

DAFTAR PUSTAKA

- Abdurrahman, M. (2003). *Pendidikan bagi anak berkesulitan belajar*. Rineka
- Abosalem, Y. (2016). Assessment techniques and students' higher-order thinking skills. *International Journal of Secondary Education*, 4(1), 1–11. <https://doi.org/10.11648/j.ijsedu.20160401.11>
- Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Wiley.
- Ahmad Mukhibin, A., Suyanto, E., & Widodo, S. (2024). Teachers' challenges in implementing diagnostic assessment in secondary education. *Journal of Educational Evaluation*, 15(2), 145–160.
- Ainsworth, Shaaron, Russell Tytler, and Vaughan Prain. "Learning by construction of multiple representations." *Handbook of learning from multiple representations and perspectives*. Routledge, 2020. 92-106.
- Airasian, P. W., & Gullickson, A. R. (1994). *Assessment and evaluation in the classroom*. McGraw-Hill.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Arslan, H. O., Cigdemoglu, C., & Moseley, C. (2012). A three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. *International Journal of Science Education*, 34(11), 1667–1686.
- Ashlock, R. B. (1986). *Error patterns in computation: Using error patterns to improve instruction*. Merrill.
- Ayeh, I. G. (2025). Students' mathematics conceptual challenges: Exploring students' thinking, understanding, and misconceptions in functions and graphs. *European Journal of Science and Mathematics Education*, 13(3), 191–206. <https://doi.org/10.30935/scimath/16596>
- Bandura, A. (1997). *Self-efficacy: The exercise of control* (Vol. 11). Freeman.
- Bariyah, N., Setyaningrum, W., & Fatwana, R. (2024). Students' difficulties and misconceptions in exponential functions. *Journal of Mathematics Education*, 15(1), 33–45.
- Biggs, J. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.
- Bird, T., et al. (2005). *Assessment practices and student learning*. Routledge.
- Birenbaum, M., Kimron, H., Shilton, H., & Shahor, N. (2015). Cyclic processes of assessment for learning: Exploring teachers' perspectives. *Assessment in Education: Principles, Policy & Practice*, 22(1), 58–76. <https://doi.org/10.1080/0969594X.2014.988208>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Bollen, J. S. (1998). Long (Eds.) *Testing Structural Equation Models* (pp. 136-162). Sage

- Bradshaw, L., Izsák, A., Templin, J., & Jacobson, E. (2014). Diagnosing teachers' understandings of rational numbers: Building a multidimensional test within the diagnostic classification framework. *Educational Measurement: Issues and Practice*, 33(1), 2–14. <https://doi.org/10.1111/emip.12020>
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2004). *How people learn: Brain, mind, experience, and school*. National Academy Press.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281.
- Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education*, 32(7), 939–961.
- Caleon, I., & Subramaniam, R. (2010). Do students know what they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*, 40(3), 313–337.
- Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education*, 12(1), 39–54.
- Çelik, H., & Özdemir, O. F. (2020). Development of six-tier diagnostic tests in science education. *Journal of Science Education*, 31(2), 145–160.
- Chadha, N. K. (2009). *Applied psychometrics*. Sage.
- Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, 8(3), 293–307.
- Chiu, M. S. (2010). Using multiple-choice questions to diagnose students' misconceptions in science. *Educational Research Review*, 5(2), 131–145. Cipta.
- Cornoldi, C., & Lucangeli, D. (2004). Mathematics and learning disabilities. *Learning Disability Quarterly*, 27(1), 21–31.
- Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage Publications.
- de la Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, 33(3), 163–183. <https://doi.org/10.1177/0146621608320523>
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational Measurement*, 46(1), 115–130. <https://doi.org/10.1111/j.1745-3984.2009.01078.x>
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179–199. <https://doi.org/10.1007/s11336-011-9207-7>
- de la Torre, J., & Chiu, C.-Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika*, 81(2), 253–273. <https://doi.org/10.1007/s11336-015-9467-8>

- de la Torre, J., & Chiu, C.-Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika*, 81(2), 253–273. <https://doi.org/10.1007/s11336-015-9467-8>
- de la Torre, J., & Lee, Y.-S. (2013). Evaluating the Wald test for item-level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, 50(4), 355–373. <https://doi.org/10.1111/jedm.12022>
- de la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa*, 20(2), 89–97. <https://doi.org/10.1016/j.pse.2014.11.001>
- De la Torre, J., & Sorrel, M. A. (2023). Cognitive diagnosis models. *New handbook of mathematical psychology*, 3, 385–420.
- Dindar, A. C., & Geban, O. (2011). Development of a three-tier test to assess high school students' understanding of acids and bases. *Procedia Social and Behavioral Sciences*, 15, 600–604.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. SAGE Publications.
- Ergul, R., Sen, M., & Okur, M. (2022). Seven-tier diagnostic assessment for metacognitive skills. *Educational Measurement*, 41(3), 215–230.
- Fitriani, A., Suryadi, D., & Turmudi. (2022). Students' conceptual understanding of exponential functions in senior high school. *Journal of Mathematics Learning*, 6(2), 120–132.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231–235). Lawrence Erlbaum Associates.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8, Article 443. <https://doi.org/10.3389/fnhum.2014.00443>
- Gierl, M. J., Zheng, Y., & Cui, Y. (2008). Using cognitive diagnosis models to scale educational tests. *Journal of Educational Measurement*, 45(3), 227–249.
- Goldin, G. A. (2008). Perspective on representation in mathematical learning and problem solving. *Handbook of international research in mathematics education*, 2, 176–201.
- Gonen, S., & Kocakaya, S. (2010). A four-tier diagnostic test for assessing students' misconceptions in physics. *Procedia Social and Behavioral Sciences*, 2, 2383–2388.
- Gradini, E. (2019). Analisis kemampuan numerasi siswa Indonesia berdasarkan hasil PISA. *Jurnal Pendidikan Matematika*, 13(1), 45–56.
- Green, S. (2018). *Assessment for learning in the classroom*. Routledge.
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 989–1008.
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26(4), 301–321.
- Haring, N. G., & Eaton, M. D. (1978). *Systematic instructional procedures*. Merrill.

- Harlen, W. (2005). Teachers' summative practices and assessment for learning. *Assessment in Education*, 12(2), 169–182.
- Hasan, S., Bagayoko, D., & Kelley, E. L. (1999). Misconceptions and the certainty of response index (CRI). *Physics Education*, 34(5), 294–299. <https://doi.org/10.1088/0031-9120/34/5/304>
- Himmah, F., & Wiyarno, Y. (2023). Development of CBT-based diagnostic assessment using Moodle. *Journal of Educational Technology*, 17(3), 210–222.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. Basic Books.
- Jitendra, A. K., & Woodward, J. (2019). The role of visual representations in mathematical word problems. *Cognitive foundations for improving mathematical learning*, 269–294. Pengganti arcavi
- Johnson, D. J., & Myklebust, H. R. (1967). *Learning disabilities: Educational principles and practices*. Grune & Stratton.
- Julaeha, S., Hidayat, T., & Rustaman, N. Y. (2020). Development of three-tier multiple choice test using Google Classroom. *Journal of Science Education*, 4(2), 98–109.
- Junker, B. W., & Sijsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258–272. <https://doi.org/10.1177/01466210122032064>
- Kaiser, G. (2020). Mathematical modelling and applications in education. In *Encyclopedia of mathematics education* (pp. 553–561). Cham: Springer International Publishing.
- Kaltakci-Gurel, D., Eryilmaz, A., & McDermott, L. C. (2017). Development and application of a four-tier diagnostic test. *Physical Review Physics Education Research*, 13(1), 010113.
- Kemendikbud. (2020). *Panduan asesmen diagnostik*. Kementerian Pendidikan dan Kebudayaan Republik Indonesia.
- Kemendikbudristek. (2023). *Laporan hasil Asesmen Nasional 2023*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
- Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2021). Projecting the potential impacts of COVID-19 school closures on academic achievement. *Educational Researcher*, 49(8), 549–565.
- Kuhn, D. (2008). *Education for thinking*. Harvard University Press.
- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). A practical illustration of multidimensional diagnostic skills profiling. *Educational Research and Evaluation*, 18(1), 3–21. <https://doi.org/10.1080/13803611.2011.640874>
- Kusaeri. (2012). *Pengukuran dan penilaian pendidikan*. Graha Ilmu.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>

- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Leighton, J. P., & Gierl, M. J. (2007). *Cognitive diagnostic assessment for education*. Cambridge University Press.
- Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy method for cognitive assessment. *Journal of Educational Measurement*, 41(3), 205–237.
- Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
- Lerner, J. W. (1988). *Learning disabilities: Theories, diagnosis, and teaching strategies*. Houghton Mifflin.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385.
- Ma, W., & de la Torre, J. (2020). GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software*, 93(14), 1–26. <https://doi.org/10.18637/jss.v093.i14>
- Ma, W., & de la Torre, J. (2020). GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software*, 93(14), 1–26. <https://doi.org/10.18637/jss.v093.i14>
- Ma, W., Iaconangelo, C., & de la Torre, J. (2016). Model similarity, model selection, and attribute classification. *Applied Psychological Measurement*, 40(3), 200–217. <https://doi.org/10.1177/0146621615621717>
- Mansell, W., & James, M. (2009). Assessment, learning and social justice. *Assessment in Education*, 16(1), 1–5.
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Mitra Cendekia.
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology*, 46(1), 4–11.
- McMillan, J. H. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction*. Pearson.
- Nasution, F. R., Yuanita, P., & Hutapea, N. M. (2025, September). Analisis Kesalahan Siswa SMA dalam Menyelesaikan Soal Fungsi Eksponensial Ditinjau dari Teori Kastolan. In *Seminar Nasional Pendidikan Sarjanawiyata Tamansiswa* (Vol. 2, No. 1, pp. 19-27).
- Nasution, R. A., Yuanita, P., & Hutapea, N. (2025). Students' difficulties in connecting multiple representations of exponential functions. *Journal of Mathematics Education Research*, 9(1), 55–68.
- Neill, J. (2011). *Designing and using surveys*. Sage.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60053-5](https://doi.org/10.1016/S0079-7421(08)60053-5)
- Nitko, A. J. (1996). *Educational assessment of students*. Merrill.
- Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Merrill.
- Nurzaenah, & Peni, N. R. N. (2025). Miskonsepsi dalam pembelajaran matematika: Tinjauan literatur pada materi eksponen. *Polinomial: Jurnal Pendidikan Matematika*, 4(4), 902–912. <https://doi.org/10.56916/jp.v4i4.2236>

- Nurzaenah, N., & Nofikusumawati, N. (2025). Persistent misconceptions in exponential concepts among senior high school students. *International Journal of Mathematics Education*, 18(1), 75–90.
- OECD. (2019). *OECD future of education and skills 2030*. OECD Publishing.
- OECD. (2023). *PISA 2022 results: What students know and can do (Volume I)*. OECD Publishing.
- Okoyere, C. Y., Kuranchie, A., Larbi, E., & Twene, F. (2018). Assessment practices and students' learning outcomes. *International Journal of Educational Research*, 90, 15–25.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press.
- Peşman, H., & Eryılmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *The Journal of Educational Research*, 103(3), 208–222. <https://doi.org/10.1080/00220670903383002>
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>
- Polit, D. F., & Beck, C. T. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459–467. <https://doi.org/10.1002/nur.20199>
- Popham, W. J. (2008). *Transformative assessment*. ASCD.
- Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, 29(4), 717–761.
- Rau, M. A., & Matthews, P. G. (2017). How to make 'more' better? Principles for effective use of multiple representations to enhance students' learning about fractions. *ZDM*, 49(4), 531–544.
- Reys, R., Lindquist, M., Lambdin, D., & Smith, N. (2007). *Helping children learn mathematics*. Wiley.
- Roberts, M. R., & Gierl, M. J. (2010). Developing score reports for cognitive diagnostic assessments. *Educational Measurement: Issues and Practice*, 29(3), 25–38.
- Roid, G. H., & Haladyna, T. M. (1982). *A technology for test-item writing*. Academic Press.
- Rosy, B., & Pahlevi, T. (2015). Pengembangan instrumen tes diagnostik pilihan ganda satu tingkat. *Jurnal Pendidikan*, 16(2), 101–112.
- Roussos, L. A., Templin, J., & Henson, R. A. (2007). Skills diagnosis using IRT-based latent class models. *Journal of Educational Measurement*, 44(4), 293–311.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.
- Schellekens, L. H., et al. (2021). Assessment as learning and student metacognition. *Educational Psychology Review*, 33(1), 89–112.
- Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4(1), 33–45. <https://doi.org/10.1007/s11409-008-9031-3>

- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351–371. <https://doi.org/10.1007/BF02212307>
- Schukajlow, S., Kaiser, G., & Stillman, G. (Eds.). (2025). *Mathematical Modelling from a Cognitive Perspective*. Routledge.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. *Practical Assessment, Research & Evaluation*, 22(4), 1–13.
- Sekercioglu, A., & Kose, S. (2015). Five-tier diagnostic tests in science education. *Journal of Science Education*, 39(2), 145–158.
- Serani, G., & Hairida. (2024). Teachers' perception and implementation of diagnostic assessment. *Journal of Educational Practice*, 11(1), 66–78.
- Slavin, R. E. (2009). *Educational psychology: Theory and practice*. Pearson.
- Stiggins, R. J. (2010). *Essential formative assessment competencies for teachers and school leaders*. Corwin.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. Routledge.
- Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, 30(2), 251–275. <https://doi.org/10.1007/s00357-013-9129-4>
- Templin, J., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11(3), 287–305. <https://doi.org/10.1037/1082-989X.11.3.287>
- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Indiana University.
- Thorndike, R. L., & Hagen, E. (1969). *Measurement and evaluation in psychology and education*. Wiley.
- Torrance, H. (2007). Assessment as learning? *Assessment in Education*, 14(3), 281–298.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159–169.
- Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2003). The role of diagnostic tests in identifying students' misconceptions. *International Journal of Science Education*, 25(3), 325–343.
- UNESCO. (2025). *Global education monitoring report: Learning recovery after COVID-19*. UNESCO Publishing.
- Uno, H. B. (2006). *Orientasi baru dalam psikologi pembelajaran*. Bumi Aksara.
- Utomo, A., Prasetyo, Z. K., & Widodo, A. (2024). Wordwall-based cognitive diagnostic assessment in junior secondary education. *Journal of Educational Innovation*, 8(2), 134–148.
- Vazquez-Alonso, A., Manassero-Mas, M. A., & Bennàssar-Roig, A. (2018). Multiple-tier diagnostic tests: Advantages and limitations. *Science Education International*, 29(3), 151–160.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61(2), 287–307. <https://doi.org/10.1348/000711007X193957>

- Vorhölter, K., Greefrath, G., Borromeo Ferri, R., Leiß, D., & Schukajlow, S. (2019). Mathematical modelling. In *Traditions in German-speaking mathematics education research* (pp. 91-114). Cham: Springer International Publishing.
- Wang, C., & Gierl, M. J. (2011). Using cognitive diagnostic models to analyze test data. *Educational Measurement*, 30(1), 3–14.
- Wilson, M. (2009). *Measuring progressions: Assessment structures underlying a learning progression*. *Journal of Research in Science Teaching*, 46(6), 716–730.
- Zhou, Y., Gierl, M. J., & Cui, Y. (2009). Using cognitive diagnosis models to analyze student performance. *Journal of Educational Measurement*, 46(4), 395–418.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. *Handbook of Self-Regulation*, 13–39.

