

國立臺灣師範大學
音樂學系碩士班音樂教育組
碩士論文

臺北市國民中學音樂教師教學評量素養
與教學評量實施之調查

**A Survey on the Assessment Literacy and Implementation in
Music Classes -Taken Junior High School Music Teachers in
Taipei City as Example**

研究生：黃貞瑜
指導教授：吳舜文 博士
中華民國 104 年 6 月

誌謝

回想起這本論文從構思主題、文獻整理、內容編寫、論文口試至不斷地來回修改，這從無到有的過程，除了學到許多知識外，要感謝的人實在是太多了！

首先，非常謝謝我的指導教授吳舜文老師，每次我忐忑不安地拿著修正的檔案去找您時，您總是像媽媽一樣，耐心地指出方向與修改的地方並同時給我鼓勵，讓我燃起了希望，繼續面對論文，這本論文最終能順利完成，我對老師有滿滿的感謝！

此外，感謝幫我進行專家效度及口試的老師們，以及幫我發問卷及填寫問卷的同學們，有您們的寶貴意見，讓我獲得了鼓勵及正向能量。同時感謝自師大研究所時曾經教導我的老師：賴美鈴教授、陳曉霽教師、潘宇文教授及莊惠君教授，在你們身上我學到了許多知識及對學問認真的精神。

也要感謝萬里國小的一群好同事們，有你們的陪伴跟幫忙，總能讓我在寫作論文的過程中得到歡笑。僑善國小的杜主任，你是我每次無助時最大的救星，常常解答我所有的疑問。

最後，要謝謝我的爸爸媽媽，你們關心我的研究所生活跟論文進度，有了你們的支持，我才能無後顧之憂的去當研究生。

有那麼多人的幫忙跟鼓勵，在此致上我最大的感激，謝謝你們！

貞瑜

2015. 07. 26

摘要

本研究旨在調查臺北市國民中學音樂教師教學評量素養與教學評量實施情形，同時分析其背景變項之差異，最後探討兩者之間是否具有相關。本研究採用「調查研究法」，以臺北市國民中學音樂教師為研究對象，設計「國民中學音樂教師教學評量素養與教學評量實施調查問卷」，內容改編「評量素養調查問卷」(Campbell & Mertler, 2005)與自行編製「國中音樂教學評量實施情形之調查問卷」作為研究工具，發出 120 份正式問卷，共回收 112 份，問卷回收率達 93%，有效問卷為 110 份。本研究之主要發現如下：

- 壹、臺北市國民中學音樂教師教學評量素養於七項標準中，共六項達判定標準值，僅「標準五」未達，多數教師具備了此次所欲評估之教學評量素養知能。
- 貳、就教學評量素養整體層面而言，在教師性別、學歷、工作職務、參加相關研習以及其任教學校規模等變項上無顯著差異；但在年齡與年資部分，30 歲以下之音樂教師優於年齡介於 41-50 歲之音樂教師，年資 5 年以下之音樂教師優於年資 11-15 年之音樂教師。
- 參、音樂教師教學評量實施整體情形及其分項之「實施教學評量前之預備情形」、「實施教學評量時之情形」、「實施教學評量後之結果運用情形」層面良好。
- 肆、不同性別、年齡、學歷、學科專長、年資以及學校規模之音樂教師在教學評量實施整體層面皆無顯著差異，但在工作職務與參加相關研習部分，兼任行政工作之音樂教師優於擔任班級導師之音樂教師，曾參與教育測驗與評量相關研習之音樂教師優於不曾參與過教育測驗與評量相關研習之音樂教師。
- 伍、音樂教師之教學評量素養及其七項標準分層面與教學評量實施整體層面皆未達顯著相關。

最後，根據研究結果，對教育行政單位、國中音樂教師及後續研究提出建議。

關鍵詞：教學評量素養、教學評量實施

Abstract

The purpose of this study was to investigate the teaching assessment literacy and music teaching implementation of the junior high school music teachers. In addition, the study tried to explore the influence of music teacher's background variables and if the relationship between them. The survey method was utilized in this study. The subjects were the junior high school music teachers in Taipei City. The instrument was revised Assessment Literacy Inventory (Campbell & Mertler, 2005) and self-designed questionnaires. The survey was sent out 120 questionnaires and 112 copies were returned. The recover rate was 93%. There were 110 effective questionnaires. The major findings of this study were as follows:

1. Music teachers' teaching assessment literacy in terms of standard 1.2.3.4.6 and 7 were over 60% of the standard valuation. Most teachers had the teaching assessment literacy of this evaluation.
2. In overall teaching assessment literacy results, there were no significant differences in the variable of teachers' backgrounds, which involved genders, educational backgrounds, positions. Experience of participating related studies, and the scale of school. However, the ages and experience of teaching that under 30-year were better than 41-50 year. Teachers who had less than 5 years of teaching experience are better than teachers who had taught more than 11-15 years.
3. The plane of the music teachers' implementation between holistic and discrete was good.
4. There were no significant differences in the variable of music teachers' backgrounds, which involved genders, ages, educational backgrounds, subject expertise, years of teaching and the scale of schools. However, the positions and experience of participating related studies, the music teachers who had administrative experience were better than being the homeroom teachers. The music teachers who had experience of participating related studies were better than who had none.
5. The music teachers' teaching assessment literacy and the seven planes of standards and implementation were no significant difference.

Finally, based on the research findings above, the researcher proposed relevant suggestions to the education administration, junior high school music teachers and future studies.

Keywords: teaching assessment literacy, teaching assessment implementation

目次

目次	i
表次	ii
圖次	iv
第一章 緒論	1
第一節 研究動機	1
第二節 研究目的與待答問題	4
第三節 名詞解釋	6
第四節 研究範圍與限制	8
第二章 文獻探討	9
第一節 教學評量的理論基礎	9
第二節 音樂教學評量的理論基礎	27
第三節 教師評量素養的意涵	41
第三章 研究方法	49
第一節 研究架構	49
第二節 研究對象	51
第三節 研究工具	53
第四節 實施程序	63
第五節 資料處理	64
第四章 結果與討論	65
第一節 臺北市國民中學音樂教師教學評量素養現況與差異分析	65
第二節 臺北市國民中學音樂教學評量實施現況與差異分析	97
第三節 臺北市國民中學音樂教師教學評量素養與實施相關分析	120
第五章 結論與建議	123
第一節 結論	123
第二節 建議	129
參考文獻	132
附錄	138
附錄一 評量素養問卷(Assessment Literacy Inventory , ALI)原文	138
附錄二 專家效度問卷修正意見彙整表	145
附錄三 國民中學音樂教師教學評量素養與教學評量實施調查問卷	161

表次

表 2-1-1 《國民中學及國民中學學生成績評量準則》資料整理.....	22
表 2-1-2 教學評量相關之論文一覽表.....	26
表 2-2-1 音樂概念統整說明	31
表 2-2-2 表現等級描述.....	32
表 2-2-3 內容標準與九年一貫課程網要能力指標對應表	34
表 2-2-4 音樂教學評量相關之論文一覽表	37
表 2-3-1 STCEAS 七項標準之內容.....	45
表 3-2-1 正式問卷總人數統計表	52
表 3-3-1 「國小教師評量素養問卷」情境題之前後對照表（情境一）	54
表 3-3-2 「國小教師評量素養問卷」情境題之前後對照表（情境二）	54
表 3-3-3 「國小教師評量素養問卷」情境題之前後對照表（情境三）	54
表 3-3-4 教師評量素養各標準的 ALI 題目	56
表 3-3-5 教師評量素養各標準描述統計.....	56
表 3-3-6 教師評量素養難度分析一	57
表 3-3-7 教師評量素養難度分析二.....	57
表 3-3-8 教師評量素養標準一次數分配.....	58
表 3-3-9 教師評量素養標準二次數分配.....	58
表 3-3-10 教師評量素養標準三次數分配	59
表 3-3-11 教師評量素養標準四次數分配.....	59
表 3-3-12 教師評量素養標準五次數分配	60
表 3-3-13 教師評量素養標準六次數分配	60
表 3-3-14 教師評量素養標準七次數分配	61
表 3-3-15 教師評量素養問卷總分次數分配	61
表 3-3-16 教師評量實施情形量表信度分析摘要表	62
表 4-1-1 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準一	67
表 4-1-2 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準二.....	69
表 4-1-3 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準三.....	71
表 4-1-4 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準四	73
表 4-1-5 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準五	75
表 4-1-6 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準六	76

表 4-1-7	臺北市國民中學音樂教師教學評量素養現況分析摘要-標準七	78
表 4-1-8	不同性別音樂教師在教學評量素養量表的獨立樣本 t 考驗.....	81
表 4-1-9	不同年齡音樂教師在教學評量素養之各組人數、平均數和信賴區間	83
表 4-1-10	不同年齡音樂教師之教學評量素養單因子變異數分析.....	83
表 4-1-11	不同學歷音樂教師在教學評量素養量表的獨立樣本 t 考驗.....	85
表 4-1-12	不同學科專長音樂教師在教學評量素養量表的獨立樣本 t 考驗	87
表 4-1-13	不同修習教育測驗與評量課程經驗音樂教師在教學評量素養量表獨立 樣本 t 考驗.....	88
表 4-1-14	不同年資音樂教師在教學評量素養之各組人數、平均數和信賴區間.....	89
表 4-1-15	不同年資音樂教師之教學評量素養單因子變異數分析.....	90
表 4-1-16	不同工作職務音樂教師在教學評量素養之各組人數、平均數和信賴區間.....	92
表 4-1-17	不同工作職務音樂教師之教學評量素養單因子變異數分析	92
表 4-1-18	不同參與教育測驗與評量研習經驗音樂教師在教學評量素養量表獨立 樣本 t 考驗.....	94
表 4-1-19	不同學校規模音樂教師在教學評量素養之各組人數、平均數和信賴區間.....	95
表 4-1-20	不同學校規模音樂教師之教學評量素養單因子變異數分析	95
表 4-1-21	不同教師背景變項與教師評量素養七個標準之差異情形綜合分析	96
表 4-2-1	臺北市國民中學音樂教師教學評量實施現況分析摘要	97
表 4-2-2	臺北市國民中學音樂教師教學評量量表題項填答分析層面一	99
表 4-2-3	臺北市國民中學音樂教師教學評量量表題項填答分析層面二	101
表 4-2-4	臺北市國民中學音樂教師教學評量量表題項填答分析層面三	106
表 4-2-5	不同性別音樂教師在教學評量實施量表的獨立樣本 t 考驗.....	109
表 4-2-6	不同年齡音樂教師在教學評量實施之各組人數、平均數和信賴區間.....	110
表 4-2-7	不同年齡音樂教師之教學評量實施單因子變異數分析.....	110
表 4-2-8	不同學歷音樂教師在教學評量實施量表的獨立樣本 t 考驗.....	111
表 4-2-9	不同學科專長音樂教師在教學評量實施量表的獨立樣本 t 考驗	112
表 4-2-10	不同修習教育測驗與評量課程經驗音樂教師在教學評量實施量表獨立 樣本 t 考驗.....	112
表 4-2-11	不同年資音樂教師在教學評量實施之各組人數、平均數和信賴區間.....	114
表 4-2-12	不同年資音樂教師之教學評量實施單因子變異數分析.....	114
表 4-2-13	不同工作職務音樂教師在教學評量實施之各組人數、平均數和信賴區間.....	115
表 4-2-14	不同工作職務音樂教師之教學評量實施單因子變異數分析	116

表 4-2-15 不同參與教育測驗與評量研習經驗音樂教師在教學評量實施量表獨立 樣本 t 考驗.....	118
表 4-2-16 不同學校規模音樂教師在教學評量實施之各組人數、平均數和信賴區間.....	119
表 4-2-17 不同學校規模音樂教師之教學評量實施單因子變異數分析	119
表 4-3-1 臺北市國民中學音樂教師教學評量素養與教學評量實施之積差相關分析....	120

圖次

圖 2-1-1 教學的基本模式(Glaser , 1962).....	11
圖 3-1-1 研究架構.....	50
圖 3-4-1 研究步驟流程圖	63



第一章 緒論

本章共分四節，以下就研究動機、研究目的與待答問題、名詞解釋及研究範圍與限制等部分加以說明。

第一節 研究動機

教學評量為教學歷程中極重要的一環，具有引導與總結之功能（Linn & Gronlund, 2000）。教學評量是指教師在教學歷程中，持續蒐集學生的學習資料，加以選擇、分析、組織，之後將測驗分數或其他資料加以轉換，並根據學生個人潛能與特質等因素，對學生學習成果作價值判斷（紀雅真，2007）。在教學現場中，不論任何科目，教師進行教學中，都會選擇適當時機使用評量來瞭解學生的學習情況，教學評量不僅只是打分數或等第，它提供音樂教師教學所需的各種資料，作為瞭解學生的學習狀況並針對學生的個別差異給予補救教學，同時幫助學生清楚瞭解各階段的學習重點。因此，教學評量在教師的整個教學中，扮演了舉足輕重的角色。

從一個不斷循環的教學歷程來看，「教學活動」與「教學評量」是密不可分的關係，教學為知識訊息的傳遞，而評量是在教學歷程中提供回饋並扮演仲介的重要角色（劉向欣，2008）。教師在進行評量前應充分瞭解教學目標、學習者的先備能力、起點行為，並隨時依據學生的學習情況調整教學及評量方式，以達到設定的教學目標。因此，當教學與評量兩者可密切結合時，不僅教師與學生都能得到良好的回饋，同時讓教學變成一個完整的過程。

近年來，我國在教育方面有極大的變革，西元 2002 年於國中階段實施的九年一貫課程，衍生出許多的相關議題。以藝術與人文學習領域來說，課程綱要（教育部，2008）中指出：「本領域教學評量可併用『量』與『質』的評量，且可視

教學目標、教學範圍、教學方法、教學流程之需要，採取教師評量、學生互評、學生自評等方式，並應用：觀察、問答、晤談、問卷調查、軼事記錄、測驗、自陳法、評定量表、檢核表、討論...等方式評量，且可酌採相對解釋法與自我比較法等彈性評量措施。」由此可看出，現今的教學評量，教師擁有更多的自主性，且評量方式有多元化的趨勢。但國內的音樂教學評量，大多仍是以實作評量及紙筆測驗方式進行，最普遍的就是直笛吹奏、樂理考試，兼顧音樂技能與知識的表現，但除了關心學生的唱奏能力之外，更需要多元的方式來檢測教學成效（謝苑玫，2000）。因此，教育部在西元 2008 年大幅修正藝術與人文課程綱要之教學評量部分。在評量原則方面，應依據能力指標及教材內容，採用多元評量的方式，並兼顧形成性與總結性評量，可採實作評量、動態評量、真實評量、檔案評量等方式進行，以呈現學生多元的學習表現。另一方面，為配合「十二年國民基本教育」（教育部，2007）政策之規劃與實施，教育部修正發布了「國民中學及國民中學學生成績評量準則」，將學生學習評量方法、成績的訂定有了修正，以回應各界對國中小學生學習評量的關注。從上述政策的修訂與修正，可發現政府對評量方面的推動。

然而，「藝術與人文」領域不同於其他領域，它提供非語文的溝通形式，提供人們的直覺、推理、聯想與想像的創造思考能力，使人們分享源自生活的思想與情感，並從中獲得知識，建立價值觀（教育部，2008）。正因為藝術有無法量化的特性，如果只以傳統單一的紙筆測驗，是無法得到學生完整的學習歷程、學習態度與學習成果，音樂評量比其他學科評量更加難以達到客觀性與整體性，音樂教師必須配合不同的教學類別，如歌唱、演奏、創作、音樂欣賞等，選擇適合的評量方式來瞭解學生真實的學習成果。所以，音樂教師如果希望所進行的評量能有良好的成效，教師本身在評量方面的素養就顯得格外重要。

美國教師聯盟（American Federation of Teachers, AFT）、全國教育測量會議（National Council on Measurement in Education, NCME）與全國教育協會（National Education Association, NEA）於 1990 發展的「學生學習評量之師資能

力標準」(Standards for Teacher Competence in Educational Assessment of Students, STCEAS)，提到教師評量素養的內涵，包括：(1) 選擇適合於教學決策的評量方法(2) 發展適合於教學決策的評量方法(3) 執行、分析、評分與詮釋外界發展或教師自製之評量的結果(4) 針對個別學生進行選擇、規劃教學、發展課程以及學校發展時去使用評量能力的結果(5) 去發展學生評量之有效評分步驟(6) 對於學生、家長、一般民眾與教育工作者傳遞評量結果(7) 分辨、不道德、不合法或其他不適當之評量方法與評量資訊的使用，共七項標準。從上述可知，教師評量素養能力涵蓋許多層面以及其重要性。因此，音樂教師是否具備足夠的能力執行教學評量，是研究者想探究的方向之一。

從 2007 年至 2013 年間的相關研究中，以博碩士論文為例，國小音樂評量為主題的研究仍居多，包括：直笛教學評量、歌唱教學評量、教學評量現況及音樂教學評量觀點之調查。而國中音樂評量的相關研究有 3 篇，分別為紀雅真(2007)「國民中學音樂教學評量實施現況之調查研究」、黃于芬(2007)「臺北縣國民中學藝術與人文領域音樂教學多元評量之現況調查研究」與廖怡君(2007)「評分規準於國中直笛教學之個案研究」。另外在相關期刊文獻發表中，與音樂評量相關之議題並不多，如學者潘宇文(2009)於其著作「臺灣北部地區中小學歌唱教學評量實施狀況之調查研究」中，探討臺北市、臺北縣、基隆市的國民小學、中學及高級中學之音樂教師，進行音樂課堂歌唱教學評量的情形；以及學者黃珮瑄、莊敏仁(2012)於其著作「國小音樂教師運用教學檔案夾提升音樂學習評量專業知能之個案研究」中，探究音樂學習評量檔案夾的實際執行效益。

綜上所述可知，國內針對國中階段教師音樂教學評量相關議題所進行的實證研究篇數較少，甚至針對國中階段音樂教師教學評量素養內涵進行探討的研究亦不多見。有鑑於音樂教學評量實施對學生學習成效有其重要影響，研究者選擇以教育發展核心的臺北市國中音樂教師為研究對象，除探究其音樂教學評量素養與教學評量實施之現況外，亦進一步分析其相關及各項差異情形，並作成相關結論及建議以作為教育行政單位、國民中學音樂教師與後續研究者之參考。

第二節 研究目的與待答問題

壹、研究目的

本研究旨在探討國民中學音樂教師教學評量素養與教學評量實施之情形，同時分析其背景變項之差異，最後探討兩者之間是否具有相關。根據上述的研究動機，本研究目的如下：

- 一、瞭解臺北市國民中學音樂教師教學評量素養之現況。
- 二、探討不同背景變項臺北市國民中學音樂教師教學評量素養之差異情形。
- 三、瞭解臺北市國民中學音樂教師教學評量實施之現況。
- 四、探討不同背景臺北市國民中學音樂教師教學評量實施之差異情形。
- 五、探討臺北市國民中學音樂教師教學評量素養與教學評量實施之相關。

貳、待答問題

基於上述研究目的，本研究待答問題如下：

- 一、臺北市國民中學音樂教師教學評量素養之現況為何？
 - (一) 教師於「選擇適合於教學決策的評量方法」之素養表現為何？
 - (二) 教師於「發展適合於教學決策的評量方法」之素養表現為何？
 - (三) 教師於「執行、分析、評分與詮釋外界發展或教師自製之評量的結果」之素養表現為何？
 - (四) 教師於「針對個別學生進行選擇、規劃教學、發展課程以及學校發展時去使用評量能力的結果」之素養表現為何？
 - (五) 教師於「發展學生評量之有效評分步驟」之素養表現為何？
 - (六) 教師於「對於學生、家長、一般民眾與教育工作者傳達評量的結果」之素養表現為何？

(七) 教師於「分辨、不道德、不合法或其他不適當之評量方法與評量資訊的使用」之素養表現為何？

(八) 不同性別、年齡、學歷狀況、教職狀況之教師於教學評量素養上是否具有差異？

二、瞭解臺北市國民中學音樂教師教學評量實施之現況。

(一) 教師教學評量實施之現況為何？

(二) 教師於「實施教學評量前之預備情形」的成效為何？

(三) 教師於「實施教學評量時之情形」的成效為何？

(四) 教師於「實施教學評量後之結果運用情形」的成效為何？

(五) 不同性別、年齡、學歷狀況、教職狀況之教師於教學評量實施現況上是否具有差異？

三、探討臺北市國民中學音樂教師教學評量素養與教學評量實施之相關。

(一) 教師評量素養與教學評量實施之間有何相關？

(二) 教師評量素養與實施教學評量前之預備情形有何相關？

(三) 教師評量素養與實施教學評量時之情形有何相關？

(四) 教師評量素養與實施教學評量後之結果運用情形有何相關？

第三節 名詞解釋

本節針對研究中的重要名詞，包括教學評量素養、教學評量實施加以說明。

壹、教學評量素養（teaching assessment literacy）

美國教師聯盟（American Federation of Teachers, AFT）、全國教育測量會議（National Council on Measurement in Education, NCME）與全國教育協會（National Education Association, NEA）於 1990 發展的「學生學習評量之師資能力標準」（Standards for Teacher Competence in Educational Assessment of Students, STCEAS）將教學評量素養的內涵加以歸納，包括：（1）選擇適合於教學決策的評量方法；（2）發展適合於教學決策的評量方法；（3）執行、分析、評分與詮釋外界發展或教師自製之評量的結果；（4）針對個別學生進行選擇、規劃教學、發展課程以及學校發展時去使用評量能力的結果；（5）去發展學生評量之有效評分步驟；（6）對於學生、家長、一般民眾與教育工作者傳遞評量結果；（7）分辨、不道德、不合法或其他不適當之評量方法與評量資訊的使用，共七項標準，此為本研究教學評量素養之依據。

本研究所稱之教學評量素養指「國民中學音樂教師教學評量素養與教學評量實施調查」問卷中第貳部分，教師評量素養部分之情境題得分，受試者所獲得之得分愈高，其代表其教學評量素養愈高。

貳、教學評量實施（teaching assessment implementation）

教學評量是指教師運用各種方式收集學生的學習資料，包括量的資料，如考試成績、作業分數；質的資料，如軼事記錄、觀察表等。經過審慎分析、綜合與解釋後，對學生學習情形作價值的判斷（李坤崇，1999）。本研究所指的教學評量為「國民中學音樂教師教學評量素養與教學評量實施調查」問卷中第參部分，教師評量實施情形之得分表現作為判斷依據，此部分包含實施音樂教學評量前之預備情形、實施音樂教學評量時之情形及實施音樂教學評量後的結果運用情形等三個向度。受試者所獲得之得分愈高，則代表在其音樂教學活動中，經常融入該項教學評量概念。



第四節 研究範圍與限制

本研究旨在瞭解國中音樂教師教學評量素養與實施教學評量的現況。由於研究者的能力、時間與人力的限制，茲將本研究的範圍與限制說明如下：

壹、研究對象

本研究以臺北市公私立國民中學音樂教師為母群體，包含音樂專任教師、兼任導師或行政工作之音樂教師以及音樂代課教師，若學校無音樂教師，則以任教音樂教師為代表，以隨機抽樣方式進行。臺灣其他縣市公私立國中之音樂教師皆不在本研究範圍內。

貳、研究方法與工具

本研究採用問卷調查方式進行。依據研究目的，採用與改編 Campbell, C.與 Mertler, C. A.於 2005 年設計之「評量素養調查問卷」(Assessment Literacy Inventory, ALI)部分試題與自行編製「國中音樂教學評量實施情形之調查問卷」作為調查工具。研究內容主要針對教師基本資料、教師評量素養與教師評量實施情形三部分進行討論。其他評量相關議題及工具，未列入本研究範圍內。

第二章 文獻探討

本章依據前述之研究目的與待答問題，共分成三節，第一節為教學評量的理論基礎，第二節探討音樂教學評量的理論基礎，第三節為教師評量素養的意涵，以下於各節分別探討之。

第一節 教學評量的理論基礎

本節內容為教學評量的基本理念、各類教學評量、教學評量之相關法規與政策以及教學評量之相關研究，分別詳述如下：

壹、教學評量的基本理念

一、教學評量之意涵

教學評量在整個教學的工作中占有重要的地位，好的評量可以從中瞭解學生的學習情形，同時檢驗教師的教學成效。Linn 和 Miller（2005）強調評量能提供學生目前的學習狀態、進展以及教育品質相關的訊息。李坤崇（2006）指出教學評量為教師將所得的訊息資料加以選擇、組織，並解釋之，以幫助學生作決定或價值判斷的過程；張春興（2007）提出教學評量為有系統地蒐集有關學生學習行為的資料，加以分析處理後，再根據預定之教學目標給予價值判斷的歷程。

紀雅真（2007，頁 11）將教學評量定義為，教師在教學歷程中，持續蒐集學生的學習資料，加以選擇、分析、組織，將測驗分數或其他資料加以轉換，並根據學生個人潛能、特質等相關因素，對學生學習成果作價值判斷。

綜上所述，教師要瞭解學生的表現情形，須靠教學評量來達成。教學評量涵蓋了資料的蒐集與整理（包含量與質的資料），以及教師根據資料解釋結果所形

成價值判斷的一個歷程。

張麗麗（2002）將教學評量歸納為五項特質：（1）評量能力的多元，包含認知、情意、技巧與後設認知...等；（2）評量方式的多元，如：紙筆、實作、觀察、晤談、札記、計畫、檔案等；（3）評量情境的多元，包含結構與非結構、模擬與真實的情境；（4）評量包括蒐集與綜合資料的歷程；（5）評量涵蓋結果的解釋及教學有關的決定。歐滄和（2006）則認為教學評量有以下特徵：（1）以改進學生的學習行為為目的；（2）由授課教師主導；（3）與學科及班級文化密切配合；（4）形成性評量重於總結性評量；（5）提供豐富回饋，以增進教學相長。

從教學評量的特質可發現，目前的教學評量已漸漸脫離分數、等第的傳統包袱，從強調結果轉變為多元的評量向度與情境，以及提供教師改進學生學習的學習方式。

二、教學評量之定位

教學評量為教學過程的一部分，教學歷程包含教學目標、起點行為、教學活動與教學評量（如圖 2-1-1），一般的教學順序依循箭頭次序前進，下面的箭頭是回饋路線，藉由評量給予教師與學生回饋，以下藉由教學的基本模式（Glaser, 1962）說明評量在教學中的地位（劉向欣，2008）：

（一）教學目標

教學目標在教學活動中相當重要，教師在選擇教學目標時會根據學習者傳遞出來的訊息，加以彙整後訂定教學目標，其主要資訊為「學習之前學生能做些什麼」；「學習之後學生應該能做些什麼」；以及「可用的教學資料」。此外，教師將九年一貫課程裡的能力指標納入，並訂定可觀察的行為、行為的情境以及用明確的評量標準之行為目標，使教學結束後得以確切獲得相關資訊。

（二）起點行為

教師的教學目標和學生的起點行為相互影響，教師若明確瞭解學生的起點行為，則可精確的訂定教學目標，而教師的教學目標也要隨時檢核學生的起點行為是否具備相當的能力來達成教學目標。因此，學生的起點行為是從學生行為的觀點來瞭解和敘述學生的實際狀況。其中包括：學生已經具有的學習、智慧、能力和動機等，以此作為決定教學程式的參考和依據。在教學流程圖中所顯示的過程雖然以教學目標先於起點行為，但二者之間的關係，卻是可以交互變換的。

（三）教學流程

教學流程是教學實際的過程，其中包含的範圍很大，而教師在教學流程中將以學生為中心，依據學生的起點行為、教學目標、能力指標，發展或選擇教材，並採用適當的教學方法引起學生動機，維持學生興趣，指導學生有效的學習。

（四）教學評量

教學評量以教學目標為主要依歸，教師在教學過程中運用不同的形式與方法（包括文字與非文字的方式）來進行評量，並在評量過後蒐集學生各種量化與質化的資訊。教師在教學目標、起點行為和教學流程三個階段中，不斷的進行教學評量結果的修正，當教學評量的結果沒有達到所訂定的標準時，就表示學生尚未達成教師先前訂定的教學目標，有可能是：學生先備知識不足、教師高估學生起點行為或是在使用的教學方法、策略、教材有所缺失...等種種可能。因此，教學評量在教學流程中所擔任的工作就是協助各項歷程進行檢核，並提供回饋。

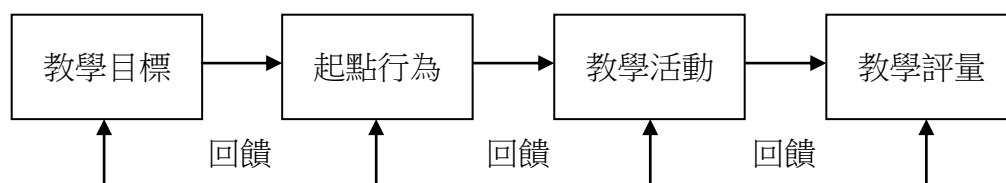


圖 2-1-1 教學的基本模式（Glaser, 1962）

從上述的教學基本模式中可發現，在教學過程中，教師須建立明確的教學目標，經由起點行為的分析及適當的教學活動，使學生學習到知識與技能，在此過程中藉由教學評量可給予教師回饋訊息以進行不斷的修正，同時可看出，教學評量在教學歷程中扮演舉足輕重的角色。

三、教學評量之功能

許多學者分別提出教學評量之功能，國內學者黃光雄（1996）認為，評量在教學過程中具有以下四種主要功能：（1）瞭解學生的潛能與學習成就，以判斷其努力程度；（2）瞭解學生學習的困難，作為補救教學及個別輔導的依據；（3）評鑑教師教學的效率，作為教師改進教材、教法的參考；（4）獲悉學習進步的情形，促發學生學習的動機。

郭生玉（2004）更提出教學評量對教師與學生都具有正面的功能，對教師分別有以下三種回饋功能：（1）瞭解學生的起點行為；（2）瞭解自身教學的效果；（3）瞭解學生的學習困難。另一方面，教學評量對學生則有：（1）增進學生瞭解教師的教學目標；（2）激發學生的學習動機；（3）增強學生的自我瞭解的正面回饋功能。

由此可看出，教學評量在整個教學工作中的重要性是無庸置疑，教學評量可作為學生瞭解自己的學習進步情形，同時作為教師補救教學及個別輔導的依據，以發揮最佳的教學效果。因此，如何有效利用評量來達到教學的目的，是每位教師需要思考且重視的課題。

四、教學評量之原則

教師如何改善現行的評量方式，發展符合評量原則的教學評量。歐滄和（2006）指出教學評量的一般性原則為：（1）事先告知評量的內容與範圍；（2）要依據教學目標選擇評量方法；（3）評量內容應綜合三大領域；（4）試題內容要公平且具代表性；（5）尋求合作改進評量技術；（6）實施上兼顧學習歷程與結果；（7）評

量結果力求客觀與精確；(8) 施測後應尊重學生隱私權；(9) 要能提供師生立即回饋；(10) 要能充分利用評量結果。

李坤崇 (2006) 提出實施教學評量時應掌握以下原則：(1) 掌握活化教學與教學評量的趨勢；(2) 兼顧認知、技能與情意；(3) 納入小組評量、同儕評量或自評；(4) 兼顧平時評量與定期評量；(5) 適切結合學習檔案與學習成果紀錄；(6) 兼容能力與努力，兼顧質化紀錄與量化紀錄。

高博銓 (2007) 則認為相關的基本原則包括：(1) 綜合性：指教學評量的內容能夠涵蓋各種不同種類的學習內容；(2) 全程性：指教學評量的實施兼重過程與結果，掌握學生全程的學習狀況；(3) 真實性：強調必須在班級教學的情境下，由教師設計，並且在教室課程的脈絡之下，針對學生的表現加以評量；(4) 有效性：評量必須確有效地測量出學生的實際表現；(5) 多樣性：指評量的方式採取多元的策略。

紀雅真 (2007) 則結合各學者的觀點，將教師進行評量時應遵守的原則分為三類：(1) 實施程序：須以教學目標為依據來規劃評量活動，在評量前向學生說明評量內容、方式、程序與範圍，評量過程要兼顧歷程與結果；(2) 評量方式的選擇：應具有彈性，兼用不同的評量方式與技術；(3) 評量結果之運用：妥善運用評量成果，將其視為蒐集資料之方法，作為改進教學之依據。

因此，研究者認為不管原則標準為何，教師在實施評量的過程中需有自己的準則，才能使教學評量發揮良好的功效。

五、教學評量新趨勢

順應教育的改革與趨勢，教學評量也須有所改變，過去的評量較重智育及知識性的背誦，忽略了高層次的問題解決能力與創意。現今的評量強調學生個人本位，評量內容多元化且方式也不再局限於筆試才能提升教學的品質，進而達到教學的目的。Line 與 Gronlund（2000）提出美國教學評量之趨勢，包括以評量作為教育改革的手段；科技進步、電腦化測驗日漸普遍、國家與社會大眾對評量的重視。

王文中、呂金燮、吳毓瑩、張郁雯、張淑慧（2005）則認為評量的趨勢為：

（1）靜態評量變為動態評量；（2）機構化評量變為個人化評量；（3）單一評量變為多元評量；（4）虛假評量變為真實評量。

高博銓（2007，頁 51）認為教學評量的革新之道，可以從以下六方面著手：

（1）擺脫傳統教學評量的束縛；（2）建立適切的教學評量指標；（3）擴大教學評量的參與面；（4）兼顧制式和非制式的評量；（5）正視教學評量的文化脈絡；（6）強化教學與評量的聯結。

從以上趨勢可發現，教學評量愈來愈受到關注，不同以往只著重記憶層面的評量，教師開始運用不同的評量方式來測量學生的學習與解決問題的能力。教師須落實教學與評量的統合概念，讓評量與課程目標、教學方法、內容及活動上做配合，以達到預期的學習成效。

貳、各類教學評量

依據不同的分類標準，可將教學評量依教學過程、分數解釋與新興之教學評量作以下的整理：

一、依教學過程

美國的兩個教師組織：美國教師聯盟（AFT）與全國教育協會（NEA）於1990年初期，擬定了「教師教育評量能力準則」（The Standards for Teacher Competence in Educational Assessment, STCEAS）。此準則指出：教學與評量的整合可依教學時程的三個階段來進行：分為教學前、教學中與教學後的評量。

（一）教學前的評量

教學前的評量包括準備度評量（readiness evaluation）與安置性評量（placement evaluation）兩種。準備度評量之目的在瞭解學生是否具備下一階段的先備知識，於課程或單元教學開始前實施，評量內容包括要達到成功學習所須具備的基本技巧與起點行為，多數採用觀察或問問題的方式；安置性評量主要作為分班或分組教學的依據（紀雅真，2006；歐滄和，2006）。

（二）教學中的評量活動

教學中的評量則可分為形成性評量（formative assessment）與診斷性評量（diagnostic assessment）。形成性評量主要目的為，在教學歷程中提供持續不斷的回饋訊息，讓學生的思考與學習歷程，藉由各種評量的方式顯現（visible）或透明化。主要是協助教師與學生監控學習進展、提昇與改進教學與學習，它強調的是持續的歷程（張麗麗，2002）；診斷性評量關注的是，學生在學習時持續或重複發生的困難，這些困難無法經由形成性評量後的矯正處置來解決。形成性評量其評量範圍較小，限於特定的教學內容，通常評量的結果不給予等第或分數；而診斷性評量較為複雜且詳細，主要是確定學習的困難之處與發生原因（紀雅真，2006）。

（三）教學後的評量

教學後的評量稱為總結性評量（**summative assessment**），它是單元及課程告一個段落或成果完成時，對學生表現所作的總結式評量，主要目的為確認學生的學習成就、正式評定學生成績、評斷教學目標是否達到。

不同於形成性評量重視歷程的特質，總結性評量是教師教學與學生學習完成之後所做的評量，關注的重點是對學生的能力及成就提出總結的憑證（張麗麗，2002）

二、依分數的解釋

以學生分數的解釋來分，評量有兩種：常模參照評量（**norm-referenced assessment**）和標準參照評量（**criterion-referenced assessment**）（郭生玉，2004；歐滄和，2006）：

（一）常模參照評量

常模參照評量，指評量後將學生的分數和其他學生的分數比較而得到學生表現的好壞，換句換說，就是根據學生分數在團體當中所得到的相對位置，來解釋分數的意義，此種評量是屬於相對的解釋。常模參照評量的目的，在於區分學生（團體）彼此之間的個別差異，以作為教師決策之用，例：編班或評定等級，因此，常模參照測驗須選擇中等難度的題目為主，以達到區分學生各種不同成就水準的目的。這種評量最大的優點，即是可清楚看出個人分數在班級中的相對地位。

（二）標準參照評量

標準參照評量，是根據學生已學到、已經熟的内容來說明分數的意義。學生的成績是和教師已事先設定好的標準來進行比較，因此標準參照評量在測驗前，教師必須要先確認成績比較的標準，之後再將評量得到的分數轉換成標準參照的形式，解釋學生是否精熟學習的内容。

三、依評量取向

依評量取向，分為實作評量（performance assessment）與檔案評量（portfolio assessment），分別做簡單的整理：

（一）實作評量

實作評量（performance assessment），指呈現工作（tasks）給學生，並要求學生以口頭、寫作或完成作品方式表現的一種評量方法（郭生玉，2004）。Linn 和 Miller（2005）認為實作評量包括文章寫作、科學實驗、語言表達，與運用數學解決問題，強調的不僅是知，還著重於做，同時強調歷程與結果並重。綜合國內外學者的觀點，實作評量可定義為：「具相當評量專業素養的教師，編擬與學習結果應用於情境頗類似的模擬測驗情境，讓學生表現所知、所能的學習結果」（李坤崇，2006）。

一般而言，實作評量具有下列三項主要特點（Popham, 2002）：（1）多元的評量標準。學生的表現不是依據一項標準判斷，而是使用多元的標準；（2）預先詳述品質標準。評量學生表現的每一項標準，都須在判斷學生表現之前，很清楚的敘述；（3）判斷的評鑑。實作評量需要人的判斷，以確定學生表現可接受的程度，但客觀測驗則可使用電腦評分。

國內學者李坤崇（2006）將實作評量特質分為：（1）強調實際生活的表現；（2）著重較高層次思考與解決問題技巧；（3）兼容跨領域或學科知識；（4）重視學生學習個別差異；（5）允許評量時間彈性化；（6）適於年齡幼小、發展較遲緩學生；（7）促進學生自我決定與負責；（8）講求評分、標準與人員的多元化；（9）強化溝通與合作學習能力；（10）兼重評量的結果與歷程；（11）著重統整化、全方位、多樣化的評量；（12）強調專業化、目標化的評量；（13）強調教學與評量的統合。

綜合郭生玉（2004）與李坤崇（2006）國內學者之看法，實作評量具有以下優點與限制。優點包含：（1）兼重評量的結果與歷程；（2）與真實生活結合；（3）可以評量實作和統整的能力；（4）評量高層認知思考與問題解決能力；（5）

促進學生自我決定與負責；(6) 引導高層次認知學習；(7) 瞭解學生學習缺點，具有診斷功能；(8) 教學與評量結合；(9) 有助於改進教學。

實作評量的限制為：(1) 費時費力的一種評量方式，施測計分時間較長；(2) 無法測量到所有學習目標；(3) 評量信度與效度較低；(4) 題數不多，評量結果的推論性較低；(5) 計分較複雜主觀；(6) 施測花費昂貴，難以大量實施。

(二) 檔案評量

檔案評量 (portfolio assessment) 是指有目的的長期收集學生作品，以顯示學生在某一個領域的努力、進步和成就。此收集須包括：學生參與檔案內容的選擇、選擇的指引、評斷優劣的標準和學生自我反省的證據等。(Paulson、Paulson & Meyer, 1991；Arter & Spandel, 1992，引自郭生玉 p.388)。

綜合國內外各學者的看法，歸納出檔案評量具有以下特性 (郭生玉，2004；引自王文中等，2005)：(1) 採用多元方式評量學生的作品，以適用個別差異的需要，同時鼓勵學生自我反省及自評；(2) 與教師教學脈絡結合；(3) 教師與學生的共同參與；(4) 強調縱貫的學習歷程；(5) 低推論性的問題；(6) 評量需要花費較多的時間。

參、教學評量之相關法規與政策

近年來，國內更加重視教學評量議題，以下是近期相關政策之建立與修改，分別為：十二年國民基本教育、國民中學學生學習成就評量標準、國民小學及國民中學學生成績評量準則、臺灣學生學習成就評量資料庫之文獻整理：

一、十二年國民基本教育

教育部推動十二年國民基本教育，預定西元 2014 年實施，其特色為非強迫入學（非義務）、普及入學且低學費的教育。其政策目標為：提升國民素質，增進國家競爭力；促進教育機會均等，實現社會公平正義；縮小教育落差，均衡城鄉發展；舒緩升學壓力，引導學生適性發展（教育部，2007）。雖然贊成者和反對者的意見均有，但教育部已著手修正《高級中學法》與《職業學校法》整合為《高級中等學校法》，作為推動延長國民基本教育的法律依據。其法源是依據西元 1999 年總統公布的《教育基本法》第十一條規定：「國民基本教育應視社會發展需要延長其年限；其實施另以法律定之。」由此可看出，教育部早已為邁向十二年國民基本教育政策做準備。

十二年國民基本教育其實施內容，與教學評量相關之政策為「精進成績評量方式。」為鼓勵教師實施多元評量，開發「國民中學學生學習成就評量標準」，提供檢視並瞭解學生在每個學習階段的學習成效；透過修訂，提高對國中畢業生的要求，以確保學生的學習品質。「國民中學及國民中學學生成績評量準則」與「國民中學學生學習成就評量標準」的編定與內容整理，後面會有詳細的整理。

二、國民中學學生學習成就評量標準

國民中學學生學習成就評量標準（以下簡稱評量標準）旨在建置與國民中小學九年一貫課程綱要（以下簡稱課綱）能力指標相對應的評量標準。此標準可幫助國中教師瞭解學生的學習表現，進而調整其教學內容、方式，或設計適當的補救教學課程與活動，以協助學生提升其學習品質，達到維繫基本學力、縮減學習成就落差的任務。

國民中學學生學習成就評量標準建立的目的，有以下五點：

（一）提供與課綱相對應的評量參照依據，轉型為標準參照的評量。

評量標準則是採「標準參照」的方式，透過絕對的學習成就表現標準來描述學生的學習成就表現。其目的在於了解學生對於學習內容的精熟程度，換句話說，標準參照並不需注意排名或與他人的分數進行比較，學生自己的學習狀況才是標準參照關注的焦點。

（二）即時提供教學回饋

課綱推動以來，對於教師評量能力的提升始終沒有加強，評量標準提供了教師更具體明確的命題方針及對學生能力層級的定義，強化學校教師的評量素養，以提升全國國中教師在評量上的專業發展。

（三）建立學力監控的機制

透過建立全國一致的評量標準，將可達到各學習領域學力監控之目標。這樣的機制，其實已在世界先進國家中推行，例如英國學生進展評量（Assessing Pupils' Progress, 簡稱 APP）、澳洲昆士蘭省課程評量與報告（Queensland Curriculum, Assessment and Reporting, 簡稱 QCAR）、美國共同核心標準（Common Core State Standards）等。

評量標準除了可作為校內和全國學生學力水平的參照標準以達到客觀評分的效果外，主管教育行政機關也可根據評量結果追蹤學生學習變化趨勢並進行分析，以檢視目前課程與教學成效，做為政策改進之參考。

（四）評量概念與國際接軌

從上述資料可發現，各國教育改革中可發現西方各國在評量結果的解釋方面，已由常模參照轉變為標準參照評量。臺灣也希望透過全國一致的評量標準，轉變

國中教育現場之評量方式，將常模參照轉為標準參照，使教師和學生皆能同步學習和成長，同時與世界各國之評量概念接軌。

國民中學學生學習成就評量標準的制定範圍為，課綱涵蓋的七大學習領域（語文、健康與體育、社會、藝術與人文、自然與生活、數學與綜合活動學習領域），分科訂定各學習領域或科目之評量標準。表現等級區分為「通過」與「未通過」兩大類，進一步再將「通過」者分三級，為「優秀（A 級）」、「良好（B 級）」與「基礎（C 級）」；尚未通過評量標準的學生，可針對其落後程度的差異，大致區分為「不足（D 級）」與「落後（E 級）」。

此標準包含內容標準與表現標準，「內容標準」（content standard）為指出學生於學習歷程中應該學會的知識與技能，包含在各個學科中，如何運用思考、溝通、理解、推理、研究等方式來學習相關的知識、概念與議題，依各領域訂定適當的學習內容；「表現標準（performance standard），則是透過一段表現描述（performance level descriptors, PLDs）來說明學生完成一個階段的課程學習後，呈現其知識與技能所到達的程度，換句話說，表現標準描述的是學生經過一段時間的學習後，其表現成果達到怎樣的程度才算「好」（how good is good enough）。當學生接受相同的學習內容，卻呈現出不同的表現程度時，必須透過「表現等級」（performance level）加以區分，再給予各表現等級不同的表現描述。內容標準與表現標準其呈現形式於下一節會有更詳細的解釋。

三、《國民中學及國民中學學生成績評量準則》

教育部為配合「十二年國民基本教育」政策之規劃與實施，於 101 年 5 月修正並公布《國民中學及國民中學學生成績評量準則》，以回應各界對國中小學生學習評量方法、成績及格定義等相關問題之關注。整理表格 2-1-1 如下：

表 2-1-1 《國民中學及國民中學學生成績評量準則》資料整理

說明	條文內容整理
評量目的與功能	將國民教育宗旨列為評量之總目的；分別列舉此準則對學生、教師、學校、家長及主管機關之功能 國民中小學學生成績評量，以協助學生德智體群美五育均衡發展為目的，並具有下列功能： 一、學生據以瞭解自我表現，並調整學習方法與態度。 二、教師據以調整教學與評量方式，並輔導學生適性學習。 三、學校據以調整課程計畫，並針對學生需求安排激勵方案或補救教學。 四、家長據以瞭解學生學習表現，並與教師、學校共同督導學生有效學習。 五、直轄市、縣(市)政府及教育部據以進行學習品質管控，並調整課程與教學政策。
評量範圍與內涵	將國民中小學九年一貫課程綱要所定之七大學習領域及其融入之重大議題，納入學習領域之評量範圍。 國民中小學學生成績評量，應依學習領域及日常生活表現，其評量範圍及內涵如下： 一、學習領域：其評量範圍包括國民中學及國民中學九年一貫課程綱要所定之七大學習領域及其所融入之重大議題；其內涵包括能力指標、學生努力程度、進步情形，並應兼顧認知、情意、技能及參與實踐等層面，且重視學習歷程與結果之分析。 二、日常生活表現：其評量範圍及內涵包括學生出缺席情形、獎懲紀錄、團體活動表現、品德言行表現、公共服務及校內外特殊表現等。
評量原則	分別依目標、對象、時機、方法、結果解釋、結果功能、結果呈現、結果管理等八個向度敘述國民中小學成績評量之原則。 國民中小學學生成績評量原則如下： 一、目標：應符合教育目的之正當性。 二、對象：應兼顧適性化及彈性調整。 三、時機：應兼顧平時及定期。 四、方法：應符合紙筆測驗使用頻率最小化。 五、結果解釋：應標準參照為主，常模參照為輔。 六、結果功能：應形成性及總結性功能並重;必要時應兼顧診斷性及安置性功能。 七、結果呈現：應質性描述及客觀數據並重。 八、結果管理：應兼顧保密及尊重隱私。

接次頁

表 2-1-1 (續)

評量方式	強調評量方式之選擇應依學習領域與日常生活表現之評量內涵及目標而定。	國民中小學學生成績評量，教師依學生身心發展及個別差異，採取下列適當之方式辦理： 一、 紙筆測驗及表單 ：依重要知識與概念性目標，及學習興趣、動機與態度等情意目標，採用學習單、習作作業、紙筆測驗、問卷、檢核表、評定量表等方式。 二、 實作評量 ：依問題解決、技能、參與實踐及言行表現性目標，採書面報告、口頭報告、口語溝通、實際操作、作品製作、展演、行為觀察等方式。 三、 檔案評量 ：依學習目標，指導學生本於目的導向系統彙整或組織表單、測驗、表現評量等資料及相關紀錄，以製成檔案，展現其學習歷程及成果。
評量時機與次數	學習領域應兼重定期評量與平時評量，日常生活表現以平時評量為原則。	國民中小學學生成績評量時機，分為 定期評量 及 平時評量 二種。學習領域評量應兼顧定期評量及平時評量，惟 定期評量 中紙筆測驗之次數，每學期至多三次， 平時評量 中紙筆測驗之次數，應符合最小化原則。
評量人員	評量人員以各科授課教師及導師為專業評量之主體；各學習領域評量人員，須於每學期初向家長與學生說明評量相關計畫。	國民中小學學生成績評量之評量人員及其實施方式： 一、各學習領域：由授課教師評量，且須於每學期初向學生及家長說明評量計畫。 二、日常生活表現：由導師參據學校各項紀錄，以及各學習領域授課教師、學生同儕及家長意見反應等加以評定。
評量結果	明定平時和定期評量依評量方法之性質，以等第、數量或文字描述紀錄評量結果；令明定學習末應以等第總結學生全學期各學習領域之評量結果，並指出學生未來學習之具體建議，以引導教師藉由評量協助學生之學習。	國民中小學學生學習領域之平時及定期成績評量結果，應依評量方法之性質以等第、數量或文字描述記錄之。 前項各學習領域之成績評量，至學期末，應綜合全學期各種評量結果紀錄，參酌學生人格特質、特殊才能、學習情形與態度等，評定及描述學生學習表現和未來學習之具體建議；並應以優、甲、乙、丙、丁之等第，呈現各學習領域學生之全學期學習表現。
評量結果之應用	明定學生成績評量紀錄應通知家長，學校不得公開呈現個別學生在班級及學校排名。	學校就國民中小學學生學習領域及日常生活表現之成績評量紀錄及具體建議，每學期至少應以書面通知家長及學生一次。 學校得公告說明學生分數之分布情形。但不得公開呈現個別學生在班級及學校排名。

另外，教育部於 103 年 4 月增加了「國民小學及國民中學成績評量準則」第 13 條。為了瞭解並確保國民中學學生學力品質，中華民國 103 年起每年 5 月針對國民中學三年級學生統一舉辦國中教育會考，強調教育會考之結果供學生、教師、學校、家長及主管機關瞭解學生學習品質及其他相關法規規定之使用。但不得納入在校學習評量成績計算。

四、臺灣學生學習成就評量資料庫

(Taiwan Assessment of Student Achievement, TASA)

臺灣學生學習成就評量資料庫是由國家教育研究院策劃督導，所進行之長期性研究計畫，該研究所蒐集之資料涵蓋國中、高中、高職此三個教育階段，用以蒐集學生國語文、英語文、數學、社會、自然此五科之學習成就表現。此資料庫建置動機為，國內長期缺乏量化指標和標準化測量工具來檢視學生學習成就的表現及其差異，以致無法確實瞭解課程實施的成效，同時也不利於課程發展之進行與相關教育政策之研擬。因此，臺灣須建置一個完整且客觀的學生學習成就資料庫，以力客觀且量化的國際比較。

臺灣學生學習成就評量資料庫其測驗目的，是希望能夠建置一個屬於本國的學習成就資料庫，提供可靠的量化資料讓國內外相關人員深入探討學生學習成就情形、並反映國內課程綱要學生學習的表現，同時也可追蹤、分析學生在學習上變遷的趨勢，進而檢視國家教育課程體制與教學政策實施之成效。施測對象與時間分別為，國中、高中及高職分年施測，時間大約每年五月底，學生課程學習將結束之際。每年施測的學習階段都不同，每名受測學生施測的科目也是隨機選擇，希望可以檢視學生學習成就的情形，進行分析探討。

肆、教學評量之相關研究

研究者以關鍵字搜尋：95~102 學年度教學評量為主題的相關論文文獻，所列共有八篇。其中跟視覺藝術有關的議題有四篇，邱慧娟(2011)、謝祈蒂(2009)

與王晶瑩（2006）之論文研究方法都是採用問卷調查法：邱慧娟（2011）目的在瞭解國中視覺藝術教師之多元評量實施作法、多元評量意義的理解與觀點，以及實施多元評量所遭遇的困擾、謝祈蒂（2009）則是探討視覺藝術教學評量在臺灣的發展歷史以及目前運用現況及王晶瑩（2006）為探討臺北市國中視覺藝術教師實施教學評量之現況，並瞭解教師實施評量時所遇到的困難。另外一篇，廖文卿（2009）採用了實驗研究法與個案研究法，對個案學校的一個特定視覺藝術課程計畫之教學評量過程進行有系統的分析，試圖瞭解學校是如何進行視覺藝術教育課程的標準化評量，之後進行視覺藝術學習成效評量，以驗證藝術與人文學習成效評量是否能夠建立班級教學之標準化評量模式。

語文部分共有兩篇：余智娟（2006）使用問卷調查法，探討臺灣華語文教師對教學評量方式及使用頻率、教學評量方式選用考量因素及教學評量知能描述；謝玲華（2006）則採用問卷調查法及訪談法，瞭解國中教師之國語文多元評量知能、實施困境與因應策略現況。

數學部分，陳怡淇（2007）採用行動研究法，呈現一位國中數學科教師施行數學遊戲融入教學評量的歷程紀錄，主要目的在探討七年級數學課中施行數學遊戲融入教學評量的模式、可能所會面臨的困難與解決策以及瞭解此教學模式對學生數學學習成效的影響。

王雅蕾（2009）探討歷史教師的教學評量是否會受基測試題影響而轉化。研究採用內容分析法，分析基測歷史試題與學校教學評量歷史試題之特色，並比較兩類試題的異同，歸納出教學評量歷史試題如何受基測影響而有所轉化；另外採訪談法，瞭解學校教師的觀點，分析教師在命題過程中如何受基測影響而轉化。

關於上述相關論文，整理表格 2-1-2 如下：

表 2-1-2 教學評量相關之論文一覽表（註：按年份排列）

西元	院校/研究生	論文標題	研究對象	研究方法	研究工具
2011	國立新竹教育大學/邱慧娟	臺中地區國中視覺藝術教師多元評量實施現況之研究	臺中地區公立國民中學視覺藝術教師	問卷調查法	自編「臺中地區國中視覺藝術教師多元評量實施現況之調查問卷」
2009	國立臺灣師範大學/王雅蓓	基測改變了什麼？從教學現場探究基測實施後歷史教學評量的轉換	1. 基測與學校各類教學評歷史試題 2. 某國中一位歷史教師	內容分析法 訪談法	1. 自編整理表格 2. 自編訪談大綱
2009	國立彰化師範大學/廖文卿	國中高年級視覺藝術教學評量之個案研究：全國性藝術與人文學習成效評量模式	1. 全國性藝術與人文學習成效評量之內容 2. 臺中市某國小六年級學生	文獻研究法 實驗研究法 個案研究法	1. 自編整理表格 2. 自行設計試題
2009	國立新竹教育大學/謝祈蒂	視覺藝術教學評量工具在臺灣的發展與現況調查研究	1. 視覺藝術教學評量之相關資料 2. 全國國中美術班術科專任教師和視覺藝術教師	內容分析法 問卷調查法	自編問卷
2007	國立彰化師範大學/陳怡淇	發展七年級數學遊戲融入教學評量之行動研究	研究者所任教的七年級學生	行動研究法	學習態度量表 自編晤談單 學習單 上課日誌 學生學習週記
2006	國立臺灣師範大學/王晶瑩	臺北市國中視覺藝術教師教學評量現況之研究	臺北市公私立國中專任視覺藝術教師	問卷調查法 訪談法	自編問卷 自編訪談大綱
2006	國立高雄師範大學/余智娟	臺灣華語文教師教學評量研究	臺灣華語文教師	文獻分析法 問卷調查法	「臺灣華語文教師採用之教學評量與相關因素問卷」
2006	國立高雄師範大學/謝玲華	國中教師國語文多元評量知能、實施困境與因應策略研究	高雄市國語文教師	問卷調查法 訪談法	自編「國中教師實施國語文多元評量調查問卷」

第二節 音樂教學評量的理論基礎

壹、音樂教學評量之意涵與實務建議

一、音樂教學評量的價值

姚世澤（2000）指出，音樂教學評量是基於對音樂工作、演奏（唱）、教室中學生的學習態度、學習效果、教師教學效率、教材內容編排之系統性或課程標準之適用性，以及行政配合措施作綜合性之評量，將各種相關資料作成一種具有建設性及參考價值的教學評鑑；紀雅真（2007）則認為，音樂教學評量是指音樂教師蒐集學生的學習資料，以作為教學與課程決定之基礎。評量能夠具體呈現最核心重要的學習內容。教師透過評量活動瞭解學生的音樂學習情形，蒐集足夠的主觀與客觀資料，針對學生的音樂行為判斷其學習成果並給予適當的評價，激發學生學習動機，作各種教學決定，其重要性不言而喻。但同時音樂藝術之時間導向、聽覺藝術特性，其無法被量化的特質，教學評量之合理性與難度，更顯得重要（林小玉，2013）。

Boyle 與 Radocy（1987）提出，音樂教學評量為音樂教師蒐集學生的學習資料，作為教學與課程決定之基礎。音樂屬於藝術形式，情感為其要素之一，因此要適切的評量學生音樂行為，須包含主觀資料與客觀資料。主觀資料為教師非正式、非系統化的觀察與詮釋，容易會受到教師本身的經驗、知識、情感、直覺與偏見影響；客觀資料則不受教師情感與偏見的影響。Brophy（2000）認為，音樂教學評量是教師蒐集學生的學習情形。廣泛的定義包含，所有被教師用來確定學生目前學習情況的活動，而學生的現況是和教育標準是否達成，以及學生如何學習音樂有關。

Brophy (2000) 指出正確運用音樂教學評量，可達到之功能有四：(1) 可幫助教師獲得學生音樂成長與進步情形之證據，作為教師處理教學相關事項之輔助；(2) 引導教師教學與教學策略之選擇；(3) 可提升家長與學生對音樂課程之重視。學生知道教師評量的方式與內容，瞭解評量結果後會通知父母，這樣的評量效果變具有意義；(4) 當教師的教學效能受到質疑時，教師可運用評量來說明學生的音樂學習情形與成果。

由此可知，音樂教學評量指音樂教師蒐集學生的主觀與客觀資料，一個好的音樂教學評量，可讓教師反思學生的作品、音樂表現情形，使教師得以決定與調整後續的教學目標、活動內容及教學計畫，以符合學生之學習需求。

二、音樂教學評量的實務建議

藝術類科的中心概念是美的、和諧的、優雅的、具有風格的；其規準是反覆與變化的、對比與勻稱的、合乎形式本質的、創意的；其研究方法是直覺、想像、感受；其意義源於生活也落實於生活；其課程目標是探索與創作、審美與理解、文化與理解。從上述可發現，藝術類科具有無法被量化的特質，評量的有效實施更顯得重要（林小玉，2001）。

Brophy (2000) 提出六個音樂教學評量的準備步驟，做為音樂教師進行教學評量的參考，內容為：(1) 以一個特定的行為或概念為目標，確定所欲評量的內容；(2) 預先決定並確認評量的標準，以確定評量的焦點；(3) 準備評分規準或其他計分方法，考量所有可能會發生的狀況；(4) 決定紀錄保存的策略；(5) 決定評量時所運用的音樂素材；(6) 決定學生評量結果的通知方式。同時建議評量規準六項原則：(1) 設定特殊行為、概念或知識為一個測量目標；(2) 思考並設立用來判斷評量結果的標準；(3) 確定並發展評分工具（例如：規準、評分技巧）；(4) 確定可靠又有效的紀錄方法；(5) 選擇評量作業的素材；(6) 通知學生。

林小玉（2001）主張，音樂藝術之本質表現於音樂教學上，應秉持下列三個原則：（1）聲音先於符號之教學；（2）音樂的元素當為音樂課程、教學與評量之主體；（3）美學原則是助長學生成為音樂創作者之關鍵。教師須將評量與音樂藝術的特性相結合，否則評量音樂如同評量其他學科，不僅無法正確掌握教學與評量，更可能窄化音樂多元的可能性。

音樂教師在實施評量前，應先考量以下方向：（1）是否以課程目標為依據來訂定評量標的、評量設計與評量策略；（2）是否已全面檢測音樂學習內涵以反映音樂課程的全貌；（3）是否以多重角度設計評量策略，以收集豐富的回饋資訊；（4）是否訂定完整而合理的評分規範以溝通共識；（5）是否善用媒體以提供真實音樂刺激（謝苑玫，2000）。另外，謝苑玫（2002）在藝術與人文學習領域的研習手冊中，提供了評量改革之實施建議：改革評量初期，教師嘗試評量與教學活動結合時，可先從有限的局部做起，首先可先選一個班級並選一段時期開始做，並設定一段六至八周的時間，教師徹底地熟悉課程目標，這樣之後便可專心在教學評量的工作。之後教師必須要決定評量什麼、決定如何來蒐集學生學習學習資訊以及決定什麼時候收集資訊，以上這些評量前的工作，教師都須在教學過程前思考完善，以達到完整的評量準備工作；教學過程中，教師依評量之目的，訂定評分規範以及如何清楚記錄學生的評量結果；評量結束後，教師則要把所收集到的資訊視情況讓學生、家長等相關人士瞭解，以及自身的反思與修改，如此一來，教學評量才能達到良好的效果與進步。

教師確定評量目的與行為後，非常重要的工作是選擇適當的評量方法，評量的方法相當多元，不同方法所達到的目的也不相同，如果教師認為評量就是使用傳統音樂教學評量的作法，進行歌唱教學、直笛吹奏或樂理考試，來評斷學生是否真正的學習，這已經無法適應現在的教育環境。音樂教師可以轉為積極態度來面對挑戰、努力提昇自己的音樂教學評量知能，進而提供學生豐富的回饋協助與溝通，相信音樂教學品質會更加提昇。

貳、音樂教學評量之相關法規與政策

根據教育部（2008）國民中小學九年一貫課程綱要，藝術與人文學習領域明訂教學評量方法如下：「本領域教學評量可併用『量』與『質』的評量，且可視教學目標、教學範圍、教學方法、教學流程之需要，採取教師評量、學生互評、學生自評等方式，並應用：觀察、問答、晤談、問卷調查、軼事記錄、測驗、自陳法、評定量表、檢核表、討論...等方式評量，且可酌採相對解釋法與自我比較法等彈性評量措施。」由此可看出，現今的教學評量，教師擁有更多的自主性，且評量方式有多元化的趨勢。以下為目前國立臺灣師範大學心理與教育測驗研究發展中心所發展的「國民中學學生學習成就評量標準」之藝術與人文學習領域部分之資料整理：

一、國民中學學生學習成就評量標準－藝術與人文學習領域

國民中學學生學習成就評量標準（以下簡稱評量標準）旨在建置與國民中小學九年一貫課程綱要（以下簡稱課綱）能力指標相對應的評量標準，由各學科之大學教授、輔導團教師、國中現場教師以及測驗評量專家組成，並以能力指標重整後的學科評量重點，作為不同表現等級應具備的學習表現標準之基礎。此標準在前面已有簡單說明，以下為藝術與人文領域之學習成就評量標準之資料整理：

（一）發展理念與原則

此評量標準採用教育部藝術與人文學習領域課程目標：探索與表現、審美與理解、實踐與應用為內容標準主題；同時依據藝術與人文學習領域第四階段能力指標及其理念，訂出表現能力、創作能力、鑑賞能力、基礎知能、生活應用與合作參與等六項內容標準當作次主題，整理表格如下：

表 2-2-1 音樂概念統整說明

學 科	主題	次主題	概念統整說明
音 樂	探索 與 表現	表現能力	能透過歌唱、演奏、指揮等技能表現音樂作品，感受藝術多元的特性及其與社會文化的關係。
		創作能力	能運用音樂要素知能與技法，進行簡短的音樂創作或指定曲式的創作，發展個人獨特表現，並體察情感特質、社會關懷與藝術創作的關聯。
	審美 與 理解	鑑賞能力	能辨識不同時代、文化、地區、族群等音樂表現之風格與特性，並描述音樂作品所呈現的文化內涵。
		基礎知能	能透過資訊科技的運用，蒐集藝術資料，認識音樂術語及相關概念，增廣對音樂文化的認知。
	實踐 與 應用	生活應用	能理解音樂與生活的關聯，選擇適合自己的音樂活動並身體力行，進而在日常生活中養成音樂表現及音樂鑑賞的興趣與習慣。
		合作參與	能集體共同規劃音樂活動與創作等，彼此相互溝通、尊重、協調並表現合作學習及團隊參與的態度。

(二) 訂定表現標準

表現等級之描述，均以學生要獲得不同表現等級所要達到的最低表現水準（門檻）為撰寫原則，共訂為 A-E 五級，其代表意義如下：A 表示「優秀」；「B」表示「良好」；C 表示「通過」；D 表示「不足」；E 表示「落後」。研發團隊以能力指標重整後的學科評量重點，做為不同表現等級應具備的學習表現標準之基礎。表格 2-2-2 如下：

表-2-2-2 表現等級描述

內容標準		表現等級				
主題	主題	A	B	C	D	E
探索與表現	表現能力	1. 能運用良好的歌唱技能表現音樂作品。 2. 能以熟練的演奏技能操作特定樂器（如：曲調樂器、節奏樂器）表現音樂作品。 3. 能運用熟練的指揮技能進行樂曲指揮。	1. 能運用基本的歌唱技能表現音樂作品。 2. 能以基本的演奏技能操作特定樂器（如：曲調樂器、節奏樂器）表現音樂作品。 3. 能運用適當的指揮技能進行樂曲指揮。	1. 能運用粗略的歌唱技能表現音樂作品。 2. 能以粗略的演奏技能操作特定樂器（如：曲調樂器、節奏樂器）表現音樂作品。 3. 能運用粗略的指揮技能進行樂曲指揮。	1. 能嘗試運用歌唱技能表現音樂作品。 2. 能嘗試運用演奏技能操作特定樂器（如：曲調樂器、節奏樂器）。 3. 能嘗試運用基本的指揮技能。	未達D級
	創作能力	能運用適切的音樂要素與技法進行音樂創作。	能運用音樂要素與技法進行音樂創作。	能運用音樂要素進行音樂創作。	能嘗試運用音樂要素進行音樂創作。	未達D級
	鑑賞能力	能理解音樂作品的内容，並表達作品所蘊含的美感。	能理解音樂作品的内容，並表達對作品的感受。	能大致理解音樂作品的内容，並約略表達對作品的感受。	能理解有限的音樂作品内容。	未達D級
審美與理解	基礎知識	能分析音樂作品的基本知識。	能應用音樂作品的基本知識。	能理解音樂作品的基本知識。	能在提示下，說出音樂作品的基本知識。	未達D級
	理解能力					

接次頁

表 2-2-2 (續)

實踐與應用	生活應用	能在生活中主動參與適合自己性向、興趣或能力的音樂活動，積極展現音樂應用的能力。	能在生活中樂於參與適合自己性向、興趣或能力的音樂活動，展現音樂應用的能力。	能生活中能參與音樂活動，展現音樂應用的能力。	能生活中嘗試參與音樂活動。	未達D級
	合作參與	能主動參與歌唱、演奏或創作等集體音樂活動，表現團隊精神與態度。	能樂於參與歌唱、演奏或創作等集體音樂活動，表現團隊精神與態度。	能參與集體音樂活動。	能嘗試參與集體音樂活動。	未達D級

(三) 音樂科內容標準與九年一貫課程綱要能力指標對應表

表 2-2-3 為音樂科內容標準與九年一貫課程綱要能力指標之對應表，以藝術與人文學習領域課程綱要為範圍，逐一條列出與第四階段（7~9 年級）學生相關的國中階段之能力指標，確保三主題包含所有可評量的國中階段能力指標，且不脫離課綱之範疇；內容標準次主題是經由重整藝術與人文領域國中階段可評量的能力指標並考量學科學習重點與特性，以結構化方式呈現，各個內容次主題皆對應一個到多個能力指標；每個能力指標都對應到一個或多個次主題，並呈現內容標準與能力指標之對應關係以及各次主題涵蓋的學科評量重點。

表 2-2-3 內容標準與九年一貫課程綱要能力指標對應表

內容標準		九年一貫課程綱要-分段能力指標		
主 次	能力指標	能力指標	能力指標	
題 主	1-4-1-1-4-4	2-4-5-2-4-8	3-4-9-3-4-11	
題				
探 索 與	1-4-2 體察人群間各種情感的特質，設計關懷社會及自然環境的主題，運用適當的媒體與技法，傳達個人或團體情感與價值觀發展獨特的表現。		3-4-10 透過有計畫的集體創作與展、演活動，表現自動、合作、尊重、秩序、溝通、協調的團隊精神與態度。	
	1-4-3 嘗試各種藝術媒體，探求傳統與非傳統藝術風格的差異。			
	1-4-4 結合藝術與科技媒體，設計生活應用及傳達訊息的作品。			
表 現 作 能 力	1-4-1 了解藝術創作與社會文化的關係，表現獨立的思考能力嘗試多元的藝術創作。		3-4-10 透過有計畫的集體創作與展、演活動，表現自動、合作、尊重、秩序、溝通、協調的團隊精神與態度。	
	1-4-3 嘗試各種藝術媒體，探求傳統與非傳統藝術風格的差異。			
	1-4-4 結合藝術與科技媒體，設計生活應用及傳達訊息的作品。			
鑑 賞 能 力	1-4-1 了解藝術創作與社會文化的關係，表現獨立的思考能力，嘗試多元的藝術創作。	2-4-5 鑑賞各種自然物、人造物與藝術作品，分析其美感與文化特質。		
	1-4-2 體察人群間各種情感的特質，設計關懷社會及自然環境的主題，運用適當的媒體與技法，傳達個人或團體情感與價值觀發展獨特的表現。	2-4-6 辨識及描述各種藝術品內容、形式與媒體的特性。		
	1-4-3 嘗試各種藝術媒體，探求傳統與非傳統藝術風格的差異。	2-4-7 感受及識別古典藝術與當代藝術、精緻藝術與大眾藝術風格的差異，體會不同時代、社會的藝術生活與價值觀。		
審 美 與				
表 基 礎 知 能		2-4-6 辨識及描述各種藝術品內容、形式與媒體的特性。		
		2-4-8 運用資訊科技，蒐集中外藝術資料，了解當代藝術生活趨勢，增廣對藝術文化的認知範圍。		

表 2-2-3 (續)

內容標準		九年一貫課程綱要-分段能力指標		
主 次	能力指標	能力指標	能力指標	
題 主	1-4-1-1-4-4	2-4-5-2-4-8	3-4-9-3-4-11	
題				
實			3-4-10 透過有計劃的集體創作與展演活動，表現自動、合作、尊重、秩序、溝通、協調的團隊精神與態度。	
生			3-4-11 選擇適合自己的性向、興趣與能力的藝術活動，繼續學習。	
踐				
活				
應				
與				
用				
應				
合			3-4-10 透過有計劃的集體創作與展演活動，表現自動、合作、尊重、秩序、溝通、協調的團隊精神與態度。	
用				
作				
參				
與				

參、音樂教學評量之相關研究

以下分為兩部份，首先針對國內外專家學者所發表音樂教學評量的研究加以探討，之後分別敘述國內外近幾年所辦理有關教學評量之音樂教育研討會，整理如下：

一、國內外有關音樂教學評量之相關研究

國內音樂教學評量之研究，研究對象為國中音樂教師的文獻共有兩篇，都是使用自編問卷：紀雅真(2007)目的在了解我國國民中學音樂教學評量實施情形、教學評量實施成效與困擾評量實施之因素，並探討音樂教學評量相關議題；黃于芬(2007)則是了解臺北縣國民中學藝術與人文教師進行音樂教學所使用的評量方式、對於實施多元評量認同度以及所面臨的困擾。

有關國小音樂教學評量之議題共有四篇，其中與音樂基本學習內容有關的調查有兩篇，分別為劉以勤(2013)與林桂妃(2009)的研究，研究方法都是使用問卷調查法：前者為了解桃園縣國民中學藝術與人文課程中歌唱教學評量實施現

況，針對歌唱教學評量實施情形、實施成效及困擾評量實施因素進行調查；後者則是瞭解基隆市國民中學教師直笛教學評量實施現況，針對直笛教學評量實施情形、實施成效，及困擾評量實施之因素三方面進行探討。

另外兩篇都是現況調查之研究，陳臆如（2013）探討雲林縣國民中學音樂教師實施教學評量之現況，調查雲林縣國中音樂教師實施音樂教學評量之頻率、音樂教學評量之方式與內涵，並探究實施音樂教學評量之困難與所需之協助；吳盈麗（2011）則調查臺中市國民中學音樂教師對九年一貫藝術與人文領域課程綱要音樂教學評量之認知、看法、實踐情形及需求。

唯一使用內容分析法為黃翠瓊（2008）之研究，研究者使用自編整理表格，分析 1994-2007 臺灣地區音樂評量相關學位論文共 138 篇，藉由內容分析法試圖瞭解臺灣地區音樂評量之概況，期望探究臺灣地區音樂評量之現況與趨勢。

國外研究部分，研究對象大部分為國小音樂教師，舉例：Hepworth-Osiowy (2004)調查了 190 位小學音樂教師對音樂評量的觀點及實踐情形，調查結果顯示受訪音樂教師認為音樂評量是有價值的，他們會運用不同的評量工具和教學策略進行評量，但在實際運用時所面臨的困難，讓受訪音樂教師認為自己在音樂教學評量的實踐仍顯不足。

Talley (2005)調查了密西根 200 位小學音樂教師對評量的使用頻率、方法、目標與應用情形，研究發現許多音樂教師並不常對學生進行評量，有些教師甚至完全不進行評量，填答的教師並不使用出版的音樂成就測驗，也很少進行音樂性向測驗：在音樂評量中，最常用的教師自編測驗為評定量表或評分規準、檢核表、個別談話、書寫測驗或學習單，以及個別演奏。受訪教師表示每一年級最常使用的評量內容為拍子、歌唱、音高配對、節奏、直笛、譜表辨識、樂器家族、節奏視譜以及視譜演奏。研究建議教師們積極為學生進行評量，以達到有效能及健全的音樂教學。

Peppers (2010)調查了密西根 232 位小學音樂教師對評量的態度與調查統計結果之間的關係，受訪教師均認為評量的目的是為了改善教學，但他們提出評

量最大的疑義是會造成教學的負擔，學生數太多、班級數太多以及缺乏時間是進行評量時最大的困難。教學年資 20 年以上的受訪者對於和評量有關的任務提出困惑，而且對評量抱持否認的態度；教學年資在 10-19 年之間的受訪者認為他們在大學階段並未對評量做充分的準備，並提出促進教學評量最重要的方式是在學區內參加研習會、教學分享會議以及教學觀察。

由上述三篇國外音樂教師對音樂教學評量之調查研究可看出，音樂教師普遍認為評量的目的是為了改善教學，但在實施的層面卻面臨了諸多困難與疑惑，教師們也思考了解決方案與需求，可見音樂教師仍願意為改善教學而積極努力。

綜合國內外音樂教師對音樂教學評量之調查研究，國內研究重點在國中音樂教師於教學評量實踐層面的探討，國外則已開始關注小學音樂教師對教學評量的觀點及態度。本研究希望在小學音樂教師對教學評量的認知、看法、實踐情形及需求進行探究，企盼對國內音樂教學評量提出結論與建議。關於上述國內外相關論文，整理表 2-2-4 如下：

表 2-2-4 音樂教學評量相關之論文一覽表（註：按年份排列）

西元	院校/研究生	論文標題	研究對象	研究方法	研究工具
2012	臺北市立教育大學/劉以勤	桃園縣國民中學藝術與人文課程中歌唱教學評量實施現況之調查研究	桃園縣國民中學藝術與人文課程中教授音樂課程之教師	問卷調查	自編問卷「桃園縣國民中學藝術與人文課程中歌唱教學評量實施現況之調查研究」
2013	國立臺中教育大學/陳臆如	雲林縣國民中學音樂教師實施教學評量現況之調查	雲林縣國中音樂教師	問卷調查	自編「雲林縣國民中學音樂教師實施教學評量現況之調查問卷」
2011	國立臺中教育大學/吳盈麗	臺中市國民中學音樂教師對九年一貫藝術與人文領域課程綱要音樂教學評量之觀點調查	臺中市國民中學音樂教師	問卷調查	自編「臺中市國民中學音樂教師對 97 課綱音樂教學評量之觀點調查問卷」

接次頁

表 2-2-4 (續)

2010	Hepworth-Osiowy, K.	Assessment in elementary music education: Perspectives and practices of teachers in Winnipeg public schools	加拿大 190 位 國小音樂教師	問卷調查	自編問卷
2010	Peppers, M.	An examination of teachers' attitudes toward assessment and their relationship to demographic factors in Michigan elementary general music classrooms	密西根 232 位 國小音樂教師	問卷調查	自編問卷
2009	國立臺北教育大學/林桂妃	基隆市國民中學教師直笛教學評量實施現況之調查研究	基隆市國民中學 學授予音樂課程之教師	問卷調查	自編「基隆市國民中學教師直笛教學評量現況之調查問卷」
2008	臺北市立教育大學/黃翠瓊	1994-2007 臺灣地區音樂評量相關學位論文之內容分析研究	1994-2007 臺灣地區音樂評量相關學位論文	內容分析	自編整理表格
2007	國立臺灣師範大學/紀雅真	國民中學音樂教學評量實施現況之調查研究	臺北縣、市公立國中音樂教師	問卷調查	自編「國中音樂教學評量實施情形調查問卷」
2007	國立臺灣師範大學/黃于芬	臺北縣國民中學藝術與人文領域音樂教學多元評量之現況調查研究	臺北縣國中音樂教師	問卷調查	自編「臺北縣國民中學藝術與人文領域音樂教學多元評量之現況調查問卷」
2005	Talley, K. E.	An investigation of the frequency, methods, objectives, and applications of assessment in Michigan elementary general music classrooms	密西根 200 位 小學音樂教師		自編問卷

二、音樂教育研討會有關教學評量之議題

近年來，中華民國音樂教育協會與師資培育大學致力於舉辦音樂教學評量相關學術研討會，對此議題之推廣極為重視。研討會包括：「2010 音樂教育學術研討會－邁向卓越之音樂教育理論與實務：從課程、教學策略、教材、音樂行為、評量與科技探討」、「2011 第八屆亞太音樂教育研究研討會」、「101 學年度藝術與人文學習領域輔導群教學之能專輯（3）－評量與教學」、「2012 藝術與人文學習領域評量標準知能工作坊」、「第四屆國際音樂教育評量研討會」、「2013 音樂教育學術研討會－因應十二年國教的有效音樂教學」、「2013 第一屆師資培育國際學術研討會－各科教材教法工作坊」。從上述主題可發現，音樂教學評量是國內外近幾年學者們關注的議題。

因應 12 年國教，高雄市國教輔導團開始在 100 年度辦理藝文評量工作坊，其中最重要的就是建立模組化評量規準之建立。不同於一般評量示例多偏重於課程內容與工具表格，此輔導團聚焦課程目標與課程綱要中能力指標之間的轉化過程，完成「藝文課程評量七帖」。理論部分，臺南大學林玫君教授分享了如何從課程綱要內的能力指標轉化為教師的教學目標；明華國中的張克享教師則介紹其在校內推動的藝文評量歷程，將評量等第的文字說明配合學生作品，形成模組化概念。工作坊教師的課程實作部分，呈現了七個藝術與人文領域之課程評量示例紀錄教師們所完成的課程與評量紀實。期待透課程目標與課程綱要中能力指標之間的轉化過程，能讓教學現場的教師更能完整落實教學評量。

2013 年 4 月第四屆國際音樂教育評量研討會（ISAME4）於國立臺灣師範大學舉辦，因應「評量」議題在國際之間備受重視及與會者的期盼，此研討會以「音樂評量與全球多樣性：實務、測量與政策」(Music Assessment and Global Diversity: Practice, Measurement, and Policy)為主題，強調音樂教育評量和全球化的重要性。其中主講者 Dr. Paul Lehman 提到，一個有效的評量是會面臨到哲學與實際的難題的挑戰，而這些挑戰是我們幾乎每天都會遇到的情況。評量的目的是讓老師去

瞭解與反映出學生的學習狀況，它不是在教學情況的結尾才會出現，而是發生在學生回答問題或是師生互動的每個情況中，以制式的、非制式的形式進行著，且持續在每一天發生。

2013 年 12 月臺北市立教育大學舉辦「2013 音樂教育學術研討會－因應十二年國教的有效音樂教學」中，潘宇文教授透過一個歌曲教唱活動的實務分享，進行音樂教學策略與評量等多方面的交流討論；陳映蓉教師則透過三個課程「從聲音到符號」、「從符號到賞析」及「從賞析到創作」的設計，進行有效音樂教學的實施，並嘗試從多元評量的方式來檢視學生的學習成效。

101 學年度國民中小學九年一貫課程推動工作「課程與教學輔導組－藝術與人文學習領域輔導群」的期末研討會，針對當前十二年國教重要議題：「教學評量」，分別邀請視覺藝術、音樂與表演藝術領域的教師，進行分享討論。在音樂部分，張超倫教師強調「音樂創新教學與評量」；廖明玲教師深究「學習共同體應用於中國風流行歌曲創作教學暨多元評量工作」；黃世傑教師則以跨領域的方式進行「資訊科技在音樂評量上的應用」。

上述可知，愈來愈多的學者強調音樂教學評量之重要性，積極進行關於此議題之研究，但是，研究者也發現對於教師評量知能方面的探討並不多見，故此論文除了瞭解評量現況外，同時針對教師評量知能進行討論。

第三節 教師評量素養的意涵

壹、教師評量素養之意涵

教師評量素養是，針對有效的評量行為所具備的基本的知識，包括術語學、發展和使用評量方法和技術的能力、熟悉評量的品質等級以及熟悉傳統的學習評量選題...等（Paterno, 2001）。美國北部中心地區教育實驗室（North Central Regional Educational Laboratory）也有相關定義，教師評量素養是教育者設計、實施以及討論評量策略所具備的能力。波宜斯州立大學學校改進與政策研究中心（Center for School Improvement and Policy Studies, Boise State University）則列點：

（1）瞭解好的評量、評鑑和傳達的實務（2）理解哪一種評量方法可以用來蒐集可靠的資料與學生的學習成就（3）無論使用成績單成績、考試分數、歷程檔案或是論壇，均可有效的傳達評量的結果（4）藉由與學生完全參與評量、保存紀錄與溝通傳達，使評量能最佳化學生的動機與學習。（王子華，2005）。

Stiggins（1995）在 *Assessment Literacy For The 21st Century* 論文中，給了以下定義：具有評量素養的人會知道有效的評量與無效的評量之不同。他們不會因為現代技術的落後而影響評量的品質。他認為具有評量素養的教育者，當他們進入評量的領域時，就已經知道他們要評量什麼、他們為什麼要評量、怎樣才能最有效地評出學生的知識與技能、怎樣讓學生產生優秀的例子、在評量過程中存在哪些潛在的風險以及如何防止這些風險的發生等。他們同樣清楚差的、不準確的評量會造成怎樣的不良的後果。

Airasian（1991）提到，教師評量素養是指，教師將所得的訊息資料加以選擇組織並解釋之，以助學生做決定或價值判斷的過程，而資訊指的是教師所收集到的種種量的或質的資訊。

美國教師聯盟（American Federation of Teachers, AFT）、全國教育測量會議（National Council on Measurement in Education, NCME）與全國教育協會（National Education Association, NEA）於 1990 發展的「學生學習評量之師資能力標準」（Standards for Teacher Competence in Educational Assessment of Students, STCEAS），將教師應具有的評量素養提供一個完整的概念，共列出七個標準，以下簡要說明之（整理自劉向欣，2008；王子華，2005；Linn & Gronlund，2000）：

一、教師應具備能力去選擇適合於教學決策的評量方法

Stiggins（1995）曾指出，懂得評量的人可以描述在特定學期的特定情境下的成就目標，他可以描述該學的知識、該精熟的思考模式、必須展現的特定行為以及必須要創作的作品。從上述可知，教師第一個應展現的評量知能，是熟悉從教學計劃角度來選擇評量的方法及評鑑。教師可察覺和認知到不同評量方法與某些教學目標是不相容的，也可能會對教學有不同的影響，同時考慮學生的文化、社會、經濟和語言背景來選擇適合的評量方法。重視每一種評量方法在對學生學習時之適當性。

二、教師應具備能力去發展適合於教學決策的評量方法

雖然教師經常使用外界出版社的測驗，但教師仍要對訊息蒐集、發展和使用評量的原則加以熟悉，並依據自己的教學目標選擇適當的評量技術，並藉著使用學生資料來分析所用的評量技術之品質。

三、教師應具備能力去執行、分析、評分與詮釋外界發展或教師自製之評量的結果

教師應具備解釋各種評量結果之能力，當獲得不一致的結果時，可自身去尋找、解釋這些不一致現象的方法或是尋找其他資料。教師在進行決策前，需先解

決部分不確定性的問題，例：確認學生之長處和錯誤...等，而不要不當使用評量而增加學生的焦慮。

四、在針對個別學生進行選擇、規劃教學、發展課程以及學校發展時，教師應具備能力去使用評量能力的結果

教師能使用累積的評量訊息，組成一個合理的教學計畫以協助學生發展；或是在參與各層級課程決定時，提供有效的評量結果，藉以作為規劃課程、改善課程的依據，以促使有效改善各層級之教育決定。

五、教師應具備能力去發展學生評量之有效評分步驟

教師應具備設計、實施和解釋發展評分的程序之能力。建立客觀的評分標準，不但能理解並詳細對學生說明等第的客觀解釋，並符合合理、公平公正的要求，同時也確認、評鑑及修正評分程序，以避免錯誤的評分程序。

六、教師應具備能力對於學生、家長、一般民眾與教育工作者傳達評量的結果

教師能夠與學生及家長溝通評量的發展、理解學生各種背景脈絡並給予評量結果之適當說明及解釋各種評量方法的限制。

七、教師應具備能力去分辨、不道德、不合法或其他不適當之評量方法與評量資訊的使用

教師應熟悉自己在評量上的倫理與法律責任，不宜讓評量程序被誤用或過度使用而造成有害的後果，例：讓學生難堪、不當使用成就測驗成績來測量教學效能...等。另外，應參與更廣泛的教育社群，以界定評量中適當的專業行為。

貳、評量素養調查表 (Assessment Literacy Inventory, ALI)

ALI (Campbell & Mertler, 2005) 是以「學生學習評量之師資能力標準」(Standards for Teacher Competence in Educational Assessment of Students, STCEAS) 為基礎所發展出來的，主要是透過五個情境來檢測教師的評量素養能力，藉由教師對五個情境下之測驗試題的回答情形，分析教師在學生學習評量之師資能力標準 (Standards for Teacher Competence in Educational Assessment of Students, STCEAS) 的七個標準上的表現情形 (調查問卷原文於附錄一)。以下為 STCEAS 七個標準的整理：

一、學生學習評量之師資能力標準 (Standards for Teacher Competence in Educational Assessment of Students, STCEAS)

STCEAS 是 AFT (American Federation of Teachers)、NCME (National Council on Measurement in Education) 與 NEA (National Education Association)，為了因應提升教育者評量能力的需求，於 1990 年共同發展出來的工具。雖然 STCEAS 是多年前發展出來的標準，但在近幾年仍受到重視與引用，像是依據此標準所展出來，應用於在職教師評量素養調查工具 (Teacher Assessment Literacy Questionnaire, TALQ) (Plake, 1993)；Campbell et al. (2002) 將 TALQ 修改為 Assessment Literacy Inventory (ALI)，應用在職前教師的評量素養調查，以及 Mertler (2004) 修改 TALQ 與 ALI 為 Classroom Assessment Literacy Inventory (CALI) 應用於職前與在職教師的評量素養調查。由此可見，STCEAS 在教師評量素養上有一定的貢獻與參考價值。

STCEAS 一共有七個標準之概念性與應用性之技巧，涵蓋的評量素養能力非常廣 (引自王子華，2005)，整理表格如表 2-3-1：

表 2-3-1 STCEAS 七項標準之內容

標準	說明
標準一： 選擇適合於教學決策的評 量方法	◎ 教師在教室中選擇和發展對學生進行的評量方法時使用「評量誤用」與「評量效度」的概念。 ◎ 能理解有效的評量資料以協助教師進行教學活動，如：引起學生學習動機、規劃個別化教育方案...等。 ◎ 能理解無效的資訊對於進行學生教學決策的影響性。 ◎ 當使用與評鑑教師自行決定之評量時兼具考慮學生的文化、社會、經濟、語言等因素。 ◎ 能覺察不同的評量方法無法符合某些特定的教學目標，並且可能對自己的教學有所影響。 ◎ 能適當的決策與學生有關的每一個使用的評量技術。
標準二： 發展適合於教學決策的評 量方法	◎ 教師能具備規劃與蒐集的能力，以協助自己做決定。 ◎ 在教師教學中發展和使用評量會遵循適當的規則，以避免在學生評量上犯錯。 ◎ 能依據教師的教學目標選擇適當的評量技術。 ◎ 能使用學生的資料來分析教師所使用的每個評量技術的品質。
標準三： 執行、分析、評分與詮釋外 界發展或教師自製之評量 的結果	◎ 能詮釋教師自製正式與非正式評量的結果。 ◎ 能依據指引手冊對申論題或專案進行評分，能使用量表對實作評量進行評分，並且能夠正確使用這些方法獲得一致性的結果。 ◎ 能執行標準化的成就測驗，能詮釋一般的成績報告，如百分等級。 ◎ 能對評量結果報告中常提到的總結性指數（ Summary indexes ），有概念性理解。 ◎ 能運用分數或總結性指數來強化自製評量的使用。
標準四： 針對個別學生進行選擇、規 劃教學、發展課程以及學校 發展時，教師應具備能力去 使用評量能力的結果	◎ 能利用自己累積的評量資訊來設計一個好的教學計畫，並以此計畫來促進學生教育方面的發展。 ◎ 在採用評量結果來計畫與（或）評鑑教學與課程時，能正確的詮釋結果和避免一般的詮釋錯誤。 ◎ 能得到當地、學區、州與全國性的評量結果資訊，以及能適當的應用在學生身上的課室、學校、學區、州與全國性教育改革資訊。

接次頁

表 2-3-1 (續)

標準五： 去發展學生評量之有效評 分步驟	◎ 能設計、執行與說明發展評分的程序，而此評分是由各種作業、專案、課室活動、測驗以及其他可能使用的評量所組成。 ◎ 能理解和清楚表達自己所給的成績是否合理、正當與公平，並且承認這些成績反應著自己的好惡與偏見。 ◎ 能了解與避免錯誤的評分步驟，例：利用成績來懲罰學生。 ◎ 能評鑑與修改評分步驟，以改善對學生學習成果之詮釋的有效性。
標準六： 對於學生、家長、一般民眾 與教育工作者傳達評量的 結果	◎ 能理解並有能力的依據學生之社經地位、文化、語言與其他背景因素，調整對學生的評量方式，並依此給予適當的解釋，且能說明這些評量的結果。這並不代表這些背景因素會限制學生最終的教育發展。 ◎ 能與學生、家長或監護人溝通如何評量學生在教育方面的進步。 ◎ 能理解與說明，在針對學生個人進行選擇時，將測驗誤差列入考量的重要性。 ◎ 能說明不同的正式與非正式評量方法之限制。 ◎ 能詮釋學生的課室測驗、學區測驗、州測驗與全國性測驗結果報告書。
標準七： 分辨、不道德、不合法或其 他不適當之評量方法與評 量資訊的使用	◎ 能知道影響自己的教室、學區與州評量實務之法規與個案選擇。 ◎ 能注意到各種評量步驟均可能被誤用或過度使用，進而產生傷害性的後果，例：使學生尷尬、侵犯學生隱私權、不適當的使用學生的標準化成就測驗成績去評鑑教學效益。

評量素養調查問卷 (ALI) 將上述 STCEAS 之七個標準，透過五個情境來考察教師的評量素養。藉由教師對於五個情境下之測驗試題的回答情況，分析教師在 STCEAS 的七個標準上的表現情形。此調查研究問卷共有 35 個題目，分為七個部份，每個部分分別有五個子題，每一子題分別有四個答案，其中一個為正確答案，該題答對計 1 分，答錯不計分。(問卷原文詳見附錄一)

參、國內外相關研究

Campbell 等人 (2002) 實施一項研究。此研究將教師評量素養調查表發放了 220 位職前教師。實施過程包括創造與評判各種評量的方法、討論評量的道德性、理解課堂評量與標準化評量的結果並討論評估心理測量學的品質。評量的結果顯示，職前教師的回答優於在職教師。

Mertler (2004) 指出，教學經驗與教師評量素養是有關的，職前教師與在職教師在評量素養上也確實有顯著的差異，主要差異點在「標準一：教師應具備能力去選擇適合於教學決策的評量方法」、「標準二：教師應具備能力去發展適合於教學決策的評量方」、「標準三：教師應具備能力去執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：在針對個別學生進行選擇、規劃教學、發展課程以及學校發展時，教師應具備能力去使用評量能力的結果」、「標準七：教師應具備能力去分辨、不道德、不合法或其它不適當之評量方法與評量資訊的使用」等五個類別，此五類別在職教師優於職前教師；而「標準五：教師應具備能力去發展學生評量之有效評分步驟」與「標準六：教師應具備能力對於學生、家長、一般民眾與教育工作者傳達評量的結果」則為在職教師與職前教師都顯現出較差的部分。

國內評量素養相關研究部份，王子華、范雅晴、王國華（2008）針對教師評量素養培育模式進行探究，採用 e-learning 方式進行在職教師評量素養培育，對於在職教師在評量知識學習上有顯著效益。

劉向欣（2008）則採用評量素養調查表 ALI（Campbell & Mertler, 2005）與評量觀點調查表為工具，透過五個情境來瞭解國小教師對於評量素養及評量觀點的認知情形，此研究發現不同地區之教師不論在評量素養及評量觀點整體或細項上均無呈現顯著差異；在整體教學評量素養上，教學年資 10 年以下之教師優於教學年資 10 年以上之教師，而曾經參加相關研習或修習相關之課程之教師優於不曾參加相關研習或修習相關課程之教師。

但國內對於教師評量素養現況的調查之研究數量仍偏少，教師評量素養為教師評量能力的一個關鍵，教師在進行教學評量所面臨的各種決定之適當與否，都與教師評量素養能力息息相關。故研究者針對臺北市國中音樂教師評量素養進行調查，希望藉此瞭解教師評量素養之能力。



第三章 研究方法

本研究以問卷調查的方式蒐集資料，並將資料彙整與分析，以瞭解目前國中音樂教師評量素養能力與教學評量方式之現況。本章共分五節，第一節為研究架構；第二節為研究對象；第三節為研究工具；第四節為實施程序；第五節為資料處理。

第一節 研究架構

本研究先透過文獻探討，瞭解教師教學評量素養、教學評量、音樂教學評量之意涵與國內外音樂教學評量之相關內容。之後根據文獻分析的整理，改編與自編問卷作為研究工具，採取問卷調查法瞭解國中音樂教師教學評量素養與教學評量實施之現況。

壹、研究設計

本研究採用問卷調查方式進行。研究者依據研究目的，設計「國民中學音樂教師教學評量素養與教學評量實施調查問卷」，問卷內容改編「評量素養調查問卷」(ALI) 與自行編製「國中音樂教學評量實施情形之調查問卷」作為調查工具。針對臺北市 79 所學校之音樂教師進行調查，以瞭解國中音樂教師評量素養與教學評量實施之現況，並根據研究結果提出建議。

貳、研究架構

根據第一章研究動機、研究目的與待答問題以及第二章相關文獻之整理，提出研究架構如圖 3-1-1。本研究之變項包括：依變項（教師教學評量素養與教學評量實施之現況）與自變項（教師個人背景變項）。架構詳細內容如下：

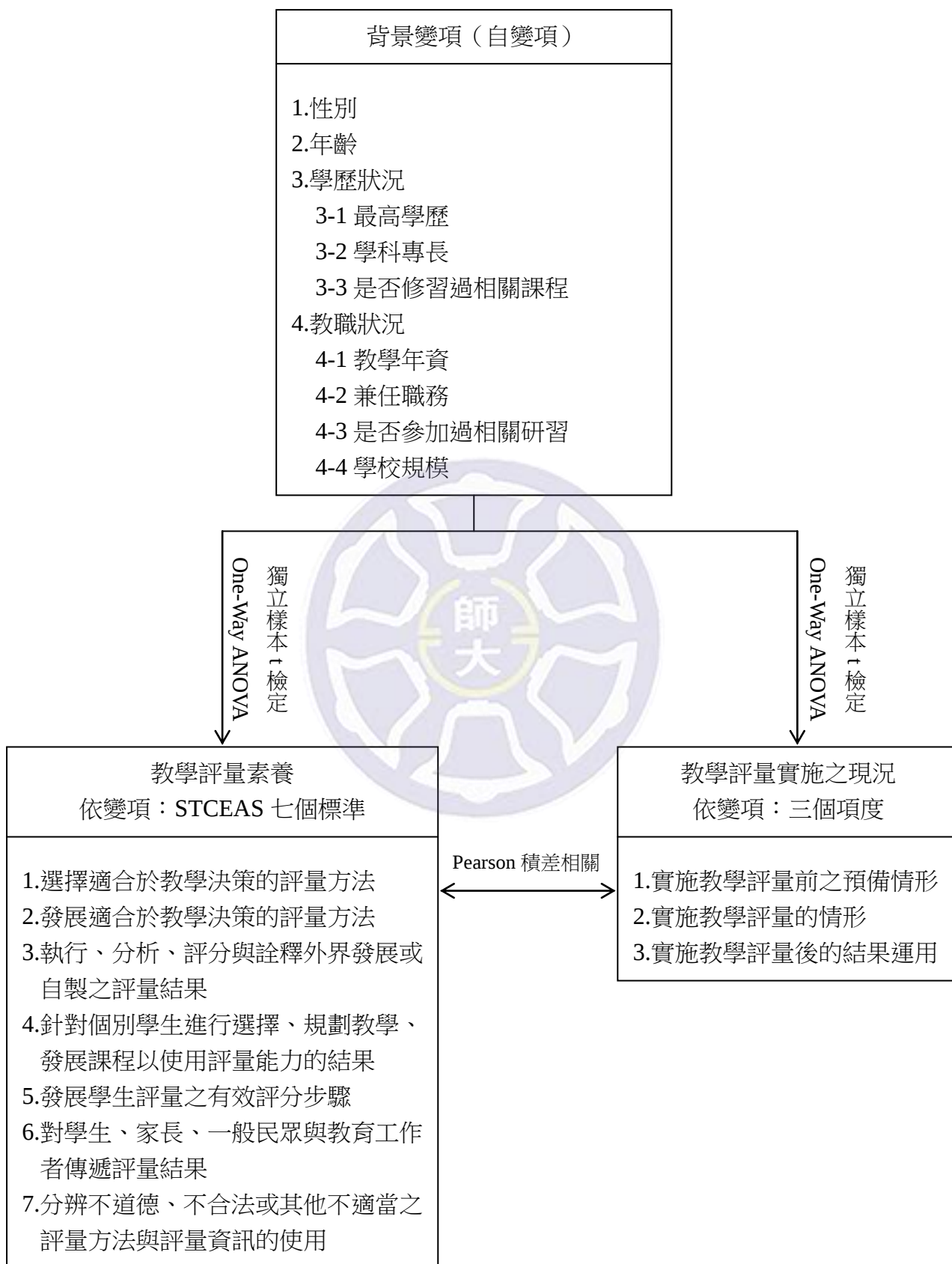


圖 3-1-1 研究架構

第二節 研究對象

臺北市為臺灣北部的重要都會區，由於人力、資源豐富與交通便利，使得教育發展極為快速，成為發展的核心。研究者考量人力與時間限制，決定以臺北市作為本研究問卷調查的範圍，並以隨機抽樣方式進行，以瞭解我國音樂教師教學評量素養與教學評量實施之情形。

本研究以任教臺北市公私立國民中學音樂教師為研究對象，包含音樂專任教師、兼任導師或行政工作之音樂教師以及音樂代課教師，若學校無音樂教師，則以任教音樂教師為代表。臺北市公私立國民中學共計 79 所，依班級數預估共有 149 位音樂教師。母群體人數統計根據台北市教育資源網路資料（網站：<http://www.tp.edu.tw/>），以臺北市學校班級數規模概算，且根據教育部國民中小學教師授課節數，國民中學專任教師每週以十八節至二十二節為原則，分為 48 班以上為大型規模學校，預估音樂教師至少有一人；25 至 48 班級為中型規模學校，預估音樂教師至少有一人；25 班級以下為小型規模學校，預估音樂教師至少有一人。

之後，研究者以電話逐一與各校音樂教師聯絡，若連絡不到則與學校教務處聯繫，調查各校可進行問卷調查之教師人數，最後於 103 年 12 月底寄發正式問卷共 120 份(詳見表 3-2-1)，104 年 1 月 9 日以電話方式進行問卷催收，1 月 12 日以信件進行第二次催收，最後回收問卷 112 份，回收率達 93%，扣除 2 份無效卷，實際可用之有效問卷 110 份，可用率為 91%。

表 3-2-1 正式問卷總人數統計表

分區	學校班級數		
	48 班以上	25~48 班	25 班以下
北投區	2 校 北投、明德、	0 校	1 校 桃源
大同區	0 校	3 校 建成、民權、重慶	3 校 忠孝、蘭州、靜修
中山區	0 校	2 校 大直、長安	4 校 大同、新興、五常、濱江
萬華區	1 校 萬華	1 校 龍山	1 校 雙園
中正區	3 校 南門、弘道	1 校 古亭	1 校 螢橋
大安區	3 校 仁愛、大安、金華	2 校 龍門、延平	5 校 師大附中、和平、芳和 民族、復興實中
士林區	2 校 士林、蘭雅	1 校 百齡	3 校 陽明、至善、福安
松山區	2 校 介壽、敦化	2 校 西松、中山	1 校 中崙
內湖區	3 校 麗山、 東湖、明湖	1 校 三民	0 校
信義區	2 校 興雅、信義	1 校 瑠公	0 校
南港區	1 校 誠正	1 校 南港	0 校
文山區	1 校 景興	4 校 萬芳、木柵 實踐、景美	3 校 景文、北政 政大附中
音樂教師人數	20 校×3 人=60 (人)	19 校×2 人=38 (人)	22 校×1 人=22 (人)
總人數	60+38+22=120 (人)		

資料來源：臺北市政府教育局網站 <http://www.tp.edu.tw/>

第三節 研究工具

本研究依研究目的與文獻探討，自行編製「國民中學音樂教師教學評量素養與教學評量實施」之調查問卷，作為研究工具。問卷編製過程及主要內容說明如下：

壹、問卷設計

問卷內容分為基本資料、教師教學評量素養及教學評量實施之現況，各部份之內容編製說明如下，詳細問卷內容放於附錄三：

一、基本資料

個人基本資料方面，包含教師性別、年齡、學歷狀況與教職狀況。學歷狀況又分為最高學歷、學科專長與師培階段曾修習過教育測驗與評量之相關課程；教職狀況部分，分為年資、目前工作職務、曾參加過教育測驗與評量相關之研習以及所任教學校之規模。

二、音樂教師評量素養

評量素養調查方面，改編「評量素養調查問卷」(ALI)。「評量素養調查問卷」(ALI)根據STCEAS之七個標準設計了五個情境題且情境設定為一般學科。研究者考量到受試者作答時間，只採用三個情境題並將其改編為音樂教學情境，改編情境題之前後對照表於表3-3-1、3-3-2、3-3-3。藉由教師於三個情境下之測驗試題的回答情況，分析教師在STCEAS的七個標準上的評量素養表現情形。

此調查研究問卷共有21個題目，分為三個情境題，每個情境題分別有七個子題，每一子題分別有四個答案，其中一個為正確答案，該題答對計1分，答錯不計分。

表 3-3-1 「國小教師評量素養問卷」情境題之前後對照表（情境一）

問卷	原始問卷	改編後問卷
內容		
情境題	情境一	情境一
教師設定	國中數學教師	國中音樂教師
科目設定	數學科	音樂科
評量情境設定	測量數學概念理解之習題	測量音樂元素概念理解之習題
測驗形式	標準化數學測驗	標準化音樂測驗

表3-3-2 「國小教師評量素養問卷」情境題之前後對照表（情境二）

問卷	原始問卷	改編後問卷
內容		
情境題	情境四	情境二
教師設定	中學的英語教師 對班級評量議題有豐富經驗	中學的音樂教師 對班級評量議題有豐富經驗
學生設定	國小六年級	國中七年級
評量情境設定	評量英文寫作技巧 數學理解	評量音樂創作技巧 歌唱能力
測驗形式	本位故事數學測驗 標準化英語、科學、閱讀測驗	歌曲評定量表 標準化音樂演奏、音感認譜測驗

表3-3-3 「國小教師評量素養問卷」情境題之前後對照表（情境三）

問卷	原始問卷	改編後問卷
內容		
情境題	情境五	情境三
教師設定	四年級自然教師	八年級音樂教師
學生設定	對水的變化研究學習吃力	對歌唱學習吃力
科目設定	自然科	音樂科
評量情境設定	水的變化教學內容	歌唱教學內容
測驗形式	全國性標準化測驗 物理學部分	全國性標準化測驗 音感認譜部分

三、音樂教學評量實施情形

此項內容分成三部分。第一部份為實施教學評量前之預備情形，內容為教師訂定評量目標之依據；第二部份為實施音樂教學評量之情形，包括：是否會向學生解釋評量規準的意義、學生配合度、評量項目及實施評量困難處之調查；最後為實施教學評量後的結果運用情形，包括：學生可否知道自己成績、教師評量後是否會檢視教學目標及修正教學內容、評量後的結果運用及成績紀錄方式。

貳、建立問卷專家效度

研究者根據研究目的、研究問題以及文獻探討編制「國民中學音樂教師教學評量素養與教學評量實施之調查」問卷，問卷初稿完成後，聘請三位音樂教育專家學者及一位國中音樂任教教師，協助填答並提供意見，進行問卷效度評鑑。最後，根據專家效度審查結果，修正問卷初稿內容，完成預試問卷。

專家效度用卷採「適合」、「修正後適合」、「不適合」三種選項，請專家學者就問卷題目提供修正意見，其專家效度問卷修正意見彙整之內容詳見附錄二。最後，根據專家效度審查內容及兩次預先施測之結果，進行多次修改與增刪，將問卷原本的五個情境改為三個情境，讓教師可以減少問卷作答時間，最後完成預試問卷。

參、實施問卷預試

本研究預試以苗栗縣國中教師為對象，實施時間為民國 103 年 11 月，共發出 35 份，回收 35 份，回收率為 100%。

第貳部分教師評量素養量表共分為 3 個情境，每個情境 7 題，共 21 題。每答對一題給一分，每個標準最高分為 3 分，總分 21 分。21 個題目與七個標準的對照如表 3-3-4 教師評量素養各標準的 ALI 題目。

表 3-3-4 教師評量素養各標準的 ALI 題目

標準	內 容	題 號	滿分
一	選擇適合於教學決策的評量方法	1、8、15	3
二	發展適合於教學決策的評量方法	2、9、16	3
三	執行、分析、評分與詮釋外界發展或教師自製之評量的結果	3、10、17	3
四	針對個別學生進行選擇、規劃教學、發展課程以及學校發展時去使用評量能力的結果	4、11、18	3
五	去發展學生評量之有效評分步驟	5、12、19	3
六	對於學生、家長、一般民眾與教育工作者傳達評量的結果	6、13、20	3
七	分辨、不道德、不合法或其他不適當之評量方法與評量資訊的使用	7、14、21	3

本研究預試結果如表 3-3-5 教師評量素養各標準描述統計，每個標準滿分為 3 分，以標準一得到滿分的人數最多，平均得分最高為標準一的 2.57 分，平均得分最低為標準三的 1.83 分。量表總分為 21 分，得分最高為 19 分，最低為 11 分，平均分數為 15.00 分，預試 35 位受試者中，以得到 15 分的人數為最多。

表 3-3-5 教師評量素養各標準描述統計

	標準一	標準二	標準三	標準四	標準五	標準六	標準七	素養總分
個 數	35	35	35	35	35	35	35	35
平均數	2.5714	2.1714	1.8286	2.2000	1.9143	2.2571	2.0571	15.0000
中位數	3.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	15.0000
眾 數	3.00	2.00	2.00	2.00	2.00	2.00	2.00	15.00
標準差	0.50210	0.70651	0.89066	0.71948	0.78108	0.78000	0.68354	1.94029
最小值	2.00	1.00	0.00	1.00	0.00	0.00	0.00	11.00
最大值	3.00	3.00	3.00	3.00	3.00	3.00	3.00	19.00

本研究以預試 35 位受試者的答題狀況，針對教師評量素養問卷進行難度分析，結果如表 3-3-6 教師評量素養難度分析一。其中難度小於.50 的題目有 5 題，難度介於.51~.79 的題目有 6 題，難度大於.80 的題目有 8 題，如表 3-3-7 教師評量素養難度分析二所示。

表 3-3-6 教師評量素養難度分析一

題 號	答 對 人 數	難 度
01	23	.657
02	29	.829
03	26	.743
04	29	.829
05	25	.714
06	31	.886
07	34	.971
08	34	.971
09	33	.943
10	22	.629
11	15	.429
12	12	.343
13	22	.629
14	8	.229
15	33	.943
16	14	.400
17	16	.457
18	33	.943
19	30	.857
20	26	.743
21	30	.857

表 3-3-7 教師評量素養難度分析二

難 度	題 號	題 數
.50<	11、12、14、16、17	共 5 題
.51~.79	01、03、05、10、13、20	共 6 題
>.80	02、04、06、07、08、09、15、18、19、21	共 10 題

預試 35 位受試者在各標準的答題情況分別如下：標準一平均分數 2.57 分，為七個標準中最高，得 0 分和 1 分的受試者皆為 0 人，得 2 分的受試者有 15 人，得 3 分的受試者有 20 人，如表 3-3-8 教師評量素養標準一次數分配所示。

表 3-3-8 教師評量素養標準一次數分配

得分	人數	百分比	有效百分比	累積百分比
2.00	15	42.9	42.9	42.9
3.00	20	57.1	57.1	100.0
總和	35	100.0	100.0	

標準二平均分數 2.17 分，得 0 分的受試者為 0 人，得 1 分的受試者有 6 人，得 2 分的受試者有 17 人，得 3 分的受試者有 12 人，如表 3-3-9 教師評量素養標準一次數分配所示。

表 3-3-9 教師評量素養標準二次數分配

得分	人數	百分比	有效百分比	累積百分比
1.00	6	17.1	17.1	17.1
2.00	17	48.6	48.6	65.7
3.00	12	34.3	34.3	100.0
總和	35	100.0	100.0	

標準三平均分數 1.82 分，為七個標準中最低，得 0 分的受試者有 3 人，得 1 分的受試者有 8 人，得 2 分的受試者有 16 人，得 3 分的受試者有 8 人，如表 3-3-10 教師評量素養標準三次數分配所示。

表 3-3-10 教師評量素養標準三次數分配

得分	人數	百分比	有效百分比	累積百分比
0.00	3	8.6	8.6	8.6
1.00	8	22.9	22.9	31.4
2.00	16	45.7	45.7	77.1
3.00	8	22.9	22.9	100.0
總和	35	100.0	100.0	

標準四平均分數 2.20 分，得 0 分的受試者為 0 人，得 1 分的受試者有 6 人，得 2 分的受試者有 16 人，得 3 分的受試者有 13 人，如表 3-3-11 教師評量素養標準四次數分配所示。

表 3-3-11 教師評量素養標準四次數分配

得分	人數	百分比	有效百分比	累積百分比
1.00	6	17.1	17.1	17.1
2.00	16	45.7	45.7	62.9
3.00	13	37.1	37.1	100.0
總和	35	100.0	100.0	

標準五平均分數 1.91 分，得 0 分的受試者有 1 人，得 1 分的受試者有 9 人，得 2 分的受試者有 17 人，得 3 分的受試者有 8 人，如表 3-3-12 教師評量素養標準五次數分配所示。

表 3-3-12 教師評量素養標準五次數分配

得分	人數	百分比	有效百分比	累積百分比
0.00	1	2.9	2.9	2.9
1.00	9	25.7	25.7	28.6
2.00	17	48.6	48.6	77.1
3.00	8	22.9	22.9	100.0
總和	35	100.0	100.0	

標準六平均分數 2.25 分，得 0 分的受試者有 1 人，得 1 分的受試者有 4 人，得 2 分的受試者有 15 人，得 3 分的受試者有 15 人，如表 3-3-13 教師評量素養標準六次數分配所示。

表 3-3-13 教師評量素養標準六次數分配

得分	人數	百分比	有效百分比	累積百分比
0.00	1	2.9	2.9	2.9
1.00	4	11.4	11.4	14.3
2.00	15	42.9	42.9	57.1
3.00	15	42.9	42.9	100.0
總和	35	100.0	100.0	

標準七平均分數 2.05 分，得 0 分的受試者有 1 人，得 1 分的受試者有 4 人，得 2 分的受試者有 22 人，得 3 分的受試者有 8 人，如表 3-3-14 教師評量素養標準七次數分配所示。

表 3-3-14 教師評量素養標準七次數分配

得分	人數	百分比	有效百分比	累積百分比
0.00	1	2.9	2.9	2.9
1.00	4	11.4	11.4	14.3
2.00	22	62.9	62.9	77.1
3.00	8	22.9	22.9	100.0
總和	35	100.0	100.0	

教師評量素養問卷最低得分者為 11 分，最高得分者為 19 分。得 11 分的受試者有 1 人，得 12 分的受試者有 3 人，得 13 分的受試者有 6 人，得 14 分的受試者有 1 人，得 15 分的受試者有 9 人，得 16 分的受試者有 9 人，得 17 分的受試者有 2 人，得 18 分的受試者有 3 人，得 19 分的受試者有 1 人，整體問卷平均分數為 15 分，人數最多的分數為 15 分和 16 分，如表 3-3-15 教師評量素養問卷總分次數分配所示。

表 3-3-15 教師評量素養問卷總分次數分配

得分	人數	百分比	有效百分比	累積百分比
11.00	1	2.9	2.9	2.9
12.00	3	8.6	8.6	11.4
13.00	6	17.1	17.1	28.6
14.00	1	2.9	2.9	31.4
15.00	9	25.7	25.7	57.1
16.00	9	25.7	25.7	82.9
17.00	2	5.7	5.7	88.6
18.00	3	8.6	8.6	97.1
19.00	1	2.9	2.9	100.0
總和	35	100.0	100.0	

針對第參部分進行信度分析，在第參部分得到各分量表的 Cronbach α 係數分別如下：實施教學評量前為 0.762，實施教學評量時為.727，實施教學評量後為.749，第參部分總量表的整體信度為.840（如表 3-3-16 所示）。

表 3-3-16 教師評量實施情形量表信度分析摘要表

分量表	題目	Cronbach α 係數
實施教學評量前	共 4 題	.762
實施教學評量時	共 15 題	.727
實施教學評量後	共 8 題	.749
總量表	共 27 題	.840

肆、完成正式問卷

預試完成後，研究者根據預試結果及填答教師意見，進行「國中音樂教師教學評量素養與評量實施之調查問卷」正式問卷（詳見附錄三）之施測。

第四節 實施程序

研究者先蒐集相關資料以擬定研究題目、目的與問題，接著進行文獻探討編製初試問卷，經過專家效度、預試與修改後，完成正式問卷。最後將回收之問卷進行統計與分析，綜合結果與討論後，撰寫研究報告。研究步驟流程圖如下：

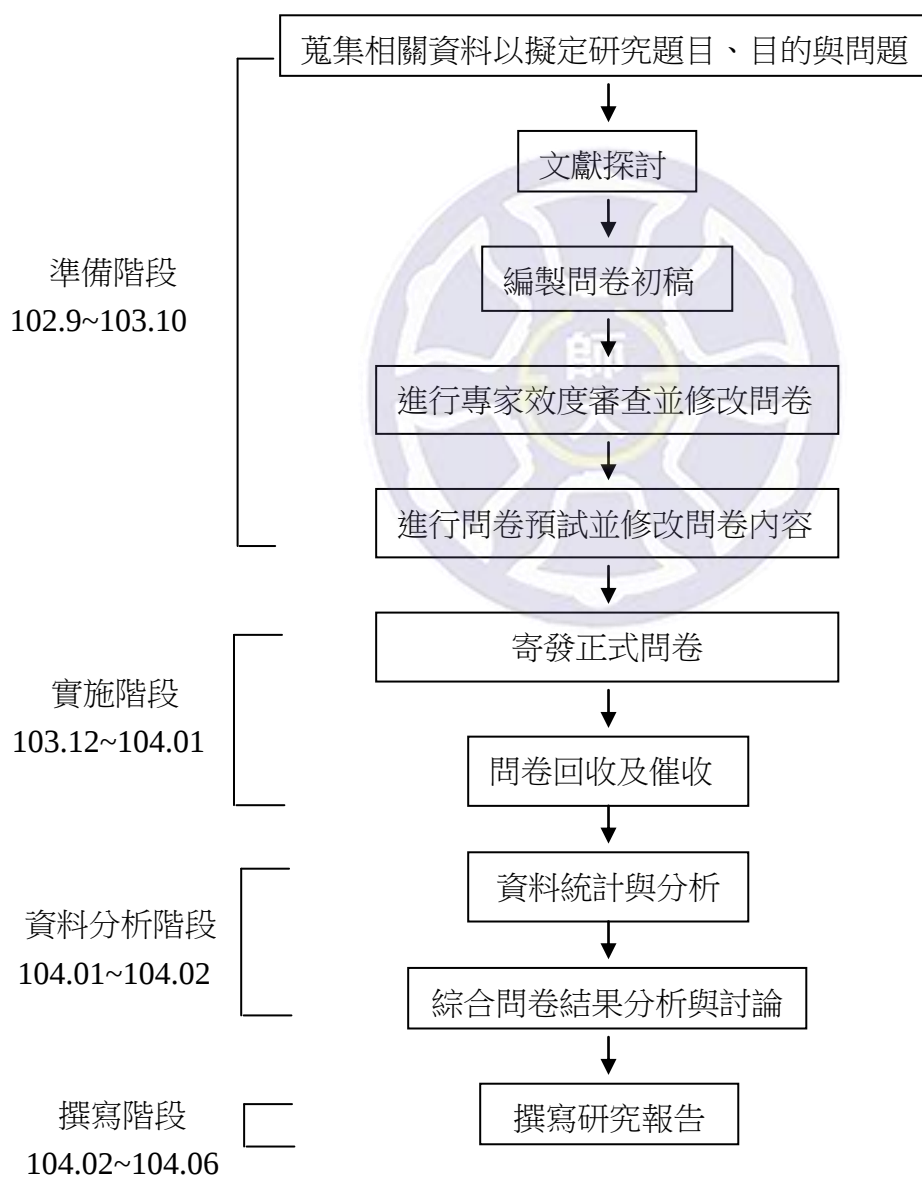


圖 3-4-1 研究步驟流程圖

第五節 資料處理

壹、本研究以 SPSS for Windows 19.0 套裝軟體進行統計分析。

貳、本研究採用之統計方法如下：

一、描述性統計

本研究採用次數累計、次數百分比、算術平均數、最大值、最小值、標準差等統計量數，用以瞭解臺北市國民中學音樂教師教學評量素養與教學評量實施之現況。

二、t 檢定

以 t 檢定考驗不同性別、學歷、學科專長、修習教育測驗與評量課程經驗、參與教育測驗與評量研習經驗之臺北市國民中學音樂教師教學評量素養與教學評量實施的差異情形。

三、單因子變異數分析（One-Way ANOVA）及 Scheff'e 事後比較

以單因子變異數分析考驗不同年齡、年資、工作職務、學校規模之臺北市國民中學音樂教師教學評量素養與教學評量實施的差異情形。若單因子變異數分析之考驗結果達顯著時，再進行 Scheff'e 事後比較以瞭解各背景變項分組間之差異情形。

四、Pearson 積差相關

以 Pearson 積差相關考驗臺北市國民中學音樂教師教學評量素養與教學評量實施間之相關情形。

第四章 結果與討論

本章內容乃依據調查問卷所得資料進行統計分析與討論，以探討臺北市國民中學音樂教師之教學評量素養與實施現況，並進而探究不同背景變項教師教學評量素養與實施之差異。本章第一節為「臺北市國民中學音樂教師教學評量素養現況與差異分析」，第二節為「臺北市國民中學音樂教師教學評量實施現況與差異分析」，第三節為「臺北市國民中學音樂教師教學評量素養與實施相關分析」。

第一節 臺北市國民中學音樂教師教學評量素養現況與差異分析

本節共分為兩部份，分別為臺北市國民中學音樂教師教學評量素養現況分析與臺北市國民中學音樂教師教學評量素養差異分析，資料如下：

壹、臺北市國民中學音樂教師教學評量素養現況分析

本節透過描述性統計分析臺北市國民中學音樂教師教學評量素養的現況，量表試題透過三種情境探究臺北市國民中學音樂教師在七項教學評量素養標準之表現。本研究於進行描述性統計時，以逐題分析方式進行得分次數百分比統計，希冀透過研究受試者於量表各試題的答題狀況來瞭解受試者的教學評量素養現況；其次，為便於統計結果的分析解釋，本研究參照大學入學考試中心在設定大學入學指定科目考試試題分析標準（網站：<http://www.ceec.edu.tw/default.htm>），以答對率 50% 為取決，兼及本研究以教師為對象，將本研究有關教師評量素養答對率之決斷定為 60%，以期能在更嚴謹的統計考驗標準下，獲取較為精確與周延之統計結果與分析結論。因此，本研究設定試題答對百分比之判定標準為 60%，答對百

分比若超過 60%，則認為多數臺北市國民中學音樂教師具備該題所欲評估之教學素養知能，若低於 60%則否，且本研究將針對答對人數百分比低於 60%之題目加以解析，以試圖瞭解受試著的可能答題意向。分析結果簡述如下：

一、臺北市國民中學音樂教師教學評量素養現況分析

（一）標準一：選擇適合於教學決策的評量方法

標準一各試題之分析結果如表 4-1-1 所示，茲將其分析結果簡述如下：

1.臺北市國民中學音樂教師於試題 1 之答對人數為 51 人，答對人數百分比為 46.36%，未超過判定標準值 60%，故多數臺北市國民中學音樂教師未具備試題一所欲評估之教學評量素養知能。

從受試者對其它錯誤題項的選填統計結果發現：總計有 50 人選填題項(A)「實作評量」，共佔總受試者人數的 45.45%，此人數幾乎與答對人數的比例相同。本研究推測造成此種現象的原因，應與部分受試者無法正確釐清(A)題項「實作評量」與(B)題項「真實評量」的意涵差異有關。其二者之主要差異，在於實作評量主要在測量學生之特定程序性知識與能力，而真實評量則在於瞭解學生是否實際具備某種能力或在實際生活應用該種能力的程度。因此，題項(A)為與題項(B)為相似性頗高的誘答題項，受試者若無法正確判知二者內涵的實質差異，可能傾於選填較常接觸或聽聞的「實作評量」題項，因而造成填答錯誤的現象。

2.臺北市國民中學音樂教師於試題 8 之答對人數為 99 人，答對人數百分比為 90.00%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 8 所欲評估之教學評量素養知能。

3.臺北市國民中學音樂教師於試題 15 之答對人數為 99 人，答對人數百分比為 90.00%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 15 所欲評估之教學評量素養知能。

4.整體而言，臺北市國民中學音樂教師於標準一之答對累積人次為 249 人次，累積答對人次百分比為 75.45%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備「標準一：選擇適合於教學決策的評量方法」所欲評估之教學評量素養知能。

表 4-1-1 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準一（n=110）

層面	試題	答對人數	答對人數百分比	各選項答對人數與百分比
標準一：選擇適合於教學決策的評量方法	試題 1	51	46.36%	(A) 50 人 （45.45%）
				(B) *51 人 （46.36%）
				(C) 7 人 （6.36%）
				(D) 0 人 （0%）
				無法作答： 2 人 （1.81%）
	試題 8	99	90.00%	(A) *99 人 （90.00%）
				(B) 0 人 （0%）
				(C) 0 人 （0%）
				(D) 8 人 （7.27%）
				無法作答： 3 人 （2.72%）
試題 15	99	90.00%	(A) *99 人 （90.00%）	
			(B) 3 人 （2.72%）	
			(C) 8 人 （7.27%）	
			(D) 0 人 （0%）	
			無法作答： 0 人 （0%）	
整體層面		249	75.45%	

註：符號說明：*代表此試題之正確題項

（二）標準二：發展適合於教學決策的評量方法

標準二各試題之分析結果如表 4-1-2 所示，茲將其分析結果簡述如下：

1.臺北市國民中學音樂教師於試題 2 之答對人數為 89 人，答對人數百分比為 80.91%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 2 所欲評估之教學評量素養知能。

2.臺北市國民中學音樂教師於試題 9 之答對人數為 97 人，答對人數百分比為 88.18%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 9 所欲評估之教學評量素養知能。

3.臺北市國民中學音樂教師於試題 16 之答對人數為 50 人，答對人數百分比為 45.45%，未超過判定標準值 60%，故多數臺北市國民中學音樂教師未具備試題 16 所欲評估之教學評量素養知能。

從受試者對其它錯誤題項的選填統計結果發現：較多填答錯誤的受試者選填題項「無法作答」(共 22 人，佔總受試者人數的 20%)，其它較多選填錯誤的選項為題項(A)「難度值在 0.50 和 0.75 之間」(共 17 人，佔總受試者人數的 15.45%)與題項(D)「難度值為 0.90」(共 12 人，佔總受試者人數的 10.90%)。本研究推測造成此種現象的原因，應與國民中學音樂師資之專業養成背景的內涵較缺乏評量理論或試題分析素養的納入，因而對「試題難度」與「試題鑑別度」之概念無法瞭解與釐清，因此，即使本研究在題目的設計上已提示受試者「難度 0.50 以上及鑑別度 +0.30 以上為優良試題」，但仍有部分的受試者直接選填「無法作答」的題項，也可能為受試者受既有刻板印象的影響，直接認定難度過大的試題就屬於不當或不良試題而須予以修正或刪除，故而選填題項(A)與題項(D)，因而造成填答錯誤的現象。

4.整體而言，臺北市國民中學音樂教師於標準二之答對累積人次為 236 人次，累積答對人次百分比為 71.52%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備「標準二：發展適合於教學決策的評量方法」所欲評估之教學評量素養知能。

表 4-1-2 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準二（n=110）

層面	試題	答對人數	答對人數百分比	各選項答對人數與百分比
標準二：發展適合於教學決策的評量方法	試題 2	89	80.91%	(A)* 89 人 (80.91%) (B) 13 人 (11.81%) (C) 5 人 (4.54%) (D) 3 人 (2.72%) 無法作答： 0 人 (0%)
	試題 9	97	88.18%	(A) 4 人 (3.63%) (B) 6 人 (5.45%) (C) 3 人 (2.72%) (D) *97 人 (88.18%) 無法作答： 0 人 (0%)
	試題 16	50	45.45%	(A) 17 人 (15.45%) (B) 9 人 (8.18%) (C) *50 人 (45.45%) (D) 12 人 (10.90%) 無法作答： 22 人 (20%)
	整體 層面	236	71.52%	

註：符號說明：*代表此試題之正確題項

(三) 標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果

標準三各試題之分析結果如表 4-1-3 所示，茲將其分析結果簡述如下：

1.臺北市國民中學音樂教師於試題 3 之答對人數為 79 人，答對人數百分比為 71.82%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 3 所欲評估之教學評量素養知能。

2.臺北市國民中學音樂教師於試題 10 之答對人數為 71 人，答對人數百分比為 64.55%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 10 所欲評估之教學評量素養知能。

3.臺北市國民中學音樂教師於試題 17 之答對人數為 48 人，答對人數百分比為 43.64%，未超過判定標準值 60%，故多數臺北市國民中學音樂教師未具備試題 17 所欲評估之教學評量素養知能。

從受試者對其它錯誤題項的選填統計結果觀之，較多填答錯誤的受試者選填題項(C)「一個評分檢核表」(共 37 人，佔總受試者人數的 33.63%)，此填答錯誤人數與比例甚至較選填其它錯誤題項受試者之總和為多。本研究推測造成此種現象的原因，應與部分受試者無有正確釐清(C)題項「一個評分檢核表」與(D)題項「分析式評分基準」的意涵差異有關。究其二者之主要差異，評分檢核表之主要架構僅為評量者所欲測量的各項具體學習行為目標，而分析式評分基準則除條列了各項學習行為目標外，亦清楚標明評分等及其所欲評估的行為特點或範疇，因此，分析式評分基準將較有助於評量者瞭解學生的特定學習問題及其行為表現。因此，題項(C)為與題項(D)相似性頗高的誘答題項，受試者若無法正確判知二者內涵的實質差異，可能傾向於選填題項(C)，因而造成填答錯誤的現象。

4. 整體而言，臺北市國民中學音樂教師於標準三之答對累積人次為 198 人次，累積答對人次百分比為 60.00%，符合判定標準值 60%，故多數臺北市國民中學音樂教師具備「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」所欲評估之教學評量素養知能。

表 4-1-3 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準三（ $n=110$ ）

層面	試題	答對人數	答對人數百分比	各選項答對人數與百分比
標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果	試題 3	79	71.82%	(A) 8 人 （7.27%）
				(B) 2 人 （1.81%）
				(C) 15 人 （13.63%）
				(D) *79 人（71.82%）
				無法作答： 6 人（5.45%）
	試題 10	71	64.55%	(A) 11 人 （10.00%）
				(B) *71 人（64.55%）
				(C) 7 人 （6.36%）
				(D) 6 人 （5.45%）
				無法作答： 15 人（13.63%）
試題 17	48	43.64%	(A) 6 人 （5.45%）	
			(B) 11 人（10.00%）	
			(C) 37 人（33.63%）	
			(D) 48 人（43.64%）	
			無法作答： 8 人（7.27%）	
整體層面		198	60.00%	

註：符號說明：*代表此試題之正確題項

(四) 標準四：針對個別學生進行選擇、規劃教學、發展課程以及學校發展時去使用評量能力的結果

標準四各試題之分析結果如表 4-1-4 所示，茲將其分析結果簡述如下：

1.臺北市國民中學音樂教師於試題 4 之答對人數為 71 人，答對人數百分比為 71.82%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 4 所欲評估之教學評量素養知能。

2.臺北市國民中學音樂教師於試題 11 之答對人數為 57 人，答對人數百分比為 51.82%，未超過判定標準值 60%，故多數臺北市國民中學音樂教師未具備試題 11 所欲評估之教學評量素養知能。

從受試者對其它錯誤題項的選填統計結果觀之，總計有 48 人選填題項(D)「評估學生平時成績」，佔總受試者人數的 43.63%，此人數幾與答對人數的比例相同。本研究推測造成此種現象的原因，應與部分受試者無法正確釐清「形成性評量」概念內涵有關。「形成性評量」之主要目的在於「評估」，即教學者透過形成性評量的實施以瞭解其實際教學成效，並依此作為調整後續教學內容與實施補救教學之依據，甚至也可將此評量成果回饋給學生，以作為其瞭解自我學習狀況與修正學習錯誤的憑藉。但由於我國教育現場長久以來受考試引導教學現象影響，導致教師教學活動過度重視評量活動的形式功能，即透過評量結果區隔學生學習等級或高下，以致於部分受試者在填答此題項時，受其既有刻板印象影響，直覺認為教學活動後所實施之評量活動的主要目的即為評定學生的學習成績，因而造成填答錯誤的現象。

3.臺北市國民中學音樂教師於試題 18 之答對人數為 101 人，答對人數百分比為 91.82%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 18 所欲評估之教學評量素養知能。

4.整體而言，臺北市國民中學音樂教師於標準四之答對累積人次為 229 人次，累積答對人次百分比為 69.39%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備「標準四：針對個別學生進行選擇、規劃教學、發展課程以及學校發展時去使用評量能力的結果」所欲評估之教學評量素養知能。

表 4-1-4 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準四（n=110）

層面	試題	答對人數	答對人數百分比	各選項答對人數與百分比
標準四：針對個別學生進行選擇、規劃教學、發展課程以及學校發展時去使用評量能力的結果	試題 4	71	64.55%	(A) 13 人 (11.81%) (B) *71 人 (64.55%) (C) 6 人 (5.45%) (D) 18 人 (16.36%) 無法作答： 2 人 (1.81%)
	試題 11	57	51.82%	(A) 5 人 (4.54%) (B) 0 人 (0%) (C) *57 人 (51.82%) (D) 48 人 (43.63%) 無法作答： 0 人 (0%)
	試題 18	101	91.82%	(A) 3 人 (2.72%) (B) *101 人 (91.82%) (C) 2 人 (1.81%) (D) 0 人 (0%) 無法作答： 4 人 (3.63%)
	整體 層面	229	69.39%	

註：符號說明：*代表此試題之正確題項

（五）標準五：發展學生評量之有效評分步驟

標準五各試題之分析結果如表 4-1-5 所示，茲將其分析結果簡述如下：

1.臺北市國民中學音樂教師於試題 5 之答對人數為 78 人，答對人數百分比為 70.91%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 5 所欲評估之教學評量素養知能。

2.臺北市國民中學音樂教師於試題 12 之答對人數為 26 人，答對人數百分比為 23.64%，未超過判定標準值 60%，故多數臺北市國民中學音樂教師未具備試題 12 所欲評估之教學評量素養知能。

從受試者對其它錯誤題項的選填統計結果發現：填答錯誤題項(C)「規劃一個統整的評分方法，減少教師在評分上的主觀」的人數為 56 人，佔總受試者人數的 50.90%，選擇此錯誤題項的受試者人數超過總受試者人數的一半；其它較多選填錯誤的選項為題項(D)「為了公平起見，最好限制使用多重的自陳法題目之考試」，其人數為 12 人，佔總受試者人數的 25.45%，選擇此錯誤題項的受試者人數同樣也多於選擇正確題項的人數。本研究推測造成此種現象的原因，應與本題所設計之各題項中，除題項(B)外，其餘之題項(C)與題項(D)皆具備高度的誘答力，題項(C)與題項(D)有出現「為減少主觀」或「為公平起見」，此更讓填答受試者將注意力放置在此兩個題項上，而忽略對其它題項的思考與比較，因而造成填答錯誤的現象。

本題所設計各題項其實在某種程度上皆能有助於減少評分者的評分不一致現象，但由於學生自陳的心向與反應多元，教師不易規劃一個周延統整的評分方法，且若限制使用多重的自陳法題目，也會對學生真實反應的瞭解產生限制。因此，較為適當之方法則為在評定全體第一題後，再進行第二題的評分，除了可獲得較為全面與一致性的評分結果，同時可避免後續評量題目所產生混淆評分現象。

3.臺北市國民中學音樂教師於試題 19 之答對人數為 74 人，答對人數百分比為 67.27%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 19 所欲評估之教學評量素養知能。

4.整體而言，臺北市國民中學音樂教師於標準五之答對累積人次為 178 人次，累積答對人次百分比為 53.94%，未超過判定標準值 60%，故多數臺北市國民中學音樂教師未具備「標準五：發展學生評量之有效評分步驟」所欲評估之教學評量素養知能。

表 4-1-5 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準五（ $n=110$ ）

層面	試題	答對人數	答對人數百分比	各選項答對人數與百分比
標準五：發展學生評量之有效評分步驟	試題 5	78	70.91%	(A) 20 人 (18.18%) (B) 2 人 (1.81%) (C) *78 人 (70.91%) (D) 8 人 (7.27%) 無法作答：2 人 (1.81%)
	試題 12	26	23.64%	(A) *26 人 (23.64%) (B) 0 人 (0%) (C) 56 人 (50.90%) (D) 28 人 (25.45%) 無法作答：0 人 (0%)
	試題 19	74	67.27%	(A) *74 人 (67.27%) (B) 20 人 (18.18%) (C) 10 人 (9.09%) (D) 2 人 (1.81%) 無法作答：4 人 (3.63%)
	整體層面	178	53.94%	

註：符號說明：*代表此試題之正確題項

(六) 標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果

標準六各試題之分析結果如表 4-1-6 所示，茲將其分析結果簡述如下：

1.臺北市國民中學音樂教師於試題 6 之答對人數為 85 人，答對人數百分比為 77.27%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 6 所欲評估之教學評量素養知能。

2.臺北市國民中學音樂教師於試題 13 之答對人數為 74 人，答對人數百分比為 67.27%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 13 所欲評估之教學評量素養知能。

3.臺北市國民中學音樂教師於試題 20 之答對人數為 70 人，答對人數百分比為 63.64%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 20 所欲評估之教學評量素養知能。

4.整體而言，臺北市國民中學音樂教師於標準六之答對累積人次為 229 人次，累積答對人次百分比為 69.39%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」所欲評估之教學評量素養知能。

表 4-1-6 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準六 (n=110)

層面	試題	答對人數	答對人數百分比	各選項答對人數與百分比
標準六：對於學生、 家長、一般民眾與教 育工作者傳達評量 的結果	試題 6	85	77.27%	(A) 8 人 (7.27%)
				(B) 5 人 (4.54%)
				(C) 12 人 (10.90%)
				(D) *85 人 (77.27%)
				無法作答： 0 人 (0%)
	試題 13	74	67.27%	(A) 8 人 (7.27%)
				(B) 3 人 (2.72%)
				(C) *74 人 (67.27%)
				(D) 2 人 (1.81%)
				無法作答： 23 人 (20.90%)

接次頁

表 4-1-6 (續)

			(A) 5 人 (4.54%)
			(B) *70 人 (63.64%)
試題 20	70	63.64%	(C) 19 人 (17.27%)
			(D) 7 人 (6.36%)
			無法作答： 9 人 (8.18%)
整體 層面	229	69.39%	

註：符號說明：*代表此試題之正確題項

(七) 標準七：分辨、不道德、不合法或其他不適當之評量方法與評量資訊使用

標準七各試題之分析結果如表 4-1-7 所示，茲將其分析結果簡述如下：

1.臺北市國民中學音樂教師於試題 7 之答對人數為 98 人，答對人數百分比為 89.09%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 7 所欲評估之教學評量素養知能。

2.臺北市國民中學音樂教師於試題 14 之答對人數為 21 人，答對人數百分比為 19.09%，未超過判定標準值 60%，故多數臺北市國民中學音樂教師未具備試題 14 所欲評估之教學評量素養知能。

從受試者對其它錯誤題項的選填統計結果發現：填答錯誤題項(B)「上述陳述是增加測驗效度的可行方法」的人數為 63 人，佔總受試者人數的 57.27%，選擇此錯誤題項的受試者人數超過總受試者人數的一半。本研究推測造成此種現象的原因，應與部分受試者對「標準化測驗」的意涵產生混淆有關。「標準化測驗」是指依科學方法編製而成的測驗，多由測驗專家根據測驗的編製標準化程序而編成，其標準化包括建立常模、統計信度與效度，並對測驗之施測程序、計分方法、資料解釋...等訂定標準化程序。因此，本題題幹敘述既明確指出「標準化測驗」之施測，施測者應當嚴守既定的各項施測程序與規定，以免造成施測結果的誤差，進而造成標準化測驗結果無法正確反應學生的學習能力與程度。因此，本研究認為受試者在填答此題時，可能缺乏對「標準化測驗」的正確瞭解外，也可能是受平

時編製與施測「非標準化測驗」經驗的影響，而造成對「標準化測驗」的認知混淆，造成填答錯誤的現象。

3.臺北市國民中學音樂教師於試題 21 之答對人數為 79 人，答對人數百分比為 71.82%，超過判定標準值 60%，故多數臺北市國民中學音樂教師具備試題 21 所欲評估之教學評量素養知能。

4.整體而言，臺北市國民中學音樂教師於標準七之答對累積人次為 198 人次，累積答對人次百分比為 60.00%，符合判定標準值 60%，故多數臺北市國民中學音樂教師具備「標準七：分辨、不道德、不合法或其他不適當之評量方法與評量資訊的使用」所欲評估之教學評量素養知能。

表 4-1-7 臺北市國民中學音樂教師教學評量素養現況分析摘要-標準七（ $n=110$ ）

層面	試題	答對人數	答對人數百分比	各選項答對人數與百分比
標準七：分辨、不道德、不合法或其他不適當之評量方法與評量資訊的使用	試題 7	98	89.09%	(A)* 98 人 (89.09%)
				(B) 4 人 (3.63%)
				(C) 6 人 (5.45%)
				(D) 0 人 (0%)
				無法作答： 2 人 (1.81%)
				(A) 7 人 (6.36%)
	試題 14	21	19.09%	(B) 63 人 (57.27%)
				(C) 3 人 (2.72%)
				(D)* 21 人 (19.09%)
				無法作答： 16 人 (14.54%)
				(A)* 79 人 (71.82%)
				(B) 14 人 (63.64%)
	試題 21	79	71.82%	(C) 19 人 (17.27%)
				(D) 7 人 (6.36%)
				無法作答： 9 人 (8.18%)
	整體層面	198	60.00%	

註：符號說明：*代表此試題之正確題項

二、綜合討論

整體而言，臺北市國民中學音樂教師教學評量素養之標準一、二、三、四、六及七均達判定標準值 60%，多數教師具備了此次所欲評估之教學評量素養知能。其可能原因，應與臺北市所擁有之教育資源相較其它縣市來得豐厚有關，臺北市政府相關主管教育行政機關與大學學術單位針對教師所辦理之教學評量知能發展的在職進修研習、工作坊、學位或學分進修班等提供教師多元化的進修選擇，故教師能提升教學評量素養的機會亦相對增加。

七項標準中，標準六之三個子題皆符合判定標準值 60%，臺北市國民中學音樂教師在「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」表現良好。其餘標準皆有一個子題未達判定標準值 60%，分別為試題 1、試題 11、試題 12、試題 14、試題 16 及試題 17，臺北市國民中學音樂教師在「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以及學校發展時去使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」與「標準七：分辨、不道德、不合法或其他不適當之評量方法與評量資訊的使用」之層面未全面符合判定標準值 60%，表現尚可。

本研究臺北市國中音樂教師在標準一：選擇適合於教學決策的評量方法答對人數百分比最高（75.45%），標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果與標準七：分辨、不道德、不合法或其他不是當之評量方法與評量使用答對人數百分比最低（60%）。此與劉向欣（2008）的研究有所不同，其研究對象為國小教師，結果顯示於標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果得分最高（平均數 1.8657），在標準五：去發展學生評量之有效評分步驟的得分上為最低分（平均數 1.0182）。

貳、臺北市國民中學音樂教師教學評量素養差異分析

本節針對不同背景變項之臺北市國民中學音樂教師之教學評量素養差異進行分析，其背景變項包括「性別」、「年齡」、「學歷狀況」與「教職狀況」等；其中，「學歷狀況」可區分為「最高學歷」、「學科專長」與「師培階段是否修習過教育測驗與評量之相關課程」等三項，而「教職狀況」亦包含「年資」、「目前工作職務」、「曾參加過教育測驗與評量相關之研習」與「任教學校規模」等四項。為探究不同背景變項之臺北市國民中學音樂教師教學評量素養的差異情形，本研究採 t 檢定及單因子變異數分析（one-way ANOVA）進行分析。

一、不同性別音樂教師之教學評量素養差異分析

本研究將教師性別變項區分為男性、女性兩組，並以 t 檢定來考驗不同性別臺北市國民中學音樂師之教學評量素養差異情形，分析結果如表 4-1-8 所示。

依分析結果顯示，不同性別音樂教師在教學評量素養整體層面及大部分個別層面並無顯著差異，包括「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」等層面並無顯著差異；但在「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」層面有顯著差異($t=-2.75, p<.01$)，顯示女性音樂教師($M=1.76$)在此層面的表現優於男性音樂教師($M=1.00$)。

表 4-1-8 不同性別音樂教師在教學評量素養量表的獨立樣本 t 考驗

層面	男性 ($n=7$)		女性 ($n=103$)		t	p	95%CI	
	M	SD	M	SD			LL	UL
標準一	2.57	.535	2.18	.668	1.52	.133	-.12	.91
標準二	2.00	1.000	2.08	.753	-.28	.781	-.68	.51
標準三	1.86	1.069	1.69	.888	.46	.646	-.54	.86
標準四	2.29	.756	1.99	.792	.96	.341	-.32	.91
標準五	1.43	1.134	1.52	.770	-.28	.780	-.71	.53
標準六	2.29	.756	1.99	.917	.83	.407	-.41	1.00
標準七	1.00	.817	1.76	.695	-2.75	.007**	-1.30	-.21
整體教學評 量素養	13.43	2.225	13.21	2.95	.19	.849	-2.04	2.48

* $p < .05$ ** $p < .01$ *** $p < .001$

討論：

此結果與劉向欣（2008）研究發現部分類似，即教師整體的教學評量素養及其他各標準均不因教師性別的不同而有所差異，唯在標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果上有所不同，顯示男性教師在此項標準上優於女性教師。

二、不同年齡音樂教師之教學評量素養差異分析

本研究將教師年齡變項區分為「30 歲以下」、「31-40 歲」、「41-50 歲」與「50 歲以上」等四組，並以單因子變異數分析來考驗不同年齡臺北市國民中學音樂師之教學評量素養差異情形，分析結果如表 4-1-9 與表 4-1-10 所示。

依分析結果顯示，不同年齡音樂教師在教學評量素養整體層面達顯著差異（ $F=10.52, p < .001$ ），經 Scheff'e 法事後比較分析得知，年齡為 30 歲以下的音樂老師（ $M=14.82$ ）其整體的教學評量素養高於年齡介於 41-50 歲的音樂老師（ $M=11.38$ ）。在個別標準分層部分，不同年齡音樂師之差異情形簡述如下：

（一）不同年齡音樂教師在個別之「標準一：選擇適合於教學決策的評量方法」層面達顯著差異（ $F=6.32, p < .01$ ），經 Scheff'e 法事後比較分析得知，年齡為 30

歲以下 ($M=2.42$) 與 50 歲以上 ($M=2.67$) 的音樂老師之教學評量素養高於年齡介於 41-50 歲的音樂老師 ($M=1.84$)。

(二) 不同年齡音樂教師在個別之「標準二：發展適合於教學決策的評量方法」層面達顯著差異 ($F = 3.13, p < .05$)，經 Scheff'e 法事後比較分析，並未發現各年齡分組間有明顯差距。

(三) 不同年齡音樂教師在個別之「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」層面達顯著差異 ($F = 3.03, p < .05$)，經 Scheff'e 法事後比較分析，並未發現各年齡分組間有明顯差距。

(四) 不同年齡音樂教師在個別之「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」層面達顯著差異 ($F = 3.74, p < .05$)，經 Scheff'e 法事後比較分析得知，年齡為 30 歲以下 ($M=1.92$) 的音樂老師之教學評量素養高於年齡介於 41-50 歲的音樂老師 ($M=1.38$)。

(五) 不同年齡音樂教師在個別之「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」層面並無顯著差異。

表 4-1-9 不同年齡音樂教師在教學評量素養之各組人數、平均數和信賴區間 ($n=110$)

變項	30 歲以下 ($n=40$)			31-40 歲 ($n=31$)			41-50 歲 ($n=32$)			50 歲以上 ($n=7$)		
	<i>M</i>			<i>M</i>			<i>M</i>			<i>M</i>		
	(<i>SD</i>)	<i>LL</i>	<i>UL</i>	(<i>SD</i>)	<i>LL</i>	<i>UL</i>	(<i>SD</i>)	<i>LL</i>	<i>UL</i>	(<i>SD</i>)	<i>LL</i>	<i>UL</i>
標準一	2.42 (.64)	2.21	2.63	2.23 (.59)	1.99	2.47	1.84 (.63)	1.62	2.07	2.67 (.52)	2.12	3.20
標準二	2.37 (.68)	2.15	2.59	1.92 (.89)	1.56	2.28	1.88 (.71)	1.62	2.13	2.00 (.63)	1.34	2.66
標準三	1.92 (.91)	1.62	2.22	1.73 (.78)	1.42	2.04	1.50 (.92)	1.17	1.82	1.33 (1.03)	.25	2.42
標準四	2.11 (.69)	1.88	2.33	2.08 (.93)	1.70	2.45	1.81 (.78)	1.53	2.09	2.17 (.75)	1.38	2.96
標準五	1.74 (.64)	1.53	1.95	1.54 (.82)	1.21	1.87	1.22 (.83)	.92	1.52	1.50 (1.05)	.40	2.60
標準六	2.34 (.71)	2.11	2.57	1.85 (.92)	1.47	2.22	1.75 (1.05)	1.37	2.13	2.00 (.63)	1.37	2.66
標準七	1.92 (.63)	1.71	2.13	1.77 (.65)	1.51	1.03	1.38 (.79)	1.09	1.66	1.83 (.75)	1.56	1.85
教學評量素養 整體層面	14.82 (2.04)	14.15	15.49	13.12 (2.85)	11.9 7	14.2 7	11.38 (2.84)	10.3 5	12.40	13.50 (2.66)	10.70	16.30

* $p < .05$ ** $p < .01$ *** $p < .001$

表 4-1-10 不同年齡音樂教師之教學評量素養單因子變異數分析 ($n=110$)

層面	<i>SV</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	事後 比較	η^2	$1 - \beta$
標準一	組間	7.25	3	2.42	6.32**	.001	1 > 3 4 > 3	.137	.961
	組內	37.43	98	.38					
	總和	44.68	101						
標準二	組間	5.18	3	1.73	3.13*	.029		.059	.712
	組內	54.19	98	.55					
	總和	59.37	101						
標準三	組間	3.97	3	1.32	1.68	.177			
	組內	77.21	98	.79					
	總和	81.18	101						

接次頁

表 4-1-10 (續)

標準四	組間	1.86	3	.62	.99	.400			
	組內	61.13	98	.62					
	總和	62.99	101						
標準五	組間	4.69	3	1.56	2.61	.056			
	組內	58.80	98	.60					
	總和	63.49	101						
標準六	組間	7.05	3	2.35	3.03*	.033		.057	.697
	組內	75.94	98	.78					
	總和	82.99	101						
標準七	組間	5.47	3	1.82	3.74*	.014	1 > 3	.075	.795
	組內	47.71	98	.49					
	總和	53.18	101						
教學評量素養 整體層面	組間	206.45	3	68.82	10.52***	.000	1 > 3	.220	1.000
	組內	641.36	98	6.55					
	總和	847.81	101						

* $p < .05$ ** $p < .01$ *** $p < .001$

註：1 表示「30 歲以下」，2 表示「31-40 歲」，3 表示「41-50 歲」，4 表示「50 歲以上」

討論：

綜合上述分析，研究結果發現，不同年齡音樂教師在教學評量素養整體層面達顯著差異，年齡為 30 歲以下的音樂老師其整體的教學評量素養高於年齡介於 41-50 歲的音樂老師。造成此種情形之可能原因，應與 30 歲以下教師因剛修畢師資培育課程與通過教師檢定及考試有關。由於教學評量乃為近年來新興之教學理論，對於 30 歲以下之教師而言，其所修習之師資培育課程內容應有涵括此教學理論的教授，甚至對於剛通過教師檢定及考試之新手教師而言，由於其有準備相關教師甄試的需求，故其對新興教學理論的關注程度乃相對高於已是正式教師之其他年齡層的教師。

三、不同學歷音樂教師之教學評量素養差異分析

本研究將教師學歷變項區分為「專科畢業」、「學士」與「碩博士」等三組，並擬以單因子變異數分析來考驗不同學歷臺北市國民中學音樂師之教學評量素養差異情形，但在正式問卷回收後，並無受試者歸屬在學歷變項中之「專科畢業」類別中，因此，在考量教師學歷變項僅餘「學士」與「碩博士」兩組，故本研究乃改採 t 檢定來考驗不同學歷臺北市國民中學音樂師之教學評量素養差異情形，分析結果如表 4-1-11 所示。

依分析結果顯示，不同學歷音樂教師在教學評量素養整體層面及大部分個別層面並無顯著差異，包括：「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」、「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」等層面並無顯著差異；但在「標準一：選擇適合於教學決策的評量方法」層面有顯著差異（ $t=-2.12, p<.05$ ），顯示學歷為學士之音樂教師（ $M=2.36$ ）在此層面的表現優於學歷為碩博士之音樂教師（ $M=2.09$ ）。

表 4-1-11 不同學歷音樂教師在教學評量素養量表的獨立樣本 t 考驗

層面	學士 ($n=45$)		碩博士 ($n=65$)		t	p	95%CI	
	M	SD	M	SD			LL	UL
標準一	2.36	.650	2.09	.657	2.12*	.036	.02	.54
標準二	2.11	.722	2.05	.804	.40	.688	-.24	.37
標準三	1.64	.917	1.76	.885	-.68	.498	-.48	.23
標準四	2.02	.820	2.00	.772	.14	.886	-.29	.34
標準五	1.57	.818	1.47	.777	.65	.520	-.21	.42
標準六	1.89	.993	2.10	.831	-1.20	.233	-.58	.14
標準七	1.82	.691	1.62	.745	1.37	.175	-.09	.48
整體教學評量素養	13.41	2.847	13.09	2.952	.56	.580	-.83	1.48

* $p<.05$ ** $p<.01$ *** $p<.001$

討論：

綜合上述分析，國中音樂教師在「標準一：選擇適合於教學決策的評量方法」層面顯示，學歷為學士之音樂教師在此層面的表現優於學歷為碩博士之音樂教師。究其可能原因，雖然碩、博學歷之音樂教師擁有較高的學歷，但此部分之差異應顯現於其在音樂領域素養高於學士學歷的教師，此無法與教學評量領域素養等同視之。

四、不同學科專長音樂教師之教學評量素養差異分析

本研究將教師學科專長變項區分為「音樂相關科系」、「非音樂相關科系」兩組，並以 t 檢定來考驗不同學科專長臺北市國民中學音樂師之教學評量素養差異情形，分析結果如表 4-1-12 所示。

依分析結果顯示，不同學科專長音樂教師在教學評量素養整體層面及大部分個別層面並無顯著差異，包括：「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」、「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」等層面並無顯著差異；但在「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」層面有顯著差異（ $t=-3.32, p<.01$ ），顯示非音樂相關科系之音樂教師（ $M=2.00$ ）在此層面的表現優於音樂相關科系之音樂教師（ $M=1.70$ ）。

表 4-1-12 不同學科專長音樂教師在教學評量素養量表的獨立樣本 t 考驗

層面	音樂相關科系 ($n=107$)		非音樂相關科系 ($n=3$)		t	p	95%CI	
	M	SD	M	SD			LL	UL
標準一	2.19	.665	2.67	.577	-1.22	.225	-1.25	.300
標準二	2.10	.763	1.33	.577	1.73	.088	-.12	1.65
標準三	1.70	.909	2.00	.000	-3.32**	.001	-.48	-.12
標準四	1.99	.789	2.67	.577	-1.47	.145	-1.590	.236
標準五	1.51	.800	1.67	.577	-.35	.730	-1.087	.764
標準六	2.00	.915	2.33	.577	-.63	.533	-1.39	.724
標準七	1.69	.723	2.33	.577	-1.53	.129	-1.48	.192
整體教學評 量素養	13.17	2.910	15.00	2.000	-1.08	.284	-5.19	1.54

* $p < .05$ ** $p < .01$ *** $p < .001$

討論：

不同學科專長音樂教師在教學評量素養「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」層面顯示非音樂相關科系之音樂教師在此層面的表現優於音樂相關科系之音樂教師。究其可能原因，應與非音樂相關科系之教師除其自身擁有之一般大學科系專長外，同時在專業師資培育的過程中，學習教育領域之各項教育理論知能，因此，在其執行、分析、評分與詮釋外界發展或教師自製之評量結果時，除能透過自身教育專業的角度來進行思考外，同時能跳脫教育專業的框架限制，利用一般大學科系專長來輔助教育專業的進行。因此，非音樂相關科系之音樂教師在此層面的表現優於音樂相關科系之音樂教師。

五、不同修習教育測驗與評量課程經驗音樂教師之教學評量素養差異分析

本研究將修習教育測驗與評量課程經驗變項區分為「曾在師培階段修習過教育測驗與評量之相關課程」與「不曾在師培階段修習過教育測驗與評量之相關課程」兩組，並以 t 檢定來考驗不同修習教育測驗與評量課程經驗臺北市國民中學音樂師之教學評量素養差異情形，分析結果如表 4-1-13 所示。

依分析結果顯示，不同修習教育測驗與評量課程經驗音樂教師在教學評量素養整體層面及大部分個別層面上並無顯著差異，包括：「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」、「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」等層面皆無顯著差異。

表 4-1-13 不同修習教育測驗與評量課程經驗音樂教師在教學評量素養量表獨立樣本 *t* 考驗

層面	曾有修習經驗 (<i>n</i> =83)		不曾有修習經驗 (<i>n</i> =27)		<i>t</i>	<i>p</i>	95% <i>CI</i>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			<i>LL</i>	<i>UL</i>
標準一	2.21	.596	2.19	.849	.12	.905	-.28	.32
標準二	2.09	.769	2.04	.774	.31	.760	-.29	.40
標準三	1.76	.862	1.54	.989	1.10	.272	-.18	.63
標準四	1.92	.812	2.27	.666	-1.97	.052	-.70	.00
標準五	1.59	.802	1.27	.724	1.81	.073	-.03	.68
標準六	2.11	.888	1.73	.919	1.84	.069	-.03	.78
標準七	1.74	.737	1.62	.697	.74	.464	-.21	.45
整體教學評量素養	13.42	2.927	12.65	2.785	1.17	.246	-.54	2.07

p* < .05 *p* < .01 ****p* < .001

討論：

上述研究發現，不同修習教育測驗與評量課程經驗音樂教師在音樂教學評量素養整體層面及個人之七項標準皆無顯著差異。究其可能原因，雖然有修習教育測驗與評量課程經驗之教師有其相當程度的音樂教學評量素養，但對於未有修習教育測驗與評量課程經驗之教師而言，仍有其它專業成長的管道與方式來提升自身音樂教學評量的知能，如透過相關主管教育行政機關或學校單位辦理之研習進修活動、教學工作坊、教學講座、教師專業成長社群...等，此皆有助於教師提升自

身的教育專業知能。因此，不同修習教育測驗與評量課程經驗之音樂教師在音樂教學評量素養整體層面及各分項標準乃無顯著差異。

六、不同年資音樂教師之教學評量素養差異分析

本研究將教師年資變項區分為「5 年以下」、「6-10 年」、「11-15 年」、「16-20 年」與「20 年以上」等五組，並以單因子變異數分析來考驗不同年資臺北市國民中學音樂師之教學評量素養差異情形，分析結果如表 4-1-14 與表 4-1-15 所示。

依分析結果顯示，不同年資音樂教師在教學評量素養整體層面達顯著差異（ $F=4.79, p<.01$ ），經 Scheff'e 法事後比較分析得知，年資在 5 年以下的音樂老師（ $M=14.67$ ）其整體的教學評量素養高於年資在 11-15 年的音樂老師（ $M=11.27$ ）。在個別標準分層部分，不同年資音樂教師在大部分個別層面上無顯著差異，包括：「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」、「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」等層面皆無顯著差異。

表 4-1-14 不同年資音樂教師在教學評量素養之各組人數、平均數和信賴區間（ $n=110$ ）

變項	5 年以下 ($n=37$)			6-10 年 ($n=21$)			11-15 年 ($n=13$)			16-20 年 ($n=17$)			20 年以上 ($n=22$)		
	<i>M</i> (<i>SD</i>)	95% <i>CI</i>		<i>M</i> (<i>SD</i>)	95% <i>CI</i>		<i>M</i> (<i>SD</i>)	95% <i>CI</i>		<i>M</i> (<i>SD</i>)	95% <i>CI</i>		<i>M</i> (<i>SD</i>)	95% <i>CI</i>	
標準一	2.39 (.64)	2.17	2.61	2.24 (.56)	1.95	2.52	2.00 (1.00)	1.33	2.67	1.94 (.44)	1.70	2.17	2.18 (.66)	1.89	2.48
標準二	2.33 (.63)	2.12	2.55	1.94 (.83)	1.52	2.37	1.73 (.90)	1.12	2.33	1.94 (.85)	1.48	2.39	2.05 (.72)	1.73	2.37
標準三	1.92 (.91)	1.61	2.22	1.59 (.80)	1.18	2.00	1.27 (.90)	.67	1.88	1.88 (.89)	1.40	2.35	1.55 (.91)	1.14	1.95

接次頁

表 4-1-14 (續)

標準四	2.17 (.70)	1.93	2.40	2.06 (.83)	1.63	2.48	1.73 (.90)	1.12	2.33	1.94 (1.00)	1.41	2.47	1.91 (.68)	1.61	2.21
標準五	1.75 (.65)	1.53	1.97	1.53 (.94)	1.04	2.01	1.55 (.69)	1.08	2.01	1.31 (.87)	.85	1.78	1.23 (.81)	.87	1.59
標準六	2.22 (.72)	1.98	2.47	2.12 (1.05)	1.58	2.66	1.36 (1.12)	.61	2.12	1.88 (1.09)	1.30	2.45	2.00 (.69)	1.69	2.31
標準七	1.89 (.62)	1.68	2.10	1.71 (.99)	1.20	2.21	1.64 (.67)	1.18	2.09	1.56 (.73)	1.17	1.95	1.55 (.67)	1.25	1.84
教學評量 素養整體	14.67 (2.07)	13.9 7	15.37	13.18 (3.30)	11.4 8	14.87	11.27 (3.32)	9.04	13.5 0	12.44 (2.85)	10.9 2	13.9 6	12.45 (2.69)	11.26	13.6 5

* $p < .05$ ** $p < .01$ *** $p < .001$ 表 4-1-15 不同年資音樂教師之教學評量素養單因子變異數分析 ($n=110$)

層面	SV	SS	df	MS	F	p	事後 比較	η^2	1 -
標準一	組間	2.85	4	.71	1.65	.167			
	組內	41.83	97	.43					
	總和	44.68	101						
標準二	組間	4.36	4	1.09	1.92	.113			
	組內	55.02	97	.57					
	總和	59.38	101						
標準三	組間	4.92	4	1.23	1.57	.190			
	組內	76.25	97	.786					
	總和	81.17	101						
標準四	組間	2.11	4	.53	.84	.502			
	組內	60.88	97	.63					
	總和	62.99	101						
標準五	組間	4.48	4	1.12	1.84	.127			
	組內	59.01	97	.61					
	總和	63.49	101						
標準六	組間	6.71	4	1.68	2.13	.083			
	組內	76.28	97	.79					
	總和	82.99	101						
標準七	組間	2.15	4	.54	1.02	.399			
	組內	51.02	97	.53					
	總和	53.17	101						
教學評量素養 整體層面	組間	139.77	4	34.94	4.79**	.001	1 > 3	.165	.946
	組內	708.04	97	7.30					
	總和	847.81	101						

* $p < .05$ ** $p < .01$ *** $p < .001$

註：1 表示「5 年以下」，2 表示「6-10 年」，3 表示「11-15 年」，4 表示「16-20 年」
5 表示「20 年以上」

討論：

不同年資音樂教師在教學評量素養整體層面達顯著差異，年資在 5 年以下的音樂老師其整體的教學評量素養高於年資在 11-15 年的音樂老師。此部分研究結果可與上述「不同年齡音樂教師在教學評量素養整體層面達顯著差異」之結果相互呼應，其可能原因，由於教學年資較淺之教師通常為年齡較輕之新進教師，因此，在其修習師資培育課程或準備教師檢定、甄試時，對教學評量理論的內涵而有所接觸，此對教學年資較久之資深教師而言，除非其透過相關在職進修或研習活動來強化對教學評量知能的瞭解，否則在其過往接受師資培訓的過程中，較無機會接觸此種新興理論的學習。

七、不同工作職務音樂教師之教學評量素養差異分析

本研究將教師工作職務變項區分為「音樂專任教師」、「級任導師」、「兼行政」與「代課教師」等四組，並以單因子變異數分析來考驗不同工作職務臺北市國民中學音樂師之教學評量素養差異情形，分析結果如表 4-1-16 與表 4-1-17 所示。

依分析結果顯示，不同工作職務音樂教師在教學評量素養整體層面及大部分個別層面無顯著差異，包括：「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」等層面皆無顯著差異。但在「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」層面有顯著差異（ $F = 3.50, p < .05$ ），經 Scheff'e 法事後比較分析得知，各組間之差異並未達顯著差異。

表 4-1-16 不同工作職務音樂教師在教學評量素養之各組人數、平均數和信賴區間
($n=110$)

變項	音樂專任教師 ($n=61$)			級任導師 ($n=12$)			兼行政 ($n=23$)			代課教師 ($n=14$)		
	<i>M</i> (<i>SD</i>)	95% <i>CI</i> <i>LL UL</i>		<i>M</i> (<i>SD</i>)	95% <i>CI</i> <i>LL UL</i>		<i>M</i> (<i>SD</i>)	95% <i>CI</i> <i>LL UL</i>		<i>M</i> (<i>SD</i>)	95% <i>CI</i> <i>LL UL</i>	
標準一	2.20 (.72)	2.00	2.39	2.33 (.65)	1.92	2.75	2.33 (.58)	2.07	2.60	1.92 (.49)	1.62	2.22
標準二	2.02 (.70)	1.83	2.21	1.83 (.94)	1.24	2.43	2.33 (.86)	1.94	2.72	2.15 (.69)	1.74	2.57
標準三	1.59 (.87)	1.36	1.82	1.50 (.80)	.99	2.01	2.05 (1.02)	1.58	2.51	1.85 (.80)	1.36	2.33
標準四	1.98 (.82)	1.76	2.20	1.92 (.90)	1.34	2.49	2.19 (.60)	1.92	2.46	1.92 (.86)	1.40	2.44
標準五	1.54 (.83)	1.31	1.76	1.83 (.83)	1.30	2.36	1.24 (.70)	.92	1.56	1.54 (.66)	1.14	1.94
標準六	1.89 (.93)	1.64	2.14	2.17 (.72)	1.71	2.62	2.24 (.89)	1.83	2.64	2.00 (1.00)	1.40	2.60
標準七	1.63 (.73)	1.43	1.82	2.17 (.72)	1.71	2.62	1.48 (.68)	1.17	1.79	2.00 (.58)	1.65	2.35
教學評量素養 整體層面	12.84 (3.14)	12.00	13.68	13.75 (2.22)	12.34	15.16	13.86 (2.48)	12.73	14.98	13.38 (3.01)	11.56	15.21

* $p < .05$ ** $p < .01$ *** $p < .001$

表 4-1-17 不同工作職務音樂教師之教學評量素養單因子變異數分析 ($n=110$)

層面	SV	SS	df	MS	<i>F</i>	<i>p</i>	事後 比較	η^2	$1 - \beta$
標準一	組間	1.58	3	.53	1.20	.315			
	組內	43.10	98	.44					
	總和	44.68	101						
標準二	組間	2.37	3	.79	1.36	.261			
	組內	57.01	98	.58					
	總和	59.38	101						
標準三	組間	3.98	3	1.33	1.68	.186			
	組內	77.20	98	.79					
	總和	81.18	101						
標準四	組間	.93	3	.31	.49	.690			
	組內	62.06	98	.63					
	總和	62.99	101						
標準五	組間	2.86	3	.95	1.54	.210			
	組內	60.64	98	.62					
	總和	63.50	101						

接次頁

表 4-1-17 (續)

標準六	組間	2.16	3	.72	.87	.459		
	組內	80.83	98	.83				
	總和	82.99	101					
標準七	組間	5.15	3	1.72	3.50*	.018	.069	.765
	組內	48.03	98	.49				
	總和	53.18	101					
教學評量素養 整體層面	組間	23.36	3	6.79	.80	.495		
	組內	827.45	98	8.44				
	總和	847.81	101					

* $p < .05$ ** $p < .01$ *** $p < .001$

八、不同教育測驗與評量研習經驗音樂教師之教學評量素養差異分析

本研究將參與教育測驗與評量研習經驗變項區分為「曾參與過教育測驗與評量相關之研習」與「不曾參與過教育測驗與評量相關之研習」兩組，並以 t 檢定來考驗不同參與教育測驗與評量研習經驗臺北市國民中學音樂師之教學評量素養差異情形，分析結果如表 4-1-18 所示。

依分析結果顯示，不同參與教育測驗與評量研習經驗的音樂教師在教學評量素養整體層面及大部分個別層面無顯著差異，包括：「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」、「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」等層面並無顯著差異；但在「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」層面有顯著差異（ $t=-2.60, p<.05$ ），顯示不曾參與過教育測驗與評量相關研習之音樂教師（ $M=2.28$ ）在此層面的表現優於曾參與過教育測驗與評量相關研習之音樂教師（ $M=1.86$ ）。

表 4-1-18 不同參與教育測驗與評量研習經驗音樂教師在教學評量素養量表獨立樣本 t 考驗

層面	曾參與相關研習 ($n=69$)		不曾參與相關研習 ($n=41$)		t	p	95%CI	
	M	SD	M	SD			LL	UL
標準一	2.23	.63	2.17	.74	.44	.662	-.21	.34
標準二	2.02	.75	2.19	.79	-1.13	.261	-.49	.14
標準三	1.64	.92	1.83	.85	-1.06	.291	-.57	.17
標準四	1.86	.76	2.28	.79	-2.60*	.011	-.73	-.10
標準五	1.53	.83	1.47	.74	.35	.726	-.27	.39
標準六	1.98	.94	2.06	.86	-.38	.709	-.44	.30
標準七	1.64	.74	1.83	.70	-1.32	.192	-.49	.10
整體教學評量素養	12.89	2.87	13.83	2.88	-1.58	.118	-2.12	.24

* $p < .05$ ** $p < .01$ *** $p < .001$

九、不同學校規模音樂教師之教學評量素養差異分析

本研究將學校規模變項區分為「12 班以下」、「13-36 班」與「37 班以上」等三組，並以單因子變異數分析來考驗不同學校規模臺北市國民中學音樂師之教學評量素養差異情形，分析結果如表 4-1-19 與表 4-1-20 所示。

依分析結果顯示，不同工作職務音樂教師在教學評量素養整體層面及大部分個別層面無顯著差異，包括：「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」、「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」等層面皆無顯著差異。但在「標準一：選擇適合於教學決策的評量方法」層面有顯著差異（ $F = 3.69, p < .05$ ），經 Scheff'e 法事後比較分析得知，各組間之差異並未達顯著差異。

表 4-1-19 不同學校規模音樂教師在教學評量素養之各組人數、平均數和信賴區間

變項	12 班以下 (n=15)			13-26 班 (n=55)			37 班以上 (n=40)		
	95%CI			95%CI			95%CI		
	M (SD)	LL	UL	M (SD)	LL	UL	M (SD)	LL	UL
標準一	2.50 (.52)	2.20	2.80	2.00 (.74)	1.76	2.24	2.28 (.61)	2.11	2.45
標準二	2.00 (.78)	1.55	2.45	2.18 (.77)	1.93	2.44	2.02 (.77)	1.80	2.24
標準三	1.65 (.93)	1.11	2.18	1.74 (1.00)	1.41	2.07	1.70 (.81)	1.47	1.93
標準四	2.29 (.73)	1.87	2.71	1.87 (.81)	1.60	2.14	2.04 (.78)	1.82	2.26
標準五	1.71 (.61)	1.36	2.07	1.29 (.77)	1.04	1.54	1.62 (.83)	1.38	1.86
標準六	2.14 (1.03)	1.55	2.74	1.76 (1.05)	1.42	2.11	2.16 (.71)	1.96	2.36
標準七	1.64 (.74)	1.21	2.07	1.66 (.78)	1.40	1.91	1.76 (.69)	1.56	1.96
教學評量素養 整體層面	13.93 (2.09)	12.72	15.14	12.50 (3.30)	11.41	13.59	13.58 (2.69)	12.82	14.34

* $p < .05$ ** $p < .01$ *** $p < .001$

表 4-1-20 不同學校規模音樂教師之教學評量素養單因子變異數分析 ($n=110$)

層面	SV	SS	df	MS	F	p	事後 比較	η^2	1 -
標準一	組間	3.10	2	1.55	3.69*	.029		.051	.665
	組內	41.58	99	.42					
	總和	44.68	101						
標準二	組間	.68	2	.34	.58	.564			
	組內	58.69	99	.59					
	總和	59.37	101						
標準三	組間	.09	2	.05	.06	.944			
	組內	81.08	99	.82					
	總和	81.17	101						
標準四	組間	1.87	2	.94	1.52	.225			
	組內	61.12	99	.62					
	總和	62.99	101						
標準五	組間	3.04	2	1.52	2.49	.088			
	組內	60.45	99	.61					
	總和	63.49	101						
標準六	組間	3.69	2	1.84	2.30	.105			
	組內	79.30	99	.80					
	總和	82.99	101						
標準七	組間	.29	2	.15	.27	.763			
	組內	52.89	99	.53					
	總和	53.18	101						
教學評量素養 整體層面	組間	33.21	2	.15	.27	.763			
	組內	814.61	99	.53					
	總和	847.82	101						

下表為不同教師背景變項與教師評量素養七個標準之差異情形綜合分析：

表 4-1-21 不同教師背景變項與教師評量素養七個標準之差異情形綜合分析

STCEAS	標準 一	標準 二	標準 三	標準 四	標準 五	標準六	標準 七	整體
教師性別	---	---	---	---	---	---	女性 >男 性	---
年齡	30 以下 與 50 歲 以上> 41-50 歲	---	---	---	---	---	30 歲 以下 > 41-50 歲	30 歲以 下> 41-50 歲
學歷	學士> 碩博士	---	---	---	---	---	---	---
學科專長	---	---	非音樂 相關科 系>音 樂相關 科系	---	---	---	---	---
修習相關 課程	---	---	---	---	---	---	---	---
年資	---	---	---	---	---	---	---	5 年以 下> 11-15 年
工作職務	---	---	---	---	---	---	有顯 著差 異	---
修習相關 研習	---	---	---	無>有	---	---	---	---
學校規模	有顯著 差異	---	---	---	---	---	---	---

註：符號說明：---代表未達顯著水準

第二節 臺北市國民中學音樂教師教學評量實施現況與差異分析

本節共分為兩部份，分別為臺北市國民中學音樂教師教學評量實施現況分析與臺北市國民中學音樂教師教學評量實施差異分析，資料如下：

壹、臺北市國民中學音樂教師教學評量實施現況分析

一、臺北市國民中學音樂教師整體教學評量實施現況分析

本節透過描述性統計來分析臺北市國民中學音樂教師教學評量實施的現況，並從而對填答現況進行分析。本研究問卷採 Likert 式五點量表，五個填答選項分別為「總是」、「經常」、「偶爾」、「很少」與「從未」，在進行統計分析時，各選項依序分別採計「5分」、「4分」、「3分」、「2分」與「1分」。若其平均得分高於 3.4 分，則代表受試者在其音樂教學活動中，經常融入該項教學評量概念；反之，若其平均得分低於 3.4，則代表受試者在其音樂教學活動中，較少融入該項教學評量概念。臺北市國民中學音樂教師教學評量實施之整體與各分層面現況分析結果表 4-2-1 所示，簡述如下。

表 4-2-1 臺北市國民中學音樂教師教學評量實施現況分析摘要

變項	層面	最小值	最大值	平均數	標準差
教師 教學 評量 實施	1.實施教學評量前之預備情形	1.50	5.00	3.63	.645
	2.實施教學評量時之情形	2.67	4.67	3.56	.364
	3.實施教學評量後之結果運用情形	2.63	5.00	3.89	.528
	整體層面	2.70	4.70	3.67	.373

臺北市國民中學音樂教師教學評量實施量表之各題項平均得分最小值為 2.70 分，最大值為 4.70 分，其各題項平均得分為 3.67 分（標準差為.373），顯示臺北市國民中學音樂教師在其音樂領域教學的各階段學習活動中，能經常融入教學評量的各項實施概念與策略。

二、臺北市國民中學音樂教師教學評量實施各分層現況分析

（一）實施教學評量前之預備情形

臺北市國民中學音樂教師之「實施教學評量前之預備情形」層面各題項平均得分最小值為 1.50 分，最大值為 5.00 分，其各題項平均得分為 3.63 分（標準差為.645），顯示臺北市國民中學音樂教師在實施教學評量前之預備活動中，能經常依據九年一貫課程綱要、學校自訂之音樂評量標準與教科書提供之評量標準來預先訂定教學活動的各項具體學習評量目標，並能利用自編學習單或測驗工具來彈性與適性評量學生學習能力。

（二）實施教學評量時之情形

臺北市國民中學音樂教師之「實施教學評量時之情形」層面各題項平均得分最小值為 2.67 分，最大值為 4.67 分，其各題項平均得分為 3.56 分（標準差為.364），顯示臺北市國民中學音樂教師在實施教學評量時，能經常向學生說明教學評量的各項規準，以使學生瞭解課程學習活動的方向與重點，並能運用各項音樂教學項目建立多元化的教學評量規準，及利用多元評量方法以適切評估學生的各項實際學習能力表現。

（三）實施教學評量後之結果運用情形

臺北市國民中學音樂教師之「實施教學評量後之結果運用情形」層面各題項平均得分最小值為 2.63 分，最大值為 5.00 分，其各題項平均得分為 3.89 分（標準

差為.528)，顯示臺北市國民中學音樂教師在實施教學評量後，不僅能適切記錄學生的學習成果，亦能經常運用評量結果來檢視課程教學目標的達成程度與學生實際學習成效，並據以彈性修正後續課程教學內容、進度與教學方法，甚或據以規劃合宜適切的補救教學措施以建立學生應有之基本學習知能。

三、臺北市國民中學音樂教師教學評量量表試題填答分析

本段針對受試者對「臺北市國民中學音樂教師教學評量量表」的填答結果進行填答分析，以瞭解受試者的填答狀況。

(一) 實施教學評量前之預備情形

「臺北市國民中學音樂教師教學評量量表」層面一之各題項填答分析摘要表如下所示。

表 4-2-2 臺北市國民中學音樂教師教學評量量表題項填答分析層面一

層面	題向	從未 (%) (n)	很少 (%) (n)	偶爾 (%) (n)	經常 (%) (n)	總是 (%) (n)	題平均數 (標準差)
實施教學 評量前之 預備情形	1.您會依據九年一貫課程綱要訂定評量目標	2% (2)	9% (10)	34% (37)	45% (50)	10% (11)	3.53 (.841)
	2.您會依據學校自訂之音樂評量標準訂定評量目標	3% (3)	11% (12)	16% (18)	54% (59)	16% (18)	3.70 (.948)
	3.您會依據教科書所提供之評量目標訂定評量目標	5% (6)	10% (11)	35% (38)	44% (48)	6% (7)	3.36 (.928)
	4.您使用自編學習單來評量學生的表現	2% (2)	7% (8)	16% (18)	47% (52)	28% (30)	3.94 (.921)

1.臺北市國民中學音樂教師於試題 1 中，選填選項「從未」的人數為 2 人，佔總填答人數的 2%，選填選項「很少」的人數為 10 人，佔總填答人數的 9%，選填選項「偶爾」的人數為 37 人，佔總填答人數的 34%，選填選項「經常」的人數為

50 人，佔總填答人數的 45%，選填選項「總是」的人數為 11 人，佔總填答人數的 10%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.53（標準差為.841）。

2.臺北市國民中學音樂教師於試題 2 中，選填選項「從未」的人數為 3 人，佔總填答人數的 3%，選填選項「很少」的人數為 12 人，佔總填答人數的 11%，選填選項「偶爾」的人數為 18 人，佔總填答人數的 16%，選填選項「經常」的人數為 59 人，佔總填答人數的 54%，選填選項「總是」的人數為 18 人，佔總填答人數的 16%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.70（標準差為.948）。

3.臺北市國民中學音樂教師於試題 3 中，選填選項「從未」的人數為 6 人，佔總填答人數的 5%，選填選項「很少」的人數為 11 人，佔總填答人數的 10%，選填選項「偶爾」的人數為 38 人，佔總填答人數的 35%，選填選項「經常」的人數為 48 人，佔總填答人數的 44%，選填選項「總是」的人數為 7 人，佔總填答人數的 6%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.36（標準差為.928）。

4.臺北市國民中學音樂教師於試題 4 中，選填選項「從未」的人數為 2 人，佔總填答人數的 2%，選填選項「很少」的人數為 8 人，佔總填答人數的 7%，選填選項「偶爾」的人數為 18 人，佔總填答人數的 16%，選填選項「經常」的人數為 52 人，佔總填答人數的 47%，選填選項「總是」的人數為 30 人，佔總填答人數的 28%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.94（標準差為.921）。

(二) 實施音樂教學評量時之情形

「臺北市國民中學音樂教師教學評量量表」層面二之各題項填答分析摘要表如下所示。

表 4-2-3 臺北市國民中學音樂教師教學評量量表題項填答分析層面二

層面	題向	從未 (%) (n)	很少 (%) (n)	偶爾 (%) (n)	經常 (%) (n)	總是 (%) (n)	題平均數 (標準差)
實施 教學 評量 前之 預備 情形	1.您在實施教學評量時，會先與學生解釋此考試的評量規準	0% (0)	4% (4)	17% (19)	40% (44)	39% (43)	4.15 (.784)
	2.您在實施評量時，班上學生願意配合評量之進行	1% (1)	5% (5)	18% (21)	33% (36)	43% (47)	4.12 (.902)
	3.您使用音樂知識作為評量的項目	1% (1)	4% (4)	27% (30)	47% (52)	21% (23)	3.84 (.797)
	4.您使用音感認譜作為評量的項目	3% (3)	14% (15)	32% (35)	35% (39)	16% (18)	3.49 (1.000)
	5.您使用歌唱作為評量的項目	1% (1)	9% (10)	24% (26)	42% (47)	24% (26)	3.79 (.920)
	6.您使用樂器演奏作為評量的項目	0% (0)	6% (7)	15% (17)	47% (51)	32% (35)	4.04 (.817)
	7.您使用創作作為評量的項目	4% (4)	21% (23)	43% (48)	24% (26)	8% (9)	3.12 (.935)
	8.您使用欣賞作為評量的項目	3% (3)	7% (8)	22% (24)	54% (59)	14% (16)	3.70 (.883)
	9.您使用紙筆測驗來評量學生的表現	4% (4)	19% (21)	39% (42)	31% (34)	8% (9)	3.21 (.952)
	10.您使用實作評量來評量學生的表現	4% (4)	10% (11)	23% (25)	45% (50)	18% (20)	3.65 (1.000)
	11.您使用檔案評量來評量學生的表現	5% (6)	20% (22)	22% (24)	31% (34)	22% (24)	3.44 (1.198)
	12.您使用評定量表來評量學生的表現	5% (6)	12% (13)	30% (33)	31% (34)	22% (24)	3.52 (1.123)
	13.您在實施評量時，會遇到學生沒有練習或準備的困難	3% (3)	21% (23)	53% (58)	18% (20)	5% (6)	3.04 (.814)

接次頁

表 4-2-3 (續)

14.您在實施評量時，會遇到班上秩序難掌握的困難	1% (1)	25% (27)	37% (41)	32% (35)	5% (6)	3.16 (.868)
15.您在實施評量時，會遇到學生及家長不重視評量的過程和結果	4% (4)	25% (28)	38% (42)	23% (25)	10% (11)	3.10 (.998)

1.臺北市國民中學音樂教師於試題 1 中，選填選項「從未」的人數為 0 人，佔總填答人數的 0%，選填選項「很少」的人數為 4 人，佔總填答人數的 4%，選填選項「偶爾」的人數為 19 人，佔總填答人數的 17%，選填選項「經常」的人數為 44 人，佔總填答人數的 40%，選填選項「總是」的人數為 43 人，佔總填答人數的 39%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 4.15（標準差為.784）。

2.臺北市國民中學音樂教師於試題 2 中，選填選項「從未」的人數為 1 人，佔總填答人數的 1%，選填選項「很少」的人數為 5 人，佔總填答人數的 5%，選填選項「偶爾」的人數為 21 人，佔總填答人數的 18%，選填選項「經常」的人數為 36 人，佔總填答人數的 33%，選填選項「總是」的人數為 47 人，佔總填答人數的 43%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 4.12（標準差為.902）。

3.臺北市國民中學音樂教師於試題 3 中，選填選項「從未」的人數為 1 人，佔總填答人數的 1%，選填選項「很少」的人數為 4 人，佔總填答人數的 4%，選填選項「偶爾」的人數為 30 人，佔總填答人數的 27%，選填選項「經常」的人數為 52 人，佔總填答人數的 47%，選填選項「總是」的人數為 23 人，佔總填答人數的 21%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.84（標準差為.797）。

4.臺北市國民中學音樂教師於試題 4 中，選填選項「從未」的人數為 3 人，佔總填答人數的 3%，選填選項「很少」的人數為 15 人，佔總填答人數的 14%，選填選項「偶爾」的人數為 35 人，佔總填答人數的 32%，選填選項「經常」的人數

為 39 人，佔總填答人數的 35%，選填選項「總是」的人數為 18 人，佔總填答人數的 16%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.49（標準差為 1.000）。

5.臺北市國民中學音樂教師於試題 5 中，選填選項「從未」的人數為 1 人，佔總填答人數的 1%，選填選項「很少」的人數為 10 人，佔總填答人數的 9%，選填選項「偶爾」的人數為 26 人，佔總填答人數的 24%，選填選項「經常」的人數為 47 人，佔總填答人數的 42%，選填選項「總是」的人數為 26 人，佔總填答人數的 24%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.79（標準差為.920）。

6.臺北市國民中學音樂教師於試題 6 中，選填選項「從未」的人數為 0 人，佔總填答人數的 0%，選填選項「很少」的人數為 7 人，佔總填答人數的 6%，選填選項「偶爾」的人數為 17 人，佔總填答人數的 15%，選填選項「經常」的人數為 51 人，佔總填答人數的 47%，選填選項「總是」的人數為 35 人，佔總填答人數的 32%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 4.04（標準差為.817）。

7.臺北市國民中學音樂教師於試題 7 中，選填選項「從未」的人數為 4 人，佔總填答人數的 4%，選填選項「很少」的人數為 23 人，佔總填答人數的 21%，選填選項「偶爾」的人數為 48 人，佔總填答人數的 43%，選填選項「經常」的人數為 26 人，佔總填答人數的 24%，選填選項「總是」的人數為 9 人，佔總填答人數的 8%，從填答狀況可知，選填「偶爾」選項的受試者人數最多，且本題之得分平均數為 3.12（標準差為.935）。

8.臺北市國民中學音樂教師於試題 8 中，選填選項「從未」的人數為 3 人，佔總填答人數的 3%，選填選項「很少」的人數為 8 人，佔總填答人數的 7%，選填選項「偶爾」的人數為 24 人，佔總填答人數的 22%，選填選項「經常」的人數為 59 人，佔總填答人數的 54%，選填選項「總是」的人數為 16 人，佔總填答人數的

14%，從填答狀況可知，逾半數的受試者選填「經常」選項的受試者人數最多，且人數逾受試者總人數的一半；其次，本題之得分平均數為 3.70（標準差為.883）。

9.臺北市國民中學音樂教師於試題 9 中，選填選項「從未」的人數為 4 人，佔總填答人數的 4%，選填選項「很少」的人數為 21 人，佔總填答人數的 19%，選填選項「偶爾」的人數為 42 人，佔總填答人數的 39%，選填選項「經常」的人數為 34 人，佔總填答人數的 31%，選填選項「總是」的人數為 9 人，佔總填答人數的 8%，從填答狀況可知，選填「偶爾」選項的受試者人數最多，且本題之得分平均數為 3.21（標準差為.952）。

10.臺北市國民中學音樂教師於試題 10 中，選填選項「從未」的人數為 4 人，佔總填答人數的 4%，選填選項「很少」的人數為 11 人，佔總填答人數的 10%，選填選項「偶爾」的人數為 25 人，佔總填答人數的 23%，選填選項「經常」的人數為 50 人，佔總填答人數的 45%，選填選項「總是」的人數為 20 人，佔總填答人數的 18%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.65（標準差為 1.000）。

11.臺北市國民中學音樂教師於試題 11 中，選填選項「從未」的人數為 6 人，佔總填答人數的 5%，選填選項「很少」的人數為 22 人，佔總填答人數的 20%，選填選項「偶爾」的人數為 24 人，佔總填答人數的 22%，選填選項「經常」的人數為 34 人，佔總填答人數的 31%，選填選項「總是」的人數為 24 人，佔總填答人數的 22%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.44（標準差為 1.198）。

12.臺北市國民中學音樂教師於試題 12 中，選填選項「從未」的人數為 6 人，佔總填答人數的 5%，選填選項「很少」的人數為 13 人，佔總填答人數的 12%，選填選項「偶爾」的人數為 33 人，佔總填答人數的 30%，選填選項「經常」的人數為 34 人，佔總填答人數的 31%，選填選項「總是」的人數為 24 人，佔總填答人數的 22%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的

選項，且本題之得分平均數為 3.52（標準差為 1.123）。

13.臺北市國民中學音樂教師於試題 13 中，選填選項「從未」的人數為 3 人，佔總填答人數的 3%，選填選項「很少」的人數為 23 人，佔總填答人數的 21%，選填選項「偶爾」的人數為 58 人，佔總填答人數的 53%，選填選項「經常」的人數為 20 人，佔總填答人數的 18%，選填選項「總是」的人數為 6 人，佔總填答人數的 5%，從填答狀況可知，選填「偶爾」選項的受試者人數最多，且人數逾受試者總人數的一半；其次，本題之得分平均數為 3.04（標準差為.814）。

14.臺北市國民中學音樂教師於試題 14 中，選填選項「從未」的人數為 1 人，佔總填答人數的 1%，選填選項「很少」的人數為 27 人，佔總填答人數的 25%，選填選項「偶爾」的人數為 41 人，佔總填答人數的 37%，選填選項「經常」的人數為 35 人，佔總填答人數的 32%，選填選項「總是」的人數為 6 人，佔總填答人數的 5%，從填答狀況可知，選填「偶爾」選項的受試者人數最多，且本題之得分平均數為 3.16（標準差為.868）。

15.臺北市國民中學音樂教師於試題 15 中，選填選項「從未」的人數為 4 人，佔總填答人數的 4%，選填選項「很少」的人數為 28 人，佔總填答人數的 25%，選填選項「偶爾」的人數為 42 人，佔總填答人數的 38%，選填選項「經常」的人數為 25 人，佔總填答人數的 23%，選填選項「總是」的人數為 11 人，佔總填答人數的 10%，從填答狀況可知，選填「偶爾」選項的受試者人數最多，且本題之得分平均數為 3.52（標準差為 1.123）。

(三) 實施音樂教學評量後的結果運用情形

「臺北市國民中學音樂教師教學評量量表」層面三之各題項填答分析摘要表如下所示。

表 4-2-4 臺北市國民中學音樂教師教學評量量表題項填答分析層面三

層面	題向	從未 (%) (n)	很少 (%) (n)	偶爾 (%) (n)	經常 (%) (n)	總是 (%) (n)	題平均數 (標準差)
實施 音樂 教學 評量 後的 結果 運用 情形	1.您在實施教學評量後，會讓學生知道 自己的成績	3% (3)	11% (12)	20% (22)	46% (51)	30% (32)	3.79 (1.060)
	2.您會在評量後檢視教學目標的達成 程度	0% (0)	11% (12)	20% (22)	45% (50)	24% (26)	3.82 (.898)
	3.您會在教學後修正教學的內容與進 度	2% (2)	8% (9)	12% (13)	41% (45)	37% (41)	4.04 (.972)
	4.您會根據評量結果進行學生分組	2% (2)	12% (13)	32% (35)	45% (50)	9% (10)	3.49 (.864)
	5.您會根據評量結果做為補救教學的 依據	1% (1)	5% (6)	12% (13)	57% (63)	25% (27)	3.99 (.777)
	6 您認為評量結果有助於了解學生學習 情形	0% (0)	11% (12)	15% (16)	45% (50)	29% (32)	3.93 (.911)
	7.您會使用多等級〈如：優良佳可劣〉 記錄學生成績	2% (2)	13% (14)	22% (24)	38% (42)	25% (28)	3.73 (1.031)
	8.您會使用文字描述記錄學生成績	0% (0)	5% (5)	16% (18)	41% (45)	38% (42)	4.13 (.801)

1.臺北市國民中學音樂教師於試題 1 中，選填選項「從未」的人數為 3 人，佔總填答人數的 3%，選填選項「很少」的人數為 12 人，佔總填答人數的 11%，選填選項「偶爾」的人數為 22 人，佔總填答人數的 20%，選填選項「經常」的人數為 51 人，佔總填答人數的 46%，選填選項「總是」的人數為 32 人，佔總填答人數的 30%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.79（標準差為 1.060）。

2.臺北市國民中學音樂教師於試題 2 中，選填選項「從未」的人數為 0 人，佔總填答人數的 0%，選填選項「很少」的人數為 12 人，佔總填答人數的 11%，選填選項「偶爾」的人數為 22 人，佔總填答人數的 20%，選填選項「經常」的人數為 50 人，佔總填答人數的 45%，選填選項「總是」的人數為 26 人，佔總填答人數的 24%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.82（標準差為.898）。

3.臺北市國民中學音樂教師於試題 3 中，選填選項「從未」的人數為 2 人，佔總填答人數的 2%，選填選項「很少」的人數為 9 人，佔總填答人數的 8%，選填選項「偶爾」的人數為 13 人，佔總填答人數的 12%，選填選項「經常」的人數為 45 人，佔總填答人數的 41%，選填選項「總是」的人數為 41 人，佔總填答人數的 37%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 4.04（標準差為.972）。

4.臺北市國民中學音樂教師於試題 4 中，選填選項「從未」的人數為 2 人，佔總填答人數的 2%，選填選項「很少」的人數為 13 人，佔總填答人數的 12%，選填選項「偶爾」的人數為 35 人，佔總填答人數的 32%，選填選項「經常」的人數為 50 人，佔總填答人數的 45%，選填選項「總是」的人數為 10 人，佔總填答人數的 9%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.49（標準差為.864）。

5.臺北市國民中學音樂教師於試題 5 中，選填選項「從未」的人數為 1 人，佔總填答人數的 1%，選填選項「很少」的人數為 6 人，佔總填答人數的 5%，選填選項「偶爾」的人數為 13 人，佔總填答人數的 12%，選填選項「經常」的人數為 63 人，佔總填答人數的 57%，選填選項「總是」的人數為 27 人，佔總填答人數的 25%，從填答狀況可知，選填「經常」選項的受試者人數最多，且人數逾受試者總人數的一半；其次，本題之得分平均數為 3.99（標準差為.777）。

6.臺北市國民中學音樂教師於試題 6 中，選填選項「從未」的人數為 0 人，佔總填答人數的 0%，選填選項「很少」的人數為 12 人，佔總填答人數的 11%，選填選項「偶爾」的人數為 16 人，佔總填答人數的 15%，選填選項「經常」的人數為 50 人，佔總填答人數的 45%，選填選項「總是」的人數為 32 人，佔總填答人數的 29%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.93（標準差為.911）。

7.臺北市國民中學音樂教師於試題 7 中，選填選項「從未」的人數為 2 人，佔總填答人數的 2%，選填選項「很少」的人數為 14 人，佔總填答人數的 13%，選填選項「偶爾」的人數為 24 人，佔總填答人數的 22%，選填選項「經常」的人數為 42 人，佔總填答人數的 38%，選填選項「總是」的人數為 28 人，佔總填答人數的 25%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 3.73（標準差為 1.031）。

8.臺北市國民中學音樂教師於試題 8 中，選填選項「從未」的人數為 0 人，佔總填答人數的 0%，選填選項「很少」的人數為 5 人，佔總填答人數的 5%，選填選項「偶爾」的人數為 18 人，佔總填答人數的 16%，選填選項「經常」的人數為 45 人，佔總填答人數的 41%，選填選項「總是」的人數為 42 人，佔總填答人數的 38%，從填答狀況可知，合計逾半數的受試者選填「經常」與「總是」的選項，且本題之得分平均數為 4.13（標準差為.801）。

貳、臺北市國民中學音樂教師教學評量實施差異分析

本節針對不同背景變項之臺北市國民中學音樂教師之教學評量實施差異進行分析，其背景變項包括「性別」、「年齡」、「學歷狀況」與「教職狀況」等；其中，「學歷狀況」可區分為「最高學歷」、「學科專長」與「師培階段是否修習過教育測驗與評量之相關課程」等三項，而「教職狀況」亦包含「年資」、「目前工作職務」、「曾參加過教育測驗與評量相關之研習」與「任教學校規模」等四項。為探

究不同背景變項之臺北市國民中學音樂教師教學評量實施的差異情形，本研究採 t 檢定及單因子變異數分析（one-way ANOVA）進行分析。

一、不同性別音樂教師之教學評量實施差異分析

本研究將教師性別變項區分為男性、女性兩組，並以 t 檢定來考驗不同性別臺北市國民中學音樂師之教學評量實施差異情形，分析結果如表 4-2-5 所示。

依分析結果顯示，不同性別音樂教師在教學評量實施整體層面及個別之「實施教學評量前之預備情形」、「實施教學評量時之情形」、「實施教學評量後之結果運用情形」等層面皆無顯著差異。

綜合上述分析，不同性別之臺北市國民中學音樂教師的教學評量實施情形未達顯著差異。

表 4-2-5 不同性別音樂教師在教學評量實施量表的獨立樣本 t 考驗

層面	男性 ($n=7$)		女性 ($n=103$)		t	p	95%CI	
	M	SD	M	SD			LL	UL
實施前	13.86	3.58	14.58	2.51	-.71	.477	-2.73	1.29
實施中	51.29	5.56	53.56	5.45	-1.06	.290	-6.51	1.96
實施後	28.14	4.06	31.35	4.17	-1.97	.052	-6.44	.03
整體教學評量實施	93.29	11.91	99.48	9.85	-1.59	.116	-13.96	1.56

* $p < .05$ ** $p < .01$ *** $p < .001$

二、不同年齡音樂教師之教學評量實施差異分析

本研究將教師年齡變項區分為「30 歲以下」、「31-40 歲」、「41-50 歲」與「50 歲以上」等四組，並以單因子變異數分析來考驗不同年齡臺北市國民中學音樂教師之教學評量實施差異情形，分析結果如表 4-2-6 與表 4-2-7 所示。

依分析結果顯示，不同年齡音樂教師在教學評量實施整體層面及個別之「實施教學評量前之預備情形」、「實施教學評量時之情形」、「實施教學評量後之結果運用情形」等層面皆無顯著差異。

表 4-2-6 不同年齡音樂教師在教學評量實施之各組人數、平均數和信賴區間 (n=110)

變項	30 歲以下 (n=40)			31-40 歲 (n=31)			41-50 歲 (n=32)			50 歲以上 (n=7)		
	M		95%CI		M		95%CI		M		95%CI	
	(SD)	LL	UL	(SD)	LL	UL	(SD)	LL	UL	(SD)	LL	UL
實施前	14.68 (2.87)	13.74	15.63	14.35 (2.42)	13.37	15.32	14.31 (2.51)	13.41	15.22	15.50 (1.87)	13.54	17.46
實施中	54.03 (5.29)	52.29	55.76	52.54 (5.47)	50.33	54.75	53.19 (5.70)	51.13	55.24	54.33 (5.89)	48.15	60.51
實施後	30.97 (3.98)	29.67	32.28	31.12 (3.64)	29.65	32.58	31.28 (5.04)	29.46	33.10	31.33 (4.32)	29.80	35.87
整體教學 評量實施	99.68 (9.92)	96.42	102.95	98.00 (10.00)	93.96	102.0	98.78 (10.57)	94.97	102.5	101.17 (10.38)	90.27	112.06

* $p < .05$ ** $p < .01$ *** $p < .001$

表 4-2-7 不同年齡音樂教師之教學評量實施單因子變異數分析 (n=110)

層面	SV	SS	df	MS	F	p
實施前	組間	8.94	3	2.98	.44	.724
	組內	662.47	98	6.76		
	總和	671.41	101			
實施中	組間	40.88	3	13.63	.45	.718
	組內	2965.64	98	30.26		
	總和	3006.52	101			
實施後	組間	1.91	3	.64	.04	.991
	組內	1797.43	98	18.34		
	總和	1799.34	101			
整體教學 評量實施	組間	73.13	3	24.38	.24	.871
	組內	10144.51	98	103.52		
	總和	10217.64	101			

* $p < .05$ ** $p < .01$ *** $p < .001$

三、不同學歷音樂教師之教學評量實施差異分析

本研究將教師學歷變項區分為「專科畢業」、「學士」與「碩博士」等三組，並擬以單因子變異數分析來考驗不同學歷臺北市國民中學音樂師之教學評量實施差異情形，但在正式問卷回收後，並無受試者歸屬在學歷變項中之「專科畢業」類別中，因此，在考量教師學歷變項僅餘「學士」與「碩博士」兩組，故本研究乃改採 t 檢定來考驗不同學歷臺北市國民中學音樂師之教學評量實施差異情形，分析結果如表 4-2-8 所示。

依分析結果顯示，不同學歷音樂教師在教學評量實施整體層面及個別之「實施教學評量前之預備情形」、「實施教學評量時之情形」、「實施教學評量後之結果運用情形」等層面皆無顯著差異。

表 4-2-8 不同學歷音樂教師在教學評量實施量表的獨立樣本 t 考驗

層面	學士 ($n=45$)		碩博士 ($n=65$)		t	p	95%CI	
	M	SD	M	SD			LL	UL
實施前	14.77	2.26	14.34	2.80	.83	.409	-.60	1.45
實施中	52.23	5.27	54.29	5.47	-1.92	.058	-4.20	.07
實施後	30.75	4.11	31.41	4.32	-.79	.434	-2.34	1.01
整體教學評量實施	97.75	9.16	100.05	10.66	-1.15	.254	-6028	1.68

* $p < .05$ ** $p < .01$ *** $p < .001$

四、不同學科專長音樂教師之教學評量實施差異分析

本研究將教師學科專長變項區分為「音樂相關科系」、「非音樂相關科系」兩組，並以 t 檢定來考驗不同學科專長臺北市國民中學音樂師之教學評量實施差異情形，分析結果如表 4-2-9 所示。

依分析結果顯示，不同學科專長音樂教師在教學評量實施整體層面及個別之「實施教學評量前之預備情形」、「實施教學評量時之情形」、「實施教學評量後之結果運用情形」等層面皆無顯著差異。

表 4-2-9 不同學科專長音樂教師在教學評量實施量表的獨立樣本 t 考驗

層面	音樂相關科系 ($n=107$)		非音樂相關科系 ($n=3$)		t	p	95%CI	
	M	SD	M	SD			LL	UL
實施前	14.61	2.49	12.00	4.58	1.74	.085	-.36	5.57
實施中	53.54	5.48	49.	1.00	1.43	.157	-1.78	10.85
實施後	31.18	4.27	29.33	1.15	.75	.458	-3.07	6.77
整體教學評量實施	99.32	10.06	90.33	6.03	1.54	.128	-2.63	20.61

* $p < .05$ ** $p < .01$ *** $p < .001$

五、不同修習教育測驗與評量課程經驗音樂教師之教學評量實施差異分析

本研究將修習教育測驗與評量課程經驗變項區分為「曾在師培階段修習過教育測驗與評量之相關課程」與「不曾在師培階段修習過教育測驗與評量之相關課程」兩組，並以 t 檢定來考驗不同修習教育測驗與評量課程經驗臺北市國民中學音樂師之教學評量實施差異情形，分析結果如表 4-2-10 所示。

依分析結果顯示，不同修習教育測驗與評量課程經驗音樂教師在教學評量實施整體層面及個別之「實施教學評量時之情形」、「實施教學評量後之結果運用情形」等層面皆無顯著差異；但在「實施教學評量前之預備情形」層面有顯著差異（ $t=2.13, p<.05$ ），顯示曾在師培階段修習過教育測驗與評量相關課程之音樂教師（ $M=14.84$ ）在實施教學評量前之預備情形高於不曾在師培階段修習過教育測驗與評量相關課程的音樂老師。

表 4-2-10 不同修習教育測驗與評量課程經驗音樂教師在教學評量實施量表獨立樣本 t 考驗

層面	曾有修習經驗 ($n=83$)		不曾有修習經驗 ($n=27$)		t	p	95%CI	
	M	SD	M	SD			LL	UL
實施前	14.84	2.38	13.62	2.94	2.13*	.036	.08	2.37
實施中	53.38	5.32	53.46	5.94	-.06	.949	-2.55	2.39
實施後	31.24	4.30	30.81	4.03	.446	.657	-1.48	2.34
整體教學評量實施	99.46	9.65	97.88	11.30	.69	.493	-2.97	6.12

* $p < .05$ ** $p < .01$ *** $p < .001$

討論：

不同修習教育測驗與評量課程經驗音樂教師在「實施教學評量前之預備情形」有顯著差異，顯示曾在師培階段修習教育測驗與評量相關課程音樂教師在實施教學評量前預備情形高於不曾在師培階段修習教育測驗與評量相關課程音樂老師。

教學評量強調教學評量結果對整個教學過程的回饋性，但同時也重視教師在實施教學評量前針對學生差異進行教學模式、評量方法的規劃與選擇，此能增加教學評量的實施成果，更能進而提升學生的學習成效。因此，依本研究結果顯示，有修習教育測驗與評量課程經驗之音樂教師，其實施教學評量前之預備情形高於無修習過教育測驗與評量相關課程的音樂老師，顯示教學評量課程之修習經驗將有助於提高教師實際教學評量實施成效，並對學生學習產生正面助益。

六、不同年資音樂教師之教學評量實施差異分析

本研究將教師年資變項區分為「5 年以下」、「6-10 年」、「11-15 年」、「16-20 年」與「20 年以上」等五組，並以單因子變異數分析來考驗不同年資臺北市國民中學音樂師之教學評量實施差異情形，分析結果如表 4-2-11 與表 4-2-12 所示。

依分析結果顯示，不同年資音樂教師在教學評量實施整體層面及個別之「實施教學評量前之預備情形」、「實施教學評量時之情形」、「實施教學評量後之結果運用情形」等層面皆無顯著差異。

表 4-2-11 不同年資音樂教師在教學評量實施之各組人數、平均數和信賴區間 ($n=110$)

變項	5 年以下 ($n=37$)			6-10 年 ($n=21$)			11-15 年 ($n=13$)			16-20 年 ($n=17$)			20 年以上 ($n=22$)		
	<i>M</i>			<i>M</i>			<i>M</i>			<i>M</i>			<i>M</i>		
	(<i>SD</i>)	<i>LL</i>	<i>UL</i>	(<i>SD</i>)	<i>LL</i>	<i>UL</i>	(<i>SD</i>)	<i>LL</i>	<i>UL</i>	(<i>SD</i>)	<i>LL</i>	<i>UL</i>	(<i>SD</i>)	<i>LL</i>	<i>UL</i>
實施前	14.64 (2.89)	13.66	15.62	14.82 (2.83)	13.37	16.28	14.09 (1.22)	13.27	14.91	14.06 (2.32)	12.82	15.30	14.68 (2.64)	13.51	15.85
實施中	53.56 (5.47)	51.71	55.41	54.41 (6.03)	51.31	57.51	52.00 (3.03)	49.96	54.04	51.69 (5.63)	48.69	54.69	54.32 (.72)	51.75	56.89
實施後	31.00 (4.41)	29.51	32.49	31.12 (3.85)	29.14	33.10	30.55 (3.01)	28.52	32.57	29.63 (4.47)	27.24	32.01	32.73 (4.32)	30.81	34.64
教學評量 實施整體	99.14 (10.34)	95.69	102.69	100.35 (11.07)	94.66	106.05	96.64 (5.82)	92.73	100.5 4	95.38 (10.59)	89.73	101.0 2	101.73 (9.81)	97.38	106.0 8

* $p < .05$ ** $p < .01$ *** $p < .001$

表 4-2-12 不同年資音樂教師之教學評量實施單因子變異數分析 ($n=110$)

層面	SV	SS	df	MS	F	p
實施前	組間	8.02	4	2.00	.29	.882
	組內	663.40	97	6.84		
	總和	671.42	101			
實施中	組間	105.30	4	26.33	.88	.479
	組內	2901.22	97	29.91		
	總和	3006.52	101			
實施後	組間	96.74	4	24.18	1.38	.247
	組內	1702.61	97	17.55		
	總和	1799.35	101			
教學評量實施 整體層面	組間	467.47	4	116.87	1.16	.332
	組內	9750.18	97	100.52		
	總和	10217.65	101			

* $p < .05$ ** $p < .01$ *** $p < .001$

七、不同工作職務音樂教師之教學評量實施差異分析

本研究將教師工作職務變項區分為「音樂專任教師」、「級任導師」、「兼行政」與「代課教師」等四組，並以單因子變異數分析來考驗不同工作職務臺北市國民中學音樂師之教學評量實施差異情形，分析結果如表 4-2-13 與表 4-2-14 所示。

依分析結果顯示，不同工作職務音樂教師在教學評量實施整體層面的表現達顯著差異（ $F = 3.12, p < .05$ ），經 Scheff'e 法事後比較分析得知，兼任行政工作之音樂教師（ $M = 102.19$ ）在教學評量實施整體層面的表現優於擔任班級導師的音樂教師（ $M = 91.58$ ）。在個別層面部分，不同工作職務音樂教師「實施教學評量前之預備情形」與「實施教學評量後之結果運用情形」層皆無顯著差異，但在「實施教學評量時之情形」層面達顯著差異（ $F = 4.37, p < .01$ ），經 Scheff'e 法事後比較分析得知，兼任行政工作之音樂教師（ $M = 55.57$ ）在教學評量實施時的表現優於擔任班級導師的音樂教師（ $M = 48.83$ ）。

表 4-2-13 不同工作職務音樂教師在教學評量實施之各組人數、平均數和信賴區間
($n = 110$)

變項	音樂專任教師 ($n=61$)			級任導師 ($n=12$)			兼行政 ($n=23$)			代課教師 ($n=14$)		
	<i>M</i>	95%CI		<i>M</i>	95%CI		<i>M</i>	95%CI		<i>M</i>	95%CI	
	(SD)	LL	UL	(SD)	LL	UL	(SD)	LL	UL	(SD)	LL	UL
實施前	14.43 (2.59)	13.74	15.12	13.83 (2.76)	12.0	15.5	14.62 (2.75)	13.3	15.87	15.46 (2.07)	14.21	16.71
實施中	53.50 (5.24)	52.10	54.90	48.83 (4.37)	46.0	51.6	55.57 (5.96)	52.8	58.29	53.69 (4.35)	51.07	56.32
實施後	31.34 (4.14)	30.23	32.45	28.92 (3.90)	26.4	31.3	32.00 (4.90)	29.7	34.23	30.85 (3.34)	28.83	32.86
教學評量實施 整體層面	99.27 (9.15)	96.82	101.72	91.58 (9.06)	85.8	97.3	102.19 (12.22)	96.6	107.75	100.00 (8.35)	94.96	105.0

* $p < .05$ ** $p < .01$ *** $p < .001$

表 4-2-14 不同工作職務音樂教師之教學評量實施單因子變異數分析 ($n=110$)

層面	SV	SS	df	MS	F	p	事後比較	η^2	$1 - \beta$
實施前	組間	17.85	3	5.95	.89	.448			
	組內	653.56	98	6.67					
	總和	671.41	101						
實施中	組間	350.94	3	116.98	4.32**	.007	3 > 2	.090	.854
	組內	2655.58	98	27.10					
	總和	3006.52	101						
實施後	組間	78.18	3	26.06	1.48	.224			
	組內	1721.16	98	17.56					
	總和	1799.34	101						
教學評量實施 整體層面	組間	890.51	3	296.84	3.12*	.030	3 > 2	.059	.711
	組內	9327.14	98	95.18					
	總和	10217.65	101						

* $p < .05$ ** $p < .01$

註：1 表示「音樂專任教師」，2 表示「級任導師」，3 表示「兼行政」，4 表示「代課教師」

討論：

不同工作職務音樂教師在教學評量實施整體層面的表現達顯著差異，兼任行政工作之音樂教師在教學評量實施整體層面的表現優於擔任班級導師的音樂教師。其可能原因，應與學校兼任行政人員常代表學校參與推動教學評量相關之研習活動，或在參與教學評量種子教師研習後，須在學校進行教學示範工作有關。因此行政工作知音樂教師其對教學評量的內涵與執行程度的瞭解較其他教師來得深入，其教學評量的實施程度亦較其他教師來的熟悉；其次，由於擔任導師之音樂教師肩負班級級務處理與經營的責任，常使導師在班級經營之餘，無力更進一步的對於教學評量活動付出額外心力，此也可能導致班級導師教學評量表現不如兼任行政人員的現象。綜合上述分析可知，兼任行政工作之音樂教師在教學評量實施整體層面及「實施教學評量時之情形」層面表現優於擔任班級導師之音樂教師。

八、不同教育測驗與評量研習經驗音樂教師之教學評量實施差異分析

本研究將參與教育測驗與評量研習經驗變項區分為「曾參與過教育測驗與評量相關之研習」與「不曾參與過教育測驗與評量相關之研習」兩組，並以 t 檢定來考驗不同參與教育測驗與評量研習經驗臺北市國民中學音樂師之教學評量實施差異情形，分析結果如表 4-2-15 所示。

依分析結果可知，不同參與教育測驗與評量研習經驗的音樂教師在教學評量實施整體層面的表現達顯著差異（ $t = 2.41, p < .05$ ），顯示曾參與過教育測驗與評量相關研習之音樂教師（ $M=100.79$ ）在教學評量實施整體層面的表現優於不曾參與過教育測驗與評量相關研習之音樂教師（ $M=95.89$ ）。在個別層面部分，不同參與教育測驗與評量研習經驗之音樂教師的差異情形如下：

（一）不同參與教育測驗與評量研習經驗音樂教師在「實施教學評量前之預備情形」層面達顯著差異，顯示曾參與過教育測驗與評量相關研習之音樂教師（ $M=14.98$ ）在教學評量實施前之預備情形優於不曾參與過教育測驗與評量相關研習之音樂教師（ $M=13.69$ ）。

（二）不同參與教育測驗與評量研習經驗之音樂教師在「實施教學評量時之情形」層面達顯著差異，顯示曾參與過教育測驗與評量相關研習之音樂教師（ $M=54.21$ ）的教學評量實施情形優於不曾參與過教育測驗與評量相關研習之音樂教師（ $M=51.29$ ）。

（三）不同參與教育測驗與評量研習經驗之音樂教師在「實施教學評量後之結果運用情形」層面未達顯著差異。

表 4-2-15 不同參與教育測驗與評量研習經驗音樂教師在教學評量實施量表獨立樣本 *t* 考驗

層面	曾參與相關研習 (<i>n</i> =69)		不曾參與相關研習 (<i>n</i> =41)		<i>t</i>	<i>p</i>	95% <i>CI</i>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			<i>LL</i>	<i>UL</i>
實施前	14.98	2.20	13.69	3.01	2.48*	.015	.26	2.32
實施中	54.21	5.59	51.92	4.94	2.06*	.042	.08	4.50
實施後	31.59	4.37	30.28	3.85	1.51	.134	.09	4.50
整體教學評量實施	100.79	10.27	95.89	8.96	2.41*	.018	.86	8.94

* *p* < .05 ** *p* < .01 *** *p* < .001

討論：

不同參與教育測驗與評量研習經驗的音樂教師在教學評量實施整體層面及個別層面的「實施教學評量前之預備情形」、「實施教學評量時之情形」的表現達顯著差異，顯示曾參與過教育測驗與評量相關研習之音樂教師在教學評量實施整體層面的表現優於不曾參與過教育測驗與評量相關研習之音樂教師。推究其原因在於，教師研習活動對於在職教師而言，為教師專業成長之途徑，故其對教師教學專業提升的影響性與價值性甚高。近年來，教學評量理念的推展在教育領域中已被重視，因此，主管教育行政機關與學校單位針對此主題所辦理的研習活動日益增加，此對教師教學評量素養的提升與教學評量實施的推動有所幫助。

九、不同學校規模音樂教師之教學評量實施差異分析

本研究將學校規模變項區分為「12 班以下」、「13-36 班」與「37 班以上」等三組，並以單因子變異數分析來考驗不同學校規模臺北市國民中學音樂師之教學評量實施差異情形，分析結果如表 4-2-16 與表 4-2-17 所示。

依分析結果顯示，不同年資音樂教師在教學評量實施整體層面及個別之「實施教學評量前之預備情形」、「實施教學評量時之情形」、「實施教學評量後之結果運用情形」等層面皆無顯著差異。

表 4-2-16 不同學校規模音樂教師在教學評量實施之各組人數、平均數和信賴區間
($n=110$)

變項	12 班以下 ($n=15$)			13-26 班 ($n=55$)			37 班以上 ($n=40$)		
	95%CI			95%CI			95%CI		
	<i>M</i> (<i>SD</i>)	<i>LL</i>	<i>UL</i>	<i>M</i> (<i>SD</i>)	<i>LL</i>	<i>UL</i>	<i>M</i> (<i>SD</i>)	<i>LL</i>	<i>UL</i>
實施前	14.71 (1.27)	13.98	15.45	14.00 (2.76)	13.09	14.91	14.88 (2.68)	14.12	15.64
實施中	53.71 (4.39)	51.18	56.25	52.55 (4.94)	50.93	54.18	53.96 (6.07)	52.23	55.69
實施後	30.71 (3.71)	28.57	32.86	30.00 (4.24)	28.61	31.39	32.10 (4.19)	30.91	33.29
教學評量實施 整體層面	99.14 (8.19)	94.41	103.87	96.55 (10.21)	93.20	99.91	100.94 (10.18)	98.05	103.83

* $p < .05$ ** $p < .01$ *** $p < .001$

表 4-2-17 不同學校規模音樂教師之教學評量實施單因子變異數分析 ($n=110$)

層面	<i>SV</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
實施前	組間	17.28	2	8.64	1.31	.275
	組內	654.14	99	6.61		
	總和	671.42	101			
實施中	組間	44.35	2	22.17	.74	.479
	組內	2962.17	99	29.92		
	總和	3006.52	101			
實施後	組間	97.99	2	48.99	2.85	.063
	組內	1701.36	99	17.19		
	總和	1799.34	101			
教學評量實施 整體層面	組間	415.72	2	207.86	2.10	.128
	組內	9801.93	99	99.01		
	總和	10217.65	101			

* $p < .05$ ** $p < .01$ *** $p < .001$

第三節 臺北市國民中學音樂教師教學評量素養與實施 相關分析

本節旨在探究臺北市國民中學音樂教師教學評量素養與教學評量實施及其各層面間之相關情形，為探求其相關狀況，本研究擬採 pearson 積差相關以進行各變項間之分析，分析結果如表 4-3-1 所示。

表 4-3-1 臺北市國民中學音樂教師教學評量素養與教學評量實施之積差相關分析

層面	教學評量 素養標準一	教學評量 素養標準二	教學評量 素養標準三	教學評量 素養標準四	教學評量 素養標準五	教學評量 素養標準六	教學評量 素養標準七	教學評 量素養 整體
教學評量 實施前	.03	.15	-.04	-.12	.19	.10	.03	.09
教學評量 實施時	.05	.04	.08	-.13	-.04	.06	-.19	-.03
教學評量 實施後	-.06	.00	.13	.02	.06	.06	-.00	.07
教學評量 實施整體	.01	.06	.09	-.09	.05	.09	-.10	.04
* $p < .05$ ** $p < .01$ *** $p < .001$								

壹、音樂教師教學評量素養與教學評量實施之相關分析

音樂教師之教學評量素養及其七項標準分層面與教學評量實施整體層面皆未達顯著相關。因此，臺北市國民中學音樂教師評量素養與教學評量實施未達顯著相關。

貳、音樂教師教學評量素養與教學評量實施各分層面之相關分析

一、教學評量素養與「實施教學評量前之預備情形」層面的相關分析

音樂教師之教學評量素養及其七項標準分層面與教學評量實施整體之「實施教學評量前之預備情形」層面皆未達顯著相關。

二、教學評量素養與「實施教學評量時之情形」層面的相關分析

音樂教師之教學評量素養及其七項標準分層面與教學評量實施整體之「實施教學評量時之情形」層面皆未達顯著相關。

三、教學評量素養與「實施教學評量後之結果運用情形」層面的相關分析

音樂教師之教學評量素養及其七項標準分層面與教學評量實施整體之「實施教學評量後之結果運用情形」層面皆未達顯著相關。

討論：

綜合上述分析，臺北市國民中學音樂教師評量素養與「實施教學評量前之預備情形」、「實施教學評量時之情形」、「實施教學評量後之結果運用情形」層面未達顯著相關。依相關學理邏輯推論而言，教師音樂教學評量素養應與教師教學評量實施間應具緊密之相關性，即教師音樂教學評量素養高低應對其個人教學評量實施程度有其影響，且教師教學評量實施程度亦應對其個人教學評量素養高低的提升有所影響。但本研究之統計分析結果卻與此種假設推論不符，顯見在學校實際教學現場中，應有若干因素限制了教師教學評量素養與教學評量實施間的直接聯繫，如：教師的非教學工作過多而影響教學評量實施意願、教學現場缺乏對教學評量實施者的激勵鼓舞...等，以致於教師即使具有良好之教學評量素養，也無法有效導引出對學生學習具正面影響的教學評量行為。因此，本研究之臺北市國民中學音樂教師教學評量素養與教學評量實施之未具顯著相關的統計結果，正好提供研究者對教學現場省思之檢討方向，若能適切思考實際教學現場對教師教學評量素養與實施間侷限，相信必能建立起兩者間的有效聯結以提升學生學習成效。



第五章 結論與建議

本研究旨在調查臺北市國民中學音樂教師教學評量素養與教學評量實施情形，同時分析其背景變項之差異，最後探討兩者之間是否具有相關。本研究採用「調查研究法」，以臺北市國民中學音樂教師為研究對象，設計「國民中學音樂教師教學評量素養與教學評量實施調查問卷」，內容改編「評量素養調查問卷」（Campbell & Mertler, 2005）與自行編製「國中音樂教學評量實施情形之調查問卷」作為研究工具，並將問卷所得資料進行描述性統計、t 檢定、單因子變異數分析、積差相關及 Scheff'e 事後比較加以分析。本章依據前述資料分析結果，進一步歸納出本研究之結論與建議。

第一節 結論

本節依研究問題，分為教學評量素養現況、不同背景變項在教學評量素養之差異情形、教學評量實施情形、不同背景變項在教學評量實施之差異情形及教學評量素養與實施相關五大部分說明：

壹、臺北市國民中學音樂教師教學評量素養於七項標準中，共六項達判定標準值，僅「標準五」未達，多數教師具備了此次所欲評估之教學評量素養知能。

教師在教學評量素養於七項標準中，共六項達判定標準值 60%，包含：「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以及學校發展時去使用評量能力的結果」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」及「標準七：分辨、不道德、不合法或其他不適當之評量方法與評量資

訊的使用」，僅「標準五：發展學生評量之有效評分步驟」三個子題平均未達判定標準值 50%。

貳、不同背景教師在教師評量素養各標準大致具有差異：

一、不同修習教育測驗與評量課程經驗之音樂教師在音樂教學評量素養整體層面個人之七項標準皆無顯著差異

二、不同年齡及不同年資之音樂教師在教學評量素養整體層面達顯著差異，其中 30 歲以下音樂教師之評量素養高於 41-50 歲音樂老師，年資 5 年以下音樂老師高於年資 11-15 年音樂老師

三、不同性別及不同工作職務之音樂教師在「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」層面有顯著差異，顯示女性音樂教師在此層面的表現優於男性音樂教師

就教師性別部分，除了在「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」層面中女性教師的分數顯著高於男性教師外，其餘各項標準在性別變項上並不因教師性別不同而有所差異。

不同工作職務音樂教師在教學評量素養整體層面及大部分個別標準上皆無顯著差異，包含「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」等層面皆無顯著差異。但在「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」層面有顯著差異。

四、不同學歷及不同學校規模之音樂教師在「標準一：選擇適合於教學決策的評量方法」層面有顯著差異，顯示學歷為學士之音樂教師在此層面的表現優於學歷為碩博士音樂教師

不同學歷音樂教師在教學評量素養整體層面及大部分個別層面並無顯著差異，包含「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」、「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」；但在「標準一：選擇適合於教學決策的評量方法」層面，顯示學歷為學士之音樂教師在此層面的表現優於學歷為碩博士之音樂教師。

不同學校規模音樂教師在教學評量素養整體層面及大部分個別層面皆無顯著差異，包含：「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」及「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」等層面；在「標準一：選擇適合於教學決策的評量方法」層面有顯著差異，但各組間並未達顯著差異。

五、不同學科專長音樂教師在「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」層面有顯著差異，顯示非音樂相關科系之音樂教師在此層面的表現優於音樂相關科系之音樂教師

不同學科專長音樂教師在教學評量素養整體層面及大部分個別層面無顯著差異，包含「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準四：針對個別學生進行選擇、規劃教學、發展

課程以使用評量能力的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」、「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」；但在「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」層面顯示非音樂相關科系之音樂教師在此層面的表現優於音樂相關科系之音樂教師。

六、不同教育測驗與評量研習經驗音樂教師在「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」層面有顯著差異，顯示不曾參與過教育測驗與評量相關研習之音樂教師在此層面的表現優於曾參與過教育測驗與評量相關研習之音樂教師

不同參與教育測驗與評量研習經驗的音樂教師在教學評量素養整體層面及大部分個別層面無顯著差異，包含「標準一：選擇適合於教學決策的評量方法」、「標準二：發展適合於教學決策的評量方法」、「標準三：執行、分析、評分與詮釋外界發展或教師自製之評量的結果」、「標準五：發展學生評量之有效評分步驟」、「標準六：對於學生、家長、一般民眾與教育工作者傳達評量的結果」、「標準七：分辨不道德、不合法或其他不適當之評量方法與評量資訊的使用」；但在「標準四：針對個別學生進行選擇、規劃教學、發展課程以使用評量能力的結果」層面有顯著差異，顯示不曾參與過教育測驗與評量相關研習之音樂教師在此層面的表現優於曾參與過教育測驗與評量相關研習之音樂教師。

參、臺北市國民中學音樂教師教學評量實施之現況良好

臺北市國民中學音樂教師教學評量實施量表之各題項平均得分最小值為 2.70 分，最大值為 4.70 分，其各題項平均得分為 3.67 分（標準差為.373），顯示臺北市國民中學音樂教師在其音樂領域教學的各階段學習活動中，能經常融入教學評量的各項實施概念與策略。整體而言，臺北市國民中學音樂教師實施整體情

形及分項之「實施教學評量前之預備情形」、「實施教學評量時之情形」、「實施教學評量後之結果運用情形」層面情形良好。

肆、不同個人背景變項之教師，在教師教學評量實施上之差異情形：

一、不同性別、不同年齡、不同學歷、不同學科專長、不同年資、不同學校規模之音樂教師在教學評量實施整體層面及個別之「實施教學評量前之預備情形」、「實施教學評量時之情形」、「實施教學評量後之結果運用情形」等層面皆無顯著差異

就不同性別、不同年齡、不同學歷、不同學科專長、不同年資、不同學校規模的教師在教學評量實施整體情形及其分項問卷的填答情形可知：教師在該分項及整體表現不因性別而有所差異。

二、不同修習教育測驗與評量課程經驗音樂教師在「實施教學評量前之預備情形」層面有顯著差異，顯示曾在師培階段修習過教育測驗與評量相關課程之音樂教師在實施教學評量前之預備情形高於不曾在師培階段修習過教育測驗與評量相關課程的音樂老師

不同修習教育測驗與評量課程經驗音樂教師在教學評量實施整體層面及個別之「實施教學評量時之情形」、「實施教學評量後之結果運用情形」等層面皆無顯著差異；但在「實施教學評量前之預備情形」層面有顯著差異，顯示曾在師培階段修習過教育測驗與評量相關課程之音樂教師在實施教學評量前之預備情形高於不曾在師培階段修習過教育測驗與評量相關課程的音樂老師。

三、不同工作職務音樂教師在教學評量實施整體層面的表現達顯著差異，兼任行政工作之音樂教師在教學評量實施整體層面的表現優於擔任班級導師的音樂教師

不同工作職務音樂教師在教學評量實施整體層面及「實施教學評量時之情形」層面的表現達顯著差異，兼任行政工作之音樂教師在教學評量實施整體層面的表現優於擔任班級導師的音樂教師。但在「實施教學評量前之預備情形」與「實施教學評量後之結果運用情形」層面皆無顯著差異。

四、不同參與教育測驗與評量研習經驗的音樂教師在教學評量實施整體層面及個別層面的「實施教學評量前之預備情形」、「實施教學評量時之情形」的表現達顯著差異，顯示曾參與過教育測驗與評量相關研習之音樂教師在教學評量實施整體層面的表現優於不曾參與過教育測驗與評量相關研習之音樂教師

不同參與教育測驗與評量相關研習經驗之音樂教師在教學評量實施整體層面及個別「實施教學評量前之預備情形」、「實施教學評量時之情形」層面的表現達顯著差異，顯示有參與教育測驗與評量相關研習經驗之音樂教師在教學評量實施整體及個別「實施教學評量前之預備情形」、「實施教學評量時之情形」層面的表現優於不曾參與過教育測驗與評量相關研習之音樂教師，但在「實施教學評量後之結果運用情形」層面無顯著差異。

伍、教學評量素養與實施之相關

音樂教師教學評量素養及其七項標準分層面與教學評量實施整體層面未達顯著相關。因此，臺北市國民中學音樂教師評量素養與教學評量實施未達顯著相關。

第二節 建議

根據文獻探討與研究結果，研究者針對教育行政單位、國中音樂教師與後續研究者提出幾項建議。

壹、對教育行政單位的建議

一、鼓勵學校音樂教師參與教育行政單位所辦理之研習

由本研究發現可知，兼任行政工作之音樂教師在教學評量實施整體層面表現優於擔任班級導師之音樂教師，推測其原因應與學校兼任行政人員常代表學校參與推動教學評量相關之研習活動，或在參與教學評量種子教師研習後，須在學校進行教學示範工作有關。因此，為使擔任班級導師之音樂教師能也有機會投入教學評量工作的推動，教育行政單位可多鼓勵學校音樂教師們多參與研習，除了擔任行政之音樂教師可參加研習外，其他職務之音樂教師也有管道可以強化自身評量知能。

二、對於已進入教學職場之教師安排評量進修活動，以增進教師評量素養能力

由本研究發現可知，有參與過教育測驗與評量相關研習經驗之音樂教師，其在教學評量實施整體層面的表現優於無教育測驗與評量相關研習經驗之音樂教師。因此，本研究認為上級主管教育行政機關除應多辦理以評量為主題之教師研習活動或工作坊外，亦應透過研習時數規定（如規定於固定年限內，須參與評量相關議題之研習活動滿特定時數），增進教師教學評量素養的知能。

三、音樂師資培育機關宜強化教學評量課程

有鑑於教師教學評量實施對學生學習成效提升有其正面的助益，音樂師資培育機構在設計其師資培育課程架構時，應考量音樂理論與教學理論間之平衡比重，

使其所培育出來的音樂教師資，除擁有深厚的音樂學理素養與音樂演奏技巧外，更具備適切的教學技巧與理論知能，如此才能在運用適切教學與評量方法時，使學生能順暢的習得各項音樂知識與技能。

貳、對國中音樂教師的建議

由本研究發現可知，有參與過教育測驗與評量相關研習經驗之音樂教師，其在教學評量實施整體層面的表現優於無教育測驗與評量相關研習經驗之音樂教師。因此，為促使教師教學評量實施成效之提升，進而強化學生的學習成效，除了建議上級主管教育行政機關與學校單位應多辦理相關評量議題之教師研習活動或工作坊外，也鼓勵音樂教師踴躍參與教師研習活動，以增加自身評量知能。

參、對後續研究者的建議

一、研究對象

本研究僅以臺北市國中音樂教師為研究對象，研究範圍與對象均較為受限，為使研究結果能較為周延與具推論性，建議後續研究之研究範圍與對象可擴大至臺北市以外地區之國中音樂教師，如此不僅可對國中音樂教師之教學評量素養與教學評量實施的狀況進行較為通盤性的瞭解，亦可進行不同縣市間之教學評量素養與教學評量實施的差異比較分析，以釐清各縣市間的現況落差，提供教育主管行政機關更為明確的參考依據，並進而訂定具差異化與實用性之相關教育政策，期望能在提升教師教學評量素養與評學評量實施的同時，亦強化學生學習成效。

二、研究方法與工具

在研究工具部分，本研究乃改編「評量素養調查問卷」(ALI)部分試題與自行編製「國中音樂教學評量實施情形之調查問卷」作為調查工具。為使研究能更為深入探究教師教學評量之現況，建議後續研究能採用量化與質化研究並行的研究方式，如此能透過量化的調查研究以瞭解國中音樂教學教學評量與教學評量實施的概況與差異情形之餘，亦能透過質性之訪談、課室觀察、資料分析...等研究方式，以深入探究本研究量化研究結果背後所隱涉之學校實質運作與非正式政治權力操作因素，並從而解構量化研究結果的可能因素，以獲取較為深化之研究結論與擬定適切之教育革新措施。



參考文獻

壹、中文部分

大學入學考試中心（2014 年 5 月 10 號）。「大學入學考試中心」指定科目考試統計資料。取自 <http://www.ceec.edu.tw/AppointExam/AppointExamStat.htm>

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（未出版之博士論文）。國立彰化師範大學，彰化市。

王子華、范雅晴、王國華（2008）。數理科在職教師評量素養培育模式之探究。科學教育學刊，16(1)，25-51。

王文中、呂金燮、吳毓瑩、張郁雯、張淑慧（2005）。教育測驗與評量－教室學習觀點。臺北市：五南。

王晶瑩（2006）。臺北市國中視覺藝術教師教學評量現況之研究（未出版之碩士論文）。國立臺灣師範大學，臺北市。

王雅蕾（2009）。基測改變了什麼？從教學現場探究基測實施後歷史教學評量的轉換（未出版之碩士論文）。國立臺灣師範大學：臺北市。

余智娟（2006）。臺灣華語文教師教學評量研究（未出版之碩士論文）。國立高雄師範大學，高雄市。

吳盈麗（2011）。臺中市國民中學音樂教師對九年一貫藝術與人文領域課程綱要音樂教學評量之觀點調查（未出版之碩士論文）。國立臺中教育大學，臺中市。

吳清山、高家斌（2005）。十二年國民基本教育政策發展的回顧與展望。教育資料與研究雙月刊，63，53-66。

李坤崇（1999）。多元化教學評量。臺北市：心理。

李坤崇（2006）。教學評量。臺北市：心理。

林小玉（2001）。由音樂藝術之本質探討多元評量於音樂教學之意涵與實踐。音樂藝術學刊，1，61-88。

- 林桂妃（2009）。基隆市國民中學教師直笛教學評量實施現況之調查研究（未出版之碩士論文）。國立臺北教育大學，臺北市。
- 邱慧娟（2011）。臺中地區國中視覺藝術教師多元評量實施現況之研究（未出版之碩士論文）。國立新竹教育大學，新竹市。
- 姚世澤（2000）。音樂教育與音樂行為理論基礎及方法論。臺北市：師大書苑。
- 紀雅真（2007）。國民中學音樂教學評量實施現況之調查研究（未出版之碩士論文）。國立臺灣師範大學，臺北市。
- 高博銓（2007）。教學評量的原則及其革新作法。中等教育，58，44-59。
- 國家教育研究（2014年3月24日）。臺灣學生學習成就資料庫。取自
<http://tasa.naer.edu.tw/15news-1.asp>
- 張春興（2007）。教育心理學：三化取向的理論與實踐（重修二版）。臺北市：東華。
- 張碧娟（2010）。免試入學的困境與活路。中等教育，61，12-19。
- 張麗麗（2002）。藝術與人文學習領域的教學評量。黃王來主編，藝術與人文教育下冊（579-636頁）。臺北市：桂冠。
- 教育部（2007）。推動十二年國民基本教育說帖。臺北市：教育部。
- 教育部（2008）。國民中小學九年一貫課程綱要藝術與人文學習領域。臺北市：教育部。
- 教育部（2012）。國民中學學生學習成就評量標準（試行版）－藝術與人文學習領域。臺北市：國立臺灣師範大學心理與教育測驗研究發展中心。
- 郭生玉（2004）。教育測驗與評量（修訂第一版）。臺北縣：精華。
- 陳怡淇（2007）。發展七年級數學遊戲融入教學評量之行動研究（未出版之碩士論文）。國立彰化師範大學，彰化市。

陳臆如（2013）。雲林縣國民中學音樂教師實施教學評量現況之調查（未出版之碩士論文）。國立臺中教育大學，臺中市。

黃于芬（2007）。臺北縣國民中學藝術與人文領域音樂教學多元評量之現況調查研究（未出版之碩士論文）。國立臺灣師範大學，臺北市。

黃光雄（1996）。教學原理。臺北市：師大書苑。

黃珮瑄、莊敏仁（2012）。國小音樂教師運用教學檔案夾提升音樂學習評量專業知能之個案研究。臺中教育大學學報，26，59-77。

黃翠瓊（2008）。1994-2007 臺灣地區音樂評量相關學位論文之內容分析研究（未出版之碩士論文）。臺北市立教育大學，臺北市。

臺北市政府教育局（2014 年 3 月 2 號）。學校資料。取自
http://www.ntpc.edu.tw/_file/2052/SG/25532/39347.html

臺灣學生學習成就評量資料庫（2014 年 3 月 10 號）。「臺灣學生學習成就評量資料庫」建置背景。取自 <http://tasa.naer.edu.tw/1about-1.asp?id=2>

廖文卿（2009）。國中高年級視覺藝術教學評量之個案研究：全國性藝術與人文學習成效評量模式（未出版之碩士論文）。國立彰化師範大學，彰化。

廖怡君（2007）。評分規準於國中直笛教學之個案研究（未出版之碩士論文）。國立臺灣師範大學，臺北市。

劉以勤（2013）。桃園縣國民中學藝術與人文課程中歌唱教學評量實施現況之調查研究（未出版之碩士論文）。臺北市立教育大學：臺北市。

劉向欣（2008）國中教師評量素養現況之調查研究（未出版之在職碩士論文）。國立新竹教育大學，新竹市。

劉安彥（2003）。教學評量的理論與應用。教育資料與研究，55，100-108。

歐滄和（2006）。教育測驗與評量（初版四刷）。臺北市：心理。

潘宇文 (2009)。臺灣北部地區中小學歌唱教學評量實施狀況之調查研究。藝術教育研究，18，1-31。

黎曉安 (2002)。檔案評量策略在視覺藝術課程實施之探討(未出版之碩士論文)。國立臺灣師範大學，臺北市。

謝玲華 (2006)。國中教師國語文多元評量知能、實施困境與因應策略之研究(未出版之碩士論文)。國立高雄師範大學，高雄市。

謝祈蒂 (2009)。視覺藝術教學評量工具在臺灣的發展與現況調查研究(未出版之碩士論文)。國立新竹教育大學：新竹市。

謝苑玫 (2000)。全觀而回饋豐富的音樂教學評量設計。新世紀優質學習的經營研討會論文集，250-278。國立臺南師範學院，臺南市。

貳、英文部分

American Federation of Teachers, National Council on Measurement in Education, & National Teacher Association (AFT, NCME, & NEA). (1990). The Standarda for teacher Competence in the Educational Assessment of Students. *Educational Measurement: Issues and Practice*, 9(4), 30-32.

Airasian, P. W. (1991). Perspectives on measurement instruction. *Educational Measurement: Issues and Practice*, 10(1), 13-16.

Brophy, J. D., & Radocy, R.E. (1987). *Measurment and evaluation of musical experiences*. New York: Schirmer.

Brophy, T. S. (2000). *Assessing the developing child musician*. Chicago, GIA.

Campbell, C., & Mertler, C. A. (2005). Validation of the Assessment Literacy Inventory (ALI) with preservice teachers. Annual meeting of the Mid-Western Educational Research Association, Columbus, Ohio.

Campbell, C., Murphy, J.A., & Holt, J. K. (2002). Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers. Annual meeting of the Mid-Western Educational Research Association, Columbus, OH.

- Center for School Improvement and Policy Studies, Boise State University, ND.
(2005, January). What is assessment literacy? Assessment Literacy. Retrieved
from <http://csi.boisestate.edu/al/>
- Glaser, R. (1962). *Psychology and Instructional Technology*. In R. Glaser (ed.),
Training, Research and Education, Pittsburgh: University of Pittsburgh Press.
- Hepworth-Osiowy, K. (2004). Assessment in elementary music education:
Perspectives and practices of teachers in Winnipeg public schools. *Master
Abstracts International*, 42(06), 1951.
- Kotora, E. J. (2001). Assessment practices in the choral music classroom: A survey of
Ohio high school choral music teachers and college choral methods teachers.
Dissertation Abstracts International, 62(12), 4102-A.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching*. (8th
ed.). Upper Saddle River, NJ: Prentice-Hall.
- Linn, R. L. & Miller, M.D. (2005). *Measurement and Assessment in teaching* (7th.
Ed.). Englewood Cliffs, NJ: Prentice Hall.
- Mertler, C.A. (2004). Secondary Teachers' Assessment Literacy: Does classroom
experience make a difference? , *American secondary Education*, 33(1), 49-64.
- North Central Regional Educational Laboratory. (n.d.)(2013, April 12). Indicator's
Assessment. Retrieved from <http://www.ncrel.org/projectsuccess/>
- Paterno, J. (2001). Measuring success: A glossary of assessment terms. *Building
cathedrals: Compassion for the 21st century*. Retrieved from
<http://www.tcsdk12.org/math/Documents/measuringsuccess.html>.
- Peppers, M. (2010). *An examination of teachers' attitudes toward assessment and
their relationship to demographic factors in Michigan elementary general music
classrooms (Master's thesis)*. Master Abstracts International, [UMI No. 1487165].
- Plake, B. S. (1993). Teacher assessment literacy: Teacher's competencies in the
educational assessment students. *Mid-Western Educational Researcher*, 61, 21-27.

Popham, W. J. (2000). *Modern educational assessment*. Boston: Allyn and Bacon.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.

Talley, K. E. (2005). An investigation of the frequency, methods, objectives, and applications of assessment in Michigan elementary general music classrooms. *Master Abstracts International*, 44(02), 627.



Assessment Literacy Inventory



Cynthia Campbell, Ph.D.
Northern Illinois University
and

Craig A. Mertler, Ph.D.
Bowling Green State University

Description of the ALI:

The *Assessment Literacy Inventory (ALI)* consists of five scenarios, each followed by seven questions. The items are related to the seven "Standards for Teacher Competence in the Educational Assessment of Students." Some of the items are intended to measure general concepts related to testing and assessment, including the use of assessment activities for assigning student grades and communicating the results of assessments to students and parents; other items are related to knowledge of standardized testing, and the remaining items are related to classroom assessment.

Directions:

Read each scenario followed by each item carefully; select the response you think is the best one and *mark your response on the answer sheet*. Even if you are not sure of your choice, but, *mark the response you believe to be the best*.

Scenario #1

Ms. O'Connor, a math teacher, questions how well her 10th grade students are able to apply what they have learned in class to situations encountered in their everyday lives. Although the teacher's manual contains numerous items to test understanding of mathematical concepts, she is not convinced that giving a paper-and-pencil test is the best method for determining what she wants to know.

1. Based on the above scenario, the type of assessment that would *best* answer Ms. O'Connor's question is called a/an
 - A. performance assessment.
 - B. authentic assessment.
 - C. extended response assessment.
 - D. standardized test.
2. In order to grade her students' knowledge accurately and consistently, Ms. O'Connor would be well advised to
 - A. identify criteria from the unit objectives and create a scoring rubric.
 - B. develop a scoring rubric after getting a feel for what students can do.
 - C. consider student performance on similar types of assignments.
 - D. consult with experienced colleagues about criteria that has been used in the past.
3. To get a general impression of how well her students perform in mathematics in comparison to other 10th graders, Ms. O'Connor administers a standardized math test. This practice is acceptable *only* if
 - A. the reliability of the standardized test does not exceed .60.
 - B. the standardized test is administered individually to students.
 - C. the content of the standardized test is well known to students.
 - D. the comparison group is comprised of grade level peers.
4. Which of the following is an *inappropriate* use of the results from this standardized math test?
 - A. planning instruction
 - B. assigning student grades
 - C. determining students' strengths and weaknesses
 - D. developing curriculum
5. Throughout instruction, Ms. O'Connor assesses how well her students are grasping the material. These assessments range from giving short quizzes following introduction to a new topic, to administering an end-of-the-unit final exam. In order to improve the validity of this grading procedure, Ms. O'Connor should
 - A. make the grading scale the same for all assessments.
 - B. consider students' prior performance before assigning a final grade.
 - C. weight assessments according to their relative importance.
 - D. take into consideration each student's effort when calculating grades.
6. During a parent teacher conference, one of the parents of a student in Ms. O'Connor's class wants to know what it means that his daughter scored in the 80th percentile in mathematics. Which of the following provides the *best* explanation of this student's score?
 - A. She got 80% of the items on the math test correct.
 - B. She is likely to earn a grade of 'B' in her math class.
 - C. She is demonstrating above grade level performance in math.
 - D. She scored the same or better than 80% of the norm group.
7. Which of the following is an appropriate use of assessment information?
 - A. Utilize information from a variety of assessments when making decisions about student learning.
 - B. Use scores from standardized tests to determine teacher instructional effectiveness.
 - C. Use scores from a standardized test as the primary indicator of student retention.
 - D. Post final grades in order to provide normative information to students in the class.

Scenario #2

Mr. Okawa, a fifth-grade teacher, is planning his instruction for the next grading period, aware of the fact that his students will be taking the statewide achievement test near the end of the grading period.

8. Mr. Okawa's mathematics unit for this grading period will focus on multi-step problem-solving. He wants to assess his students' problem-solving abilities at the end of the unit to determine if any reinstruction will be necessary prior to the statewide test. Which of the following assessment strategies would be the most appropriate choice?
 - A. He should choose the assessment included in the teacher's manual from the textbook he uses.
 - B. He should choose an assessment which is consistent with the content and skills he taught.
 - C. He should choose a different standardized assessment that provides a score on similar skills.
 - D. He should choose an assessment which covers single-step problem-solving skills.
9. Mr. Okawa decides to develop his own assessment in order to determine if any reinstruction will be necessary. He also wants to use his assessment as a means of anticipating how his students will perform on the statewide assessment. In order for him to accurately approximate his students' performance, which of the following would be the most appropriate type of assessment for him to develop?
 - A. a performance assessment
 - B. a multiple-choice test
 - C. a portfolio assessment
 - D. an essay test
10. Julie, one of Mr. Okawa's students, receives a percentile rank of 60 on the problem-solving skills subtest of the statewide assessment. This score is most appropriately interpreted as which of the following?
 - A. Julie scored above average.
 - B. Julie scored below average.
 - C. Julie scored at the national average.
 - D. Not enough information to determine.
11. Juan, another student in Mr. Okawa's class, receives a scaled score of 196 on the reading comprehension portion of the statewide assessment. The cut score is 200; therefore, Juan does not pass this subtest. However, the subtest has a standard error of measurement equal to 6. Which of the following is the best decision for Mr. Okawa to make regarding instruction appropriate to meet Juan's needs?
 - A. Juan has clearly not achieved the minimum level of reading comprehension and should receive remedial reading instruction.
 - B. Mr. Okawa knows that Juan could have scored higher, so the results of the test should be ignored.
 - C. Juan may likely have achieved the minimum level of reading comprehension and nothing different or additional should be done.
 - D. Mr. Okawa knows that Juan should have scored much lower, so the results of the test should be ignored.
12. Which grading practice being considered by Mr. Okawa would result in grades that would least reflect achievement?
 - A. grades based on daily homework and chapter tests
 - B. grades based on daily homework and chapter tests, with points deducted for poor effort
 - C. grades based on daily homework and chapter tests, where students are permitted to redo assignments in order to meet higher standards
 - D. grades based on chapter tests, where daily homework is not formally graded
13. Barbara scores at the 60th percentile on mathematics problem-solving and at the 56th percentile on reading comprehension. The percentile bands for each test are five percentile ranks wide. What advice should Mr. Okawa give to Barbara's parents?
 - A. They should ignore the difference; her performance was essentially the same on the two tests.
 - B. They should seek additional tutoring help for Barbara in reading.
 - C. They should force Barbara to read more at home.
 - D. They should provide enrichment experiences for Barbara in math, which is her better performance area.
14. Mr. Okawa was worried that his students would not perform well on the statewide assessment. He did all of the following to help increase students' scores. Which was unethical?
 - A. He instructed students in strategies for taking multiple-choice tests, such as how to use answer sheets.
 - B. He planned his instruction so that it focused on concepts and skills to be covered on the test.
 - C. He encouraged the students to do their best, and provided them with a reward after testing was complete.
 - D. He allowed students to practice with items from an alternate form of the test.

Scenario #3

Ms. Green is an eighth-grade American History teacher. She has just finished teaching a unit on the Industrial Revolution and wishes to make decisions about her students regarding their higher-order thinking skills. Ms. Green has decided to give her students a single assessment in the form of an end-of-unit multiple-choice test. She anticipates that most of her students will perform well on the test.

15. Based on her goal, what can you conclude about her decision to administer a multiple-choice test?
 - A. This is an appropriate choice for a unit assessment.
 - B. The test scores may not be valid for this purpose.
 - C. The test scores may not be reliable for this purpose.
 - D. A true-false test would be more appropriate.
16. To determine the quality of her multiple-choice test, Ms. Green should conduct an item analysis and examine all of the following except
 - A. item difficulty values.
 - B. item discrimination values.
 - C. reliability coefficients.
 - D. validity coefficients.
17. Ms. Green decides to score the tests using a 100-percent correct scale. Generally speaking, what is the proper interpretation of a student score of 85 on this scale?
 - A. The student answered 85% of the items on the test correctly.
 - B. The student knows 85% of the content covered by this instructional unit.
 - C. The student scored higher than 85% of other students who took this test.
 - D. The student scored lower than 85% of other students who took this test.
18. Some of Ms. Green's students do not score well on the multiple-choice test. She decides that the next time she teaches this unit, she will begin by administering a pretest to check for students' prerequisite knowledge. She will then adjust her instruction based on the pretest results. What type of information is Ms. Green using?
 - A. norm-referenced information
 - B. criterion-referenced information
 - C. both norm- and criterion-referenced information
 - D. neither norm- nor criterion-referenced information
19. The Industrial Revolution test is the only student work that Ms. Green grades for the current grading period. Therefore, grades are assigned only on the basis of the test. What is the major criticism of this practice?
 - A. The test, and therefore the grades, reflect too narrow a curricular focus.
 - B. These grades, since based on tests alone, is probably biased against some minority students.
 - C. She should add extra points to the scores of students who scored low on the test.
 - D. Decisions like grades should be based on more than one piece of information.
20. Mr. Simpson, another American History teacher, bases his grades primarily on his observations of students during class. The primary distinction between his system of assigning grades and that used by Ms. Green is best characterized as which of the following?
 - A. Ms. Green uses formal assessment; Mr. Simpson uses informal assessment.
 - B. Ms. Green uses formative assessment; Mr. Simpson uses summative assessment.
 - C. Ms. Green uses standardized assessment; Mr. Simpson uses nonstandardized assessment.
 - D. Ms. Green uses traditional assessment; Mr. Simpson uses alternative assessment.
21. Based on their grades from last year, Ms. Green believes that some of her low-scoring students are brighter than their test scores indicate. Based on this knowledge, she decides to add some points to their test scores, thus raising their grades. Which of Ms. Green's actions was unethical?
 - A. examining her student's previous academic performance
 - B. adjusting grades in her course
 - C. using previous grades to adjust current grades
 - D. adjusting some students' grades and not others'

Scenario #4

Mr. Valdez is an English teacher in the newly built middle school. Experienced in issues of classroom assessment, Mr. Valdez is often asked to respond to the district's questions concerning best practices for evaluating student learning.

22. Ms. Franklin, also an English teacher, asks what type of assessment is best for evaluating her 6th graders' writing skills. Which of the following methods is likely to provide the *best* response to her question?
 - A. selected response methods
 - B. true/false statements
 - C. completion items
 - D. essay prompts
23. One of the middle school math teachers is redesigning her tests to make greater use of "story problems" as a way to check students' math understanding. She consults with Mr. Valdez to see what, if any, concerns she should be aware of when constructing assessments of this type. Which statement is *not* an appropriate recommendation when designing story-based math tests?
 - A. make sure that the reading level is grade appropriate
 - B. avoid scenarios more familiar to certain groups over others
 - C. check for clarity of sentence construction
 - D. incorporate scenarios used during instruction
24. Isabel, a student in Mr. Valdez's class, scored 78 points on a standardized English test which had a mean of 80 and a standard deviation of 4. She scored 60 points on the science portion of this test which had a mean of 50 and a standard deviation of 3. Based on the above information, in comparison to her peers, which statement provides the most accurate interpretation?
 - A. Isabel is better in English than in science.
 - B. Isabel is better in science than in English.
 - C. Isabel is below average in both subjects.
 - D. Isabel is close to average in both subjects.
25. At the end of each class period, Mr. Valdez does a quick "check in" with his students to get an impression of their understanding. In this example, the primary purpose for conducting formative assessment is to
 - A. identify cumulative knowledge.
 - B. determine content for the final exam.
 - C. plan classroom instruction.
 - D. evaluate curriculum appropriateness.
26. To prepare students for state testing and identify areas of school improvement, all 6th grade English teachers give a common final exam which contains a series of essay items. Recently, however, several teachers have expressed concern that the time and effort necessary to complete grading on a timely basis may result in inconsistent scoring. They consult with Mr. Valdez. Which of the following provides the *best* response to the teachers' concern for consistency?
 - A. grade all responses to essay #1 before grading responses to essay #2
 - B. during grading, adjust rubric criteria to reflect exemplary student work
 - C. utilize a holistic scoring method to minimize teacher subjectivity in scoring
 - D. all things being equal, it is best to limit the use of multiple essay exams
27. Jeremy, a 6th grade student in Mr. Valdez's class, received a grade equivalent score of 7.2 on a standardized reading test. Jeremy's parents wonder what this means. Based on the above information, which of the following statements provides the most appropriate interpretation of this student's score?
 - A. Jeremy is reading at the 7th grade level.
 - B. Jeremy is reading better than the majority of students in his class.
 - C. Jeremy is reading 6th grade material as expected.
 - D. Jeremy should be placed in a 7th grade reading class.
28. "To ensure that standardized test results provide an accurate picture of what students *really* know, it is recommended that teachers clarify items that are confusing to students."
Based on best practices of assessment, which of the following is an appropriate response to the above statement?
 - A. The above statement is an acceptable way to reduce error in testing.
 - B. The above statement is an acceptable way to increase test validity.
 - C. The above statement is unacceptable because it labels students as poor readers.
 - D. The above statement is unacceptable because it breaks standardization.

Scenario #5

Ms. Hawkins is responsible for teaching science at the 4th grade level. Over the past couple of years, her students have really seemed to struggle with investigations of how water changes from one state to another (i.e., freezing, melting, condensing, and evaporating), but she is unsure of where the specific difficulties lie. She is aware that her students need to improve their conceptual understanding of this content standard.

29. Ms. Hawkins wishes to conduct some sort of assessment in order to identify the specific difficulties her students are experiencing. Which of the following would *best* meet her needs?
 - A. a diagnostic assessment
 - B. an informal assessment
 - C. a standardized assessment
 - D. a summative assessment
30. In an effort to refine both her instruction and assessment of this content, Ms. Hawkins conducts an item analysis of student scores from last year's final unit test over this material. She should definitely discard or substantially revise a test item that
 - A. has a difficulty value between .50 and .75.
 - B. has a discrimination value equal to +.30.
 - C. has a discrimination value equal to -.50.
 - D. has a difficulty value equal to .90.
31. Ms. Hawkins' unit test also includes a restricted-response essay item. She is concerned with the demonstrated level of understanding of several specific criteria in her students' responses. Which of the following would *best* facilitate her scoring of these responses?
 - A. an objective answer key
 - B. a holistic rubric
 - C. a checklist
 - D. an analytic rubric
32. Following the completion of the unit, Ms. Hawkins determines that her students have satisfactorily mastered these concepts. However, when her students take the statewide standardized assessment in the spring, she notices that her students perform very poorly on items addressing these same concepts. Considering the discrepancy between students' classroom performance and their standardized test results, what action is most appropriate when making decisions concerning school improvement?
 - A. recommend that classroom instruction be consistent among 4th grade science teachers
 - B. ensure alignment between instruction and what is measured on the standardized test
 - C. select a standardized test that is more likely to yield higher scores in science
 - D. identify the percentage of students predicted to perform well in advanced science classes
33. Ms. Hawkins wants to be sure that the term grades she assigns to her students' performance in science reflect each student's respective level of content mastery for that unit. Which of the following grading systems would *best* accomplish this goal?
 - A. a criterion-referenced grading system
 - B. a norm-referenced grading system
 - C. a pass-fail grading system
 - D. a portfolio grading system
34. Nolan is a student in Ms. Hawkins' class. He receives a raw score of 12 items answered correctly out of a possible 15 on the physical science portion of a standardized test. This raw score equates to a percentile rank of 45. His parents are confused about how he could answer so many items correctly, but receive such a low percentile rank. They approach Ms. Hawkins for a possible explanation. Which of the following is the appropriate explanation to offer to his parents?
 - A. "I don't know...there must be something wrong with the way the test company figured the scores."
 - B. "Although Nolan answered 12 correctly, numerous students answered more than 12 correctly."
 - C. "Raw scores are purely criterion-referenced and percentile ranks are merely one form of norm-referenced scoring."
 - D. "Raw scores are purely norm-referenced and percentile ranks are merely one form of criterion-referenced scoring."
35. In an attempt to try to encourage and motivate her students who are struggling academically, Ms. Hawkins decides to share her gradebook, especially test scores, with them in order to demonstrate how well others are performing. Another teacher advises her not to do this, as it is a clear violation of
 - A. *The Code of Fair Testing Practices in Education.*
 - B. *The Family and Education Rights and Privacy Act.*
 - C. *The Standards for Teacher Competence in the Educational Assessment of Students.*
 - D. *The No Child Left Behind Act.*

Assessment Literacy Inventory

Scoring Key

1. B
2. A
3. D
4. B
5. C
6. D
7. A
8. B
9. B
10. A
11. C
12. B
13. A
14. D
15. B
16. D
17. A
18. B
19. D
20. A
21. D
22. D
23. D
24. B
25. C
26. A
27. C
28. D
29. A
30. C
31. D
32. B
33. A
34. B
35. B

Alignment of Standards with items on ALL:

Standard 1—Items 1, 8, 15, 22, 29

Standard 2—Items 2, 9, 16, 23, 30

Standard 3—Items 3, 10, 17, 24, 31

Standard 4—Items 4, 11, 18, 25, 32

Standard 5—Items 5, 12, 19, 26, 33

Standard 6—Items 6, 13, 20, 27, 34

Standard 7—Items 7, 14, 21, 28, 35



附錄二 專家效度問卷修正意見彙整表

「國民中學音樂教師教學評量素養與教學評量實施之調查」問卷

敬愛的國中音樂教師，您好：

感謝您在百忙中撥空填寫此問卷！

本研究之目的，係為瞭解我國國中音樂教師評量素養，以及音樂教學評量實施之現況。本問卷純屬學術研究之用且採不記名方式，請安心作答。您的意見非常寶貴，將有助本研究之進行，懇請惠予填答！

敬祝 教安

國立臺灣師範大學音樂教育研究所

研究生 黃貞瑜 敬啟

指導教授 吳舜文 博士

連絡方式：02-24922034-30

anitahuang0324@gmail.com

專家建議	專家 A	可強調本研究之重要性
	專家 B	將我國改為「新北市」
	專家 C	適合
	專家 D	無意見

修改後題目：本研究之目的，係為瞭解新北市國中音樂教師評量素養，以及音樂教學評量實施之現況。(以粗體強調)

本調查問卷共分為三部分〈教師基本資料、教師評量素養、教師評量實施情形〉，請仔細閱讀以下情境題，並在適當的□中打「v」。

專家建議	專家 A	情境題的特性應簡單說明，而「無法作答」選項也應提到；教師評量實施情形之答題方式也應說明
	專家 B	刪除「仔細閱讀以下情境題」之字句
	專家 C	增加填答的指導語
	專家 D	問卷三個部份的指導語要明確

修改後題目：將問卷三個部份分別加入指導語

壹、教師基本資料

【填答說明】：您的資料僅供統計分析之用，請放心填答。

貳、教師評量素養

【填答說明】：此部分為單選題，每題有四個選項，其中一個為正確答案，請您根據題目勾選出適當的答案。如不知答案請勾選「無法作答」。

參、教師評量實施情形

【填答說明】：請您根據實際情形在適當的□中打「○」

壹、教師基本資料

原問卷題目	專家建議				修改後題目
	專家 A	專家 B	專家 C	專家 D	
1. 性別	適合	適合	適合	適合	保留
2. 年齡	加入「46 歲以上」之選項	適合	建議以 10 歲為單位	適合	將選項改為 〈1〉30 歲以下； 〈2〉31-40 歲； 〈3〉41-50 歲； 〈4〉50 歲以上
3. 最高學歷	專士改為「專科畢」	博碩士可合一	適合	適合	將選項改為 〈1〉專科畢 〈3〉博碩士
學科專長	適合	適合	刪除「任一階段學歷均可」用語	適合	刪除「任一階段學歷均可」
師培階段修習過評量之課程	適合	適合	適合	適合	保留
4. 年資	適合	適合	適合	適合	保留
兼任職務	兼任職務改為「目前工作職務」	適合	刪除「兼任」兩字，直接問職務即可	適合	將題目改為「目前工作職務」
曾參加評量相關之研習	適合	適合	適合	適合	保留
任教學校規模	依照相關法規列舉	適合	確認要調查的地區對小型、中型及大型學校的班級數規定	適合	將選項改為 〈1〉12 班以下； 〈2〉13-24 班； 〈3〉25-36 班； 〈4〉37 班以上

貳、教師評量素養

情境一

林老師是位國中音樂老師，她想瞭解學生是否能將課堂所學應用於日常生活中。教學手冊內雖有許多測量學生對音樂元素概念理解之習題，但她不認為紙筆測驗能發掘出她知道的答案。

專家建議	專家 A	適合
	專家 B	適合
	專家 C	將課堂所學應用於日常生活中，以粗體表示，有助第一題的回答
	專家 D	題目敘述應再更精簡順暢
修改後題目：林老師是位國中音樂老師，她想瞭解學生是否能將課堂所學應用於日常生活中，但紙筆測驗卻無法評量學生此項能力。		

1. 根據上述情境，以下何種評量最能幫助林老師找出答案呢？

- ☐ A 實作評量 (performance assessment)
☐ B 真實評量 (authentic assessment)
☐ C 檔案評量 (portfolio assessment)
☐ D 標準化測驗 (standardized test)
☐ 無法作答

專家建議	專家 A	適合
	專家 B	強調題目方向
	專家 C	將最能以粗體表現，原題目中的 <i>best</i> 亦以斜體表示
	專家 D	適合
修改後題目：將題目敘述加上強調記號		
1. 根據上述情境，以下何種評量 最能 幫助林老師找出答案呢？		

2. 為了兼顧評分之準確性及一致性，林老師對學生的知識進行評分時，應選擇以下何種方式？

- ☐ A 從全部課程中建立出一套評分標準
☐ B 先了解學生能夠做什麼之後，再發展一套評分標準
☐ C 在類似的作業上考量學生的表現
☐ D 向有經驗的同事請教過去曾使用過的評分標準
☐ 無法作答

專家建議	專家 A	應將答案方向多加引導
	專家 B	題意不明
	專家 C	在實際教學中，A 及 B 會同時考量。題目中所指的「音樂知識」是哪一類？

	專家 D	題目敘述更流暢
修改後題目：將題目敘述修改為 2. 為了兼顧評分之準確性及一致性， <u>林</u> 老師對學生的音符節奏理解部分進行評分時， 最重要 應該使用何種方式？ ☐ A 從全部課程中建立一套評分標準 ☐ B 根據學生當下的表現給予評分後，才發展一套評分標準 ☐ C 在類似的作業上評量學生的表現 ☐ D 向有經驗的同事請教曾使用過的評分標準 ☐ 無法作答		

3. 為了解她的學生與其他班級學生（對照組），在音樂方面的表現有何差異，林老師設計了一個標準化音樂測驗。這個測驗必須符合下列何種條件？
- ☐ A 此標準化測驗之信度不超過 0.60
 - ☐ B 此標準化測驗是為了提供給個別學生而設計的
 - ☐ C 學生應看懂此標準化測驗之內容
 - ☐ D 對照組是由相同年級之同儕組成
 - ☐ 無法作答

專家建議	專家 A	D 選項，將同儕改為「學生」
	專家 B	不適合
	專家 C	根據原文題義，應是「實施」一份標準化測驗，而非設計；選項 C 依照原文解釋，是指「學生應熟知標準化測驗內容」
	專家 D	題目強調方向及敘述精簡
修改後題目： 3. 為了解她的學生與其他班級學生（對照組），在音樂方面的表現有何差異， <u>林</u> 老師實施了一個標準化音樂測驗。這個測驗必須 符合 下列何種條件？ ☐ A 此標準化測驗之信度不超過 0.60 ☐ B 此標準化測驗是為了提供給個別學生而設計的 ☐ C 學生應熟知標準化測驗的內容 ☐ D 對照組是由相同年級之學生組成 ☐ 無法作答		

4. 下列關於林老師所設計實施的標準化音樂測驗，其結果之應用，何者是不適當的？
- ☐ A 用來修改教學方式
 - ☐ B 用來評定學生成績
 - ☐ C 用來認定學生之長處與短處
 - ☐ D 用來發展課程計畫
 - ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	刪除「設計」兩字；「不適當」請以粗體表示；選項 A 的原文 planning instruction 翻譯為修改教學方式並不適當，因為教師在教學時，並不以標準化測驗的內容為目標，這樣翻譯會使 A 跟 B 兩個答案都是錯的。
	專家 D	題目與選項敘述應更清楚

修改後題目：

4. 下列關於林老師實施的標準化音樂測驗，其結果之應用，何者為不適當？

- ☐ A 瞭解學生的困難處，針對困難處在教學時做加強
- ☐ B 評定學生成績
- ☐ C 了解學生學習情況
- ☐ D 發展課程計畫
- ☐ 無法作答

5. 林老師可以透過教學來評定學生理解教材的狀況，評量的方式由介紹新單元後的隨堂測驗，到實施期末考試皆有。為了改善這些評分流程的效度，林老師應該？

- ☐ A 統一所有評分量表
- ☐ B 在評定最後總成績前，考量學生起點行為
- ☐ C 依據相對的重要性來加重計分
- ☐ D 在計算成績時，考量學生付出之努力
- ☐ 無法作答

專家建議	專家 A	應將答案方向多加引導
	專家 B	不適合
	專家 C	適合
	專家 D	題目敘述不夠流暢

修改後題目：

5. 林老師評量的方式由隨堂測驗到期末考試皆有。為了使這些評量有效的呈現學生所學的情況，林老師**最重要**應該注意以下何者？

- ☐ A 統一所有評分量表
- ☐ B 在評定最後總成績前才考量學生起點行為
- ☐ C 依據評量項目的重要性來分配評分比例
- ☐ D 在計算成績時，主要考量學生付出的努力
- ☐ 無法作答

6. 在一場親師會中，林老師班級的一位家長，想要了解他女兒在音樂科得到的 80 百分等級（80 percentile）所代表的意義。下列何者是對於該學生分數的最好解釋？

- ☐ A 她在音樂測驗中答對了 80 %的試題
- ☐ B 她的音樂成績應該可以得到” B”
- ☐ C 她在音樂方面，表現出高於該年級應有的表現
- ☐ D 她的得分和 80 %之常模一樣或更好
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	適合
	專家 D	敘述應再調整
修改後題目：		
6. 一位家長，想要了解他女兒在音樂科得到的百分等級為 80（80 percentile）所代表的意義。下列何者是對於該學生分數的 <u>最好解釋</u> ？		
☐ A 她在音樂測驗中答對了 80 %的試題		
☐ B 她的音樂成績應該可以得到” B”		
☐ C 她在音樂方面，表現出高於該年級應有的表現		
☐ D 她在班上的成績優於百分之 80 的同學		
☐ 無法作答		

7. 下列哪一項是對於評量資訊適當之使用？

- ☐ A 教師在對學生的學習下決定時，要利用多元化的評量方式收集資訊
- ☐ B 使用標準化測驗分數來判定教師教學的有效性
- ☐ C 使用標準化測驗分數來當作學生對學習能夠持久的指標
- ☐ D 公佈最後總成績以提供常模資訊給班上學生
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	適合
	專家 D	敘述應再調整
修改後題目：		
7. 下列哪一項是對於評量資訊 <u>適當</u> 之使用？		
☐ A 教師對學生的學習下決定時，要利用多元化的評量方式收集資訊		
☐ B 使用標準化測驗分數來判定教師教學的有效性		
☐ C 使用標準化測驗分數來當作學生對學習能否認真的指標		
☐ D 公佈最後總成績給班上學生		
☐ 無法作答		

情境二

周老師是一位國中音樂教師，對於音樂評量議題有豐富經驗，常常有人諮詢他關於評量學生學習的最好方法之相關問題。

專家建議	專家 A	適合
	專家 B	適合
	專家 C	適合
	專家 D	適合
修改後題目：保留		

8. 陳老師也是位音樂教師，她詢問何種評量最適合用來評量七年級學生創作的技巧。下列何者是最好的答案？

- ☐ A 實作評量 (performance assessment)
- ☐ B 問答 (questions)
- ☐ C 檢核表 (checklist)
- ☐ D 測驗 (tests)
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	題目修改為創作後，答案應改為 A
	專家 D	題目敘述應更精簡
修改後題目：		
8. 何種評量最適合用來評量七年級學生創作的技巧。下列何者是最好的答案？		
☐ A 實作評量 (performance assessment)		
☐ B 問答 (questions)		
☐ C 檢核表 (checklist)		
☐ D 測驗 (tests)		
☐ 無法作答		

9. 一位國中音樂教師正在重新設計她的測驗，希望更能善用「評定量表」來檢核學生的歌唱能力，她向周老師請教，在設計這類評量時，是否還有其他注意事項。針對一首歌曲的評定量表設計，下列何者並非合適的建議？

- ☐ A 確認歌唱曲是否適合學生的程度
- ☐ B 應該要避免歌唱情境對於某些族群而言，較其他族群熟悉的情況發生
- ☐ C 檢查曲子構造是否清楚易懂
- ☐ D 採用課堂教學使用過的情境
- ☐ 無法作答

專家建議	專家 A	D 選項敘述不清
	專家 B	不適合
	專家 C	D 選項敘述不適合
	專家 D	題目敘述應更精簡
修改後題目：		
9. 請問針對一首歌曲的評定量表設計，下列何者並非合適的建議？ ☐ A 確認歌唱曲是否適合學生的程度 ☐ B 避免歌唱情境對於某些族群而言，較其他族群熟悉的情況發生 ☐ C 檢查曲子構造是否清楚易懂 ☐ D 一定只能採用課堂歌唱教學使用過的情境 ☐ 無法作答		

10. 王同學是周老師的學生，在一項標準化音樂演奏測驗中得到 78 分，測驗的平均分數為 80、標準差為 4。她在另一項音感認譜測驗部分得到 60 分，平均分數為 50、標準差為 3。依據以上資訊，與她的同儕相較之下，下列何項敘述最能解釋她的表現？

- ☐ A 王同學的演奏比音感認譜表現好
- ☐ B 王同學的音感認譜比演奏表現好
- ☐ C 王同學兩個科目的表現都低於平均
- ☐ D 王同學兩個科目的表現都接近平均
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	適合
	專家 D	語句敘述應再修正，宜更精簡
修改後題目：		
10. 有位學生在一項標準化音樂演奏測驗中得到 78 分，測驗的平均分數為 80、標準差為 4。她在另一項音感認譜測驗部分得到 60 分，平均分數為 50、標準差為 3。依據以上資訊，與她的同儕相較之下，下列何項敘述最能解釋她的表現？ ☐ A 學生的演奏表現較好 ☐ B 學生的音感認譜表現較好 ☐ C 學生兩個科目的表現都低於平均 ☐ D 學生兩個科目的表現都接近平均 ☐ 無法作答		

11. 在每堂課課近尾聲時，周老師會對學生做個快速的「確認」動作，以了解他們的理解狀況。在此情況下，這個形成性評量基本目的為何？

- ☐ A 確認累積性的知識
- ☐ B 決定期末考試的內容
- ☐ C 計畫班級教學課程
- ☐ D 評鑑課程的適當性
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	適合
	專家 D	敘述宜更精簡，選項更明確

修改後題目：

11. 在每堂課接近尾聲時，周老師會對學生的理解狀況進行快速的確認。請問這個形成性評量**主要目的**為何？

- ☐ A 確認學生的專心度
- ☐ B 決定期末考試的內容
- ☐ C 計畫下節課的教學內容
- ☐ D 評定學生平時成績
- ☐ 無法作答

12. 為了幫助學生準備全國性考試，與確認學校的進步空間，所有七年級音樂教師實施一個統一的期末考試。本考試包含一系列自陳法之題目。然而，有幾位老師表示，為了快速改完考試，及每位老師所花費的時間和心力的不同，可能造成評分上的不一致，因此他們向周老師請教。下列哪項敘述最能解決老師們對不一致性的憂慮？

- ☐ A 在評分第二題題目之前，先評分全數的第一題
- ☐ B 在評分時，調整評分標準以顯現傑出學生的表現
- ☐ C 規劃一個統整的評分方法，減少教師在評分上的主觀
- ☐ D 為了公平起見，最好限制使用多重的自陳法題目之考試
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	什麼是「確認學校的進步空間」？
	專家 D	題目敘述應再做修正

修改後題目：

12. 這學期七年級音樂的期末考試包含一系列自陳法的題目。但是，為了快速改完題目，及每位老師花費的時間和心力的不同，可能造成評分上的不一致，他們向周老師請教。下列哪項敘述**最能公正解決**老師們對不一致性的

憂慮？

- ☐ A 在評分第二題題目之前，先評分全數的第一題
- ☐ B 在評分時，調整評分標準以顯現傑出學生的表現
- ☐ C 規劃一個統整的評分方法，減少教師在評分上的主觀
- ☐ D 為了公平起見，最好限制使用多重的自陳法題目之考試
- ☐ 無法作答

13. 高同學是周老師的七年級學生，在一個標準化音感認譜測驗上得到 7.2 的年級相等分數 (grade equivalent score)。高同學的雙親想要了解其代表意義。下列何者能提供最佳的解釋？

- ☐ A 高同學具有七年級的音感認譜程度
- ☐ B 高同學的音感認譜能力比其他班上大部分的學生都優秀
- ☐ C 高同學符合一般期望，適合七年級的教材
- ☐ D 高同學應該被安排在七年級的班級
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	適合
	專家 D	語句再精簡

修改後題目：

13. 高同學為七年級學生，在一個標準化音感認譜測驗上得到 7.2 的年級等值分數 (grade equivalent score，表示受試者的年級程度)。高同學的雙親想要了解其代表意義。下列何者能提供最佳的解釋？

- ☐ A 高同學具有八年級的音感認譜程度
- ☐ B 高同學的音感認譜能力比所有的學生都優秀
- ☐ C 高同學的程度適合七年級的教材
- ☐ D 高同學應該被安排在八年級的班級
- ☐ 無法作答

14. 「為了確保標準化測驗的結果能正確反應學生真正所學的東西，建議老師們要釐清一些會造成學生困惑的試題。」為達到最佳評量實施的標準，下列何者是對上述陳述的適當回應？

- ☐ A 上述陳述是減少測驗錯誤的可行方法
- ☐ B 上述陳述是增加測驗效度的可行方法
- ☐ C 上述陳述是不可行的，因為它將學生預設為音樂能力不良者
- ☐ D 上述陳述是不可行的，因為它打破了標準化
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	適合
	專家 D	選項 D 敘述要再做修正
修改後題目： ☐ D 上述陳述是不可行的，因為標準化測驗不該讓學生事先知道試題		

情境三

金老師負責教八年級的音樂，過去幾年中，她的學生似乎都對歌唱學習得很吃力，但她不確定特定的難處在哪裡。她希望改善學生此內容上的能力

專家建議	專家 A	語意不順
	專家 B	語意不清
	專家 C	將原本題目中的自然科學改為音樂的歌唱學習，歌唱學習多採實作評量，並不適合後面的題目。
	專家 D	歌唱學習應再修改
修改後題目：將歌唱學習調整為直笛演奏技巧 金老師負責教八年級的音樂，過去幾年中，學生似乎都對直笛演奏技巧部分學習得很吃力，但她不清楚學生的困難處在哪裡。她希望能夠了解學生的困難來改善學生此方面的能力		

15. 金老師希望實施評量，來辨別學生所遭遇的特定困難。下列何者最能符合她的需要？

- ☐ A 診斷性測驗
☐ B 非正式評量
☐ C 標準化評量
☐ D 總結性評量
☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	適合
	專家 D	強調是選「符合」的選項
修改後題目：下列何者 最能符合 她的需要？		

16. 為了改進她在歌唱內容的教學與評量，下列何者是她絕對應該去除或改革的測驗題目？

- ☐ A 難度值在 0.50 和 0.75 之間
- ☐ B 鑑別值為 + 0.30
- ☐ C 鑑別值為 - 0.50
- ☐ D 難度值為 0.90
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	適合
	專家 D	可附上難度與鑑別度的提示
修改後題目： 題目後附上（難度 0.50 以上及鑑別度 +0.30 以上為優良試題）		

17. 金老師的單元測驗包括有範圍限制的自陳法論述題（self-report method），學生的答案在一些特定標準上所顯示的理解程度，讓她感到有些擔憂。下列何者最有助於她對那些回答進行評分？

- ☐ A 一個客觀的答案
- ☐ B 一個全面性的評分標準
- ☐ C 一個評分檢核表
- ☐ D 分析式評分基準
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	選項建議附上原文
	專家 D	題目敘述太長
修改後題目： 17. 金老師的單元測驗為封閉式題目，下列何者最有助於她了解學生 <u>特定問題</u> 之想法？		
☐ A 一個客觀的答案 ☐ B 一個全面性的評分標準（holistic rubric） ☐ C 一個評分檢核表（checklist） ☐ D 分析式評分基準（analytic rubric） ☐ 無法作答		

18. 隨著教學單元的結束，金老師認為她的學生們已經充分的精熟這些概念與技巧。然而，當她的學生們參加全國性標準化評量時，卻在相同概念的試題上表現欠佳。針對此情況，下列哪項做法最適合用來改進教學？

- ☐ A 建議八年級每位音樂教師的教學方式應該一效
- ☐ B 確保課堂教學與標準化評量所測量的內容一致
- ☐ C 選擇一個容易得到較高音樂成績的標準化測驗
- ☐ D 確認在程度好的班級中，預期可以表現傑出之學生的比例
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	評量設定的情境需做修改；選項 B 敘述不夠精確
	專家 D	題目情境設定應修改

修改後題目：

18. 金老師認為學生們已經充分精熟課程所教的內容。然而，當他們參加全國性標準化評量時，卻在相似的技巧上表現欠佳。針對此情況，下列哪項做法**最適合**用來改進教學？

- ☐ A 建議八年級每位音樂教師的教學方式應該一效
- ☐ B 確保課堂教學中的技巧與標準化評量所要測的內容是相符的
- ☐ C 選擇一個容易得到較高音樂成績的標準化測驗
- ☐ D 確認在程度好的班級中，預期可以表現傑出之學生的比例
- ☐ 無法作答

19. 金老師要確認她在學生音樂表現上所評定的成績，能反應出個別學生對單元內容的精熟程度。下列的評分方式中，何者最能達到她的目標？

- ☐ A 標準參照測驗
- ☐ B 常模參照測驗
- ☐ C 通過—不通過評分（pass-fail grading）
- ☐ D 學習歷程檔案
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	適合
	專家 D	適合

修改後題目：保留

20. 許同學是金老師班上的學生。他在一個標準化測驗的音感認譜部分，得到的原始分數，是來自於 15 題中答對了 12 題。這個原始分數為 45 百分等級。他的雙親對於他答對這麼多試題，卻只得到這麼低的百分等級感到不解。所以向金老師尋求合理解釋。下列何者是能提供他雙親的適當解釋？
- ☐ A 「我也不清楚...考試主辦單位在評定分數時一定出了什麼差錯。」
- ☐ B 「雖然許同學答對 12 題，但有許多學生答對多於 12 題。」
- ☐ C 「原始分數是單純的標準參照，而百分等級只是常模參照的一種模式。」
- ☐ D 「原始分數是單純的常模參照，而百分等級只是標準參照的一種模式。」
- ☐ 無法作答

專家建議	專家 A	適合
	專家 B	不適合
	專家 C	適合
	專家 D	題目選項要再精簡些
修改後題目：		
20. 下列何者能提供適當的解釋？		
☐ A 「我也不清楚...考試主辦單位在評定分數時一定出了什麼差錯。」		
☐ B 「雖然許同學答對 12 題，但有許多學生答對多於 12 題。」		
☐ C 「原始分數與百分等級皆是常模參照的一種模式。」		
☐ D 「原始分數與百分等級皆是標準參照的一種模式。」		
☐ 無法作答		

21. 為了激勵努力向上的學生，金老師決定要公開她的成績分數紀錄表，用來顯示部分學生的用心。另一位教師勸阻她的這個做法，因為這已違反下列何者法規？
- ☐ A 「國民中小學學生成績評量準則」法規
- ☐ B 「家庭和教育權利及隱私權」
- ☐ C 「學生教育評量之教師能力標準」
- ☐ D 「絕不放棄任何一位孩子」政策
- ☐ 無法作答

專家建議	專家 A	不適合
	專家 B	不適合
	專家 C	用美國的法規測量台灣教師的音樂教學評量素養，不適合
	專家 D	選項 D 改為臺灣的法規，比較符合試題情境
修改後題目：		
☐ D 「高級中等學校學生學習辦法」		

參、教師評量實施情形

一、實施音樂教學評量前之預備情形

題目			總是	經常	偶爾	很少	從未
1. 您會依據九年一貫課程綱要訂定評量目標							
2. 您會依據學校自訂之音樂評量標準訂定評量目標							
3. 您會依據教科書所提供之評量目標訂定評量目標							
4. 您使用自編學習單來評量學生的表現							
專家建議	專家 A	適合					
	專家 B	適合					
	專家 C	適合					
	專家 D	適合					
修改後題目：保留							

二、實施音樂教學評量時之情形

題目	總是	經常	偶爾	很少	從未
1. 您在實施教學評量時，會先與學生解釋此考試的評量標準					
2. 您在實施評量時，班上學生願意配合評量之進行					
3. 您使用音樂知識作為評量的項目					
4. 您使用音感認譜作為評量的項目					
5. 您使用歌唱作為評量的項目					
6. 您使用樂器演奏作為評量的項目					
7. 您使用創作作為評量的項目					
8. 您使用欣賞作為評量的項目					
9. 您使用紙筆測驗來評量學生的表現					
10. 您使用實作評量來評量學生的表現					
11. 您使用檔案評量來評量學生的表現					
12. 您使用評定量表來評量學生的表現					
13. 您在實施評量時，會遇到學生沒有練習或準備的困難					
14. 您在實施評量時，會遇到班上秩序難掌握的困難					
15. 您在實施評量時，會遇到學生及家長不重視評量的過程和結果					

專家建議	專家 A	適合
	專家 B	適合
	專家 C	適合
	專家 D	適合
修改後題目：保留		

三、實施音樂教學評量後的結果運用情形

題目	總是	經常	偶爾	很少	從未
1. 您在實施教學評量後，會讓學生知道自已的成績					
2. 您會在評量後檢視教學目標的達成程度					
3. 您會在教學後修正教學的內容與進度					
4. 您會根據評量結果進行學生分組					
5. 您會根據評量結果做為補救教學的依據					
6 您認為評量結果有助於了解學生學習情形					
7. 您會使用多等級〈如：優良佳可劣〉記錄學生成績					
8. 您會使用文字描述記錄學生成績					

專家建議	專家 A	適合
	專家 B	適合
	專家 C	適合
	專家 D	適合
修改後題目：保留		

附錄三 國民中學音樂教師教學評量素養與教學評量實施調查問卷

敬愛的音樂教師，您好：

感謝您在百忙中撥空填寫此問卷！

本研究之目的，係為瞭解臺北市國中音樂教師評量素養，以及音樂教學評量實施之現況。本問卷純屬學術研究之用且採不記名方式，請安心作答。您的意見非常寶貴，將有助本研究之進行，懇請惠予填答！

敬祝 教安

國立臺灣師範大學音樂教育研究所
研究生 黃貞瑜 敬啟
指導教授 吳舜文 博士
連絡方式：02-24922034-30
anitahuang0324@gmail.com

壹、教師基本資料

【填答說明】：您的資料僅供統計分析之用，請放心填答。

- 1.性別：☐ (1) 男 ☐ (2) 女
- 2.年齡：☐ (1) 30 歲以下 ☐ (2) 31-40 歲 ☐ (3) 41-50 歲
☐ (4) 50 歲以上
- 3.學歷狀況
 - 3-1 最高學歷：☐ (1) 專科畢 ☐ (2) 學士 ☐ (3) 碩博士
 - 3-2 學科專長：☐ (1) 音樂相關科系 ☐ (2) 非音樂相關科系
 - 3-3 師培階段曾修習過教育測驗與評量之相關課程：☐ (1) 是 ☐ (2) 否
- 4.教職狀況
 - 4-1 年資（教學年總計）：☐ (1) 5 年以下 ☐ (2) 6-10 年 ☐ (3) 11-15 年
☐ (4) 16-20 年 ☐ (5) 20 年以上
 - 4-2 目前工作職務：
☐ (1) 音樂專任教師 ☐ (2) 級任導師
☐ (3) 兼行政 ☐ (4) 代課教師 ☐ (5) 其他【_____】
 - 4-3 曾參加過教育測驗與評量相關之研習：☐ (1) 是 ☐ (2) 否
 - 4-4 所任教學校之規模：☐ (1) 12 班以下
☐ (2) 13~36 班
☐ (3) 37 班以上

貳、教師評量素養

【填答說明】：此部分為單選題，每題有四個選項，其中一個為正確答案，請您根據題目勾選出適當的答案。如不知答案請勾選「無法作答」。

情境一

林老師是位國中音樂老師，她想瞭解學生能否將課堂所學應用於日常生活，但紙筆測驗卻無法評量學生此項能力。

1. 根據上述情境，以下何種評量最能幫助林老師測出此能力？
 - ☐ A 實作評量（performance assessment）
 - ☐ B 真實評量（authentic assessment）
 - ☐ C 檔案評量（portfolio assessment）
 - ☐ D 標準化測驗（standardized test）
 - ☐ 無法作答
2. 為了兼顧評分之準確性及一致性，林老師對學生的音符節奏理解部分進行評分時，最重要應該使用何種方式？
 - ☐ A 從全部課程中建立一套評分標準
 - ☐ B 根據學生當下的表現給予評分後，才發展一套評分標準
 - ☐ C 在類似的作業上評量學生的表現
 - ☐ D 向有經驗的同事請教曾使用過的評分標準
 - ☐ 無法作答
3. 為了解她的學生與其他班級學生（對照組），在音樂方面的表現有何差異，林老師實施了一個標準化音樂測驗。這個測驗必須符合下列何種條件？
 - ☐ A 此標準化測驗之信度不超過 0.60
 - ☐ B 此標準化測驗是為了提供給個別學生而設計的
 - ☐ C 學生應熟知標準化測驗的內容
 - ☐ D 對照組是由相同年級之學生組成
 - ☐ 無法作答
4. 下列關於林老師實施的標準化音樂測驗，其結果之應用，何者為不適當？
 - ☐ A 瞭解學生的困難處，針對困難處在教學時做加強
 - ☐ B 評定學生成績
 - ☐ C 了解學生學習情況
 - ☐ D 發展課程計畫
 - ☐ 無法作答

5. 林老師評量的方式由隨堂測驗到期末考試皆有。為了使這些評量有效的呈現學生所學的情況，林老師**最重要**應該注意以下何者？
- ☐ A 統一所有評分量表
 - ☐ B 在評定最後總成績前才考量學生起點行為
 - ☐ C 依據評量項目的重要性來分配評分比例
 - ☐ D 在計算成績時，主要考量學生付出的努力
 - ☐ 無法作答
6. 一位家長，想要了解他女兒在音樂科得到的百分等級為 80（80 percentile）所代表的意義。下列何者是對於該學生分數的**最好解釋**？
- ☐ A 她在音樂測驗中答對了 80 %的試題
 - ☐ B 她的音樂成績應該可以得到”B”
 - ☐ C 她在音樂方面，表現出高於該年級應有的表現
 - ☐ D 她在班上的成績優於百分之 80 的同學
 - ☐ 無法作答
7. 下列哪一項是對於評量資訊**適當**之使用？
- ☐ A 教師對學生的學習下決定時，要利用多元化的評量方式收集資訊
 - ☐ B 使用標準化測驗分數來判定教師教學的有效性
 - ☐ C 使用標準化測驗分數來當作學生對學習能否認真的指標
 - ☐ D 公佈最後總成績給班上學生
 - ☐ 無法作答

情境二

周老師是一位國中音樂教師，對於音樂評量議題有豐富經驗，常常有人諮詢他關於評量學生學習的最好方法之相關問題。

8. 何種評量最適合用來評量七年級學生**創作**的技巧。下列何者是最好的答案？
- ☐ A 實作評量（performance assessment）
 - ☐ B 問答（questions）
 - ☐ C 檢核表（checklist）
 - ☐ D 測驗（tests）
 - ☐ 無法作答

9. 請問針對一首歌曲的評定量表設計，下列何者並非合適的建議？
- ☐ A 確認歌唱曲是否適合學生的程度
 - ☐ B 避免歌唱情境對於某些族群而言，較其他族群熟悉的情況發生
 - ☐ C 檢查曲子構造是否清楚易懂
 - ☐ D 一定只能採用課堂歌唱教學使用過的情境
 - ☐ 無法作答
10. 有位學生在一項標準化音樂演奏測驗中得到 78 分，測驗的平均分數為 80、標準差為 4。她在另一項音感認譜測驗部分得到 60 分，平均分數為 50、標準差為 3。依據以上資訊，與她的同儕相較之下，下列何項敘述最能解釋她的表現？
- ☐ A 學生的演奏表現較好
 - ☐ B 學生的音感認譜表現較好
 - ☐ C 學生兩個科目的表現都低於平均
 - ☐ D 學生兩個科目的表現都接近平均
 - ☐ 無法作答
11. 在每堂課接近尾聲時，周老師會對學生的理解狀況進行快速的確認。請問這個形成性評量主要目的為何？
- ☐ A 確認學生的專心度
 - ☐ B 決定期末考試的內容
 - ☐ C 計畫下節課的教學內容
 - ☐ D 評定學生平時成績
 - ☐ 無法作答
12. 這學期七年級音樂的期末考試包含一系列自陳法的題目。但是，為了快速改完題目，及每位老師花費的時間和心力的不同，可能造成評分上的不一致，他們向周老師請教。下列哪項敘述最能公正解決老師們對不一致性的憂慮？
- ☐ A 在評分第二題題目之前，先評分全數的第一題
 - ☐ B 在評分時，調整評分標準以顯現傑出學生的表現
 - ☐ C 規劃一個統整的評分方法，減少教師在評分上的主觀
 - ☐ D 為了公平起見，最好限制使用多重的自陳法題目之考試
 - ☐ 無法作答

13. 高同學為七年級學生，在一個標準化音感認譜測驗上得到 7.2 的年級等值分數（grade equivalent score，表示受試者的年級程度）。高同學的雙親想要了解其代表意義。下列何者能提供最佳的解釋？
- ☐ A 高同學具有八年級的音感認譜程度
 - ☐ B 高同學的音感認譜能力比所有的學生都優秀
 - ☐ C 高同學的程度適合七年級的教材
 - ☐ D 高同學應該被安排在八年級的班級
 - ☐ 無法作答
14. 「為了確保標準化測驗的結果能正確反應學生真正所學的東西，建議老師們要釐清一些會造成學生困惑的試題。」為達到最佳評量實施的標準，下列何者是對上述陳述的適當回應？
- ☐ A 上述陳述是減少測驗錯誤的可行方法
 - ☐ B 上述陳述是增加測驗效度的可行方法
 - ☐ C 上述陳述是不可行的，因為它將學生預設為音樂能力不良者
 - ☐ D 上述陳述是不可行的，因為標準化測驗不該讓學生事先知道試題
 - ☐ 無法作答

情境三

金老師負責教八年級的音樂，過去幾年中，學生似乎都對直笛演奏技巧部分學習得很吃力，但她不清楚學生的困難處在哪裡。她希望能夠了解學生的困難來改善學生此方面的能力

15. 金老師希望實施評量，來辨別學生所遭遇的特定困難。下列何者最能符合她的需要？
- ☐ A 診斷性測驗
 - ☐ B 非正式評量
 - ☐ C 標準化評量
 - ☐ D 總結性評量
 - ☐ 無法作答
16. 為了改進她在歌唱內容的教學與評量，下列何者是她絕對應該去除或修改的測驗題目？（難度 0.50 以上及鑑別度 +0.30 以上為優良試題）
- ☐ A 難度值在 0.50 和 0.75 之間
 - ☐ B 鑑別值為 + 0.30
 - ☐ C 鑑別值為 - 0.50
 - ☐ D 難度值為 0.90
 - ☐ 無法作答

17. 金老師的單元測驗為封閉式題目，下列何者最有助於她了解學生特定問題之想法？
- ☐ A 一個客觀的答案
 - ☐ B 一個全面性的評分標準（holistic rubric）
 - ☐ C 一個評分檢核表（checklist）
 - ☐ D 分析式評分基準（analytic rubric）
 - ☐ 無法作答
18. 金老師認為學生們已經充分精熟課程所教的內容。然而，當他們參加全國性標準化評量時，卻在相似的技巧上表現欠佳。針對此情況，下列哪項做法最適合用來改進教學？
- ☐ A 建議八年級每位音樂教師的教學方式應該一效
 - ☐ B 確保課堂教學中的技巧與標準化評量所要測的內容是相符的
 - ☐ C 選擇一個容易得到較高音樂成績的標準化測驗
 - ☐ D 確認在程度好的班級中，預期可以表現傑出之學生的比例
 - ☐ 無法作答
19. 金老師要確認她在學生音樂表現上所評定的成績，能反應出個別學生對單元內容的精熟程度。下列的評分方式中，何者最能達到她的目標？
- ☐ A 標準參照測驗
 - ☐ B 常模參照測驗
 - ☐ C 通過與不通過評分（pass-fail grading）
 - ☐ D 學習歷程檔案
 - ☐ 無法作答
20. 許同學在一個標準化測驗的音感認譜部分，得到的原始分數，是來自於 15 題中答對了 12 題。這個原始分數的百分等級為 45。他的雙親對於他答對這麼多試題，卻只得到這麼低的百分等級感到不解。所以向金老師尋求合理解釋。下列何者能提供適當的解釋？
- ☐ A 「我也不清楚...考試主辦單位在評定分數時一定出了什麼差錯。」
 - ☐ B 「雖然許同學答對 12 題，但有許多學生答對多於 12 題。」
 - ☐ C 「原始分數與百分等級皆是常模參照的一種模式。」
 - ☐ D 「原始分數與百分等級皆是標準參照的一種模式。」
 - ☐ 無法作答

21. 為了激勵努力向上的學生，金老師決定要公開她的成績分數紀錄表，用來顯示部分學生的用心。另一位教師勸阻她的這個做法，因為這已違反下列何者法規？

- ☐ A「國民中小學學生成績評量準則」
☐ B「家庭和教育權利及隱私權」
☐ C「學生教育評量之教師能力標準」
☐ D「高級中等學校學生學習辦法」
☐ 無法作答

參、教師評量實施情形

【填答說明】：請您根據實際情形在適當的□中打「○」

一、實施音樂教學評量前之預備情形

題目	總是	經常	偶爾	很少	從未
1. 您會依據九年一貫課程綱要訂定評量目標					
2. 您會依據學校自訂之音樂評量標準訂定評量目標					
3. 您會依據教科書所提供之評量目標訂定評量目標					
4. 您使用自編學習單來評量學生的表現					

二、實施音樂教學評量時之情形

題目	總是	經常	偶爾	很少	從未
1. 您在實施教學評量時，會先與學生解釋此考試的評量規準					
2. 您在實施評量時，班上學生願意配合評量之進行					
3. 您使用音樂知識作為評量的項目					
4. 您使用音感認譜作為評量的項目					
5. 您使用歌唱作為評量的項目					
6. 您使用樂器演奏作為評量的項目					
7. 您使用創作作為評量的項目					
8. 您使用欣賞作為評量的項目					
9. 您使用紙筆測驗來評量學生的表現					
10. 您使用實作評量來評量學生的表現					
11. 您使用檔案評量來評量學生的表現					

題目	總是	經常	偶爾	很少	從未
12. 您使用評定量表來評量學生的表現					
13. 您在實施評量時，會遇到學生沒有練習或準備的困難					
14. 您在實施評量時，會遇到班上秩序難掌握的困難					
15. 您在實施評量時，會遇到學生及家長不重視評量的過程和結果					

三、實施音樂教學評量後的結果運用情形

題目	總是	經常	偶爾	很少	從未
1. 您在實施教學評量後，會讓學生知道自己的成績					
2. 您會在評量後檢視教學目標的達成程度					
3. 您會在教學後修正教學的內容與進度					
4. 您會根據評量結果進行學生分組					
5. 您會根據評量結果做為補救教學的依據					
6. 您認為評量結果有助於了解學生學習情形					
7. 您會使用多等級〈如：優良佳可劣〉記錄學生成績					
8. 您會使用文字描述記錄學生成績					

~ 問卷調查結束 ~

請檢視每題是否皆已作答，非常感謝您的協助