師範院校量化取向科學教育碩士論文之內容分析研究

許良榮[1] 李永進[2]
[1]台中教育大學 科學應用與推廣系 stevehsu@ms3.ntcu.edu.tw
[2]台中市永安國小 leeug@ms32.hinet.net

摘 要

本研究旨在針對國內師範院校量化取向之科學教育碩士論文進行內容分析，研究工具為自行發展之檢核表，分析 2001 年至 2003 年之碩士論文共 134 篇。研究結果發現：研究工具發展方式以研究者自行發展者為最多，但是有 14.2%的論文沒有說明研究工具之信度或效度之檢驗方式。信度檢驗方式以「Cronbach α」為最多數，效度檢驗方式則以「內容關聯的效度證據」最多。考驗方式以「t 考驗」被使用次數最多，且統計考驗之次數有過多的現象。論文的缺失分析顯示，除了「論文沒有電子全文於國家圖書館上網公開」有 88 篇（65.7%）為最多，其他尚有四項的缺失超過 50 篇次。

關鍵詞：師範院校、量化研究、科學教育

一、緒論

我國科學教育的研究，近十年來博碩士論文有漸增的趨勢，各師範院校陸續成立科學教育相關研究所，以及夜間與暑期教學碩士班如雨後春筍般的出現，大量的研究人員投入科學教育，使科學教育研究發展上有著極大的助力。蔡秉宸（2004）研究發現，目前國內已經有超過十所以上的科學教育相關研究所，每年約有兩百多名研究生力軍加入，此外再加上暑期與夜間進修的人數，每年新投身於科學教育研究領域的人數可達五百人以上。

學位論文可說是學術研究中相當嚴謹又具有系統的研究，其花費時機也較一般研究來得長，同時亦有多位教授進行審查，可說是學術研究中相當重要的進步動力來源，陳伯璋（2001）曾指出一個國家或社會學術研究水準大都可由此一窺全貌。然而，面對越來越多的人力投入於科學教育研究領域，科學教育類博碩士論文增加快速，但是針對我國量化科學教育論文之相關內容的分析研究卻付之闕如。因此，研究者的研究動機之一即是想透過內容分析的方法來探究科學教育學門中，量化研究論文之研究工具發展、信效度檢驗方式與統計考驗方式狀況，以了解近年來量化科學教育研究之研究工具與統計考驗方式的趨向。

本研究旨在針對近年來國內師範院校量化取向之科學教育碩士論文進行內容分析，研究問題為：(一)、台灣地區 2001 至 2003 年師範院校量化科學教育碩士論文研究工具之信、效度檢驗方式為何？(二)、上述量化科學教育碩士論文統計考驗方式之狀況為何？(三)、量化科學教育碩士論文中的缺失為何？

二、文獻探討

(一)、量化研究的特性

教育研究的方法一般而言可粗略地分為量化方法論（quantitative methodology）與質化方法論（qualitative methodology）（姜得勝，1997；洪仁進，2000；朱柔若，2000；Martyn，1992）。在量化與質化的研究中，存在著不同的學術傳統，也因而形成了相異的研究取向。量化研究是實徵研究（empirical investigation）的一種，旨在透過資料或經驗蒐集以獲得資訊、知識、理解現象（Black，1999）。學者對於量化研究的定義寬鬆不一，最簡單的定義是「處理數字」，凡所蒐集的資料為數字，所分析的資料為數字時，就是量化研究（Weiss，1998；Punch，1998），可以說量化研究的核心工作就是數數字（count）與區分程度（scaling）（Punch，1998）。

鄔昆如（1977）與徐宗林（1985）指出量化研究理論依據主要為經驗主義（empiricism）和實證主義（positivism）的觀點。李端明（2002）認為量化研究相信社會現象與物理現象一樣，具有客觀的實存，不因個人主觀的意念而改變，並假定社會現象量化的可能性。其運用科學的方法獲得研究資料，並假定資料來源具有客觀性。王麗雲（2005）則指出量化研究採取邏輯實證論的觀點，依據實證與經驗的累積，主張避免研究者過度介入，有幾分證據說幾分話，欲以價值中立、依據經驗--分析的法則，追求社會科學的普遍原理，是其具體的貢獻。

在量的研究中，研究問題著重在其意義性與重要性，並能具體的提出研究結果與建議。研究方法一般多標榜是科學的方法，然何謂科學方法？傳統上多以能控制、預測、重複…等視為科學方法的特質。李端明（2002）認為，在量的研究中，研究一問題是從現象界歸納出一套假設（hythoesis），再藉由實證的方式來推論其真實程度。但是如果將方法與問題異化，操弄研究方法、統計工具，卻未能深入問題的核心，如未去考慮統計方法的適宜性，工具的信效度…等，則將只是一種「術」，而非「學」。另一方面，拜電腦科技進步之神速，使得電腦的運算能力益加快速，所能處理資料的量也更加龐大，特別是統計套裝程式（如 SPSS、SAS）不斷推陳出新，更為研究者添加資料分析之利器。

(二)、國內外科學教育趨勢分析與量化研究問題之相關研究

蔡秉宸（2004）的研究旨在分析台灣地區 1990～2002 十二年的 488 份師範院校科學教育相關系所之學位論文。其研究發現有：（1）國內科教研究者在修業身份上有著顯著改變；在 1999 年之前國內科學教育研究者主要以一般生為主，但在 1999 年以後轉變為以在職生為主。（2）國內科教研究在數量上從 1999 年開始有明顯的成長，此成長主要是在於碩士論文的大量增加。（3）國內科教研究在主題上，1993 年到 2000 年之間，以教學為主要研究主題，但在 2001 年起學習（概念）已成為最主要的研究主題。（4）科教研究的研究取向當前仍是以實徵性的研究取向為主，而如此的研究取向對其研究設計、研究主題、樣本類型、研究設計、資料收集方式、研究工具均具有影

響。（5）國內與國際的科教研究發展有著一定的相似性，且發現國內研究有追隨國際發展的現象。

蔡今中（2002）的研究主要針對 SE、JRST、IJSE 進行分析 1998 年到 2002 年間的所發表論文共 798 篇，其中有 693 篇（86.3%）為實徵性研究。至於在主題上，在有關於學生概念學習的研究，近五年來仍是很熱門的研究主題之一，其比例有逐漸下降之趨勢，有關文化、性別議題之研究，也頗獲得科教學者關注。令人驚訝的是，有關於師資培育、教學及教學科技相關研究，數量上所佔的比例不多。

而 White（1997）的研究則是以 ERIC 中蒐錄的文章為研究對象，分析《Research in science education》期刊 1965 到 1995 年的論文，分析結果顯示，研究形式由 1975 年的實驗為主轉變為 1995 年的描述為主；在教學或學習模組上，由簡單非特定變為綜合複雜；研究設計則由複雜變為簡單；對研究的介入，則是從先決、詳細定義、控制、短期轉變為發展、模糊定義、引導、長期；測量方式則由測驗、簡單、表面變成面談、觀察、複雜、深入；統計方式由推論轉換成描述或不使用等等。

由上述三篇研究顯示科學教育研究對概念學習、學習情境、學習環境等議題的重視，但是並沒有研究是將研究問題聚焦於近年來台灣地區量化科學教育研究之內容分析。因此，本研究聚焦於量化科學教育研究其研究結果，相信研究結果應具有相當程度的參考價值。

三、研究方法

(一)、分析向度

本研究採用內容分析法，分析近年來師範院校科學教育相關系所之量化碩士論文，內容分析分為四大向度，包括研究工具信度檢驗方式、研究工具效度檢驗方式、統計考驗方式與量化研究的可能缺失。各分析細目，主要參考自林清山（1993）、郭生玉（1996）。「量化科學教育論文的缺失」則參考了林清山（1992）；許良榮（2003）；Collette & Chiappetta（1994）；Best & Kahn（1993）；Fisher（1993）；McComas & Almazroa（1998）等文獻，分析細目與編號請參閱附錄。

(二)、樣本論文選擇範圍與方法

資料收集主要透過國家圖書館立之「全國博碩士論文摘要檢索系統」，並以各師範院校圖書館藏論文相互檢視。本研究進行兩階段之過濾，第一階段將論文主題不是量化取向的論文去除，例如：以「質性」方法（個案研究、行動研究、詮釋性研究等）、純科學研究以及數學教育之論文去除。第二階段則去除摘要中提及以質性方法蒐集資料之論文。研究對象之蒐尋結果整理如表 1 所示。2001 年至 2003 年十一所師範院校之碩士論文總計有 954 篇，經過上述步驟篩選檢核，最後確定進行分析的量化取向科學教育碩士論文共 134 篇。

表 1　師範院校科學教育碩士論文篇數

	2001 年		2002 年		2003 年		合　計	
	總計	量化	總計	量化	總計	量化	總計	量化
1.台灣師大	31	8	23	4	14	2	68	14
2.彰化師大	32	1	27	0	31	2	90	3
3.高雄師大	2	0	9	0	8	2	19	2
4.國北師院	75	8	60	4	81	9	215	21
5.新竹師院	10	1	23	4	49	3	82	8
6.台中師院	49	9	22	5	44	8	115	22
7.嘉義大學	1	0	12	6	9	4	22	10
8.台南師院	25	0	29	11	17	2	71	13
9.屏東師院	45	4	41	7	18	4	104	15
10.花蓮師院	16	2	18	2	11	0	45	4
11.市北師院	32	4	42	9	49	9	123	22
總計	318	37	306	52	331	45	954	134

(三)、研究工具

本研究自行發展「研究工具發展檢核表」、「信效度檢核表」、「統計考驗方式檢核表」與「可能缺失」檢核表共四份。除了進行以進行文獻探討外，經過一位具科學教育博士之教授，一位進修科學教育碩士班五年教學年資之國小自然科教師的檢核後修訂。修訂後，研究者隨機選取十篇論文，邀請一位評分員（進修科學教育碩士班五年教學年資之國小自然科教師）進行獨立檢核，評分結果經過評分之等級轉換後，以肯德爾和諧係數分析，結果顯示各項工具之分析結果皆達顯著相關。此外，經過兩個星期，研究者再次針對相同十篇論文重新進行登錄，計算檢核結果後前後兩次之相關係數 .959（P< .01），顯示研究工具的檢核分析結果頗為穩定一致。

四、研究結果與討論

(一)、研究工具信效度檢驗之分析

分析的134篇論文中使用的量化研究工具共計268件，其中有190件研究工具(71.9%）是論文作者（共 120 篇）自行發展的研究工具。在信、效度檢驗方面，268 件研究工具中信效度檢驗皆具的有 230 件（85.8%），但是合計有 12 篇論文（8.9%）、38 件研究工具（14.2%）沒有進行信度或效度的檢驗（或是於論文中沒有說明）。

(二)、研究工具之信效度檢驗方式分析

信度的檢驗方式總計出現了 284 次，其中以「Cronbach α 係數」為最多（97 篇，166 次），其次為「再測信度」(37 篇，58 次)。值得注意的是，由 Cronbach α 與庫李信度（16 篇，17 次）相差之懸殊，顯示有不少作者對於二者之區分，有混淆不清的現象。表 3 的分析顯示，在 B10 的缺失（成就測驗的信度考驗方法敘述錯誤，將庫李信度誤認為 Cronbach α 信度）共有 38 篇論文。

研究工具效度檢驗方式總計有 367 次，平均每篇使用了 2.74 次的效度檢驗。其中以「內容效度證據」為最多（127 篇，275 次）。「效標關聯的效度證據」次之（28 篇，51 次）。研究者深入分析發現，採「專家效度」則共計有 181 次，但只有 22 篇論文（18.2％）說明學者專家的修改意見或學者專家的學經歷背景，這些資料可以說明專家的 credit，但是經常被論文作者忽略。

(三)、統計考驗方式之分析

統計結果如表 2，排序最高的「t 考驗」，共計 87 篇，3481 次，單篇論文使用 t 考驗最高達 422 次。其次「單因子變異數分析」共計 84 篇，2389 次，單篇論文使用最高達 288 次。居第三位的「皮爾遜積差相關」共計 71 篇，1913 次，單篇論文使用最多達 376 次。

表2　統計考驗方式統計結果

統計考驗方式	*篇數（%）*	*次數（%）*
t 考驗	87 (64.9)	3481 (35.7)
皮爾遜積差相關	71 (53.0)	1913 (19.6)
卡方考驗	22 (16.4)	883 (9.1)
單因子變異數分析	84 (62.7)	2389 (24.5)
二因子變異數分析	6 (4.5)	26 (0.3)
三因子變異數分析	8 (6.0)	177 (1.8)
單因子共變數分析	29 (21.6)	287 (2.9)
二因子共變數分析	5 (3.7)	60 (0.6)
迴歸分析	22 (16.4)	100 (1.0)
列聯相關	7 (5.2)	77 (0.8)
其他方式	16 (11.9)	350 (3.6)
總計	357	9743

部分論文使用了數以百計的統計考驗，其中共有 58 篇（43.3%）論文使用了 50 次以上的統計考驗，而超過 100 次的有 24 篇（17.9%），甚至有一篇論文共出現了 949 次統計考驗，相當令人驚訝。

(四)、論文之缺失分析

本研究分析各論文之缺失如表 3 所示，分別說明如下。

1.一致性缺失

「A01」是統計考驗之第一類型錯誤概率（α 值）設定不一致，共有 33 篇（24.6％）。林清山（1992），在其著作《心理與教育統計學》中，曾說明在進行假設考驗時第一類型錯誤之概率（α 值），若設定不一致是不切實際且不夠嚴謹的，顯著水準一旦在實驗觀察前決定好，就不宜再變更（p.236）。不少論文對於第一類型錯誤概率（α 值）通常以 SPSS 軟體所預設 .05 為臨界值，但是考驗結果若達到 .01 或更小的顯著水準時，就會呈現更小的顯著水準。此種呈現或許能解釋為：目的在表達各項統計考驗的最低 α 值，以提供參考。準此而言，此種呈現不宜稱之為「缺失」，但是從另一角度而言，此種變動臨界值的現象，是否也反映了論文作者不重視「錯誤率觀念單位」的問題？對照每篇樣本論文平均使用了近 73 次的統計考驗的數據，以及使用超過了 100 次統計考驗的有 24 篇（17.9%），本研究有理由認為，樣本論文變動臨界值的現象，缺乏理性分析的基礎。

表3　論文之缺失檢核結果統計表

編號	篇數	(%)	編號	篇數	(%)	編號	篇數	(%)	編號	篇數	(%)	編號	篇數	(%)
A01	33	24.6	**B01**	1	0.7	**C01**	54	40.3	**D01**	68	47.0	**E01**	88	65.7
A02	11	8.2	**B02**	54	40.3	**C02**	15	11.2	**D02**	41	30.6	**E02**	49	36.6
A03	28	20.9	**B03**	38	28.4	**C03**	4	3.0	**D03**	58	43.3	**E03**	7	5.2
A04	13	9.7	**B04**	10	7.5	**C04**	4	3.0	**D04**	3	2.2			
A05	3	2.2	**B05**	2	1.5	**C05**	25	18.7	**D05**	30	22.4			
			B06	12	9.0	**C06**	8	6.0	**D06**	8	6.0			
			B07	12	9.0	**C07**	0	0.0	**D07**	23	17.2			
			B08	2	1.5				**D08**	10	7.5			
			B09	2	1.5				**D09**	7	5.2			
			B10	38	28.4				**D10**	15	11.2			
			B11	16	11.9									
			B12	4	3.0									
			B13	9	6.7									

註：細格中數據以「篇」為單位，未計算每篇出現次數。

2.精確性缺失

最常見的是（B02）「忽略分量表第一類型錯誤的累積問題」，共 54 篇（40.3%），此現象為論文作者對於分量表以及總量表皆以相同的顯著水準（α 值）進行考驗。根據 Keppel（1991）指出分量表可能犯的第一類型錯誤機率之累積為「$\alpha_{EW}=1-(1-\alpha)^n$」（n 為比較次數）；例如進行八個分量表的顯著性分析，且都以 α= .05 為臨界值，則考驗結果的八個結論中，其中一個會發生 Type I error 的機率為 33.7%。（B03）「使用”顯著差異”之描述，卻未進行統計考驗」的缺失共 38 篇（28.4%）。本項問題出現於論文作者直接以結果數據進行直覺的比較，例如經過調查發現 A 特徵的為 12.1%，B 特徵的為 25.5%，描述為「B 明顯多於 A」或「B 顯著多於 A」，此種宣稱缺乏客觀基礎。「明顯」的語意過於模糊，沒有客觀分界線，而「顯著」則屬於約定成俗的統計用語，如果未進行統計考驗，不應使用這些語詞。

3.客觀性缺失

（C01）「論文內容中沒有說明統計數據就使用主觀敘述」之缺失共出現 54 篇（40.3%），例如論文作者使用了例如「大多數」、「絕大多數」等等缺乏客觀性之用語，此類程度用詞，其程度上的分界標準模糊不清，應避免直接使用，因為每個人對”絕大多數”、”很多”的界定都不一樣（試問其界線是 80%、90%或 99%？），既然有統計數據，應客觀描述統計數據，不應再冠上個人主觀認定的形容詞。數據的大小意涵與其來源有關，其價值沒有絕對性的程度。例如 2%的數據，如果是「全班 2%學生對重力有迷思概念」，此 2%我們可能認為不多。但是「全班表示搶劫殺人無罪的有 2%」，此 2%可能就值得我們重視了。因此對於數據的解讀，不宜由研究者主觀的直覺宣稱，必須經過理性論證或是統計考驗。

將統計考驗結果「沒有顯著差異」解釋為「一致」（編號 C05）共出現 25 篇（18.7

％）。統計考驗的結果「沒有顯著差異」是指第一類型錯誤高於 5%，研究者必須接受「虛無假說」，而非指比較結果「沒有差異」或「一致（相同）」。居第三位的是編號（C02）「對態度量表所得結果，使用優劣、好壞之主觀評判」。王文科（2002）指出「量表（scales or inventories）與測驗（test）不同。後者如成就測驗係用來測量個人之最佳或最大的表現，前者如態度量表則用於測量典型的表現（typical performance）。此種量表乃是分派於受試者、物體或行為的數值，旨在測量其素質，並予以量化，它不在標示成敗或優劣，只在評量個人擁有其感興趣之特徵的程度，如接納或拒絕、喜歡或不喜歡（p.177）」。由此可知態度量表並不是使用於評判優劣、好壞，而是研究對象之態度較為「正向或負向」。

4.完備性缺失

出現最多篇數的是編號（D01）「在進行變異數分析時沒有先進行同質性考驗」（共 63 篇，47.0％），，林清山（1992）在其著作《心理與教育統計學》（p.333）中指出；進行變異數分析時須滿足各組變異數同質性的假定，如果違反此一假定。本研究發現有 47.0％的樣本論文並沒有說明是否進行了「變異數同質性考驗」。其他出現比率較高的「完備性」缺失，包括（D03）「沒有說明抽樣誤差（Sampling error）」（43.3％）與（D02）「沒有說明抽樣的原則或依據」（30.6％）。在各式機率抽樣中，由抽樣而產生的誤差，主要是由抽樣的隨機性所產生，此種誤差是一種隨機性的誤差。在進行機率抽樣時，應設定母群體平均數上下各多少百分點作為估計的容忍誤差（tolerated error），因為調查者事先設定容忍範圍，用以交待樣本估計值來推論母群體實際值大小的信賴水準（confidence level）是有其必要的。研究分析發現 65 篇（48.5%）採用機率抽樣方式的論文，只有 7 篇有計算並說明抽樣誤差。

5.公開性缺失

本研究進行論文資料蒐集工作時，分析的 134 篇論文只有 46 篇（34.3％）在「全國博碩士論文資訊網」有電子全文提供下載。本研究逐本詳細查驗樣本論文，發現除了 2 篇之外，每位論文作者都在論文授權同意書上勾選同意聲明。研究者與國家圖書館相關工作人員電訪後，獲知某些論文作者雖然有簽署同意書授權，但是並沒有上傳全文電子檔，而國家圖書館現目前並無人力可掃描全國博碩士論文之紙本全文，因而導致部分論文僅有摘要資料，而無電子全文可供下載的情形。按國家圖書館之網頁說明是由各校應屆畢業生，透過各研究所之行政人員與國家圖書館聯繫，取得線上建檔之帳號之後，才可自行上網建檔。此步驟讓本研究省思，若研究者同意授權公開，如果各系所能要求並協助畢業生上傳論文，此種電子全文缺乏的情形應可大大改善，也才能有效達成科學研究的公開性要求並促進學術交流。

另一方面，(E02)「論文沒有電子全文於所屬學校圖書館上網公開」也有 49 篇（36.6%），顯示我們需要再加強研究生在辦理離校時，是否確實完成了論文的公開性程序。

五、結語

本研究分析了 2001 年至 2003 年量化取向之科學教育碩士論文後，發現有不少值得我們注意與思考的問題，這些問題大多屬於科學性研究或統計考驗的基本理念，而非艱深的理論，只要仔細閱讀研究法、教育統計的教科書，應該都可以避免。而某些論文有使用過多統計考驗的傾向，建議量化研究者應在解答研究問題之後，即可停止統計考驗，以避免第一類型錯誤率累積過高的問題，畢竟科學性研究應首重減少犯錯率，而非結論的多寡。

參考文獻（略）

附錄　各項缺失編號

（一）、一致性缺失：包括 A01：「統計考驗之第一類型錯誤概率（α 值）設定不一致」。A02：「P 值寫法不統一；例如出現 P＜.05 以及 P＜0.05」。A03：「數據的有效位數沒有一致」。A04：「施測樣本數與總和樣本數（自由度總和）不一致」。A05：「其它一致性缺失」。

（二）、精確性缺失：包括 B01：「變量種類區分錯誤，例如誤將“名義量尺”定義為“次序量尺”」。B02：「分量表的第一類型錯誤（α 值）產生累積」。 B03：「使用“顯著差異”之描述，但未進行統計考驗」。 B04：「對考驗結果之解釋非為二分法，例如敘述上稱其“並不那麼顯著”」。 B05：「相關係數之 r 值，未考慮樣本數（N）的多寡，而比較其大小」。 B06：「信度考驗方法不當，例如成就測驗使用重測信度、題目數過少而使用折半信度」。 B07：「使用不適切的考驗方法，例如 t-test、單因子變異數分析、多因子變異數分析的使用時機不當」。B08：「對於信、效度的區別不當；例如“可信度”之敘述」。B09：「使用卡方檢定理論次數太小時，沒有進行相關校正」。B10：「成就測驗的信度考驗方法敘述錯誤，將庫李信度誤認為 Cronbach α 信度」。B11：「假說之考驗敘述錯誤（統計考驗是拒絕虛無假說而非證實研究假說）」。B12：「自變項只有兩組（level），變異數分析之後又進行沒有必要的事後比較」。B13：「其它精確性缺失」。

（三）、客觀性缺失：包括 C01：「使用主觀敘述，例如：大多數、絕大多數等」。C02：「對態度量表所得結果，使用優劣、好壞之主觀評判」。C03：「對統計結果之相關分析，解釋為因果關係」。C04：「使用“證實”語詞描述“拒絕虛無假說”」。C05：「將統計考驗結果未達顯著差異解釋為“一致”或“相同”」。C06：「不認同統計考驗的結果，而主觀敘述差異性」。C07：「其它客觀性缺失」。

（四）、完備性缺失：包括 D01：「在進行變異數分析時沒有先進行同質性考驗」。D02：「沒有說明抽樣的原則或依據」。 D03：「沒有說明抽樣誤差」。 D04：「沒有進行事後比較但有比較結論」。D05：「研究工具沒有進行信度考驗」。D06：「使用卡方檢定時未進行事後比較而提出不同組別之間的差異」。D07：「問卷所蒐集之自變項資料沒有完全與依變項進行統計考驗」。D08：「進行共異數分析前沒有進行迴歸同質性考驗」。D09：「進行事後比較但未說明採取何種方式」。D10：「其它完備性之缺失」。

（五）、公開性缺失：包括 E01：「論文沒有電子全文於國家圖書館上網公開」。E02：「論文沒有電子全文於所屬學校圖書館上網公開」。E03：「其它公開性之缺失」。