

100 年度「教學評量專題研究」教學綱要

余民寧 教授

一、教學目標

培養學生專精各種教學評量的研究專題。

二、背景知識

主修教育心理學、諮商輔導、心理計量學的教育碩士。

曾修習過「教育測驗與評量」、「教學評量概論」、「心理與教育測驗」（大學課程，必備）、或「測驗理論研究」、「測驗編製與量表發展研究」（研究所課程，選修）

三、討論專題

1. 紙筆測驗評量
2. 實作評量
3. (數位化) 檔案評量
4. 過程(動態)評量
5. 圖形(S-P表、概念圖、心智圖)評量
6. 知識結構評量(認知診斷評量)
7. 證照認證評量、標準設定
8. (電腦化) 適性(測驗)評量
9. 學生評鑑教學
10. 評量方法學(如: CTT vs. IRT assessment、e-Assessment、fuzzy assessment、neuro-network assessment、artificial intelligence assessment、social network assessment、web-based assessment、evaluation for instruction)

四、閱讀材料

1. 紙筆測驗評量：

余民寧(2011)。教育測驗與評量：成就測驗與教學評量(第三版)。台北：心理。

2. 實作評量：

Hunter, K., & Docherty, P. (2011). Reducing variation in the assessment of student writing. *Assessment and Evaluation in Higher Education*, 36(1), 109- last page.

Reb, J., & Greguras, G. J. (2010). Understanding performance ratings: Dynamic performance, attributions, and rating purpose. *Journal of Applied Psychology*, 95(1), 213-last page.

3. (數位化) 檔案評量：

Brinke, D. J., Sluijsmans, D. M. A., & Jochems, W. M. G. (2010). Assessors' approaches to portfolio assessment in assessment of prior learning procedures. *Assessment and Evaluation in Higher Education*, 35(1), 59-74.

4. 過程 (動態) 評量：

Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment. *Assessment and Evaluation in Higher Education*, 36(1), 95- 108.

Nestel, D., Kneebone, R., Nolan, C., Akhtar, K., & Darzi, A. (2011). Formative assessment of procedural skills: Students' responses to the Objective Structural Clinical Examination and the Integrated Performance Procedural Instrument. *Assessment and Evaluation in Higher Education*, 36(2), 171- 184.

5.AME 專題：Evidence-centered assessment design in practice. Special issue in AME.

Huff, K., & Plake, B. S. (2010). Evidence-centered assessment design in practice. *Applied Measurement in Education*, 23(4), 307-309.

Huff, K., Steinberg, L., & Matts, T. (2010). The promise and challenges of implementing evidence-centered design in large-scale assessment. *Applied Measurement in Education*, 23(4), 310-324.

Ewing, M., Packman, S., Hamen, C., & Thurber, A. C. (2010). Representing targets of measurement within evidence-centered design. *Applied Measurement in Education*, 23(4), 325-341.

Plake, B. S., Huff, K., & Reshetar, R. (2010). Evidence-centered assessment design as a foundation for achievement-level descriptor development and for standard setting. *Applied Measurement in Education*, 23(4), 342-357.

Hendrickson, A., Huff, K., & Luecht, R. (2010). Claims, evidence, and achievement-level descriptors as a foundation for item design and test specifications. *Applied Measurement in Education*, 23(4), 358-377.

Bejar, I. I. (2010). Application of evidence-centered assessment design to the advanced placement redesign: A graphic restatement. *Applied Measurement in Education*, 23(4), 378-391.

Brennan, R. L. (2010). Evidence-centered assessment design and the advanced placement program[®]: A psychometrician's perspective. *Applied Measurement in Education*, 23(4), 392-last page.

6.知識結構評量(認知診斷評量)：

de la Torre, J., & Karelitz, T. M. (2009). Impact of diagnosticity on the adequacy of models for cognitive diagnosis under a linear attribute structure: A simulation study. *Journal of Educational Measurement, 46*(4), 450-469.

7.證照認證評量、標準設定：

Clauser, B. E., Mee, J., Baldwin, S. G., Margolis, M. J., & Dillon, G. F. (2009). Judges' use of examinee performance data in an Angoff standard-setting exercise for a medical licensing examination: An experimental study. *Journal of Educational Measurement, 46*(4), 390-407.

Pfeiffer, N., Hagemann, D., & Backenstrass, M. (2011). A new method for estimating the variance overlap between the short and the long form of a psychological test. *Educational and Psychological Measurement, 71*(2), 380-388.

8. (電腦化)適性(測驗)評量：

Vaughn, B. K., & Wang, Q. (2010). DIF trees: Using classification trees to detect differential item functioning. *Educational and Psychological Measurement, 70*(6), 941-952.

DeMars, C. E. (2010). Type I error inflation for detecting in the presence of impact. *Educational and Psychological Measurement, 70*(6), 961-972.

Frederickx, S., Tuerlinckx, F., DeBoeck, P., & Magis, D. (2011). RIM: A random item mixture model to detect differential item functioning. *Journal of Educational Measurement, 47*(4), 432-457.

9.學生評鑑教學：

Symbaluk, D. G., & Howell, A. J. (2010). Web-based student feedback: Comparing teaching-award and research-award recipients. *Assessment and Evaluation in Higher Education, 35*(1), 75-86.

Davies, M., Hirschberg, J., Lye, J., & Johnston, C. (2010). A systematic analysis of quality of teaching surveys. *Assessment and Evaluation in Higher Education, 35*(1), 87-100.

Venette, S., Sellnow, D., & McIntyre, K. (2010). Charting new territory: Assessing the online frontier of student ratings of instruction. *Assessment and Evaluation in Higher Education, 35*(1), 101-116.

Patrick, C. L. (2011). Student evaluation of teaching: Effects of the Big Five personality

traits, grades and the validity hypothesis. *Assessment and Evaluation in Higher Education*, 36(2), 239-last page.

Liu, O. L. (2011). Measuring value-added in higher education: Condition and caveats – Results from using the Measure of Academic Proficiency and Progress (MAPP™). *Assessment and Evaluation in Higher Education*, 36(1), 81- 94.

10. 評量方法學：

Granberg-Rademacker, J. S. (2010). An algorithm for converting ordinal scale measurement data to interval/ratio scale. *Educational and Psychological Measurement*, 70(1), 74-90.

五、上課日期表（每週四 14:00~17:00PM）

週次	日期	研討內容及進度	負責報告人
1	9/15	導論：紙筆測驗評量	老師上課
2	9/22	實作評量：口試	
3	9/29	實作評量：情意及技能	
4	10/6	檔案評量	
5	10/13	檔案評量	
6	10/20	動態評量	
7	10/27	AME 專題	
8	11/3	AME 專題	
9	11/10	AME 專題	
10	11/17	認知診斷評量	
11	11/24	認知診斷評量	
12	12/1	證照評量&標準設定	
13	12/8	證照評量&標準設定	
14	12/15	電腦化適性評量	
15	12/22	電腦化適性評量	
16	12/29	學生評鑑教學	
17	1/5	學生評鑑教學	
18	1/12	期末考試（不上課）	開始繳交期末報告

六、成績考核

- 1.每人認領兩篇專題文章，帶領同學一起研讀及課堂討論，佔 50%學期分數。
- 2.撰寫一篇自訂主題的期末報告，佔 50%學期分數。
- 3.期末報告繳交日期，課堂上決定。