

應用網路評量系統提昇 教師評量知識之探究

王子華¹ 王國華² 王瑋龍³ 黃世傑³

¹ 國立新竹教育大學 教育學系 助理教授

² 國立彰化師範大學 科學教育研究所 副教授

³ 國立彰化師範大學 生物學系 教授

thwang@mail.nhcue.edu.tw

摘要

教師評量素養會影響教學的成敗，但是令人憂心的是，許多教師評量素養相關的調查研究都指出，教師的評量素養是有待改善的。本研究的主要目的，是希望由「師資培育課程」方向提出具體策略，由提升教師評量知識的角度來改善職前教師的評量素養。本研究之具體策略為，採用 WATA 系統 (Web-based Assessment and Test Analysis system) 為基礎，融入文獻中所建議之師資培育模式，發展出一個具備「個人化」與「情境化」特性的 e-Learning 型態之教師評量素養培育模式 - 「P2R-WATA (Practicing, Reflecting and Revising with WATA system)」。

這個模式讓參與的職前教師，可以在 WATA 系統的 Triple-A Model (Assembling, Administering and Appraising) 中進行個別化的評量知能學習。WATA 系統除可以隨時提供個別化的學習支援外，也可以讓參與的職前教師實作 (Practicing) 命題與執行考試，完成命題的試卷可以在網路上真實執行考試，由真實的學生在網路上作答，系統可以立刻協助教師將學生評量過程資訊進行分析，並且回饋各種教育統計資訊給予職前教師參考，協助其反省 (Reflecting) 自己命題的品質，分析受試者的學習缺失，並進行試題修改 (Revising)。本研究為了考驗「P2R-WATA」評量素養培育模式的效益，採用準實驗設計，將六十位職前教師，分派進入實驗組與對照組 (各三十人)，實驗組採用「P2R-WATA」評量素養培育模式，而對照組則採用不具 WATA 系統輔助之「P2R (Practicing, Reflecting and Revising)」評量素養培育模式，在進行一學期的教學後，發現實驗組之評量知識的發展情形顯著較對照組佳。

關鍵字：WATA 系統、「P2R-WATA」評量素養培育模式、評量素養、師資培育、數位學習

1. 前言

評量無論對於教師或學生而言都是相當重要的，對於學生而言，良好的評量可以提升學習效益 [2][18][24]，對於教師而言，良好的評量可以用來進行好的教學探究 (good inquiry into teaching)，與理解學生的想法、信念與推理 [20]，但是，自 1990 年代起至今，許多針對「在職教師」或是「修過評量課程的職前教師」所進行的評量素養研究均發現，

教師的評量素養有缺陷 [5][6][26][30]，由文獻當中整理出的原因主要有二，分別是「師資培育過程中與評量相關的訓練課程之缺失 [5][14][15][16][17]」與「師資認證制度不重視教師評量能力或相關訓練 [25]」所造成的。

國內方面，教師評量素養相關的研究並不多，但是目前的研究結果與國外類似，都指向教師缺乏良好的評量素養，例如：王子華等人 [1] 初步針對國內生物科職前教師評量素養的調查發現，其評量素養有一些缺陷，此外，在美國存在的兩個造成評量素養不良的因素：「師資培育過程中與評量相關的訓練課程之缺失」與「師資認證制度不重視教師評量能力或相關訓練」，也存在於臺灣。目前臺灣的師資培育制度，無論「師資培育課程」或「教師認證」方面都不大重視教師的評量能力：「師資培育課程」方面，其中的評量相關課程只有四學分，且均屬於選修課，「教師認證」方面，教師評量能力的考核並沒有獨立出來，而是與其他科目包含在一個教育專業科目中，由此足見，在臺灣可能存在與美國一樣輕忽教師評量素養的問題，以美國的經驗來看，臺灣的教師評量素養可能也有缺陷，但是相關證據資料，則有待未來持續的研究來進行蒐集與探討。

本研究主要是要針對「師資培育過程中與評量相關的訓練課程之缺失」這個面向提出改善策略，建構一個 e-Learning 型態教師評量素養培育模式 - 「P2R-WATA (Practicing, Reflecting and Revising with WATA system)」評量素養培育模式 (圖 1)。

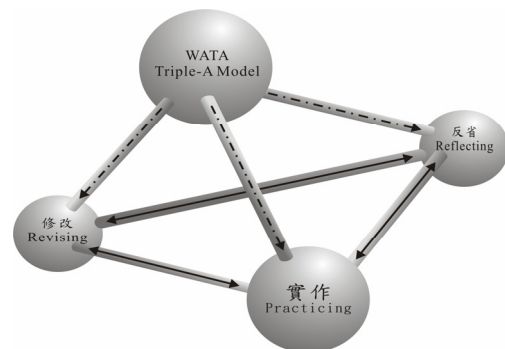


圖 1 「P2R-WATA」評量素養培育模式

感謝國科會專案計畫 NSC-90-2511-S-018-015、NSC92-2511-S-018-009 與 NSC93-2511-S-018-002 的支持，使本研究得以順利完成。

「P2R-WATA」評量素養培育模式，是基於文獻中所提及的有效教師評量素養培育模式-「Plake, & Impara [22]模式」、「NAC模式[14]」、「ALLT模式[3][27]」、「PALS模式與IPALS模式[14]」所發展出來的，此四套模式除了均採取傳統方式教授評量相關知識與技巧之外，另有一個共同特色，就是均提供學習者實作的機會，使課室經驗與評量素養訓練結合在一起，這種設計在許多相關的文獻當中都受到肯定[5][14][18][25]，而Taylor, & Nolen[28]甚至主張所有的評量概念應該在教學實際情境(instructional practice)中教導給接受評量素養訓練的教師才會有效益。因此，「將課室經驗與評量素養訓練結合」是本研究所發展之「P2R-WATA」評量素養培育模式的核心設計。

另外，關於文獻中對於師資培育課程與環境的設計方面，Fox[11]、Passig[21]與Egbert, & Thomas[10]均提出值得參考的建議。Fox[11]指出，師資培育課程設計本身必須靠近教師本身的發展程度，而Passig[21]指出，師資培育 e-Learning 環境中，「個人化」設計是一個相當重要的面向。此外，關於師資培育 e-Learning 環境的「個別化」設計方面，Egbert, & Thomas[10]指出，e-Learning 系統需隨教師的學習進度，依循個別教師的需求，隨時提供各種個別化的回饋給予教師，此外，亦應該強調教師個人主動參與學習，由教師「自己決定學習步驟」，並且應該「提供學習者更多的反省時間」，使教師依據自己的學習進度，以及遭遇的問題與狀況，隨時搜尋與獲得系統的回饋，換言之，系統提供的是「資源為基礎的學習而非教學為基礎的學習」[10]。「P2R-WATA」評量素養培育模式亦將遵循上述建議，將「個人化」設計作為重要精神。

除「個人化」外，對於師資培育環境而言，「情境化」亦是一個相當重要的設計[7][23][29]，但是因為對傳統課室教學而言，很難達到情境化[23]，所以很少有師資培育環境考慮到情境化的問題[29]。但在 e-Learning 環境中，教學者則可以藉由資訊技術模擬各種傳統教室中無法模擬的情境，讓學習者可以在這些模擬的情境中進行學習[8][9][13]。Moore et al.[19]指出，如果學習活動可以使所強調的理論與真實情境連結起來，則該理論對學習者才是真實的，因此，本研究將基於上論點，將「情境化」設計作為另一個重要精神。

本研究所發展之「P2R-WATA」評量素養培育模式，除了將「個別化」與「情境化」精神融入外，將遵循 Wiske et al. [29] 與 Bransford et al.[4] 看法，以及延續 Wang et al.[30]的研究，採用 WATA 系統的 Triple-A Model (Assembling, Administering, Appraising)為基礎，發展一個具有「完整評量流程鷹架」且以「學習者為中心」的 e-Learning 環境。此外，也會發展適當的「電子化評量教材」，而這些「電子化評量教材」則打散並融入 WATA 系統中

Triple-A Model 的每個節點，讓教師可以隨著自己的學習進度、模擬命題、執行考試與進行試後分析，而在過程中可以隨時獲得系統的支援與回饋。此外，亦依據 Wiske et al.[29]的建議，將教師於專業成長課程中的「演化出的關心事項」問題，考量進入「P2R-WATA」評量素養培育模式的設計中，讓職前教師可以依照本身的學習進度，在遭遇問題時，隨時獲得系統個別化的支援與說明。「P2R-WATA」評量素養培育模式，詳細的執行模式請參考「2.3 研究設計」。

王子華[2]指出，教師評量素養包含三個向度-「評量知識」、「評量觀點」與「評量技巧」，而本研究主要針對教師評量素養的「評量知識」向度的發展進行探究，本研究除發展一個有效的「P2R-WATA」教師評量素養培育模式外，也希望能夠探究「P2R-WATA」評量素養培育模式，對於提升職前教師評量知識的效益為何？期待本研究之研究發現，可以提供國內師資培育工作者，作為發展教師評量素養培育課程之參考。

2. 研究方法

2.1 研究樣本

本研究的參與者為某師範大學生物系之職前教師，分別為大三學生三十人(男 20 人，女 10 人)，平均年齡 21.47 歲(SD=.67)，與大四學生三十人(男 15 人，女 15 人) 平均年齡 22.53 歲(SD=.67)。大四學生屬於實驗組，採用 WATA 系統輔助教師評量素養之訓練，而大三學生採用傳統方式進行教師評量素養之訓練。所有大四學生均有上網環境，以及具備電腦網路使用能力，關於實驗組與對照組職前教師之起點行為的同質性考驗方面，請參考「3.研究結果」。

2.2 研究工具

2.2.1 WATA 系統

本研究採用「WATA 系統(Web-based Assessment and Test Analysis system, WATA system)」，建構本研究中實驗組之評量素養培育模式-「P2R-WATA」。WATA 系統具有個人化的 Triple-A(Assembling、Administering、Appraising) 架構(圖 2)，擁有「試題組成引擎(Assembling)」可以讓教學者管理考試行程，以及建立個人題庫與採用雙向細目分析表進行試卷的編製(圖 3)；「考試管理引擎(Administering)」可以讓教學者同時管理多個考試的進行(圖 4)；「試後分析引擎(Appraising)」可以進行試卷分析(圖 5)與試題分析(圖 6)，並產生各種教育統計資訊。

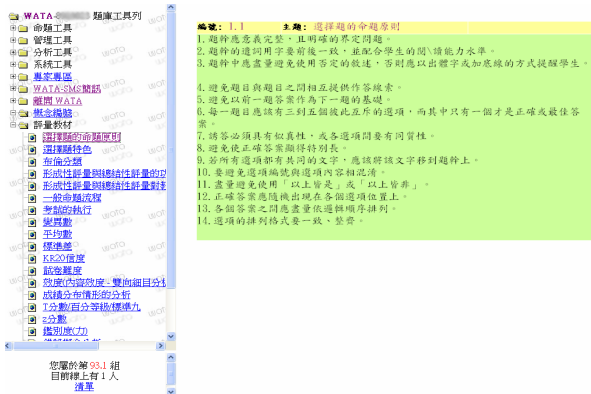


圖 2 WATA 系統個人化 Triple-A Model 介面

CK-03 - 第 3 → 3 重雙向細目分析表

部分請務必輸入

概念/年級	知識	理解	應用	分析	綜合	評鑑
0304	0	1	0	0	0	0
0305	2	0	0	0	0	0
0306	0	1	1	0	0	0
0307	0	2	0	0	0	0

提交 重新設定

圖 3 依據雙向細目分析表命題

fish 的考試行程

更新考試行程

命題	說明	日期	時間
bi-01 用排除	<ul style="list-style-type: none"> 2000 年 10 月 20 日 09:00 10:00 分 2000 年 10 月 20 日 10:00 11:00 分 2000 年 10 月 20 日 11:00 12:00 分 2000 年 10 月 20 日 12:00 13:00 分 2000 年 10 月 20 日 13:00 14:00 分 2000 年 10 月 20 日 14:00 15:00 分 2000 年 10 月 20 日 15:00 16:00 分 2000 年 10 月 20 日 16:00 17:00 分 2000 年 10 月 20 日 17:00 18:00 分 2000 年 10 月 20 日 18:00 19:00 分 2000 年 10 月 20 日 19:00 20:00 分 2000 年 10 月 20 日 20:00 21:00 分 2000 年 10 月 20 日 21:00 22:00 分 2000 年 10 月 20 日 22:00 23:00 分 2000 年 10 月 20 日 23:00 24:00 分 	0 -- 1	06 小時內可輸入 輸入密碼 [fish] 解除
bi-02 用排除	<ul style="list-style-type: none"> 2000 年 12 月 12 日 09:00 10:00 分 2000 年 12 月 12 日 10:00 11:00 分 2000 年 12 月 12 日 11:00 12:00 分 2000 年 12 月 12 日 12:00 13:00 分 2000 年 12 月 12 日 13:00 14:00 分 2000 年 12 月 12 日 14:00 15:00 分 2000 年 12 月 12 日 15:00 16:00 分 2000 年 12 月 12 日 16:00 17:00 分 2000 年 12 月 12 日 17:00 18:00 分 2000 年 12 月 12 日 18:00 19:00 分 2000 年 12 月 12 日 19:00 20:00 分 2000 年 12 月 12 日 20:00 21:00 分 2000 年 12 月 12 日 21:00 22:00 分 2000 年 12 月 12 日 22:00 23:00 分 2000 年 12 月 12 日 23:00 24:00 分 	2 -- 3	1 小時內可輸入 輸入密碼 [fish] 解除
bi-03 用排除	<ul style="list-style-type: none"> 2000 年 10 月 12 日 09:00 10:00 分 2000 年 10 月 12 日 10:00 11:00 分 2000 年 10 月 12 日 11:00 12:00 分 2000 年 10 月 12 日 12:00 13:00 分 2000 年 10 月 12 日 13:00 14:00 分 2000 年 10 月 12 日 14:00 15:00 分 2000 年 10 月 12 日 15:00 16:00 分 2000 年 10 月 12 日 16:00 17:00 分 2000 年 10 月 12 日 17:00 18:00 分 2000 年 10 月 12 日 18:00 19:00 分 2000 年 10 月 12 日 19:00 20:00 分 2000 年 10 月 12 日 20:00 21:00 分 2000 年 10 月 12 日 21:00 22:00 分 2000 年 10 月 12 日 22:00 23:00 分 2000 年 10 月 12 日 23:00 24:00 分 	1 -- 3	200 小時內可輸入 輸入密碼 [fish] 解除

圖 4 多重考試行程管理畫面

試卷分析

人數	34
總算數	134.18
標準差	11.58
標準平均	54.58
K200	0.95
難度	48.92%

班級個人成績分析

學號/姓名	原始分數	z 分數(未常態)	T 分數(未常態)
01	57.78	0.29	52.9
02	62.22	0.68	56.8
03	35.56	-1.61	33.9
04	48.89	-0.46	45.4
05	55.56	0.1	51
06	48.89	-0.46	45.4

班級個人常態化成績分析

名次	學號/姓名	原始分數	z 分數(常態化)	T 分數(常態化)
1	20	73.33	1.35	63.5
2	24	73.33	1.35	63.5
3	29	71.11	1.18	61.8
3	33	68.89	0.93	59.3
3	28	68.89	0.93	59.3
4	09	66.67	0.82	58.2
5	23	64.44	0.63	56.3
5	08	64.44	0.63	56.3
6	02	62.22	0.45	54.5
6	27	62.22	0.45	54.5
7	25	60	0.30	53

圖 5 試卷分析

題目	A	B	C	D	E	NA	Num	DF	ID
<p>題目</p> <p>下列關於「浮力」的敘述，何者正確？(選對 4 個得 4 分，選對 3 個得 3 分，選對 2 個得 2 分，選對 1 個得 1 分，選對 0 個得 0 分。)</p> <p>甲：浮力的大小與物體的體積成正比。乙：浮力的大小與物體的密度成正比。丙：浮力的大小與物體的質量成正比。丁：浮力的大小與物體的深度成正比。</p> <p>圖 1：一個裝有水的容器，其中有一個物體 A 浸在水中。物體 A 的體積為 V，密度為 ρ。物體 A 的質量為 m。物體 A 的體積為 V，密度為 ρ。物體 A 的質量為 m。</p> <p>圖 2：一個裝有水的容器，其中有一個物體 B 浸在水中。物體 B 的體積為 V，密度為 ρ。物體 B 的質量為 m。物體 B 的體積為 V，密度為 ρ。物體 B 的質量為 m。</p>	35.29% 10/9 E:31.4%	33.61% 11/9 E:17.4%	4.2% 1/9 E:1.1%	35.29% 10/9 E:17.4%	0% 0/9 E:0%	11.76% 4/9 E:6.3%	119	0.47	34%
<p>題目</p> <p>下列關於「浮力」的敘述，何者正確？(選對 4 個得 4 分，選對 3 個得 3 分，選對 2 個得 2 分，選對 1 個得 1 分，選對 0 個得 0 分。)</p> <p>甲：浮力的大小與物體的體積成正比。乙：浮力的大小與物體的密度成正比。丙：浮力的大小與物體的質量成正比。丁：浮力的大小與物體的深度成正比。</p> <p>圖 1：一個裝有水的容器，其中有一個物體 A 浸在水中。物體 A 的體積為 V，密度為 ρ。物體 A 的質量為 m。物體 A 的體積為 V，密度為 ρ。物體 A 的質量為 m。</p> <p>圖 2：一個裝有水的容器，其中有一個物體 B 浸在水中。物體 B 的體積為 V，密度為 ρ。物體 B 的質量為 m。物體 B 的體積為 V，密度為 ρ。物體 B 的質量為 m。</p>	16.81% 6/9 E:18.8%	3.32% 1/9 E:1.5%	4.22% 1/9 E:1.5%	57.14% 20/9 E:27.8%	0% 0/9 E:0%	16.67% 6/9 E:8.9%	119	0.79	57%

圖 6 試題分析、誘答力分析

2.2.2 評量知識測驗 (Assessment Knowledge Test, AKT)

AKT[2]主要用以評估職前教師的評量知識。試卷發展前，先建立試卷所應包含之概念，並將這些概念加以編號，這些概念與編號主要是由三位生物教育與評量專家，依據 Gronlund, & Linn[12]所提出之評量基本概念，選取 30 個對生物科教師較重要的教學評量概念所組成，並且依據這些概念進行 AKT 初測試卷命題，AKT 初測試卷共有 50 題，初測試卷經過某師範大學生物學系職前教師預試後，刪除鑑別度低於 .250 的題目之後，共保留 40 題作為 AKT 正式試卷，AKT 試卷的內部一致性 Cronbach α 值為 .993，難度為 .505。

2.3 研究設計

本研究採用「生物教學理論與實際」課，進行本研究所設計之所有教學活動，「生物教學理論與實際」對於大三與大四學生而言，屬於二學分的必修課，該課程的授課內容，主要是國中生物學之教學法，依照過去的授課方式，並沒有特別強調「評量」這個主題。本研究是在這門課的情境下，針對「評量」這個主題設計一個為期一個學期(三個月)的教學活動，實驗組採用「P2R-WATA (Practicing, Reflecting and Revising with WATA system)」評量素養培育模式進行學習，而對照組則採用「P2R(Practicing, Reflecting and Revising)」評量素養培育模式(圖 7)進行學習。

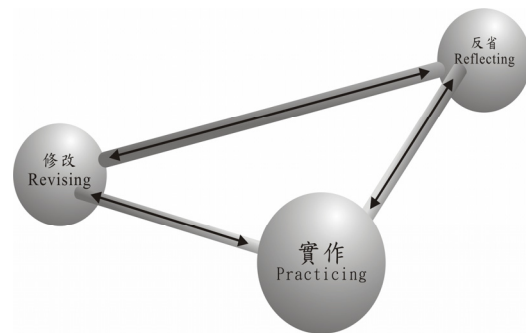


圖 7 「P2R」評量素養培育模式

實驗組職前教師在「P2R-WATA」評量素養培育模式中，由於 WATA 系統具備 Triple-A Model 可以提供「完整評量流程鷹架」，因此學習過程中 Triple-A Model 將扮演核心的角色，作為整個評量素養培育過程的支撐鷹架，在整個培育過程中 (Practicing, Reflecting, Revising, P2R)隨時提供「完整評量流程鷹架」、「電子化評量教材」、「個人化學習」與「情境化學習」的功能，運作方式簡介如下：

● 「個別化」設計

所有參與的職前教師，都可以在個人化介面中進行個別化學習，系統可以隨著職前教師的學習進度，依循個別的需求，隨時提供各種個別化的回饋與「電子化評量教材」給予職前教師。

● 「情境化」與「將課室經驗與評量素養訓練結合」設計

所有參與的職前教師，都可以在 WATA 系統中，依循 WATA 系統「完整評量流程鷹架」進行模擬命題，並且進行考試的「實作(Practicing)」，換言之，可以在 WATA 系統的輔助下，在網路上進行試卷模擬命題與實際執行考試，並由真實的學生在網路上進行考試，學生考試完畢後，WATA 系統可以立刻協助職前教師將學生考試結果進行試後分析，並且立刻回饋教育統計資訊給予職前教師，讓職前教師利用這些資訊「反省(Reflecting)」自己的命題缺失，並且找出教師教學缺失與學生學習缺失，並依據自己過去所學、「電子化評量教材」與 WATA 系統提供之試後分析資訊，針對自己的命題進行「修改(Revising)」，然後藉由 WATA 系統再次施測，來驗證自己的想法是否正確，以及自己的試題修改策略是否有效。

對照組職前教師所採用的「P2R」評量素養培育模式，與實驗組的差異主要在於「是否使用 WATA 系統」，此外，在評量教材方面也略有不同，對照組可以閱讀到的評量教材內容與實驗組之「電子化評量教材」完全相同，唯一的差別只在於，改為書面資料方式呈現評量教材。

2.4 研究流程

首先，將職前教師區分成兩大群-實驗組(大四)、對照組(大三)，然後執行 AKT 前測，然後再讓兩組職前教師針對研究者所提供之國中一年級自然與生活科技，第三章到第四章第五節課本內容與雙向細目分析表，命 10 題單選題，並形成一份試卷，題目需包含：題幹、選項、答案、章節、評量目標、預測難易度、出處等。實驗組的命題過程完全在 WATA 系統中操作，對照組則將試題打成二份 WORD 檔(一份只是單純的考題，要列印變成紙筆測驗版--給國中生施測的考卷，一份是包含所有題目詳細內容參數的資料--送交研究者保存)。然後，將所有實驗組與對照組的命題題目，亂數隨機分派到 7 位中部國中教師所任教的國一班級(共 16 班)進行施測(所有國中班均為常態分班，為了避免人為因素影響，無論實驗組與對照組職前教師的試卷，均採用亂數方式指派進入不同班級。在本研究中，所有職前教師均會進行的兩次試卷施測，其兩次施測的班級均相同，且同一班級不會同時做到實驗組與對照組的考題，兩次施測間隔約一個月，所有參與施測的學生均已經學習過與試題相關的學科內容)。

施測完畢後，實驗組在 WATA 系統進行試後分析，對照組則參考相關資料，依自己的方式進行分析，之後，實驗組與對照組均必須撰寫試後分析報告，並 E-mail 給研究者，而且會利用試後分析的結果，將有問題的題目進行修改(實驗組與對照組職前教師，均不可以增刪題目，只可以修改初次命題的

題目)。接著，同時讓實驗組與對照組進行 AKT 中測，之後，將實驗組與對照組職前教師修改過的試題，再針對原班級進行二度施測，施測完畢後，再度回收施測結果交給實驗組與對照組職前教師，再進行一次試後分析與撰寫試後分析報告，並將試後分析報告 E-mail 給研究者，然後，利用試後分析的結果，將有問題的題目進行修改(實驗組與對照組職前教師，亦均不可以增刪題目)。最後，對所有職前教師進行 AKT 後測，以瞭解其評量知識最終情形。

2.5 資料蒐集與分析

AKT 主要用以瞭解職前教師的評量知識發展情形，本研究定義，「AKT 前測成績」為職前教師在評量知識方面的學習起點行為，「AKT 中測成績」為職前教師於本研究進行過程中評量知識方面的學習效益，而「AKT 後測成績」為職前教師的最終評量知識學習效益。AKT 量化資料分析，除進行內部一致性 Cronbach α 信度分析外，亦採用 SPSS PC 10.0 中文版進行重複量數(repeated measures)分析，探討實驗組與對照組在研究的前、中與後期，其評量知識的發展情形是否有顯著差異，分析過程中，將以職前教師的年齡作為共變量(covariate)，以扣除年齡因素在評量知識上的變異(variation)來源。

3. 研究結果

3.1 實驗組與對照組評量知識起點行為分析

本研究採用 AKT 來量測教師「評量知識」。研究進行前分別針對實驗組與對照組職前教師進行 AKT 前測，並將 AKT 前測成績進行共變數分析，分析過程以職前教師的年齡為共變量，「分組」作為固定因子(共兩個水準-實驗組與對照組)，用以考驗「分組」因子對 AKT 前測成績是否具有顯著意義。進行共變數分析前，先進行同質變異數假定(homogeneity of variance)檢驗，以瞭解分析資料是否符合共變數分析的基本假定，確定無誤之後，再進行正式分析。結果發現 Levene 統計檢定量均未達顯著，換言之，所有分析資料均未違反同質變異之基本假設($F=0.003, p=.957$)。接著再進行共變數分析，結果如表 1 所示：

表 1 AKT 前測成績之共變數分析摘要表(n=60)

變異來源	離均差平方和 (SS)	自由度 (df)	均方 (MS)	F 值	P 值
年齡	399.014	1	399.014	1.889	.175
分組	21.277	1	21.277	.101	.752
誤差	12039.528	57	211.220		
總合	165806.250	60			

由表 1 可知，職前教師年齡對於 AKT 前測成績的變異而言，無顯著意義($p>.05$)；而在以職前教師「年齡」為共變量的前提下，「分組」因子對於 AKT 前測成績的變異而言，亦無顯著意義($p>.05$)，由此可知，實驗組與對照組之評量知識之起點行為為無顯著差異。

3.2 實驗組與對照組評量知識發展情形

本研究定義「AKT 前測成績」為職前教師在評量知識方面的學習起點行為，「AKT 中測成績」為職前教師於本研究進行過程中評量知識方面的學習情形，而「AKT 後測成績」為職前教師的最終評量知識學習情形，AKT 的得分越高，代表職前教師之評量知識越好。

AKT 資料收集完成後，以 SPSS PC 10.0 中文版進行敘述統計分析，表 2 呈現實驗組與對照組職前教師的 AKT 前測、中測與後測成績之各項敘述統計值：

表 2 實驗組與對照組之 AKT 前測、中測、後測成績與信度資料

AKT	Cronbach α	分組 ^a	平均分數	標準差
前測	.993	對照組	51.833	13.957
		實驗組	49.250	15.301
中測	.976	對照組	55.783	9.924
		實驗組	61.200	12.491
後測	.984	對照組	62.917	12.125
		實驗組	76.083	12.520

^a 實驗組 30 人，對照組 30 人

接著採用重複量數分析，探討實驗組與對照組在 AKT 前測、中測與後測的成績上是否有顯著差異，分析過程中將以職前教師的年齡作為共變量，以扣除年齡因素在評量知識上的變異來源，分析結果如表 3 所示：

表 3 AKT 變異數分析摘要表(n=60)

變異來源	SS	DF	MS	F
受試者間				
年齡	129.976	1	129.976	1.102
分組	541.082	1	541.082	4.588*
誤差	6722.739	57	117.943	
受試者內				
AKT	75.031	2	37.515	.586
AKT*年齡	113.187	2	56.593	.811
AKT*分組	833.740	2	416.870	5.972**
誤差	7957.435	114	69.802	

*p<.05; **p<.01

由表 3 中，受試者間變異來源的數據發現，「分組」因子對整體 AKT 成績有顯著意義(F=4.588, p<.05)，因此，接著針對「分組」因子之兩個水準(Level)-實驗組與對照組，在整體 AKT 分數的差異上，進行事後比較分析，採用 Bonferroni 法分析的結果如表 4 所示，可以發現，實驗組的整體 AKT 成績顯著較對照組為佳(p<.05)。

表 4 實驗組與對照組 AKT 成績之事後比較分析

處理模式 ^a	AKT 平均數差異 ^b	標準誤
實驗組 - 對照組	7.677*	3.584

* p<.05

^a 實驗組 30 人，對照組 30 人

^b 多重比較調整：Bonferroni 法

此外，由表 3 中，受試者內變異來源的數據發現，所有職前教師的 AKT 前測、中測與後測成績

並無顯著差異，但是「分組」因子與 AKT 前測、中測與後測成績則有交互作用(F=5.972, p<.01)，因此，進行進一步分析，表 5 與圖 8 呈現各組 AKT 成績之交叉分析比較：

表 6 組別與 AKT 成績之交叉分析表

分組	AKT	平均數 ^a	標準誤
對照組 (n=30)	前測	49.781	3.045
	中測	55.176	2.378
	後測	62.062	2.594
實驗組 (n=30)	前測	51.303	3.045
	中測	61.807	2.378
	後測	76.938	2.594

^a 在模式中所顯示的共變量評估：年齡 = 22.00

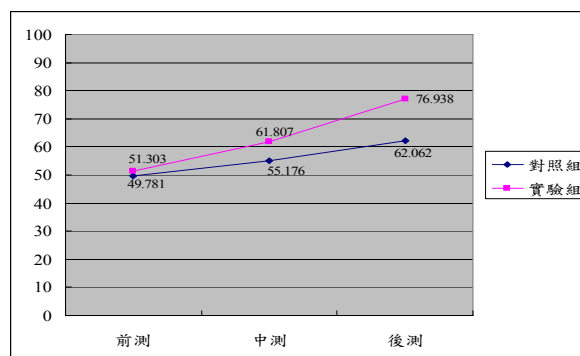


圖 8 「分組」因子與 AKT 成績之交叉分析圖

由圖 8 可以發現，實驗組與對照組在前測分數無明顯差異，但在中測分數與後測分數則有明顯偏離，由表 4 可以知道，此偏移達到顯著性(p<.05)。

由以上研究結果可以發現，實驗組與對照組職前教師的 AKT 成績，具有顯著差異，其詳細差異情形為：實驗開始階段，實驗組與對照組的 AKT 成績並無顯著差異，實驗處理(treatment)介入之後，實驗組的 AKT 成績顯著優於對照組，實驗處理介入結束後，實驗組的 AKT 成績顯著優於對照組的幅度又更大，因此，研究者推論，實驗處理介入，對於發展職前教師之評量知識具有顯著成效。

4. 結論與建議

相較於傳統培育模式(「P2R」評量素養培育模式)的職前教師，採用 WATA 系統進行評量素養培育(「P2R-WATA」評量素養培育模式)的職前教師，其「評量知識」之學習效益顯著較佳。上述研究結果可以由「P2R-WATA」評量素養培育模式的設計特色獲得解釋。本研究中的實驗組職前教師在整個訓練過程中，可以隨時使用 WATA 系統所提供之「完整評量流程鷹架(Triple-A Model)」、「電子化評量教材」、「個人化學習」與「情境化學習」等功能進行學習，系統可以隨著學習進度，依循個別的需求，隨時提供各種個別化的回饋和「電子化評量教材」給予實驗組職前教師。此外，實驗組職前教師也可以依循 WATA 系統「完整評量流程鷹架」進行命題與考試的「實作(Practicing)」，換言之，可

以在 WATA 系統的輔助下，在網路上進行試卷命題與執行考試，以及由真實的學生在網路上進行考試，考試完畢後，WATA 系統可以立刻協助實驗組職前教師，將學生考試結果進行試後分析，並且立刻回饋教育統計資訊給實驗組職前教師，讓實驗組職前教師利用這些資訊「反省(Reflecting)」自己命題的缺失，並找出學習缺失與教學缺失(施測國中學生的老師之教學缺失)。此外，實驗組職前教師可以依據自己過去所學和「電子化評量教材」的資訊，解析 WATA 系統提供的試後分析結果，針對自己的命題進行調整與「修改(Revising)」，然後藉由 WATA 系統再次施測，驗證自己的想法是否正確，以及瞭解自己的試題修改策略是否有效。

反觀對照組，雖然有同樣內容的評量教材，以及進行同樣的訓練過程(Practicing, Reflecting, Revising, P2R)，但是由於缺乏 WATA 系統的輔助，無法獲得「完整評量流程鷹架(Triple-A Model)」的支持，且在進行試後分析時也無法獲得 WATA 系統的輔助，使得對照組職前教師，在進行教育統計運算時，可能遭遇到較多阻力。上述的差異，可以用以解釋，實驗組職前教師評量知識的學習效益顯著較對照組為佳的原因。

本研究建議，本研究所建構之「P2R-WATA」評量素養培育模式，在未來宜持續進行長期的縱貫性研究，除了驗證「P2R-WATA」評量素養培育模式是否可以應用於在職教師與其他學科專長之教師外，也建議要探討經由該模式所訓練出來的教師，是否確實可以將其評量素養展現在實際課堂教學之中，用以提升學生學習效益。

5. 參考文獻

1. 王子華、王國華、王瑋龍和黃世傑。2004。網路評量系統融入師資培育之探究。資訊管理學報，11 卷，3 期，121-154。
2. 王子華。2005。建構網路評量系統發展生物科職前教師評量素養。國立彰化師範大學科學教育研究所博士論文，未出版，彰化縣。
3. Arter, J., "Learning Teams for Classroom Assessment Literacy." NASSP Bulletin, Vol. 85, No.621, pp.53-65, 2001.
4. Bransford, J. D., Brown, A., & Cocking, R., How people learn: mind, brain, experience and school, expanded edition. Washington, DC: National Academy Press, 1999.
5. Brookhart, S. M., The Standards and classroom assessment research. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Dallas, TX, 2001. (ERIC Document Reproduction Service No. ED451189)
6. Campbell, C., Murphy, J. A., & Holt, J. K., Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH, 2002.
7. Clarke, D., & Hollingsworth, H., "Elaborating a model of teacher professional growth." Teaching and Teacher Education, Vol.18, No.8, pp.913-1059, 2002.
8. Cognition and Technology Group at Vanderbilt (CTGV), "Anchored instruction and its relationship to situated cognition." Education Researcher, Vol.19, No.6, pp.2-10, 1990.
9. Cognition and Technology Group at Vanderbilt (CTGV), The Jasper project: Lesson in curriculum, instruction, assessment, and professional development. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, 1997.
10. Egbert, J., & Thomas, M., "The new frontier: A case study in applying instructional design for distance teacher education." Journal of Technology and Teacher Education, Vol.9, No.3, pp.391-405, 2001.
11. Fox, A., Change as a way of life on education. Tel Aviv: Charikover Publishers, 1995.
12. Gronlund, N. E., & Linn, R. L., Measurement and Evaluation in Teaching (6th ed.). New York, NY: Macmillan, 1990.
13. Hsu, Y. S., Liao, K. C., & Chen, C. C., A Web-Based Lesson with Situated Learning in Senior High School, 2001. (ERIC Document Reproduction Service No. ED466592)
14. Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K., Facilitating the development of assessment literacy. Educational Measurement: Issues and Practice, Vol.23, No.2, pp.26-32, 2004.
15. Maclellan, E., "Initial knowledge states about assessment: novice teachers' conceptualizations." Teaching and Teacher Education, No.20, pp.523-535, 2004.
16. McMillan, J. H., Basic Assessment Concept for Teachers and School Administrators. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA, 2000. (ERIC Document Reproduction Service No. ED447201)
17. McMillan, J. H., "Secondary Teachers' Classroom Assessment and Grading Practices." Educational Measurement: Issues and Practice, Vol.20, No.1, pp.20-32, 2001.
18. Mertler, C. A., "Secondary Teachers' Assessment Literacy: Does Classroom Experience Make a Difference?" American Secondary Education, Vol.32, No.3, pp.49-64, 2004.
19. Moore, M. G., & Kearsley, G., Distance Education: A Systems View. Belmont, CA: Wadsworth, 1996.
20. National Research Council (NRC), National Science Education Standards. Washington, DC: National Academy Press, 1996.
21. Passig, D., "Future online teachers' scaffolding: What kind of advanced technological innovations would teachers like to see in future distance training projects?" Journal of Technology and Teacher Education, Vol.9, No.4, pp.599-606, 2001.
22. Plake, B. S., & Impara, J. C., Teacher Assessment Literacy: Development of training modules, 1993. (ERIC Document Reproduction Service No. ED358131)
23. Putman, R., & Borko, H., "What do new views of knowledge and thinking have to say about research on teacher learning?" Educational Researcher, Vol.29, No.1, pp.4-15, 2000.
24. Stiggins, R. J., "Are you assessment literate?" The High School Journal, Vol.6, No.5, 20-23, 1999.
25. Stiggins, R. J., "Evaluating classroom assessment training in teacher education programs." Educational Measurement: Issues and Practice, Vol.18, No.1, 23-27, 1999.
26. Stiggins, R. J., "The Principal's leadership role in assessment." NASSP Bulletin, Vol.85, No.621, 13-26, 2001.
27. Stiggins, R. J., Student-involved Classroom Assessment (3rd ed.). Upper Saddle River, NJ: Merrill Prentice Hall, 2001.
28. Taylor, C. S., & Nolen, S. B., "What does the psychometrician's classroom look like? Reframing assessment concepts in the context of learning." Education Policy Analysis Archives, 4(17), 1996. Retrieved January 10, 2005, from the World Wide Web: <http://olam.ed.asu.edu/epaa/v4n17.html>
29. Wiske, M. S., Sick, M., & Wirsig, S., "New technologies to support teaching for understanding." International Journal of Educational Research, No. 35, pp.483-501, 2001.
30. Wang, T. H., Wang, K. H., Wang, W. L., Huang, S. C., & Chen, S. Y., "Web-based Assessment and Test Analyses (WATA) system: development and evaluation." Journal of Computer Assisted Learning, Vol.20, No.1, pp.59-71, 2004.