

المرجع العربي

الجميل محمد عبد السميع شعلة، التقييم التربوي للمنظومة التعليمية، دار الفكر العربي، القاهرة.

١٤٢١هـ/٢٠٠٠م. ص: ٢٣

زين العارفين. (٢٠١٢). تقييم التعلّم: المبادئ، والتقنيات، والإجراءات. باندونغ: شركة رماجا

روسداكاريا

السياري، منيرة. ١٤٢٩هـ. التقييم. المملكة العربية السعودية: جامعة الملك سعود، كلية التربية.

شعلة، الجميل محمد عبد السميع. ١٤٢١هـ/٢٠٠٠م. التقييم التربوي للمنظومة التعليمية . القاهرة: دار الفكر العربي.

عبد الله أحمد عبد الله، التقييم التربوي للمتعلّمين لدى العلماء المسلمين (مكة المكرمة: مكتبة الرشد، ١٤٢٣هـ)، ص.

العمرى، عوض. ١٤٢٣هـ. العوامل المؤثرة على تقييم البرامج التدريبية. الرياض: جامعة الملك سعود (رسالة ماجستير غير منشورة).

غرونلوند، نورمان إدوارد، ولين، روبرت ل. (١٩٩٠). القياس والتقييم في التدريس، الطبعة السادسة. نيويورك: ماكميلان.

كاظم، علي مهدي. ٢٠٠١م. القياس والتقييم والتعلم في التعليم. الأردن: دار الكندي.

منيرة السياري، التقييم (المملكة العربية السعودية: جامعة الملك سعود، كلية التربية، ١٤٢٩هـ)، ص. ٢-٣.

- Adinda, T., et al. (2021). Penilaian formatif dalam pembelajaran: Konsep dan implementasi. *Jurnal Pendidikan dan Evaluasi*, 5(1), 1–12.
- Ahmad, & Auliya. (2024). Efektivitas tes uraian dalam penilaian formatif untuk mengukur pemahaman konseptual siswa. *Jurnal Evaluasi Pendidikan*, 8(2), 45–58.
- Anas Sudijono. (2011). *Pengantar evaluasi pendidikan*. Jakarta: PT Raja Grafindo Persada.
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's educational objectives*. New York: Longman.
- Apriana, et al. (2023). Evaluasi pembelajaran berbasis teknik penilaian modern. *Jurnal Ilmu Pendidikan*, 10(1), 22–34.
- Arifin, Z. (2021). *Evaluasi pembelajaran: Prinsip, teknik, prosedur (Edisi Revisi)*. Bandung: PT Remaja Rosdakarya.
- Arikunto, S. (2013). *Dasar-dasar evaluasi pendidikan (Edisi Kedua)*. Jakarta: Bumi Aksara.
- Asrul, et al. (2022). *Evaluasi pembelajaran*. Bandung: Citapustaka Media.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Cohen, L., Manion, L., & Morrison, K. (2022). *Research methods in education (8th ed.)*. London: Routledge.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.)*. Thousand Oaks, CA: Sage Publications.
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.)*. Thousand Oaks, CA: Sage Publications.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement (5th ed.)*. Englewood Cliffs, NJ: Prentice Hall.
- E. Mulyasa. (2006). *Kurikulum yang disempurnakan*. Bandung: PT Remaja Rosdakarya.
- Field, A. (2022). *Discovering statistics using IBM SPSS statistics (5th ed.)*. London: Sage Publications.

- Firdaus, M., Prihantoro, W., Latifaturrohman, L., Mahmudah, R. H., & Afrida, A. I. (2025). Efektivitas alat tes uraian dibandingkan tes pilihan ganda dalam mengukur hasil belajar siswa di MAN 2 Bantul. *Jurnal Penelitian Pendidikan*, 12(1), 33–47.
- Fitriani. (2025). Konsistensi penilaian tes uraian tanpa pedoman penskoran terstruktur. *Jurnal Evaluasi dan Penelitian Pendidikan*, 7(1), 10–21.
- Fitriani, & Herdah. (2022). Analisis praktik penilaian pembelajaran bahasa Arab berdasarkan pengetahuan, keterampilan, dan sikap. *Jurnal Pendidikan Bahasa Arab*, 9(2), 78–92.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2022). *How to design and evaluate research in education* (10th ed.). New York: McGraw-Hill.
- Gronlund, N. E., & Waugh, C. K. (2009). *Assessment of student achievement* (9th ed.). Upper Saddle River, NJ: Pearson Education.
- Gunawan, A. (2025). Analisis daya pembeda dan tingkat kesukaran soal bahasa Arab di MTs Negeri Kota Cimahi. *Jurnal Penelitian Pendidikan Bahasa Arab*, 6(1), 14–26.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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), 309–333.
- Hasnawati. (2020). Perbandingan penggunaan penilaian tes objektif dan tes esai terstruktur pada mata pelajaran pendidikan agama Islam. *Jurnal Ilmu Pendidikan Islam*, 4(2), 55–68.
- Ismara, & Tanumihardja. (2015). Penilaian diagnostik dalam proses pembelajaran. *Jurnal Pendidikan dan Evaluasi*, 3(2), 44–56.
- Kaka, et al. (2024). Pemanfaatan hasil penilaian formatif sebagai dasar perbaikan pembelajaran berkelanjutan. *Jurnal Kurikulum dan Pembelajaran*, 11(1), 67–80.
- Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt College Publishers.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.
- Kusaeri. (2021). *Acuan dan teknik penilaian proses dan hasil belajar dalam kurikulum 2013*. Yogyakarta: Ar-Ruzz Media.
- Laila, et al. (2022). Evaluasi pembelajaran: Pengumpulan informasi pada ranah kognitif, afektif, dan psikomotorik. *Jurnal Pendidikan*, 9(3), 101–115.

- Mukarromah, Zamakhsari, & Muhajir. (2025). Efektivitas tes uraian dalam mengukur pemahaman gramatikal dan makna dalam pembelajaran bahasa Arab. *Jurnal Pendidikan Bahasa Arab*, 12(1), 28–42.
- Mulyadi. (2010). *Evaluasi pendidikan: Pengembangan model evaluasi pendidikan agama Islam di sekolah*. Malang: UIN Maliki Press.
- Muslimah, & Widiyanti. (2025). Instrumen evaluasi berbasis indikator pembelajaran dalam tes pilihan ganda. *Jurnal Penelitian Pendidikan*, 11(2), 88–102.
- Nitko, A. J. (2004). *Educational assessment of students* (4th ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Boston: Pearson Education.
- Novan, & Kamalia. (2025). Kualitas tes pilihan ganda: Efektivitas distractor dan kesesuaian butir soal dengan indikator pembelajaran. *Jurnal Evaluasi Pendidikan*, 9(1), 19–31.
- Popham, W. J. (2008). *Classroom assessment: What teachers need to know* (5th ed.). Boston: Allyn & Bacon.
- Prihatni, Yuli, & Kumaidi. (2016). Pengembangan instrumen diagnostik mata pelajaran IPA SMP. *Jurnal Penelitian dan Evaluasi Pendidikan*, 20(1), 88–100.
- Purwanto, M. N. (2006). *Prinsip-prinsip dan teknik evaluasi pengajaran*. Bandung: PT Remaja Rosdakarya.
- Purwanto, N. (2021). *Prinsip-prinsip dan teknik evaluasi pengajaran* (Edisi Revisi). Bandung: PT Remaja Rosdakarya.
- Putri, & Zakir. (2023). Tes pilihan ganda dalam pembelajaran bahasa Arab: Karakteristik materi dan tujuan pembelajaran. *Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 10(2), 55–69.
- Razali, N. M., & Wah, Y. B. (2022). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Rusilowati, A., Universitas Negeri Semarang. (2015). Pengembangan tes diagnostik sebagai alat evaluasi kesulitan belajar IPA. *Jurnal Penelitian dan Evaluasi Pendidikan*, 19(1), 67–81.
- Schunk, D. H. (2012). *Learning theories: An educational perspective* (6th ed.). Boston: Pearson Education.
- Shabir. (2024). Karakteristik tes pilihan ganda: Distractor dan kualitas butir soal. *Jurnal Evaluasi dan Penilaian Pendidikan*, 6(2), 33–45.
- Stufflebeam, D. L. (2003). The CIPP model for evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 31–62). Dordrecht: Kluwer Academic Publishers.

- Sudjana, N. (2016). Penilaian hasil proses belajar mengajar (Cetakan ke-20). Bandung: PT Remaja Rosdakarya.
- Sudijono, A. (2011). Pengantar evaluasi pendidikan. Jakarta: PT Raja Grafindo Persada.
- Sugiyono. (2022). Metode penelitian kuantitatif, kualitatif, dan R&D (Edisi Kedua). Bandung: Alfabeta.
- Sukardi. (2010). Evaluasi pendidikan: Prinsip dan operasionalnya. Jakarta: Bumi Aksara.
- Thorndike, R. M. (1991). Measurement and evaluation in psychology and education (5th ed.). New York: Macmillan.
- Tyler, R. W. (1949). Basic principles of curriculum and instruction. Chicago: University of Chicago Press.
- Utami, R. (2023). Penilaian sumatif sebagai dasar pengambilan keputusan hasil belajar siswa. *Jurnal Kurikulum dan Pembelajaran*, 10(2), 77–89.
- Zainal Arifin. (2012). Evaluasi pembelajaran. Bandung: PT Remaja Rosdakarya.
- Zuhri, N. Z., Syihabuddin, S., & Tatang, T. (2025). Analisis validitas, reliabilitas, dan tingkat kesukaran soal bahasa Arab berbasis kecerdasan buatan untuk jenjang menengah pertama. *Jurnal Pendidikan Bahasa Arab*, 11(1), 40–55.

