

## DAFTAR PUSTAKA

- Affiong, U. U., Nkwol, N., & Nsungu N. U. (2014). Test Types Students' Achievement in Senior Secondary School Physics and Eradication of Poverty and Hunger In Nigeria. *Journal of Research & Method in Education*, (04(01): 01-05).
- Agung, I.G.N. (2006). *Statistika Penerapan Model Rerata Sel Multivariat dan Model Ekonometri dengan SPSS*. Jakarta: Yayasan SAD Satria Bhakti.
- Ahmad, S. (2013). *Teori Belajar dan Pembelajaran di Sekolah Dasar*. Jakarta: Kencana Prenada Media Group.
- Alwi, I. (2008). Kriteria Empirik Dalam Menentukan Ukuran Sampel Pada Pengujian Hipotesis Statistika Dan Analisis Butir. *Jurnal Formatif*. 2(2): 140-148.
- Anni, C. T. (2004). *Psikologi Belajar*. Semarang: IKIP Semarang Press, h. 4.
- Ariani, T.(2020). Analysis Of Students' Critical Thinking Skills In Physics Problems. *Kasuari:Physics Education Journal (KPEJ)*. DOI: <http://doi.org/10.37891/kpej.v3i1.119>.
- Arikunto, S. (2021). *Dasar-dasar Evaluasi Pendidikan Edisi 3*. Jakarta: Bumi Aksara.
- Azwar, S. (2012). *Validitas dan Reliabilitas*. Yogyakarta: Pustaka Pelajar.
- Azizah, & Wahyuningsih, S. (2020). Penggunaan Model Rasch Untuk Analisis Instrumen Tes Pada Mata Kuliah Matematika Aktuaria. *Jurnal Pendidikan Matematika | Juni 2020 | Volume 3 Nomor 1 | Hal. 45 ± 50* DOI <https://doi.org/10.30598/jupitekv3iss1ppx45-50>
- Bagiyono. (2017) Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal Ujian Pelatihan Radiografi Tingkat 1 The Analysis of Difficulty Level and Discrimination Power of Test Items of Radiography Level 1 Examination. *Widyanuklida*, 16 (1). pp. 1-12. ISSN 1410-5357
- Bagus,H.C. (2012). Administrasi Ujian Nasional (UN) Dengan Menggunakan Model Computerized Adaptive Testing. *Jurnal Pendidikan dan Kebudayaan*, Vol.18 Nomor 1, Maret 2012.
- Baker, F.B. (1992). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker, Inc.
- Bastari, B. (2000). *Linking multiple-choice and constructed-response items to a common proficiency scale*. Doctoral dissertation. University of Massachusetts, Amherst.
- Blais, J. & Raiche, G. (2002). *Some Features of the sampling distribution of the ability estimate in computerized adaptive testing according to two stopping rules*. Makalah disajikan pada 11<sup>th</sup> International Objective Measurement Workshop, New Orleans, April 2002 (tidak dipublikasikan).

- Bock, R.D., & Mislevy, R.J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 4, 431–444.
- Bodmann, S.M. & Robinson, D.H. (2004). Speed and Performance Differences among Computer-Based and Paper-Pencil Tests. *Journal of Educational Computing Research*, 31(1), 51-60
- Bolt D. M., Cohen A. S., Wolack J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, 39, 331-348. [Google Scholar]
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition (2nd ed.)*. Psychology Press. <https://doi.org/10.4324/9781410614575>
- Borg, W.R and Gall, M.D. (2003). *Educational Research: An Introduction 4 th Edition*. London: Longman Inc.
- Boughton, K., Smith, J., & Ren, H. (2017). Using response time data to detect compromised items and/or people. In Wollack, J, and Fremer, J. (Eds.), *Handbook of Test Security*, Routledge, New York, NY.
- Bunderson, C.V., Inouye, D. K., & Olsen, J.B. (1989). The Four Generations of Computerized Educational Measurement. Dalam Robert L. Linn. *Educational Measurement*. 3<sup>rd</sup> ed. New York: American Council on Education & Macmillan Publishing Company.
- Cahyo, S. D. , Al Fariz, A. B., & Lestari, C. A. (2020) Does Internet Usage Frequency Give Impact To Student's Academic Performance?. *Indonesian Journal of Educational Assessment*. DOI: <https://doi.org/10.26499/ijea.v3i1.57>
- Chadha, N. K. (2009). *Applied psychometry*. New Delhi: Chaman Enterprises
- Choe, E.M. Kern, J. L., & Chang, H. (2017). Optimizing the Use of Response Times for Item Selection in Computerized Adaptive Testing. *Journal of Educational and Behavioral Statistics*. Vol. XX, No. X, pp. 1–24. DOI: 10.3102/1076998617723642.
- Cohen, A.S., Kane, M.T., & Seock-Ho Kim,S. (2001). The Precision of Simulation Study Results. *Applied Psychological Measurement* 25: 136
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Forth Worth: Holt, Rinehart, and Winston Inc.
- Dariyanto. (2010). *Media Pembelajaran Peranannya Sangat Penting Dalam Mencapai Tujuan Pembelajaran*. Yogyakarta: Gava Media.

- Davey, T. (2011). *A guide to computer adaptive testing systems*. Washington, DC: Council of Chief State School Officers.
- Dipayana, I. K. M., Gading, I. K., & Japa, I. G. N. (2019). Pengaruh Pendekatan Pembelajaran Matematika Realistik Indonesia terhadap Hasil Belajar Matematika Ditinjau dari Minat Komputasi. *MIMBAR PGSD Undiksha*, 7(3). <https://doi.org/10.23887/jjpgsd.v7i3.19466>
- Dodd, B.G. (1990). The effect of item selection procedure and stepsize on computerized adaptive attitude measurement using the rating scale model. *Applied Psychological Measurement*, 4, 355 – 366.
- Embretson, S.E., & Reise, S.P. (2000). *Item Response Theory for Psychologist*. New Jersey : Lawrence Erlbaum Associates, Inc.
- Entink, R.K., Fox, J.P., & van der Linden, W. J. (2007). Modeling of Responses and Response Times with the Package cirt. *Journal of Statistical Software*, Volume 20, Issue 7.
- Entink, R.K., Fox, J.-P., & van der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika*, 74(1), 21.
- Fachruddin, Imam. (2009). *Desain penelitian*. Malang: Universitas Islam Negeri.
- Fitroturrohman, M. Purwadi, & Azizah, M. (2019). Hubungan Peran Orang Tua Dengan Prestasi Belajar Siswa Kelas Tinggi SDN Kedung 01 Jepara. *Journal of Primary and Children's Education*. DOI: <http://dx.doi.org/10.35473/jnctt.v2i2.258>
- Flaugher, R.L., Melton, R. S., Myers, C.T. (1968). Item Rearrangement Under Typical Test Conditions. *Educational and Psychology measurement*, 1968. 813-824.
- Fox, J.P. & Marianti, S. (2017). Person-fit statistics for joint models for accuracy and speed. *Journal of Educational Measurement*, 54(2), 243–26
- Gelman, A., Carlin, J.B., Stern H.S., Rubin D.B. (1995). *Bayesian Data Analysis*, 2nd Edition. Chapman and Hall, London.
- Goegebeur, Y., De Boeck, P., Wollack, J. A., & Cohen, A. S. (2008). A speeded item response model with gradual process change. *Psychometrika*, 73(1), 65-87. <https://doi.org/10.1007/s11336-007-9031-2>
- Goral, D. P., & Bailey, A. L. (2019). Student self-assessment of oral explanations: Use of language learning progressions. *Language Testing*, 36(3), 391-417. <https://doi.org/10.1177/0265532219826330>
- Guisse, M., Knott, A., & Benuskova, L. (2014). A Bayesian model of polychronicity. *Neural Computation*, 26(9), 2052–2073. doi:10.1162/NECO\_a\_00620.

- Guzman, E., Conejo, R., & Garcia-Hervas, E. (2005). An authoring environment for adaptive testing. *Journal of Educational Technology & Society*, 8 (3), 66-76.
- Haar, J. H. & Hansen, M. E. (2006). *Adult Skills Assessment in Europe: Feasibility Study*. Policy and Business Analysis, Final Report”, November 2006.
- Halik, J.M. (2000). *Development of computerized middle school achievement test*. Soul, South Korea: Komesa Press.
- Hamiseno, G. S., Bagus, H. C., Fahmi, Zulaiha,Z., Susetiyo, H., & Rahman,A. (2020). *Processing National Examination Result for Junior High School, Senior High School and Vocational School in Indonesia*. Proceedings of the International Conference on Educational Assessment and Policy (ICEAP 2020)
- Hadi, F. R. (2021). Kemampuan mahasiswa dalam menyelesaikan soal matematika bertipe High Order Thinking Skills (HOTS). *Jurnal Program Studi Pendidikan Matematika*. Volume 10, No. 2, 2021, 872-879
- Halik, A. S., Mania, S., & Nur, F. (2019). Analisis Butir Soal Ujian Akhir Sekolah (Uas) Mata Pelajaran Matematika Pada Tahun Ajaran 2015/2016 Smp Negeri 36 Makassar. *Journal of Islamic Education*, 1(1), 11. <https://doi.org/10.24252/asma.v1i1.11249>
- Hall, E. T., & Hall, M. R. (1987). *Hidden differences: Doing business with the Japanese*. Garden City, NY: Doubleday.
- Hambleton, R.K. and Swaminathan, H. (1985). *Item Response Theory: principles and application*. Boston: Kluwer.
- Hambleton, R.K., Swaminathan, H, & Rogers, H.J. (1991). *Fundamental of Item Response Theory*. Volume 2. California: Sage Publications, Inc.
- Hambleton, R.K., & Swaminathan, H. (1991). *Item Response Theory : Principles and Applications*. New York : Springer Science+Business Media.
- Hambleton, R.K., & Jones, R.W (1993). An NCME Instructional Module on : Comparison of Classical Theory and Item Theory and Their Application to Test Development. *Educational Measurement : Issues and Practice*, 12 (3). p : 38-47.
- Hayat, B., Putra, M., & Suryadi, B. (2020). Comparing item parameter estimates and fit statistics of the Rasch model from three different traditions. *Jurnal Penelitian dan Evaluasi Pendidikan*, 24(1), 39-50. doi:<https://doi.org/10.21831/pep.v24i1.29871>
- Hayati, N., & Mardapi, D. (2014). Pengembangan Butir Soal Matematika SD di Kabupaten Lombok Timur Sebagai Upaya Dalam Pengadaan Bank Soal. *Jurnal Kependidikan*, 44(1), 26-38.
- He,W., & Reckase, M.D.(2014). Item Pool Design for an Operational Variable-Length Computerized Adaptive Test. *Educational and Psychological Measurement*. Vol. 74(3) 473–494.



- Herkusumo, A. P. (2011). Penyetaraan (equating) Ujian Akhir Sekolah Berstandar Nasional (UASBN) dengan teori tes klasik. *Jurnal Pendidikan Dan Kebudayaan*, 17(4), 455–471. <https://doi.org/10.24832/jpnk.v17i4.41>
- Hidayah, N. (2013). Ujian nasional dalam Perspektif Kebijakan Publik. *Jurnal Pencerahan*, 1, 35–40. <https://doi.org/https://doi.org/10.13170/jp.7.1.2053>
- Hidayat, T. (2021). Penerapan Prinsip Efektif dan Efisien dalam Pelaksanaan Monitoring Kegiatan Penelitian. *Majalah Media Perencana*, 2(1), 42-50. Retrieved from <https://mediaperencana.perencanapembangunan.or.id/index.php/mmp/article/view/15>
- Hornke, L.F. (2000). Item Response Times in Computerized Adaptive Testing. *Psicológica*. 21, 175-178.
- Howard, S. J., Woodcock, S., Ehrich, J., & Bokosmaty, S. (2016). What are standardized literacy and numeracy tests testing? Evidence of the domain-general contributions to students' standardized educational test performance. *British Journal of Educational Psychology*, 87(1), 108–122. doi:10.1111/bjep.12138
- Hullin, Lissak, H., Drasgow, P., & Parsons, C.K. (1983). *Item Respon Theory, Applications to Psychological Measurement*. Illinois: Dow Jones-Irwin.
- Husman, H. (2011). *Manajemen Teori Praktik dan Riset Pendidikan*. Jakarta: PT Bumi Aksara.cet ke-3, hlm. 2
- Idrus, L.(2019). Evaluasi Dalam Proses Pembelajaran. *Jurnal Manajemen Pendidikan Islam*. DOI : 10.35673/ajmpi.v9i2.427
- Imai, S. (2008). Development of Japanese-Computerized Adaptive Test. Didownload 18 Juli 2009 dari <http://web.cc.yamaguchi-u.ac.jp/~imai2002/j-cat/gaiyou/2008Development-of-JCAT.pdf>.
- Istiyono, E., Mardapi, D., & Suparno. (2014). Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PystHOTS) Peserta Didik SMA. *Jurnal Penilaian dan Evaluasi Pendidikan*. DOI: <https://doi.org/10.21831/pep.v18i1.2120>
- Karyadi, K. (2017). Peningkatan Hasil Belajar Melaporkan Isi Bacaan melalui Metode Diskusi. *Briliant: Jurnal Riset dan Konseptual*, 2(1), 80-91. doi:<http://dx.doi.org/10.28926/briliant.v2i1.28>
- Kayon, I. W. (2022). Upaya Optimalisasi Prestasi Belajar Matematika melalui Pemberian Tes secara Kontinu pada Siswa Kelas XI. *Journal of Education Action Research*, 6(1), 102–108. <https://doi.org/10.23887/jear.v6i1.44768>
- Kim, J. & McLean, J.E. (1995). *The Influence of Examinee Test-Taking Motivation in Computerized Adaptive Testing*. Makalah disajikan pada the Annual Meeting of the National Council on Measurement in Education. (tidak diterbitkan).

- Kirbani, Budiyo, & Sutanto. (2013). Pengembangan Model Assessment for Learning (AFL) melalui Penilaian Teman Sejawat untuk Pembelajaran Matematika pada Pokok Bahasan Persamaan Garis Lurus di Madrasah Tsanawiyah Pondok Pesantren Modern Islam Asslam Sukoharjo. *Jurnal Pembelajaran Matematika*, 1(2), 123–132.
- Kizlik, Bob. (2009). Measurement, Assessment, And Evaluation In Education. Versi Online : <http://www.adprima.com/measurement.html>.
- Kolen, M.J., & Brennan, R.L. (2004). *Tes Equating, scaling and linking : Methods and practices (3<sup>rd</sup> ed)*. New York: Springer Verlag New York Inc.
- Kubinger, K. D. (2016). Editorial: Review and forecast on research in Psychological Test and Assessment Modeling. *Psychological Test and Assessment Modeling*, 58(1), 3–13.
- Kunandar. (2013). *Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013): Suatu Pendekatan Praktis Disertai Dengan Contoh*. Jakarta: Rajawali Pers
- Lau, S.H., Hong, K.S., Lau, N.K., & Usop, H. (2005). Web based computer-Adaptive multiple choice assessment (CAAS) using the new NRET scoring method. 27 November 2009 dari [http://atcm.mathandtech.org/EP2008/papers\\_full/2412008\\_15339.pdf](http://atcm.mathandtech.org/EP2008/papers_full/2412008_15339.pdf).
- Leung, C., Chang H., & Hau K. (2005). Computerized Adaptive Testing: A Mixture Item Selection Approach for Constrained Situations. *British Journal of Mathematical & Statistical Psychology*, Nov 2005, 58, Proquest Psychology Journals, pp 239.
- Luecht, R. M., Champlain, A. D., & Nungester, R. J. (1998). Maintaining content validity in computerized adaptive testing. *Advances in Health Sciences Education*, 3, 29–41.
- Lu, Y. Sireci, S. G. (2007). Validity Issues in Test Speededness. *Journal of Education Measurement*. <https://doi.org/10.1111/j.1745-3992.2007.00106.x>
- Lord, F.M. (1977). A broad-range tailored test of verbal ability. *Applied Psychological Measurement*, 1, 95–100.
- Mahfudi, H.N. (2020). Hubungan Peran Orang Tua Terhadap Prestasi Belajar Siswa Kelas V SDN Legokulon 2. *Jurnal Pembelajaran dan Pendidikan Dasar*. ISSN 2747-0709 (online)
- Mardiana, & Handayani, F. S. (2017). Analisis Kesiapan Pelaksanaan Ujian Nasional Berbasis Komputer (UNBK) Pada SMKN Kota Palembang. *Seminar Nasional Teknologi Informasi, Bisnis, dan Desain 2017*. ISBN: 978-602-74635-1-6.
- Maris, E. (1993). Additive and multiplicative models for gamma distributed variables, and their application as psychometric models for response times. *Psychometrika*, 58, 445–469.

- Martin, L., & Mahat, M. (2017). The Assessment of Learning Outcomes in Australia: Finding the Holy Grail. *AERA Open*. <https://doi.org/10.1177/2332858416688904>
- Mehrens, W.A. & Lehmann (1986). *Measurement and evaluation in education and psychology*. (3th ed). New York: Holt Rinehart and Winston.
- Miller, M.D, Linn, R.L., & Gronlund, N.E. (2009). *Measurement and Assessment in Teaching*. Merrill/Pearson.
- Millman, J., & Arter, J.A. (1984). Issues in Item Banking. In *Journal of Educational Measurement*, Volume 21, No. 4, Winter.
- Morgan, C., & O'Reilly, M. (2001). Innovations in Online Assessment. In F. Lockwood & A. Gooley (Eds.), *Innovation in "Open and Distance Learning: Successful Development of Online and Web-Based Learning"* London: Kogan Page Limited. 2001, pp. 179-188.
- Mulyana, Deddy. (2004). *Ilmu Komunikasi : Suatu Pengantar*. Bandung: Remadja Rosdakarya
- Naga, Dali. S. (2013). *Teori Skor pada Pengukuran Mental*. Jakarta: PT. Nagarani Citrayasa.
- Ng, S. W., Kwan, Y. W., & Huey Lei, K. H. (2020). Enhancing Student Learning and Teacher Development: Does "Assessment for Learning" Matter? *International Journal of Educational Reform*, 29(3), 275–292. <https://doi.org/10.1177/1056787920918595>
- Nitko, Anthony J. (1996). *Educational Assessment of Students*, Second Edition. Ohio: Merrill an imprint of Prentice Hall Englewood Cliffs.
- Nurdin, S. (2005). *Model Pembelajaran yang Memperhatikan Keragaman Individu Siswa dalam Kurikulum Berbasis Kompetensi*, Ciputat: Quantum Teaching.
- Oshima, T. C. (1994). The Effect of Speededness on Parameter Estimation in Item Response Theory. *Journal of Education and Measurement*. <https://doi.org/10.1111/j.1745-3984.1994.tb00443.x>
- Overton, T. (2008). *Assessing Learners with Special Needs: An Applied Approach* (7th Edition), University of Texas – Brownsville.
- Partchev, I., Boeck, P.D., & Steyer, R. (2013). How Much Power and Speed Is Measured in This Test?. *Assessment* 2013 20: 242
- Pramono, A. J. B., & Retnawati, H. (2020). Implementation of CAT in Indonesia School: Current Challenges & Strategies. *Universal Journal of Educational Research*, 8(11): 5599-5609. DOI: 10.13189/ujer.2020.081164.

- Purnomo, Y. W. (2014). Assessment Based Learning: Sebuah Tinjauan untuk Meningkatkan Motivasi Belajar dan Pemahaman Matematis. *Sigma Journal*, ISSN: 1411-5166 No.01, Vol. VI
- Rapono, M., Safrial, & Wijaya, C. (2019). Urgensi Penyusunan Tes Hasil Belajar: Upaya Menemukan Formulasi Tes Yang Baik dan Benar. *Jurnal Pendidikan Ilmu-ilmu Sosial*, 11 (1) (2019): 95-104. <https://doi.org/10.24114/jupiiis.v11i1.12227>
- Rose, N. (2010). Maximum Likelihood and Bayes Modal Ability Estimation in Two-Parametric IRT Models: Derivations and Implementation. *Jurnal Institute of Psychology*.
- Reckase, M. D. (2003). Item pool design for computerized adaptive tests. *Paper presented at the National Council on Measurement in Education*, Chicago, IL.
- Retnawati, H. & Hadi, S. (2014). Sistem Bank Soal Daerah Terkalibrasi untuk Menyongsong Era Desentralisasi. *Jurnal Ilmu Pendidikan*. Jilid 20. Nomor 2. Hlm. 183-193
- Reynolds, Cecil R., Livingstone, Ronald B., & Wilson Victor (2010). *Measurement and Assessment in Education*. New Jersey: Pearson Education Inc.
- Riley, B. B., Dennis, M. L., & Conrad, K. J. (2010). A Comparison of Content-Balancing Procedures for Estimating Multiple Clinical Domains in Computerized Adaptive Testing: Relative Precision, Validity, and Detection of Persons With Misfitting Responses. *Applied Psychological Measurement*, 34(6), 410–423. <https://doi.org/10.1177/0146621609349802>
- Riza, L. S., Nurjanah, N. S., & Wihardi, Y. (2019). The development of computer adaptive test and item response theory with 4 parameters based the logistics model, *J. Phys. Conf. Ser.*, vol. 1280, no. 3, pp. 1–6, 2019. <https://doi.org/10.1088/1742-6596/1280/3/032033>.
- Robert, J. G. (2013). *Tes psikologi sejarah, prinsip dan aplikasi*. Jakarta: Erlangga. h. 218
- Roskam, E.E. (1997). *Models for Speed and Time Limit Test*, dalam W. J. Van der Linden & R.K. Hambleton (eds.), *Handbook of Modern Item Response Theory* (pp. 169 – 185), New York : Springer.
- Rukli, (2013). Generation Paket Soal dalam Penilaian Adaptif. *Prosiding Seminar Himpunan Evaluasi Pendidikan Indonesia*, Manado. ISBN 978-602-96343-1-0
- Saukah, A., & Cahyono, A,E . (2015). Ujian Nasional Di Indonesia Dan Implikasinya Terhadap Pembelajaran Bahasa Inggris. *Jurnal Penelitian dan Evaluasi Pendidikan*.p-ISSN: 1410-4725, e-ISSN: 2338-6061
- Sax, G., & Carr, A. (1962). An investigation of response sets on altered parallel forms. *Educational and Psychological Measurement*, 22(2), 371–376.



- Sax, Gilbert. (1980). *Principles of Educational and Psychological Measurement and Evaluation*. (2nd ed.). California : Wadsworth Publishing Company
- Scheiblechner, H. (1979). Specific objective stochastic latency mechanisms. *Journal of Mathematical Psychology*, 19, 18–38.
- Scrums, D. J., & Schnipke, D. L. (n.d). *A log-normal response time model*. Princeton NJ: Educational Testing Service.
- Segall, D.O. & Moreno (1999). Development of the Computerized Adaptive Testing Version of the Armed Services Vocational Aptitude Battery. Dalam Fritz Drasgow & Julie B. Olson-Buchanan. *Innovations in Computerized Assessment*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan. *Jurnal Manajemen & Pendidikan Islam*, 2(2), 192-213
- Sholihah, F., Susilaningtyas, T., & Siskawati, F. S. (2021). The Analysis of Mathematics Question Items Quality at SMPN 01 Pakusari In the Online Learning Period. *Jurnal Axioma : Jurnal Matematika Dan Pembelajaran*, 6(1), 29–37.
- Sudijono, A. (2011). Pengantar Evaluasi. Raja Grafindo Persada.
- Solomon, J. L.(1949). An Extension of Control Group Design. *Psychological Bulletin*, Vol. 46, 1949, pp. 137-150
- Srirahayu, R., & Arty, I. (2018). Pengembangan instrumen experiment performance assessment untuk menilai keterampilan proses sains dan kerja sama. *Jurnal Penelitian dan Evaluasi Pendidikan*, 22(2), 168-181.  
doi:<https://doi.org/10.21831/pep.v22i2.20270>
- Suhady, W., Roza, Y., & Maimunah, M. (2020). Pengembangan Soal untuk Mengukur Higher Order Thinking Skill (HOTS) Siswa. *Jurnal Gantang*, 5(2), 143-150.  
<https://doi.org/10.31629/jg.v5i2.2518>
- Sukardi. (2011). *Metodologi Penelitian Pendidikan: kompetensi dan Praktiknya*. Jakarta: Bumi Aksara
- Sumardyono, & Wiworo. (2011). *Pengembangan dan Pengelolaan Bank Soal Matematika di KKG/MGMP*. Kementerian Pendidikan Nasional Badan Pengembangan Sumber Daya Manusia Pendidikan dan Penjaminan Mutu Pendidikan Pusat Pengembangan dan Pemberdayaan Pendidik dan Tenaga Kependidikan Matematika.
- Surapranata, S., (2009). *Analisis, Validitas, Reliabilitas, dan interpretasi Hasil Tes Implemetasi Kurikulum 2004*. Bandung: PT. Remaja Rosdakarya.
- Suprijadi, D. (2010). Pengaruh Penggunaan Tutor Sebaya Terhadap Hasil Belajar matematika Siswa Kelas VII SMP Darussalam, Jakarta. *Jurnal Ilmiah Faktor Exacta*, Vol. 3 No. 2 Juni 2010. Univ. Indraprasta

- Susanto, E., & Retnawati, H. (2016). Perangkat pembelajaran matematika bercirikan PBL untuk mengembangkan HOTS siswa SMA. *Jurnal Riset Pendidikan Matematika*, 3(2), 189-197. doi:<http://dx.doi.org/10.21831/jrpm.v3i2.10631>
- Taylor, J., R. (1999). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books. pp.128–129. ISBN 0-935702-75-X.
- Thissen, D. (1983). *Latent Trait Scoring of Timed Ability Tests*, University of Kansas. Diambil pada tanggal 25 Nopember 2012 dari <http://www.psych.umn.edu/psylab>.
- Thissen, D., & Mislevy, R. J. (1990). Testing Algorithms. Dalam H. Wainer, N.J. Dorans, R. Flugher, & B.F. Green, *Computerized Adaptive Testing: a Primer*. New Jersey: Lawrance Erlbaum Associates, Publishers.
- Tosuncuoglu, I. (2018). Importance of Assessment in ELT. *Journal of Education and Training Studies*, v6 n9 p163-167.
- Towle, N.J. & Merrill, P.F. (1975). How Students Use Behavioral Objectives. *Educational Technology*, 15(12), 41-42
- Ulum (2008). *Akuntansi Sektor Publik*, UMM Press, Malang
- Van der Linden, J., W. (2006) A Lognormal Model for Response Times on Test Items. *Journal of Educational and Behavioral Statistics*. Summer 2006, Vol. 31, No. 2, pp. 181–204
- Van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, 72(3), 287-308.
- Van der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, 46(3), 247-272.
- Van der Linden, W. J. & Xiong, X.(2013). Speededness and Adaptive Testing. *Journal of Educational and Behavioral Statistics*. <https://doi.org/10.3102/1076998612466143>
- Van der Linden, W.J., Ariel, A., & Veldkamp, B.P. (2006). Assembling a Computerized Adaptive Testing Item Pool as a Set of Linear Tests. *Journal of Educational and Behavioral Statistics*, Spring, Vol. 31, No. 1, pp. 81–99
- Van der Linden, J.W., Scrams, D.J., & Schnipke, D.L.(1999). Using Response-Time Constraints to Control for Differential Speededness in Computerized Adaptive Testing. *Applied Psychological Measurement*, Vol. 23 No. 3, 195–210.
- Verhelst, N.D., Verstralen, H.H.F.M. & Jansen M.G.H. (1997). *A Logistic Model for Time-Limit Test*, dalam W. J. Van der Linden & R.K. Hambleton (eds.), *Handbook of Modern Item Response Theory* (pp. 169 – 185), New York : Springer

- Van Muilekom, M. M., Luijten, M. A. J., van Litsenburg, R. R. L., Grootenhuis, M. A., Terwee, C. B., & Haverman, L. (2021). Psychometric properties of the Patient-Reported Outcomes Measurement Information System (PROMIS®) Pediatric Anger Scale in the Dutch general population. *Psychological Assessment*, 33(12), 1261–1266. <https://doi.org/10.1037/pas0001051>
- Wainer, H. (2000). *Computerized Adaptive Testing: A Primer (2nd ed)*. Lawrence Erlbaum Associates, Mahwah NJ.
- Wainer, H. & Mislevy, R.J. (1990) Item Response Theory, Item Calibration, and Proficiency Estimation. Dalam H. Wainer, N.J. Dorans, R. Flueher, & B.F. Green, *Computerized Adaptive Testing: a Primer*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Wang, C., Chang, H. H., & Douglas, J. A. (2013). The linear transformation model with frailties for the analysis of item response times. *British Journal of Mathematical and Statistical Psychology*, 66(1), 144–168
- Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Acta Psychologica Sinica Statistical Psychology*, 68(3), 456–477.
- Ward, Annie W. dan Murray-Ward, Mildred. (2004). Guidelines for the Development of Item Banks. Modul pembelajaran NCME. dalam *Instructional Topics in Educational Measurement (ITEMS)*.
- Wen-shuenn Wu. (2004). Development of an online adaptive vocabulary test system. Didownload 23 November 2009 dari <http://people.chu.edu.tw/~wswu/publications/papers/conferences/06.pdf>.
- Widoyoko, E. (2015). *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Pelajar.
- Wright, B., & Stone M. (1999). *Measurement Essentials*. Wilmington: Ide Range, inc.
- Zhan, P., Jiao, H., & Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology*, 71(2), 262–286