

參考文獻

- 行政院教育改革委員會(1996)。教育改革總諮議報告書(第三章綜合建議)。2004年12月5日,取自<http://www.edu.tw/eduinfor/change/5/CH-3.html#c3>
- 考選部(2004)。考選部全球資訊網。2004年5月31日,取自<http://inter1.moex.gov.tw/statute/statute1.asp?kind=31>。
- 余民寧(1992)。試題反應理論的介紹(7):訊息函數。研習資訊,9(6),5-9。
- 余民寧(2002)。教育測驗與評量:成就測驗與教學評量(第二版)。臺北市:心理出版社。
- 余民寧、汪慧瑜(2005)。量尺分數的另類表示方法:以國中基本學力測驗為例。教育與心理研究,28期,審稿中。
- 林惠芬(1993)。通過分數設定方法在護理人員檢覈筆試測驗之研究。測驗年刊,40,253-262。
- 吳裕益(1986)。標準參照測驗通過分數設定方法之研究。國立政治大學教育研究所博士論文(未出版)。
- 吳裕益(1988)。標準參照測驗通過分數設定方法之研究,測驗年刊,35,159-166。
- 涂柏原,陳柏熹,章舜雯,林世華(2000)。基本學力分數的建立。國中基本學力測驗推動工作委員會,2004年12月6日,取自<http://www.bctest.ntnu.edu.tw/score1.htm>。
- 莊淑如(1997)。證照制度的落實—以德國經驗為借鏡。技職雙月刊,37,54-56。
- 教育部(1998)。國民中學學生基本學力指標。台北:教育部。
- 教育部國教司(2004)。國民中小學九年一貫課程暫行綱要,2004年12月5日,取自<http://140.122.120.230/9CC/temporary/temporary-all.htm>。
- 國民中學學生基本學力測驗推動工作委員會(2002a)。九十、九十一年國中基本學力測驗試題取材範圍比較。飛揚第十三期,2004年12月5日,取自<http://www.bctest.ntnu.edu.tw/>。
- 國民中學學生基本學力測驗推動工作委員會(2002b)。國中基本學力測驗自然科試題之設計理念。飛揚第十三期,2005年4月6日,取自<http://www.bctest.ntnu.edu.tw/>。
- 國民中學學生基本學力測驗推動工作委員會(2002c)。九十一年第二次國民中學學生基本學力測驗試題特色。飛揚第十六期,2005年4月6日,取自<http://www.bctest.ntnu.edu.tw/>。
- 國民中學學生基本學力測驗推動工作委員會(2003)。測驗分數的解釋(下)。飛揚第二十三期,2005年4月6日,取自<http://www.bctest.ntnu.edu.tw/>。
- 鄭明長、余民寧(1994)。各種通過分數設定方法之比較。測驗年刊,41,19-40。
- 鄭清泉(2001)。人工化與電腦化適性精熟能力判定在國小學童數學精熟分類一致性之比較研究。國立嘉義大學國民教育研究所碩士論文。
- 謝進昌、余民寧(2005)。國中基本學力測驗之DIF的實徵分析:以91年度二次

測驗為例，國立新竹師範學院學報，審稿中。

- Andrew, B. J., & Hecht, J. T. (1976). A preliminary investigation of two procedures for setting examination standards. *Educational and Psychological Measurement*, 36, 35-50.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (pp.508-600). Washington, D.C.: American Council on Education.
- Angoff, W. H. (1988). Validity: An evolving concept. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp.19-32). Hillsdale, NJ: Lawrence Erlbaum.
- Beaton, A. E., & Allen, N. L. (1992). Interpretation scales through scale anchoring, *Journal of Educational Statistics*, 17, 191-201.
- Behuniak, P., Archambault, F. X., & Gable, R. K. (1982). Angoff and Nedelsky standard setting procedures: Implication for the validity of proficiency test score interpretation, *Educational and Psychological Measurement*, 42, 247-255.
- Berk, R.A. (1976). Determination of optimal cutting scores in criterion-referenced measurement. *Journal of Experimental Education*, 45, 4-9.
- Berk, R. A. (1980). A consumers' guide to criterion-referenced test reliability. *Journal of Educational Measurement*, 17(4), 323-349.
- Berk, R. A. (1984). *A guide to criterion-referenced test construction*. Baltimore, MD: The Johns Hopkins University Press.
- Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Measurement*, 56(1), 137-172.
- Berk, R. A. (1996). Standard setting: The next generation (where few psychometricians have gone before!). *Applied Measurement in Education*, 9(3), 215-235.
- Bernknopf, S., Curry, A., & Bashaw, W. L.(1979). *A defensible model for determining a minimal cutoff score for criterion referenced tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick, *Statistical theories of mental test scores* (chapters 17-20). Reading, MA: Addison-Wesley.
- Bontempo, B. D., Marks, C. M., & Karabatsos, G. (1998). *A meta-analytic assessment of empirical differences in standard setting procedures*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.
- Brennan, R. L. & Lockwood, R. E. (1980). A comparison of the Nedelsky and Angoff

- cutting score procedures using generalizability theory, *Applied Psychological Measurement*, 4, 219-240.
- Brandon, P. R. (2002). Two versions of the contrasting-groups standard-setting method: A review. *Measurement and Evaluation in Counseling and Development*, 35(3), 167-181.
- Brandon, P. R. (2004). Conclusion about frequently studied modified Angoff standard setting topics. *Applied Measurement in Education*, 17(1), 59-88.
- Busch, J. C., & Jaeger, R. M. (1990). Influence of type of judge, normative information, and discussion on standards recommended for the national teacher examinations. *Journal of Educational Measurement*, 27(2), 145-163.
- Buckendahl, C. W., Smith, R. W., Impara, J. C., & Plake, B. S. (2002). A comparison of Angoff and Bookmark standard setting method. *Journal of Educational Measurement*, 39(3), 253-263.
- Cascio, W. F., Alexander, R.A., & Barrett, G. V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. *Personnel Psychology*, 41, 1-24.
- Chang, L., Dziuban, C. D., & Hynes, M. C.(1996). Does a standard reflect minimal competency of examinees or judge competency? *Applied Measurement in Education*, 9(2), 161-173.
- Chang, L. (1999). Judgmental item analysis of the Nedelsky and Angoff standard-setting methods. *Applied Measurement in Education*, 12(2), 151-165.
- Chang, L., van der Linder, W. J., & Vos, H. J.(2004). Setting standards and detecting intrajudge inconsistency using interdependent evaluation of response alternatives. *Educational and Psychological Measurement*, 64(5), 781-801.
- Chinn, R. N., & Hertz, N. R. (2002). Alternative approaches to standard setting for licensing and certification examinations. *Applied Measurement in Education*, 15(1), 1-14.
- Cizek, G. J. (1993). Reconsidering standard and criteria. *Journal of Educational Measurement*, 30(2), 93-106.
- Cizek, G. J. (1996). Standard-setting guidelines. *Educational Measurement: Issues and Practice*, 15(1), 13-21.
- Clauser, B. E., Swanson, D. B., & Harik, P.(2002). Multivariate generalizability analysis of the impact of training and examinee performance information on judgments made in an Angoff-style standard-setting procedure. *Journal of Educational Measurement*, 39(4), 269-290.
- Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Conover, J. W., & Iman, R. L. (1978). *The rank transformation as a method of*

- discrimination with some examples*, Albuquerque, NM: Sandia Laboratories.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. NY: CBS College Publishing.
- Cross, L. H., Impara, J. C., Frary, R. B., & Jaeger, R. M. (1984). A comparison of three methods for establishing minimum standards on the National Teachers Examinations, *Journal of Educational Measurement*, 21, 113-129.
- Cross, L. H., Frary, R. B., Kelly, P. P., Small, R. C., & Impara, J. C. (1985). Establishing minimum standards for essays: Blind versus informed review. *Journal of Educational Measurement*, 22, 137-146.
- de Gruijter, D. N. M., & Hambleton, R. K. (1984). On problems encountered using decision theory to set cutoff scores. *Applied Psychological Measurement*, 8, 1-8.
- Dillon, G. F. (1996). The expectations of standard setting judges. *CLEAR Exam Review*, 2, 22-26.
- Ebel, R. L. (1972). *Essentials of educational measurement* (2rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Ebel, R. L. (1979). *Essentials of educational measurement* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Educational Testing Service (1976). *Report on a study of the use of the National Teachers' Examination by the state of South Carolina*. Princeton, NJ: Author.
- Eignor, D. R., & Hambleton, R. K. (1979). *Effects of test length and advancement score on several criterion referenced test reliability and validity indices*. (Laboratory of Psychometric and Evaluative Research Report No. 86). Amherst, MA: University of Massachusetts, School of Education.
- Emrick, J. A. (1971). An evaluation model for mastery testing. *Journal of Educational Measurement*, 8, 321-326.
- Fehrman, M. L., Woehr, D. J., & Arthur, W. (1991). The Angoff cutoff score method: The impact of frame-of-reference rater training. *Educational and Psychological Measurement*, 51(4), 857-872.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems, *Annals of Edgenics*, 7, 179-188.
- Fitzpatrick, A. R. (1989). Social influences in standard setting: The effects of social interaction on group judgments. *Review of Educational Research*, 59(3), 315-328.
- Frick, T. W. (1992). Computerized adaptive mastery tests as expert systems. *Journal of Educational Computing Research*, 8(2), 187-213.
- Garrett, H. E. (1937). *Statistics in psychology and education*. New York: Longmans, Green.

- Gessaman, M. P., & Gessaman, P. H. (1972). A comparison of some multivariate discrimination procedures. *Journal of the American Statistical Association*, 67, 468-472.
- Giraud, G., Impara, J. C., & Buckendahl, C. (2000). Making the cut in school districts: Alternative methods for setting cut-scores. *Educational Assessment*, 6, 291-304.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*, 18, 519-521.
- Glaser, R., & Nitko, A. J. (1971). Measurement in learning and instruction . In R. L. Thorndike (Eds.). *Education Measurement* (2nd ed., pp625-670). Washington, DC: American Council on Education.
- Gray, W. M. (1978). A comparison of Piagetian theory and criterion-referenced measurement, *Review of Educational Research*, 48, 223-249.
- Green, D. R., Trimble, C. S., & Lewis, D. M. (2003). Interpreting the results of three different standard setting procedures. *Educational Measurement: Issues and Practice*, 22(1), 22-32.
- Guilford, J. P. (1942). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.
- Haertel, E. (1985). Construct validity and criterion-referenced testing, *Review of Educational Research*, 55(1), 23-46.
- Haladyna, T. M., & Roid, G. H. (1983). A comparison of two approaches to criterion-referenced test construction. *Journal of Educational Measurement*, 20(3), 271-281.
- Halpin, G., Sigmon, G., & Halpin, G.(1983). Minimum competency standards set by three divergent groups of raters using three judgemental procedures: Implication for validity, *Educational and Psychological Measurement* , 43, 185-196.
- Hambleton, R. K. (1978). On the use of cut-off scores with criterion-referenced tests in instructional settings. *Journal of Educational Measurement*, 15(4), 277-290.
- Hambleton, R. K. (1980). Test score validity and standard setting methods. In R. A. Berk(Ed.), *Criterion-referenced measurement: The State of Art*. Baltimore, Md.: John Hopkins University.
- Hambleton, R. K. (1983). Application of item response models to criterion referenced assessment. *Applied Psychological Measurement*, 7(1), 33-44.
- Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn(Eds.), *Educational measurement* (3rd ed.)(pp.147-200). New York: Macmillan.
- Hambleton, R. K. (1990). Criterion referenced testing methods and practices. In T. B.

- Gutkin, & C. R. Reynolds (2nd ed.), *The Handbook of School Psychology* (pp. 388-415). New York: John Wiley & Sons.
- Hambleton, R. K. (1998). *Enhancing the validity of NAEP achievement level score reporting*. Proceedings of achievement levels workshop (pp. 77-98). Washington, DC: National Assessment Governing Board.
- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G. J. Cizek (Ed.). *Standard setting: Concepts, methods, and perspectives* (pp. 89-116). Mahwah, NJ: Erlbaum.
- Hambleton, R. K., & de Gruijter, D. N. M. (1983). Application of item response models to criterion referenced test item selection. *Journal of Educational Measurement*, 20(4), 355-367.
- Hambleton, R. N., & Eignor, D. R.(1980). Competency test development, validation, and standard setting. In R. M. Jaeger, & C. X. Tittle (Eds.), *Minimum competency achievement testing: Motive, models, measures, and consequence* (pp.367 -396). Berkeley, CA: McCutchan.
- Hambleton, R. K., Jaeger, R. M., Plake, B. S., & Mills, C. N.(in press). *Handbook for setting performance standards*. Washington, DC: Council of Chief State School Officers.
- Hambleton, R. K., Mills, C. N., & Simon, R. (1983). Determining the lengths for criterion referenced tests. *Journal of Educational Measurement*, 20(1), 27-38.
- Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, 10, 159-170.
- Hambleton, R. K., & Plake, B. S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education*, 8(1), 41-55.
- Hambleton, R. N., Swaminathan, H., Algina, J., & Coulson, D. B. (1978). Criterion-referenced testing and measurement: A review of technical issues and developments. *Review of Educational Research*, 48, 1-47.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and application*. Boston: Kluwer Nijhoff Publishing.
- Hambleton, R. K., & Traub, R. E.(1973). Analysis of empirical data using two logistic latent trait models, *British journal of mathematical and statistical psychology*, 26, 195-211.
- Hambleton, R. K., & Zaal, J. N. (1991). *Advances in educational and psychological testing* (Eds.). Boston, MA: Kluwer.
- Harasym, P. H. (1981). A comparison of the Nedelsky and modified Angoff standard-setting procedure on evaluation outcome. *Educational and*

- Psychological Measurement*, 41(3), 725-734.
- Harwell, M. R. (1983). *A comparison of two item selection procedures in criterion referenced measurement*. Unpublished doctoral dissertation, University of Wisconsin-Madison.
- Hattie, J. A. (1985). Methodological review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.
- Hoge, R. D., & Coladarci, T.(1989). Teacher based judgments of academic achievement : A review of the literature, *Review of educational research*, 59(3), 297-313.
- Hudson, J. P. Jr., & Campion, J. E. (1994). Hindsight bias in an application of the Angoff method for setting cutoff scores. *Journal of Applied Psychology*, 79(6), 860-865.
- Hurtz, G. M., & Hurtz, N. R.(1999). How many raters should be used for establishing cutoff scores with the Angoff method? A generalizability theory study. *Educational and Psychological Measurement*, 59(6), 885-897.
- Hurtz, M. G., & Auerbach, M. A. (2003). A meta analysis of the effects of modifications to the Angoff method on cutoff scores and judgment consensus. *Educational and Psychological Measurement*, 63(4), 584-601.
- Huynh, H. (1998). On score locations of binary and partial credit items and their applications to item mapping and criterion-referenced interpretation. *Journal of Educational and Behavioral Statistics*, 23(1), 35-56.
- Huynh, H. (2000). *On item mappings and statistical rules for selecting binary items for criterion referenced interpretation and bookmark standard settings*. Paper presented at the annual meeting of the National Council on Measurement in Education.(New Orleans, LA, April), 25-27.
- Impara, J. C., & Plake, B. S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34(4), 353-366.
- Ivens, S. H. (1970). *An investigation of item analysis, reliability and validity in relation to criterion-referenced tests*. Unpublished doctoral dissertation, Florida State University.
- Jaeger, R. M. (1978). *A proposal for setting a standard on the North Carolina High School Competency Test*. Paper presented at the annual meeting of the North Carolina Association for Research in Education, Chapel Hill.
- Jaeger, R. M. (1982). An iterative structured judgment process for establishing standards on competency tests: Theory and application. *Educational Evaluation and Policy Analysis*, 4, 461-476.
- Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn(Eds.), *Educational Measurement* (3rd ed., pp. 485-514). New York: Macmillan.

- Jaeger, R. M. (1995). Setting performance standards through two-stage judgmental policy capturing. *Applied Measurement in Education*, 8(1), 15-40.
- Jaeger, R. M., & Mills, C. N. (1997, April). *A holistic procedure for setting performance standards on complex large-scale assessments*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Jaeger, R. M., & Mills, C. N. (2001). An integrated judgment procedure for setting standards on complex, large-scale assessments. In G. J. Cizek (Ed.). *Standard setting: Concepts, methods, and perspectives* (pp. 313-338). Mahwah, NJ: Erlbaum.
- Jaeger, R. M., & Tittle, C. X. (1980). *Minimum competency achievement testing*. Berkeley, CA: McCutchan.
- Kahl, S. R., Crockett, T. J., DePascale, C. A., & Rindfleisch, S. L. (1994). *Using actual student work to determine cut-scores for proficiency levels: New methods for new tests*. Paper presented at the National Conference on Large-Scale Assessment, Albuquerque, NM.
- Kahl, S. R., Crockett, T. J., DePascale, C. A., & Rindfleisch, S. L. (1995). *Setting standards for performance levels using the student-based constructed response method*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Kane, M. T. (1987). On the use of IRT models with judgmental standard setting procedures. *Journal of Educational Measurement*, 24(4), 333-345.
- Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64(3), 425-461.
- Kane, M. (1998). Choosing between examinee-centered and test-centered standard setting methods. *Educational Assessment*, 5(3), 129-145.
- Kane, M. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J. Cizek (Ed.). *Standard setting: Concepts, methods, and perspectives* (pp. 53-88). Mahwah, NJ: Erlbaum.
- Kingsbury, G. G., & Weiss, D. J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 257-283). New York: Academic Press.
- Koffler, S. L. (1980). A comparison of approaches for setting proficiency standards. *Journal of Educational Measurement*, 17, 167-178.
- Koretz D., & Deibert E. (1995). Setting standards and interpreting achievement: A cautionary tale from the National Assessment of Educational Progress. *Educational Assessment*, 3(1), 53-81.

- Kriewall, T. E. (1972). *Aspects and applications of criterion-referenced tests* (IER Tech. Paper No. 103). Downers Grove, IL: Institute for Educational Research.
- Lewis, D.M., Green, D. R., Mitzel, H. C., Baum, K., & Patz, R. J. (1998, April). *The bookmark standard setting procedure: Methodology and recent implementations*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Lewis, D. M., Mitzel, H.C., & Green, D. R. (1996). *Standard setting: A bookmark approach*. Paper presented at the Council of Chief State School Officers National Conference on Large Scale Assessment, Boulder, CO.
- Livingston, S. A., & Zieky, M. J.(1978). *Basic skills assessment program : manual for setting standards on the basic skills assessment tests*. Menlo Park, Calif. : Addison-Wesley Testing Service.
- Livingston, S. A., & Zieky, M. J.(1989). A comparison study of standard-setting methods. *Applied Measurement in Education*, 2(2), 121-141.
- Loomis, S. C., Bay, L., Yang, W., & Hanick, P. L. (1999). *Field trials to determine which rating method(s) to use in the 1998 NAEP Achievement Levels-Setting Process for Civics and Writing*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal.
- Loomis, S.C.,& Bourque, M. L. (2001). From tradition to innovation: Standard setting on the National Assessment of Educational Progress. In G. J. Cizek (Ed.). *Standard setting: Concepts, methods, and perspectives* (pp. 175-217). Mahwah, NJ: Erlbaum.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Maurer, T. J., Alexander, R. A., Callahan, C. M., Bailey, J J., & Dambrot, F. H. (1991). Methodological and psychometric issues in setting cutoff scores using the Angoff method, *Personnel psychology*, 44, 235-262.
- McGinty, D., & Neel, J. H. (1996). *Judgmental standard setting using a cognitive components model*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York.
- Melican, G. J., Mills, C. N., & Plake, B. S. (1989). Accuracy of item performance predictions based on the Nedelsky standard setting method. *Educational and Psychological Measurement*, 49, 467-478.
- Meskauskas, J. A. (1976). Evaluation models for criterion-referenced testing: Views regarding mastery and standard setting. *Review of Educational Research*, 46, 133-158.

- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33-45). Hillsdale, NJ: Lawrence Erlbaum.
- Messick, S. (1989). Validity. In R. L. Linn(Ed.), *Educational Measurement*, (pp. 13-104). New York: Macmillan.
- Millman, J. (1972). *Tables for determining number of items needed on domain-referenced tests and number of students to be tested.* (Technical Paper No. 5). Los Angeles: Instructional Objective Exchange.
- Millman, J. (1973). Passing scores and test lengths for domain-referenced measures. *Review of Educational Research*, 43, 205-216.
- Mills, C. N. (1983). A comparison of three methods of establishing cut-off scores on criterion-referenced tests. *Journal of Educational Measurement*, 20(3), 283-292.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark method: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249-281). Mahwah, NJ: Erlbaum.
- Nassif, P. M. (1978). *Standard setting for criterion referenced teacher licensing tests.* Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto.
- Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14, 3-19.
- Nitko, A. J. (1980). Distinguishing the many varieties of criterion-referenced tests, *Review of Educational Measurement*, 50, 461-485.
- Nitko, A. (1983). *Educational tests and measurement: An introduction.* New York: Harcourt Bruce Jovanovich.
- Norcini, J., Lipner, R., Langdon, L., & Strecker, C. (1987). A comparison of three variations on a standard-setting method. *Journal of Educational Measurement*, 24, 56-64.
- Norcini, J., Shea, J. A., & Grosso, L. (1991). The effect of numbers of experts and common items on cutting score equivalents based on expert judgment. *Applied Psychological Measurement*, 15(3), 241-246.
- Norcini, J. J., Shea, J. A., & Kanya, D. T. (1988). The effect of various factors on standard setting. *Journal of Educational Measurement*, 25, 57-65.
- Novick, M. R., & Lewis, C. (1974). Prescribing test length for criterion-referenced measurement. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.), *Problems in criterion-referenced measurement* (CSE Monograph Series in Evaluation, No. 3, pp. 139-158). Los Angeles: Center for the Study of Evaluation,

University of California.

- Novick, M. R., Lewis, C., & Jackson, P. H. (1973). The estimation of proportions in a groups. *Psychometrika*, 38, 19-45.
- Pitoniak, M. J. (2003). *Standard setting methods for complex licensure examinations*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.
- Plake, B. S., Hambleton, R. K., & Jaeger, R. M. (1997). A new standard-setting method for performance assessments: The dominant profile judgment method and some field-test results. *Educational and Psychological Measurement*, 57(3), 400-411.
- Plake, B. S., & Hambleton, R. K. (2001). The analytic judgment method for setting standards on complex performance assessments. In G. J. Cizek (Ed.). *Standard setting: Concepts, methods, and perspectives* (pp. 283-312). Mahwah, NJ: Erlbaum.
- Popham, W. J., & Husek, T. R. (1969). Implication of criterion-referenced measurement, *Journal of Educational Measurement*, 6, 1-9.
- Plake, B. S., Impara, J. C., & Potenza, M. T.. (1994). Content specificity of expert judgments in a standard-setting study. *Journal of Educational Measurement*, 31(4), 339-347.
- Plake, B. S., & Melican, G. J. (1989). Effects of item context on intrajudge consistency of expert judgments via the Nedelsky standard setting method. *Educational and Psychological Measurement*, 49(1), 45-51.
- Popham, W. J. (1978). *Criterion-referenced measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Putnam, S. E., Pence, P., & Jaeger, R. M. (1995). A multi-stage dominant profile method for setting standards on complex performance assessments. *Applied Measurement in Education*, 8(1), 57-83.
- Reckase, M. D. (1983). A procedure for decision making using tailored testing. In D. J. Weiss(Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 237-255). New York: Academic Press.
- Reckase, M. D. (1998). Converting boundaries between National Assessment Governing Board performance categories to points on the National Assessment of Educational Progress score scale: The 1996 science NAEP process. *Applied Measurement in Education*, 11, 9-21.
- Reilly, R. R., Zink, D. L., & Israelski, E. W. (1984). Comparison of direct and indirect methods for setting minimum passing scores. *Applied Psychological Measurement*, 8, 421-429.
- Roudabush, G. E. (1974). *Models for a beginning theory of criterion-referenced tests*. Paper presented at the annual meeting of the National Council on

- Measurement in Education, Chicago.
- Samejima, F. (1994). Estimation of reliability coefficients using the test information function and its modification. *Applied Psychological Measurement, 18*(3), 229-244.
- Saunders, J. C., & Mappus, L. L. (1984). *Accuracy and consistency of expert judges in setting passing scores on criterion-referenced tests: The South Carolina experience*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Schoon, C. G., Gullion, C. M., & Ferrara, P. (1979). Bayesian statistics, credentialing examinations, and the determination of passing points. *Evaluation and the Health Professions, 2*, 181-201.
- Shepard, L. A. (1983). Standards for placement and certification. In S. B. Anderson & J. S. Helmick (Eds.), *On educational testing* (pp. 61-90). San Francisco: Jossey-Bass.
- Sireci, S. G., Robin, F., & Patelis, T. (1999). Using cluster analysis to facilitate standard setting. *Applied Measurement in Education, 12*(3), 301-325.
- Skakun, E. N., & Kling, S. (1980). Comparability of methods for setting standards. *Journal of Educational Measurement, 17*, 229-235.
- Smith, R. L., & Smith, J. K. (1988). Differential use of item information by judges using Angoff and Nedelsky procedures. *Journal of Educational Measurement, 25*(4), 259-274.
- Spray, J. A., & Reckase, M. D. (1994). *The selection of test items for decision making with a computer adaptive test*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- Spray, J. A., & Reckase, M. D. (1996). Comparison of SPRT and sequential Bayes procedures for classifying examinees into two categories using a computerized test. *Journal of Educational and Behavioral Statistics, 21*(4), 405-414.
- Stephenson, A. S., Elmore, P. B., & Evans, J. A. (2000). Standard-setting techniques: An application for counseling programs. *Measurement and Evaluation in Counseling and Development, 32*(4), 229-244.
- Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery test. *Journal of Educational Measurement, 25*, 47-55.
- Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion-referenced tests: A decision theoretic formulation. *Journal of Educational Measurement, 11*, 262-267.
- Thorndike, E. L. (1918). *The nature, purposes, and general methods of measurements of educational products*. The seventeen yearbook of the National Society for

- the study of Education, Part II. Bloomington, III.: Public School, Publishing Company.
- van der Linden, W. J. (1981) A latent trait look at pretest-posttest validation of criterion referenced test items, *Review of Educational Research*, 51(3), 379-402.
- van der Linden, W. J. (1982). A latent trait method for determining intra-judge inconsistency in the Angoff and Nedelsky techniques of standard setting , *Journal of Educational Measurement*, 19, 25-308.
- van der Linder, W. J. (1984). Some thoughts on the use of decision theory to set cutoff scores: Comment on de Gruijter and Hambleton. *Applied Psychological Measurement*, 8, 9-17.
- Wang, N. (2003). Use of the Rasch model in standard setting: An item mapping method. *Journal of Educational Measurement*, 40(3), 231-253.
- Webb, M. W. I., & Miller, E. R. (1995). *A comparison of the paper selection method and the contrasting groups method for setting standards on constructed-response items*. U.S.; Pennsylvania: December 31, 2004, from ERIC database.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problem. *Journal of Educational Measurement*, 21(4), 361-375.
- Wilcox, R. (1976). A note on the length and passing score of a mastery test. *Journal of Educational Statistics*, 1, 359-364.
- Wilcox, R. (1979). Comparing examinees to a control. *Psychometrika*, 44, 55-68.
- Wiberg, M.(2003). An optimal design approach to criterion-referenced computerized testing. *Journal of Educational and Behavioral Statistics*, 28(2), 97-110.
- Woehr, D. J., Arthur, W., & Fehrman, M. L. (1991). An empirical comparison of cutoff score methods for content-related and criterion-related validity settings. *Educational and Psychological Measurement*, 51(4), 1029-1039.
- Zieky, M. J., & Livingston, S. A. (1977). *Manual for setting standards on the Basic Skills Assessment Tests*. Princeton, NJ: Educational Testing Service.
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). *BILOG-MG for Windows (version 3)*. Chicago, IL: Scientific Software International, Inc.