**MULTIPLE-CHOICE TEST ITEMS: SELECTED REFERENCES**

**(Last updated 20 July 2014)**

Albanese, M. A., Kent, T. H., & Whitney, D. R. (1979). Cluing in multiple-choice test items with combinations of correct responses. *Academic Medicine*, *54*(12), 948-50.

Amini, M., & Ibrahim-González, N. (2012). The washback effect of cloze and multiple choice test items on vocabulary acquisition. *Languages in India*, *12*(7), 71-91.

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.

Becker, W. E., & Johnston, C. (1999). The relationship between multiple choice and essay response questions in assessing economics understanding. *Economic Record*, *75*(4), 348-357.

Bormuth, J. R. (1967). Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading*, *10*(5), 291-299.

Brame, C. J. (2014). Writing good multiple choice test questions. Nashville, TN: Vanderbilt University. Retrieved from http://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/

Bridgeman, B. (1992). A comparison of quantitative questions in open‐ended and multiple‐choice formats. *Journal of Educational Measurement*, *29*(3), 253-271.

Bridgeman, B., & Lewis, C. (1994). The relationship of essay and multiple‐choice scores with grades in college courses. *Journal of Educational Measurement*, *31*(1), 37-50.

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment.* New York, NY: McGraw Hill.

Bruno, J. E., & Dirkzwager, A. (1995). Determining the optimal number of alternatives to a multiple-choice test item: An information theoretic perspective. *Educational and Psychological Measurement*, *55*(6), 959-966.

Buck, G., Tatsuoka, K., & Kostin, I. (1997). The subskills of reading: Rule‐space analysis of a multiple‐choice test of second language reading comprehension. *Language Learning*, *47*(3), 423-466.

Burton, R. F. (2005). Multiple‐choice and true/false tests: myths and misapprehensions. *Assessment & Evaluation in Higher Education*, *30*(1), 65-72.

Burton, S. J., Sudweeks, R. R., Merrill, P. F., & Wood, B. (1991). *How to prepare better multiple-choice test items: Guidelines for university faculty*. Provo, UT: Brigham Young University Testing Services.

Bush, M. (2001). A multiple choice test that rewards partial knowledge. *Journal of Further and Higher education*, *25*(2), 157-163.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*(4), 273.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604-616.

Celce-Murcia, M., Kooshian, G. B., & Gosak, A. J. (1974). Goal: Good multiple-choice language test items. *English Language Teaching 28*(3), 257-262.

Cizek, G. J., & O'Day, D. M. (1994). Further investigation of nonfunctioning options in multiple-choice test items. *Educational and Psychological Measurement*, *54*(4), 861-872.

Crocker, L., & Schmitt, A. (1987). Improving multiple-choice test performance for examinees with different levels of test anxiety. *The Journal of Experimental Education*, *55*(4), 201-205.

Cross, L. H., & Frary, R. B. (1977). An empirical test of Lord's theoretical results regarding formula scoring of multiple‐choice tests. *Journal of Educational Measurement*, *14*(4), 313-321.

Davis, F. B. (1959). Estimation and use of scoring weights for each choice in multiple-choice test items. *Educational and Psychological Measurement*, *19*(3), 291-298.

Delgado, A. R., & Prieto, G. (2003). The effect of item feedback on multiple‐choice test responses. *British Journal of Psychology*, *94*(1), 73-85.

Dolly, J. P., & Williams, K. S. (1986). Using test-taking strategies to maximize multiple-choice test scores. *Educational and Psychological Measurement*, *46*(3), 619-625.

Dressel, P. L., & Schmid, J. (1953). Some modifications of the multiple-choice item. *Educational and Psychological Measurement*, *13*(4), 574-595.

Ellsworth, R. A., Dunnell, P., & Duell, O. K. (1990). Multiple-choice test items: What are textbook authors telling teachers?. *The Journal of Educational Research*, *83*(5), 289-293.

Farley, J. K. (1989). The multiple-choice test: Writing the questions. *Nurse Educator*, *14*(6), 10-12.

Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple‐choice reading comprehension test. *Journal of Educational Measurement*, *27*(3), 209-226.

Frary, R. B. (1980). The effect of misinformation, partial information, and guessing on expected multiple-choice test item scores. *Applied Psychological Measurement*, *4*(1), 79-90.

Frary, R. B. (1995). More multiple-choice item writing do's and don'ts. *Practical Assessment, Research & Evaluation*, *4*(11). Retrieved from [http://pareonline.net/getvn.asp?v=4&n=11](https://mail.middlebury.edu/owa/redir.aspx?C=b2q_cSbaX02BPFwkxvdWSXom62UbztAIpG-kUADfgLX623qjim45pT6kahAkCi23obNixX-WpqM.&URL=http%3a%2f%2fpareonline.net%2fgetvn.asp%3fv%3d4%26n%3d11)

Frederick, R. I., & Foster, H. G. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3*(4), 596.

Freedle, R., & Kostin, I. (1999). Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing*, *16*(1), 2-32.

Fuhrman, M. (1996). Developing Good Multiple-Choice Tests and Test Questions. *Journal of Geoscience Education*, *44*(4), 379-84.

Haladyna, T. M., & Downing, S. M. (1989). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, *2*(1), 51-78.

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.

Haladyna, T. M., & Shindoll, R. R. (1989). Item shells a method for writing effective multiple-choice test items. *Evaluation & the Health Professions*, *12*(1), 97-106.

Hambleton, R. K., Roberts, D. M., & Traub, R. E. (1970). A comparison of the reliability and validity of two methods for assessing partial knowledge on a multiple‐choice test. *Journal of Educational Measurement*, *7*(2), 75-82.

Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *The Journal of Experimental Education*, *62*(2), 143-157.

Hansen, J. D., & Dexter, L. (1997). Quality multiple-choice test questions: Item-writing guidelines and an analysis of auditing testbanks. *Journal of Education for Business*, *73*(2), 94-97.

Heim, A. W., & Watts, K. P. (1967). An experiment on multiple-choice versus open-edned answering in a vocabulary test. *British Journal of Educational Psychology*, *37*(3), 339-346.

Horst, P. (1933). The difficulty of a multiple choice test item. *Journal of Educational Psychology*, *24*(3), 229.

Kehoe, J. (1995). Writing multiple-choice test items. *Practical Assessment, Research & Evaluation, 4*(9). Retrieved from http://PAREonline.net/getvn.asp?v=4&n=9 .

Kruglov, L. P. (1953). Qualitative differences in the vocabulary choices of children as revealed in a multiple-choice test. *Journal of Educational Psychology*, *44*(4), 229.

Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology*, *63*(5), 505.

Lehrl, S., Triebig, G., & Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurologica Scandinavica*, *91*(5), 335-345.

Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science*, *23*(11), 1337-1344.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple‐choice, constructed response, and examinee‐selected items on two achievement tests. *Journal of Educational Measurement*, *31*(3), 234-250.

Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, *14*(2), 194-199.

Mason, V. (1984). Using multiple-choice tests to promote homogeneity of class ability levels in large EGP and ESP programs. *System, 12*, 263-271.

Mason, V. (1992). A good word for multiple-choice tests. *CATESOL Journal, 5*(2), 29-44.

Masters, J. C., Hulsmeyer, B. S., Pike, M. E., Leichty, K., Miller, M. T., & Verst, A. L. (2001). Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. *The Journal of Nursing Education*, *40*(1), 25-32.

McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: a literature review. *Medical Teacher*, *26*(8), 709-712.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*(2), 142-154.

Mehrens, W.A. & Lehman, I.J. (1978). *Measurement and evaluation in education and psychology* (2nd edition). Holt, Rinehart and Winston.

Morrison, S., & Free, K. W. (2001). Writing multiple-choice test items that promote and measure critical thinking. *Journal of Nursing Education*, *40*(1), 17-24.

Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, *6*(2), 199-215.

Nicol, D. (2007). E‐assessment by design: using multiple‐choice tests to good effect. *Journal of Further and Higher Education*, *31*(1), 53-64.

Norris, S. P. (2009). Informal reasoning assessment: Using verbal reports of thinking to improve multiple-choice test validity. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 451-471). New York NY: Routledge.

Oller, J. W., Jr. (1979). *Language tests at school.* London: Longman.

Paxton, M. (2000). A linguistic perspective on multiple-choice questioning. *Assessment & Evaluation in Higher Education, 25*(2), 109-119.

Pyrczak, F. (1972). Objective evaluation of the quality of multiple-choice test items designed to measure comprehension of reading passages. *Reading Research Quarterly, 8*(1), 62-71.

Rankin, E. F., & Culhane, J. W. (1969). Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading*, *13*(3), 193-198.

Rodriguez, M. C. (2005). Three options are optimal for multiple‐choice items: A meta‐analysis of 80 years of research. *Educational Measurement: Issues and Practice*, *24*(2), 3-13.

Roediger III, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1155.

Roid, G.H., & Haladyna, T.M. (1980). The emergence of an item writing technology. *Review of Educational Research, 49*, 252-279.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin*, *106*(2), 332.

Schultheis, N. M. (1998). Writing cognitive educational objectives and multiple-choice test questions. *American Journal of Health-system Pharmacy*, *55*(22), 2397-2401.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*(4), 453-472.

Smith, J.K. (1982). Converging on correct answers: A peculiarity of multiple-choice items. *Journal of Educational Measurement, 19*, 211-220.

Tamir, P. (1971). An alternative approach to the construction of multiple choice test items. *Journal of Biological Education*, *5*(6), 305-307.

Tarrant, M., Knierim, A., Hayes, S. K., & Ware, J. (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Education in Practice*, *6*(6), 354-363.

Tarrant, M., & Ware, J. (2008). Impact of item‐writing flaws in multiple‐choice questions on student achievement in high‐stakes nursing assessments. *Medical Education*, *42*(2), 198-206.

Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC medical education*, *9*(1), 40.

Thissen, D., & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika*, *49*(4), 501-519.

Tinkelman, S. N. (1968). Checklist for reviewing local school tests. In N. E. Gronlund (Ed.), *Readings in measurement and evaluation* (pp. 103-108). New York, NY:L McMillan.

Treagust, D. (1986). Evaluating students' misconceptions by means of diagnostic multiple choice items. *Research in Science Education*, *16*(1), 199-207.

Votaw, D. F. (1936). The effect of do-not-guess directions upon the validity of true-false or multiple choice tests. *Journal of Educational Psychology*, *27*(9), 698.

Wainer, H., & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education*, *6*(2), 103-118.

Wesman, A.G. (1971). Writing the test item. In R.L. Thorndike (Ed.) *Educational measurement* (1st ed., pp. 99-111). Washington, DC: American Council on Education.

Wilhite, S. C. (1986). The relationship of headings, questions, and locus of control to multiple-choice test performance. *Journal of Literacy Research*, *18*(1), 23-40.

Willey, C. F. (1960). The three-decision multiple-choice test: A method of increasing the sensitivity of the multiple-choice item. *Psychological Reports*, *7*(3), 475-477.

Yi'an, W. (1998). What do tests of listening comprehension test?-A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing*, *15*(1), 21-44.

Zimmerman, D. W., & Williams, R. H. (1965). Chance success due to guessing and non-independence of true scores and error scores in multiple-choice tests: Computer trials with prepared distributions. *Psychological Reports*, *17*(1), 159-165.