E., Ganschow, P. S., Imbert, D., O'Connell, M., Neuner, J. M., & Schapira, M. M. (2016). Development

- and Validation of the Spanish Numeracy Understanding in Medicine Instrument. *Journal of General Internal Medicine*, 31(11), 1345–1352. <https://doi.org/10.1007/s11606-016-3759-2>
- Jauhari, S. S. (2023). *Perbandingan Skor PISA Indonesia dari Tahun ke Tahun, Alami Penurunan pada 2022*. <https://data.goodstats.id/statistic/perbandingan-skor-pisa-indonesia-dari-tahun-ke-tahun-alami-penurunan-pada-2022-TKKZ3>
- Jonathans, P. M. (2021). *Merdeka Belajar*. Kemendikbudristek.
- Jum C. Nunnally. (1975). Psychometric Theory-25 years Ago and Now. *Educational Researcher*, 4(10), 7–18.
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of educational psychology*, 30(1), 17.
- Kemdikbud. (2021). *Dorong Pemulihan Pembelajaran di Masa Pandemi, Kurikulum Nasional Siapkan Tiga Opsi*. <https://www.kemdikbud.go.id/main/blog/2021/12/dorong-pemulihan-pembelajaran-di-masa-pandemi-kurikulum-nasional-siapkan-tiga-opsi>
- Kemdikbud. (2022). Tanya Jawab Kurikulum Merdeka. In *Kementerian Pendidikan dan Kebudayaan* (hal. 1–50). Kementerian Pendidikan Kebudayaan Riset dan Teknologi. <http://ult.kemdikbud.go.id>
- Kemendikbud. (2020). Desain Pengembangan Soal Asesmen Kompetensi Minimum. *Pusat Asesmen dan Pembelajaran, Badan Penelitian dan Pengembangan dan perbukuan, Kementrian Pendidikan dan Kebudayaan*, 1–125.
- Kemendikbud Ristek. (2025). *Rapor Pendidikan*. <https://raporpendidikan.dikdasmen.go.id/>. https://pusmendik.kemdikbud.go.id/profil_pendidikan/profil-wilayah.php
- Kemendikbudristek. (2022). Buku Saku: Tanya Jawab Kurikulum Merdeka. *Kemendikbudristek*, 9–46. <http://ult.kemdikbud.go.id>
- Kementerian Pendidikan, Kebudayaan, Riset, dan T. (2021). *Modul Literasi Numerasi di Sekolah Dasar*. Kemdikbud Ristek.
- Kenzo. (2024). Analyzing Oxford Happiness Questionnaire Indonesian Version Using the Generalized Partial Credit Model. *Psyche 165 Journal*, 81–86. <https://doi.org/10.35134/jpsy165.v17i2.353>
- Kowsalya, D. N., Lakshmi, V., & Suresh, K. P. (2012). Development and Validation of a Scale to assess Self-Concept in Mild Intellectually Disabled Children. *International J. Soc. Sci. & Education*, 2(4), 2223–4934.
- Leventhal, B. C. (2018). Extreme Response Style: A Simulation Study Comparison of Three Multidimensional Item Response Models. *Applied Psychological Measurement*, 43(4), 322–335. <https://doi.org/10.1177/0146621618789392>

- Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Conference on Human Factors in Computing Systems - Proceedings*, 1–16. <https://doi.org/10.1145/3313831.3376727>
- Mac Donald, A., & Carmichael, C. (2018). Early mathematical competencies and later achievement: insights from the Longitudinal Study of Australian Children. *Mathematics Education Research Journal*, 30(4), 429–444. <https://doi.org/10.1007/s13394-017-0230-6>
- Macqueen, S., Knoch, U., Wigglesworth, G., Nordlinger, R., Singer, R., McNamara, T., & Brickle, R. (2019). The impact of national standardized literacy and numeracy testing on children and teaching staff in remote Australian Indigenous communities. *Language Testing*, 36(2), 265–287. <https://doi.org/10.1177/0265532218775758>
- Mardapi, D. (2017). Pengukuran Penilaian dan Evaluasi Pendidikan Edisi 2. In *Yogyakarta: Parama Publishing* (Vol. 45).
- Marlina, M., Rahim, A., Aziz, F., & Arsalaan, A. T. (2024). Pelatihan Analisis Kualitas Instrumen Penilaian Hasil Belajar Dengan Pendekatan Clasical Theory Dan Item Response Theory Di Sekolah Dasar. *Kacaneegara Jurnal Pengabdian Pada Masyarakat*, 7(2), 209. <https://doi.org/10.28989/kacaneegara.v7i2.1888>
- Mesra, R. (2023). *Research & Development Dalam Pendidikan*. <https://doi.org/10.31219/osf.io/d6wck>
- Meyer, J. P. (2017). jMetrik. In *Handbook of Item Response Theory* (hal. 557–566). Chapman and Hall/CRC.
- Montoya, S. (2018). Defining Literacy Standards. *Defining Literacy Standards*, 17–18. <https://doi.org/10.3726/b14582>
- Muraki, E., & Bock, R. D. (1997). PARSCALE-3: IRT based test scoring and item analysis for graded items and rating scales. In *Chicago: Scientific Software International*.
- Napoli, A. R., & Purpura, D. J. (2018). The home literacy and numeracy environment in preschool: Cross-domain relations of parent–child practices and child outcomes. *Journal of Experimental Child Psychology*, 166, 581–603. <https://doi.org/10.1016/j.jecp.2017.10.002>
- Narpila, S. D. (2023). Pengembangan Instrumen Penilaian Matematika Berbasis Literasi Numerasi Menggunakan Aplikasi Quizizz. *Math Didactic Jurnal Pendidikan Matematika*, 9(3), 446–461. <https://doi.org/10.33654/math.v9i3.2418>
- Nurgiyantoro. (2012). *Teori pengkajian fiksi perpus ummi*. Gadjah Mada University Press.
- OECD. (2009). PISA 2009 Assessment Framework. Key competencies in reading,

- mathematics and science. In *Organisation for Economic Co-operation and Development* (Vol. 20, Nomor 8). PISA. http://www.oecd-ilibrary.org/education/pisa-2009-assessment-framework_9789264062658-en
- OECD. (2019). PISA 2018 Assessment and Analytical Framework. In *OECD Publishing*. OECD publishing. <https://doi.org/10.1787/b25efab8-en>
- OECD. (2023). *PISA 2022 Results The State of Learning and Equity in Education* (I). OECD publishing.
- Ogbeche, A. T., Jude, E. I., Chidi, M. B., Eze, F. B., & Asongo, S. T. (2021). Analysis of Psychometric Qualities of National Examinations Council (NECO) Mathematics Essay Test Using Generalized Partial Credit Model. *The International Journal of Humanities & Social Studies*, 9(12). <https://doi.org/10.24940/theijhss/2021/v9/i12/hs2112-014>
- Patel, N., Baker, S. G., & Scherer, L. D. (2019). Evaluating the Cognitive Reflection Test as a Measure of Intuition/Reflection, Numeracy, and Insight Problem Solving, and the Implications for Understanding Real- World Judgments and Beliefs. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000592>
- Phye, G. D. (1997). Handbook of classroom assessment: learning, achievement, and adjustment. In *Choice Reviews Online* (Vol. 34, Nomor 10). Academic Press. <https://doi.org/10.5860/choice.34-5809>
- Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., & Cella, D. (2011). Item Banks for Measuring Emotional Distress From the Patient-Reported Outcomes Measurement Information System (PROMIS®): Depression, Anxiety, and Anger. *Assessment*, 18(3), 263–283. <https://doi.org/10.1177/1073191111411667>
- Polat, M., Turhan, N. S., & Toraman, Ç. (2022). Comparison of Classical Test Theory vs. Multi-Facet Rasch Theory. *Pegem Egitim ve Ogretim Dergisi*, 12(2), 213–225. <https://doi.org/10.47750/pegegog.12.02.21>
- Pratiwiningtyas, B. N., Susilaningsih, E., & Sudana, I. M. (2017). Pengembangan Instrumen Penilaian Kognitif untuk Mengukur Literasi Membaca. *Journal of Educational Research and Evaluation*, 6(1), 1–9. <http://journal.unnes.ac.id/sju/index.php/jere>
- Puzio, K., Colby, G. T., & Algeo-Nichols, D. (2020). Differentiated Literacy Instruction: Boondoggle or Best Practice? *Review of Educational Research*, 90(4), 459–498. <https://doi.org/10.3102/0034654320933536>
- Retnawati, H. (2014). Teori respons butir dan penerapannya Untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana. In *Yogyakarta: Nuha Medika*.
- Retnawati, H. (2016). Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, dan Psikometrian). In *Parama Publishing*. Parama Publishing.

- Riani, V. R., & Ramalis, T. R. (2020). Pengembangan Tes Keterampilan Proses Sains Siswa Sma Pada Materi Listrik Dan Magnet Dengan Analisis Teori Respons Butir. *Wapfi (Wahana Pendidikan Fisika)*, 5(2), 91–99. <https://doi.org/10.17509/wapfi.v5i2.7567>
- Saepuzaman, D., & Istiyono, E. (2022). Characteristics of Fundamental Physics Higher-Order Thinking Skills Test Using Item Response Theory Analysis. *Pegegog*, 12(4). <https://doi.org/10.47750/pegegog.12.04.28>
- Sälzer, C., & Roczen, N. (2018). Assessing global competence in PISA 2018: Challenges and approaches to capturing a complex construct. *International Journal of Development Education and Global Learning*, 10(1), 5–20. <https://doi.org/10.18546/ijdegl.10.1.02>
- Sarea, M. S., & Ruslan, R. (2019). Karakteristik Butir Soal: Classical Test Theory vs Item Response Theory? *Didaktika Jurnal Kependidikan*, 13(1), 1–16. <https://doi.org/10.30863/didaktika.v13i1.296>
- Sari, D. K., Sukaryawan, M., Ibrahim, A., & Edi, R. (2022). Pendampingan Pengembangan Soal Kimia Berbasis Berpikir Tingkat Tinggi Bidang Studi Kimia Di Kota Prabumulih. *Journal of Sriwijaya Community Service on Education (Jscse)*, 1(2), 7–16. <https://doi.org/10.36706/jscse.v1i2.435>
- Schauburger, G., & Mair, P. (2019). A Regularization Approach for the Detection of Differential Item Functioning in Generalized Partial Credit Models. *Behavior Research Methods*, 52(1), 279–294. <https://doi.org/10.3758/s13428-019-01224-2>
- Setiawati, F. A., Izzaty, R. E., & Hidayat, V. (2018). Analisis Respons Butir Pada Tes Bakat Skolastik. *Jurnal Psikologi*, 17(1), 1. <https://doi.org/10.14710/jp.17.1.1-17>
- Setyawarno, D. (2017). Penggunaan Aplikasi Software Iteman (Item and Test Analysis) Untuk Analisis Butir Soal Pilihan Ganda Berdasarkan Teori Tes Klasik. *Jurnal Ilmu Fisika Dan Pembelajarannya (Jifp)*, 1(1), 11–21. <https://doi.org/10.19109/jifp.v1i1.866>
- Sholikah, M. (2019). Early childhood literacy skills: The effect of socioeconomic status, home literacy environment, and social skills. *International Journal of Innovative Technology and Exploring Engineering*, 9(1), 3769–3775. <https://doi.org/10.35940/ijitee.A4807.119119>
- Snelling, M., Dawes, A., Biersteker, L., Girdwood, E., & Tredoux, C. (2019). The development of a South African Early Learning Outcomes Measure: A South African instrument for measuring early learning program outcomes. *Child: Care, Health and Development*, 45(2), 257–270. <https://doi.org/10.1111/cch.12641>
- Stijic, M. (2023). Numeric Rating Scale for Pain Should Be Used in an Ordinal but Not Interval Manner. A Retrospective Analysis of 346,892 Patient Reports of

- the Quality Improvement in Postoperative Pain Treatment Registry. *Pain*. <https://doi.org/10.1097/j.pain.0000000000003078>
- Subali, B. (2009). Pengembangan Tes Pengukur Keterampilan Proses Sains Pola Divergen Mata Pelajaran Biologi Sma. *Jurdik Biologi, FMIPA*, 4, 581–593.
- Sugiyono. (2019). *Metode Penelitian Pendidikan (Kuantitatif, Kualitatif, Kombinasi, R&D, dan Penelitian Pendidikan*. Alfabeta.
- Sukmadinata, N. S. (2009). Metode Penelitian Pendidikan. In *Bandung: PT. Remaja Rosdakarya*. PT Remaja Rosdakarya.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan RASCH pada assessment Pendidikan*. Trim Komunikata.
- Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, 10(2–3), 211–227. <https://doi.org/10.1007/s10833-009-9105-2>
- Susperreguy, M. I., Douglas, H., Xu, C., Molina-Rojas, N., & LeFevre, J. A. (2020). Expanding the Home Numeracy Model to Chilean children: Relations among parental expectations, attitudes, activities, and children's mathematical outcomes. *Early Childhood Research Quarterly*, 50, 16–28. <https://doi.org/10.1016/j.ecresq.2018.06.010>
- Usami, S. (2010). A polytomous item response model that simultaneously considers bias factors of raters and examinees: Estimation through a Markov Chain Monte Carlo algorithm. *Japanese Journal of Educational Psychology*, 58(2), 163–175. <https://doi.org/10.5926/jjep.58.163>
- Vasquez, V. M., Janks, H., & Comber, B. (2019). Critical Literacy as a way of Being and Doing. *Language Arts*, 96(5), 300–311.
- Vogelsmeier, L. V. D. E., Vermunt, J. K., Keijsers, L., & Roover, K. D. (2020). Latent Markov Latent Trait Analysis for Exploring Measurement Model Changes in Intensive Longitudinal Data. *Evaluation & the Health Professions*, 44(1), 61–76. <https://doi.org/10.1177/0163278720976762>
- Wagiran, W. (2014). Metodologi Penelitian Pendidikan : Teori Dan Implementasi. In *Deepublish*. Deepublish.
- Wallek, R., & Warren, A. (2012). Theory of literature. In *Theory of Literature* (Vol. 15). Harcourt, Brace & World New York. <https://doi.org/10.2307/585969>
- Wardhani, D. F., Suratno, S., & Putra, A. P. (2021). Komparasi Karakteristik Butir Soal Instrumen Penilaian Kognitif Mata Pelajaran IPA SMP. *Wahana-Bio Jurnal Biologi Dan Pembelajarannya*, 12(2), 59. <https://doi.org/10.20527/wb.v12i2.9314>
- Wells, C. S., Hambleton, R. K., & Purwono, U. (2008). Item response theory. Polytomous response IRT models and aplicatios. *Workshop of educational*

assessment and psicometry.

- Williams, R. (2019). National higher education policy and the development of generic skills. *Journal of Higher Education Policy and Management*, 41(4), 404–415. <https://doi.org/10.1080/1360080X.2019.1606690>
- Winarni. (2018). Teori dan Praktik Penelitian Kuantitatif, Kualitatif, Ptk, R&D (R.A. Kusumaningtyas (ed.)). In 2021. Bumi Aksara. <https://play.google.com/books/reader?id=Fx0mEAAQBAJ&pg=GBS.PP1&hl=id>
- Wind, S. A. (2022). Detecting Rating Scale Malfunctioning With the Partial Credit Model and Generalized Partial Credit Model. *Educational and Psychological Measurement*, 83(5), 953–983. <https://doi.org/10.1177/00131644221116292>
- Woodcock, S. (2019). A Within-Subject Experiment of Item Format Effects on Early Primary Students' Language, Reading, and Numeracy Assessment Results. *School Psychology*. <https://doi.org/10.1037/spq0000340>
- Worowirastri Ekowati, D., & Istanti Suwandayani, B. (2019). *Literasi Numerasi Untuk Sekolah Dasar* (Vol. 1). UMMPress.
- Yang, Y., Sun, Y., Zhang, Y., Tang, J., Lee, S., & Miao, D. (2013). Bifactor Item Response Theory Model of Acute Stress Response. *Plos One*, 8(6), e65291. <https://doi.org/10.1371/journal.pone.0065291>
- Yusuf, M. (2015). Asesmen Dan Evaluasi Pendidikan. In *Asesmen Dan Evaluasi Pendidikan*. Prenada Media.
- Zhang, J., Zhang, Y.-Y., Tao, J., & Chen, M. (2022). Bayesian Item Response Theory Models With Flexible Generalized Logit Links. *Applied Psychological Measurement*, 46(5), 382–405. <https://doi.org/10.1177/01466216221089343>