

P187-194

參考文獻

中文部分

- 吳明隆 (2006)。SPSS 統計應用學習實務 (3 版)。臺北市：知城。
- 曾玉琳、王暄博、郭伯臣、許天維 (2006)。不同 BIB 設計對測驗等化的影響。
測驗統計年刊，第十三輯下期，209-229。臺中市：國立臺中教育大學。
- 張郁雯 (2008)。對比效應對學業自我概念之影響—發展的觀點。*教育心理學報*，
40 (1)，23~38。
- 譚克平 (2009)。TIMSS 國際教育評比研究簡介。大型教育資料庫及相關議題學術研討會。國立臺中教育大學。
- 洪碧霞、林素微、林娟如 (2006)。認知複雜度分析架構對 TASA-MAT 六年級線上測驗試題難度的解釋力。*教育研究與發展期刊*，2 (4)，69-86。
- 張鈿富、王世英、吳慧子、周文菁 (2006)。基本能力評量跨國發展經驗之比較。
教育資料與研究，68，81-99。
- 臺灣學生學習成就評量資料庫 (2007)。2007 年 TASA 社會科成果報告，未出版。
- 臺灣學生學習成就評量資料庫 (2006)。2006 年國小四年級學生『數學』學習成就之研究調查期末報告，未出版。
- 臺灣學生學習成就評量資料庫電子報 (2009)。檢索日期：2010 年 1 月 20 日。
網址：<http://tasa.naer.edu.tw/uploadfiles/file/TASAEpaper>

英文部分

Adams, R. J., Wilson, M., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1-23.

Allen, N. L., Donoghue, J. R., & Schoeps, T. L. (2001). *The NAEP 1998 technical report*. Washington, DC: National Center for Educational Statistics.

Allen N. L., Carlson J. E., Johnson E. G. ,& Mislevy, R. J.(1999) *The NAEP 1998 technical report*. Educational Testing Service.

Andrew R. W. & Terry L. S., (2001). *The NAEP 1998 Technical Report (NCES 2001-509)*. National Assessment Governing Board, U.S. Department of Education.

Birnbaum, A. (1968). *Some latent trait model and their use in inferring an examinee's ability*. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores*, 17-20. Reading, Mass: Addison-Wesley.

Bose R. C. & Nair K. R. (1939). *Partially balanced incomplete block designs*, *Sankhya*, 4, 337-372.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213-220.

Dorans, N., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23, 355-368.

- Foy P., Galia J., & Li L. (2008). Scaling the data from the TIMSS 2007 Mathematics and Science assessments. In John F. Olson, Michael O. Martin, Ina V.S. Mullis. (Eds). *TIMSS 2007 Technical Report*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College
- Graham J. R., Christine Y. O'S., Alka A., & Ebru E. (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Huynh, H. (1994). *Some technical aspects of standard setting*. Paper presented at the Joint Conference on Standard Setting for Large-Scale Assessments. Washington, D.C.
- 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), 38-58.
- Kuehl, R. O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis*. CA: Duxbury Press.
- Lee, J., Grigg, W., & Dion, G. (2007). *The Nation's Report Card: Mathematics 2007*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence Erlbaum.

- Martin, M. O. and Kelly, D. L. (1996) *Third International Mathematics and Science Study Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.
- Mislevy, R. J. & Bock R. D. (1982)*Implementation of the EM algorithm in the estimation of item parameters: The BILOG computer program*. Item Response Theory and Computerized Adaptive Testing Conference , Wayzata, MN, July 27-30, 1982.
- Mislevy, R. J. & Sheehan, K. M. (1980). Information matrices in latent-variable models. *Journal of Educational Statistics* 14(4), 335-350. Mislevy, R. J. & Sheehan, K. M. (1987). Marginal estimation procedures, in A.E. Beaton (ed.). *The NAEP 1983-1984 Technical Report* (Report No. 15-TR-20). Educational Testing Service, Princeton, N.J.
- Mislevy, R. J., & Sheehan, K. M. (1989). Information matrices in latent-variable models. *Journal of Educational Statistics*, 14, 335-350.
- Mislevy, R. J. (1991). Randomization-based inference about latent variable from complex samples. *Psychometrika*, 56, Psychometric Society, Greensboro, pp.177-196.
- Mislevy, R. J., A. E. Beaton, B. Kaplan and K. M. Sheehan.(1992). Estimating population characteristics form sparse matrix samples of item response. *Journal of Educational Measurement*, 29, pp.133-161, National Council on Measurement in Education, Washington, D.C..

Martin, M. O., Mullis, I.V.S., & Chrostowski, S. J. (Eds.) (2004). TIMSS 2003 Technical Report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mantel N., & Haenszel, W. M. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M. O., Ruddock, G. J., O'Sullivan, C.Y., Arora, A., & Eberber, E. (2005). *TIMSS 2007 Assessment Frameworks*.
<http://timss.bc.edu/TIMSS2007/frameworks.html>

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159-176.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, 58, 690-700.

NAEP Technical Documentation (2009). *The Nation's Report Card*. Retrieved May 13, 2009, from National Center for Education Statistics:
<http://nces.ed.gov/nationsreportcard/tdw/>

Nancy L. A., James E. C., & John R. D. (2001). *The NAEP 1998 Technical Report* (NCES 2001-509). National Assessment Governing Board, U.S. Department of Education.

OECD (2005). *PISA 2003 Technical Report*. OCED, Paris.

OECD (2006). *Assessing Scientific, Reading and Mathematical Literacy*. OCED, Paris. http://www.oecd.org/document/33/0,3343,en_32252351_32236191_37462369_1_1_1_1,00.html

OECD (2006). *PISA RELEASED ITEMS - MATHEMATICS*. Retrieved February 27, 2009, from <http://www.pisa.oecd.org/dataoecd/14/10/38709418.pdf>

OECD (2007). *PISA 2006 Science Competencies for Tomorrow's World*. OCED, Paris. Retrieved January 11, 2010, from http://www.pisa.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html

OECD (2007). *The programme for international student assessment –PISA*. OECD, Paris. Form <http://www.oecd.org/dataoecd/15/13/39725224.pdf>

OECD (2009). *Top of the Class - High Performers in Science in PISA 2006*. OCED, Paris. http://www.oecd.org/document/51/0,3343,en_32252351_32236191_42642227_1_1_1_1,00.html

OECD (2009). *Green at Fifteen? How 15-year-olds perform in environmental science and geoscience in PISA 2006*. OCED, Paris. http://www.oecd.org/document/22/0,3343,en_32252351_32236191_42466966_1_1_1_1,00.html

OECD (2009). *PISA Data Analysis Manual: SPSS and SAS, Second Edition*. OCED, Paris. http://www.oecd.org/document/38/0,3343,en_32252351_32236191_42609254_1_1_1_1,00.html

OECD (2009). *PISA 2006 Technical Report*. OCED, Paris.

OECD (2009). *Equally prepared for life? How 15-year-old boys and girls perform in school*. OCED, Paris. http://www.oecd.org/document/51/0,3343,en_32252351_32236191_42837811_1_1_1_1,00.html

Rock, D. A. (1991). *Subscale dimensionality*. Paper presented at the meeting of the Design and Analysis Committee of the National Assessment of Educational Progress, Washington, DC.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, 58, 159-194.

Wang, W., & Wilson, M. (2005). Exploring local item dependence using a facet random-effects facet model. *Applied Psychological Measurement*, 29, 296-318.

Wang, X., Bradlow, E. T., & Wainer, H. (2004). *User's guide for SCORIGHT (version 3.0): A computer program for scoring tests built of testlets including a module for covariate analysis*. Princeton, NJ: Educational Testing Service; Philadelphia, PA: National Board of Medical Examiners.

- Von Davier Matthias, Gonzalez Eugenio, & Mislevy R. J. (2009). What are plausible values and why are they useful? *IERA Monograph Series: Issues and Methodologies in Large-Scale Assessment*, 2, pp.9-36.
- Wu, M. L., R. J. Adams and M. R. Wilson (1997), *ConQuest: Multi-Aspect Test Software* [computer program], Australian Council for Educational Research, Camberwell.
- Wu, M. L., Adams, R. J., & Wilson, M. R. (1998). *Acer ConQuest*. Melbourne, Victoria, Australia: Australian Council for Educational Research press.
- Yates, F. (1936). A new method of arranging variety trials involving a large number of varieties. *J. Agric. Sci.* 26, 424-455.
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary for binary items*. Mooresville IL: Scientific Software.