

## BIBLIOGRAPHY

- Alderson, J. C. (2005). *Assessing Reading*. Cambridge University Press.
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Arifin, M. (2017). Analisis Kualitas Soal Pilihan Ganda Ujian Akhir Semester. *Jurnal Pendidikan*, 5(1), 55–63.
- Arikunto, S. (2012). *Dasar-dasar Evaluasi Pendidikan* (Edisi revisi). Jakarta: Bumi Aksara.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Longman.
- Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. ASCD.
- Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. Pearson Education.
- Burton, R. (2018). Multiple-choice questions: Assessing higher-order thinking or promoting rote learning? *Journal of Educational Research*, 15(2), 78–92.

- Carr, N. T. (2011). *Designing and Analyzing Language Tests*. Oxford University Press.
- Coombe, C., Folse, K., & Hubley, N. (2007). *A Practical Guide to Assessing English Language Learners*. University of Michigan Press.
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). SAGE Publications.
- Douglas, D. (2000). *Assessing Language for Specific Purposes*. Cambridge University Press.
- Fulcher, G. (2010). *Practical Language Testing*. Routledge.
- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment*. Routledge.
- Griffin, P., Care, E., & Harding, S. (2014). Assessment and teaching of 21st-century skills. In Griffin & Care (Eds.), *Assessment and Teaching of 21st Century Skills*. Springer.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–334.
- Hopkins, D. (2008). *A Teacher's Guide to Classroom Research*. McGraw-Hill.
- Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge University Press.
- Kane, M. T. (2006). Validity. In *Educational Measurement* (4th ed.). American Council on Education.
- Kemendikbud. (2020). *Pedoman Penilaian dalam Kurikulum 2013*. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Kunnan, A. J. (2018). *Evaluating Language Assessments*. Routledge.

- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4–16.
- McMillan, J. H. (2013). *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction*. Pearson.
- McNamara, T. (2000). *Language Testing*. Oxford University Press.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). New York: Macmillan.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50(9), 741–749.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and Assessment in Teaching* (10th ed.). Pearson Education.
- Moleong, L. J. (2019). *Metodologi Penelitian Kualitatif* (Edisi Revisi). Bandung: PT Remaja Rosdakarya.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational Assessment of Students* (6th ed.). Boston: Pearson.
- Nitko, A. J., & Brookhart, S. M. (2014). *Educational Assessment of Students* (7th ed.). Boston: Pearson.
- O'Malley, J. M., & Pierce, L. V. (1996). *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. Addison-Wesley.
- Popham, W. J. (2009). *Classroom Assessment: What Teachers Need to Know* (6th ed.). Boston: Pearson Education.
- Ramdani, R. (2021). Analisis Soal Ujian Sekolah Berdasarkan Taksonomi Bloom Revisi. *Jurnal Evaluasi Pendidikan*, 12(2), 112–120.

- Reynolds, C. R., Livingston, R. B., & Willson, V. (2009). *Measurement and Assessment in Education*. Pearson.
- Rodriguez, M. C. (2016). The impact of multiple-choice question design on student performance: Validity and reliability concerns. *Assessment in Education*, 24(1), 55–72.
- Roever, C. (2021). The development of language assessments for young learners. *Language Testing*, 38(4), 554–575.
- Salkind, N. J. (2010). *Encyclopedia of Research Design*. SAGE Publications.
- Sari, D. P. (2020). Evaluasi Validitas Butir Soal Pilihan Ganda pada Penilaian Akhir Semester. *Jurnal Ilmiah Pendidikan*, 8(1), 88–95.
- Sawaki, Y. (2007). Structural equation modeling in language testing research. *Language Testing*, 24(1), 89–116.
- Shrock, S. A., & Coscarelli, W. C. (2007). *Criterion-Referenced Test Development: Technical and Legal Guidelines for Corporate Training*. Pfeiffer.
- Stobart, G. (2021). The role of assessment quality in shaping student learning: A case study approach. *Educational Measurement Journal*, 29(4), 215–231.
- Sudjana, N. (2005). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.
- Sugiyono. (2021). *Metode Penelitian Kualitatif, Kuantitatif, dan R&D*. Bandung: Alfabeta.
- Suryani, N. (2019). Analisis Distraktor dalam Soal Pilihan Ganda Ujian Sekolah. *Jurnal Pendidikan Dasar*, 10(2), 156–164.
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Palgrave Macmillan.

- Wiliam, D. (2011). *Embedded Formative Assessment*. Solution Tree Press.
- Yuliana, R. (2022). Validitas dan Reliabilitas Soal Pilihan Ganda Penilaian.
- Yulianti, R., & Suryani, N. (2019). An analysis of multiple-choice test quality in Indonesian elementary schools. *Journal of Educational Evaluation*, 6(2), 112–125.