

REFERENCES

- Agu, N. N., Onyekuba, C., & Anyichie, A. C. (2013). Measuring Teacher's Competencies in Constructing Classroom-Based Tests in Nigerian Secondary Schools: Need for a Test Construction Skill Inventory. *Academic Journals: Educational Research and Reviews* , 8 (8), 431-439.
- Ahsan, S. (2009). Classroom Assessment Culture in Secondary Schools of Dhaka City. *Teacher's World (Journal of Education and Research)* , 33, 231-244.
- Aiken, L. R. (1994). *Psychological Testing and Assessment (8th ed.)*. Boston: Allyn and Bacon.
- Airasian, P. W. (1994). *Classroom Assessment*. New York: McGraw-Hill.
- Alkharusi, H. (2011). Teachers' Classroom Assessment Skills: Influence of Gender, Subject Area, Grade Level, Teaching Experience, and In-Service Training. *Journal of Turkish Science Education* , 8 (2), 39-48.
- Anastasi, A., & Urbina, S. (1997). *Psychological Testing (7th ed.)*. New Jersey: Prentice-Hall Inc.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning Teaching and Assessing*. New York: Addison Wesley Longman.
- Angelo, T. A. (1995). Reassessing (and Defining) Assessment. *AAHE Bulletin* , 48, 7-9.
- Angrosino, M. V., & Mays de Perez, K. A. (2000). Rethinking Observation: From Method to Context. In N. K. Denzin, & Y. S. Lincoln, *Handbook of Qualitative Research (2nd ed.)* (pp. 673-702). Thousand Oaks, CA: SAGE.
- Attali, Y., & Bar-Hillel, M. (2003). Guess Where: The Position of Correct Answers in Multiple-Choice Test Items as a Psychometric Variable. *Journal of Educational Measurement* , 40 (2), 109-128.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: OUP Oxford.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. New York: Oxford University Press.

- Badgett, J. L., & Christmann, E. P. (2009). *Designing Elementary Instruction and Assessment Using the Cognitive Domain*. California: Corwin SAGE Company.
- Barnett, J. E., & Francis, A. L. (2012). Using Higher Order Thinking Questions to Foster Critical Thinking: a Classroom Study. *Educational Psychology: An International Journal of Experimental Educational Psychology* , Retrieved from <http://www.tandfonline.com/loi/cedp20>.
- Berkowitz, D., Wolkowitz, B., Fitch, R., & Kopriva, R. (2000). *The Use of Tests as Part of High-Stakes Decision-Making for Students: A Resources Guide for Educators and Policy-Makers*. Washington DC: U.S Department of Education.
- Biggs, J. (1999). *Teaching for Quality Learning at University*. Buckingham: SRHE/Open University.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: Cognitive Domain*. New York: David McKay Co.
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education. *International Journal of Social Science & Interdisciplinary Research* , 2 (2), 189-193.
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal* , 9 (2), 27-40.
- Brewster, J., Gail, E., & Denis, G. (2003). *The Primary English Teacher's Guide*. Edinburgh: Pearson Education Limited.
- Brookhart, S. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. Alexandria, VA: ASCD .
- Brown, H. D. (2004). *Language Assessment Principles and Classroom Practices*. New York: Pearson Education.
- Brown, H. D., & Abeywickrama, P. (2010). *Language Assessment: Principles and Classroom Practices (2nd ed.)*. New York: Pearson Education.
- Brown, J. D. (1996). *Testing in Language Program*. New Jersey: Prentice Hall Regents.

- Burton, S. J., Sudweeks, R. R., Merrill, P. F., & Wood, B. (1991). *How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty*. Retrieved April 04, 2016, from Brigham Young University Testing Centre: <http://testing.byu.edu/info/handbooks/betteritems.pdf>
- Chang, C. C. (2005). Developing Tailored Instruments: Item Banking and Computerized Adaptive Assessment. *Evaluation and Program Planning Journal*, 24, 215-251.
- Cheng, L. (2008). The Key to Success: English Language Testing in China. *Language Testing*, 25 (1), 15-37.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education*. New York: Routledge.
- Corbin, J., & Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (3rd ed.)*. Thousand Oaks, CA: SAGE Publications.
- Cresswell, J. C. (2012). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research 4th edition*. Boston: Pearson.
- Cresswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. New York: SAGE Publications.
- Cunningham, G. K. (1998). *Assessment in the Classroom: Constructing and Interpreting Tests*. London: The Falmer Press.
- Departemen Pendidikan Nasional. (2007). *Peraturan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 tentang Standar Penilaian Pendidikan*. Jakarta: Depdiknas
- Departemen Pendidikan Nasional. (2008). *Peraturan Pemerintah Nomor 74 Tahun 2008 tentang Guru*. Jakarta: Depdiknas
- DiBattista, D., & Kurzawa, L. (2011). Examination of the Quality of Multiple-Choice Items on Classroom Tests. *The Canadian Journal for the Scholarship of Teaching and Learning*, 2 (2), Retrieved online from <http://dx.doi.org/10.5206/cjsotl-rcacea.2011.2.4>.

- Dirjen Peningkatan Mutu Pendidikan dan Tenaga Kependidikan (PMPTK). (2009). *Analisis Data Guru*. Jakarta: Dirjen PMPTK Depdiknas
- Ebel, R. R., & Frisbie, D. A. (1991). *Essentials of Educational Measurement*. New Jersey: Prentice-Hall Inc.
- Eisner, E. W. (1991). *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice*. New York: Macmillan Publishing Company.
- Fan, J., & Jin, Y. (2013). A Survey of English Language Testing Practice in China: the Case of Six Examination Board. *Language Testing in Asia* , 3 (7), Retrieved from <http://www.language-testing-asia.com/content/3/1/7>.
- Fellenz, M. R. (2004). Using Assessment to Support Higher Level Learning: the Multiple-Choice Item Development Assignment. *Assessment & Evaluation in Higher Education* , 29 (6), 703-719.
- Fleming, M., & Chambers, B. (1983). Teacher-Made Tests: Windows in the Classroom. In W. E. Hathaway, *Testing in the Schools: New Directions for Testing and Measurement* (pp. 29-38). San Fransisco: Jossey-Bass.
- Fulcher, G. (2010). *Practical Language Testing*. London: Hodder Education.
- Gallagher, D. J. (1998). *Classroom Assessment for Teachers*. Upper Saddle River, NJ: Merrill.
- Gay, L. R., Mills, G. E., & Airasian, P. W. (2009). *Educational Research: Competencies for Analysis and Applications (9th edition)*. Upper Saddle River, New Jersey: Prentice Hall.
- Genesee, F., & Upshur, J. (1996). *Classroom-Based Evaluation in Second Language Education*. Cambridge: Cambridge University Press.
- Gronlund, N. E. (1998). *Assessment of Student Achievement (6th ed.)*. Boston: Allyn & Bacon.
- Gronlund, N. E. (1977). *Constructing Achievement Tests*. New Jersey: Prentice Hall.
- Guskey, T. R. (2003). How Classroom Assessments Can Improve Learning. *Educational Leadership* , 60 (5), 7-11.

- Gutierrez, S. L. (2014). *From National Standards to Classrooms: A Case Study of Middle Level Teachers' Assessment Knowledge and Practice*. Dissertation. Western Michigan University.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item Writing Guideliness for Classroom Assessment. *Applied Measurement in Education* , 15 (3), 309-334.
- Hamafyelto, R. S., Hamman-Tukur, A., & Hamafyelto, S. S. (2015). Assessing Teacher Competence in Test Construction and Content Validity of Teacher Made Examination Questions in Commerce in Borno State, Nigeria. *Education* , 5 (5), 123-128.
- Herman, J., & Golan, S. (1991). The Effects of Standardized Testing on Teaching and Schools. *Educational Measurement: Issues and Practices* , 12, 20-25.
- Hoshino, Y. (2013). Relationship between Types of Distractor and Difficulty of Multiple-Choice Vocabulary Tests in Sentential Context. *Language Testing in Asia* , 3 (16), 1-14 Retrieved online from <http://www.languagetestingasia.com/content/3/1/16>.
- Hughes, A. (2003). *Testing for Language Teachers*. Singapore: Cambridge University Press.
- Jacobs, & Chase. (1992). *Developing and Using Test Effectively : a Guide for Faculty*. San Fransisco: Jossey-Bass Inc.
- Jones, A. T. (2011). Comparing Methods for Item Analysis: the Impact of Different Item-Selection Statistics on Test Difficulty. *Applied Psychological Measurement* , 35 (7), 566-571.
- Kehoe, J. (1995). Basic Item Analysis for Multiple-Choice Tests. *Practical Assessment, Research, and Evaluation* , 4 (10).
- Kiuhara, S. A., Graham, S., & Hawken, L. S. (2009). Teaching Writing to High-School Students: A National Survey. *Journal of Educational Psychology* , 101 (1), 136-160.
- Lambert, C. E., & Lambert, V. A. (2012). Qualitative Descriptive Research: An Acceptable Design. *Pacific Rim International Journal of Nursing Research* , 16 (4), 255.

- Lambert, D., & Lines, D. (2000). *Understanding Assessment: Purposes, Perceptions, Practices*. London: Routledge Falmer.
- Linse, C. (2005). *Practical English Language Teaching: Young Learners*. New York: McGraw-Hill ESL/ELT.
- Luxia, Q. (2004). Has a High Stakes Test Produced the Intended Change? In L. Cheng, Y. Watanabe, & A. Curtis, *Washback in Language Testing: Research Contexts and Methods* (pp. 171-190). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mackey, A., & Gass, S. M. (2005). *Second Language Research: Methodology and Design*. Lawrence Erlbaum Associates, Inc.
- Malik, R. S., & Hamied, F. A. (2014). *Research Methods: A Guide for First Time Researchers*. Bandung: UPI Press.
- Marshall, C., & Rossman, G. B. (1999). *Designing Qualitative Research*. London: Sage Publications.
- Mazidah, M., & Abd Aziz, M. S. (2013). Juxtaposing the School Teacher's Concept of Testing with Their Test Construction Practices. *Proceeding of the 13th SoLLs.INTEC*. Bangi Selangor, Malaysia: University Kebanggaan Malaysia.
- McMillan, J. H. (2000). Fundamental Assessment Principles for Teachers and School Administrators. *Practical Assessment, Research, and Evaluation* , 7 (8).
- McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.
- McNeil, L. (2010). Beyond the Product of Higher-Order Questioning: How do Teacher and English-Language Learner Perceptions Influence Practice. *TESOL Journal* , 2, 74-90.
- Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology (4th ed.)*. New York: Longman.
- Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education*. San Fransisco, CA: Jossey-Bass.
- Messick, S. (1996). Validity and Washback in Language Testing. *Language Testing* , 13 (3), 241-256.

- Miles, M. B., & Huberman, A. M. (1994). *An Expanded Sourcebook: Qualitative Data Analysis*. California: SAGE Publications.
- Miller, D. M., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and Assessment in Teaching 10th Ed.* New Jersey: Pearson Education.
- Moon, J. (2000). *Children Learning English: A Guidebook for English Language Teachers*. London: Macmillan Heinemann.
- Nitko, A. J. (1996). *Educational Assessment of Students (2nd ed.)*. Ohio: Merrill an imprint of Prentice-Hall Englewood Cliffs.
- Nunan, D. (1992). *Research Methods in Language Learning*. Cambridge: Cambridge University Press.
- O'Malley, J. M., & Pierce, L. V. (1996). *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. New York: Addison-Wesley.
- Oosterhof, A. (2003). *Developing and Using Classroom Assessments 3rd Ed.* New Jersey: Pearson Education.
- Osterlind, S. J. (2002). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats*. New York: Kluwer Academic Publishers.
- Osterlind, S. J. (1990). Toward a Uniform Definition of a Test Item. *Educational Research Quarterly* , 14 (4), 2-5.
- Pancoro, N. H. (2011). The Item Characteristics of the Final Semester Test as A Preparation for English Item Bank. *Jurnal Penelitian dan Evaluasi Pendidikan* , 15 (1), 93-114.
- Paran, A., & Sercu, L. (2010). *Testing the Untestable in Language Education*. Canada: Short Run Press Ltd.
- Patton, M. Q. (1980). *Qualitative Evaluation Methods*. California: SAGE Publications.
- Pinter, A. (2006). *Teaching Young Language Learners*. Oxford: Oxford University Press.
- Popham, J. W. (1995). *Classroom Assessment: What Teachers Need to Know*. Boston: Allyn and Bacon.

- Popham, W. J. (2003). *Test Better, Teach Better: The Instructional Role of Assessment*. Virginia: ASCD Publication.
- Purnomo, A. (2007). Kemampuan Guru dalam Merancang Tes Berbentuk Pilihan Ganda Pada Mata Pelajaran IPS untuk Ujian Akhir Sekolah (UAS). *Lembaran Ilmu Kependidikan* , 36 (1).
- Pusat Penilaian Pendidikan, Balitbang Depdiknas. (2012). *Evaluasi Implementasi Model-Model Penilaian (Program LEA 2012)*. Jakarta: Puspendik
- Renaud, R. D., & Murray, H. G. (2007). The Validity of Higher-Order Questions as a Process Indicator of Educational Quality. *Research in Higher Education* , 48 (3), 319-351.
- Rudner, L., & Schafer, W. (2002). *What Teachers Need to Know About Assessment*. Washington DC: National Education Association.
- Scott, D., & Usher, R. (2011). *Researching Education: Data Methods and Theory in Educational Enquiry*. London: Continuum.
- Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher* , 29 (7), 4-14.
- Shermis, M. D., & Di Viesta, F. J. (2011). *Classroom Assessment in Action*. Lanham: Rowman& Littlefield Publishers.
- Shield, P., & Rangarjan, N. (2013). *A Playbook for Research Methods: Integrating Conceptual Frameworks and Project Management*. Stillwater, OK: New Forum Press.
- Shohamy, E. (1996). Test Impact Revisited: Washback Effect Over Time. *Language Testing* , 13, 298-317.
- Shulman, S. L. (1987). Knowledge and Teaching: Foundation of the New Reform. *Harvard Educational Review* , 57 (1), 1-21.
- Siri, A., & Freddano, M. (2011). The Use of Item Analysis for the Improvement of Objective Examinations. *Procedia - Social and Behavioural Sciences* , 29, 188-197.
- Smith, V. G., & Szymanski, A. (2013). Critical Thinking: More Than Test Scores. *NCPEA International Journal of Educational Leadership Preparation* , 8 (2), 16-26.

- Soureshjani, K.H. (2011). Item Sequence on Test Performance: Easy Items First? *Language Testing in Asia*, 1(3,) 46-59
- Stiggins, R. J. (1991). Relevant Classroom Assessment Training for Teachers. *Educational Measurement: Issues and Practice* , 10 (1), 7-12.
- Stiggins, R. J., & Bridgeford, N. J. (1985). The Ecology of Classroom Assessment. *Journal of Educational Measurement* , 22 (4), 271-286.
- Taylor, L. (2005). Washback and Impact. *ELT Journal* , 59 (2), 154-155.
- Thanyapa, I., & Currie, M. (2014). The Number of Options in Multiple Choice Items in Language Tests: Does It Make Any Difference? Evidence from Thailand. *Language Testing in Asia* , 4 (8), Retrieved from <http://www.languagetestingasia.com/content/4/1/8>.
- Tiemeier, A. M., Stacy, Z. A., & Burke, J. M. (2011). Using Multiple Choice Questions Written at Various Bloom's Taxonomy Levels to Evaluate Student Performance across a Therapeutics Sequence. *Inov Pharm* , 2 (2), Retrieved from <http://pubs.lib.umn.edu/innovations/vol2/>.
- Waugh, C. K., & Gronlund, N. E. (2012). *Assessment of Student Achievement 10th edition*. United States of America: Pearson Education.
- Webber, C. F., & Lupart, J. L. (2011). *Leading Student Assessment*. London: Springer.
- Weller, L. D. (2001). Building Validity and Reliability into Classroom Tests. *National Association of Secondary School Principals* , 85, 32-37. Retrieved from ProQuest Education Journals. (Document ID Number: 216049668).
- William, D. (2000). Education: The Meanings and Consequences of Educational Assessments. *Critical Quarterly* , 42 (1), 105-127.
- Wragg, E. C. (2001). *Assessment and Learning in the Secondary School*. London: Routledge-Falmer.
- Xu, Y., & Liu, Y. (2009). Teacher Assessment Knowledge and Practice: a Narrative Inquiry of a Chinese College EFL Teachers' Experience. *TESOL Quarterly* , 43 (3), 493-513.

- Zamsir. (2012). Kualitas Tes Buatan Guru Pada Mata Pelajaran Matematika di SD Negeri Kota Kendari. *VALUE, Jurnal Evaluasi & Asesmen Pendidikan* , 1 (1), 49-65.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom Assessment Practices and Teachers' Self-Perceived Assessment Skills. *Applied Measurement in Education* , 16 (4), 323-342.