

國立臺灣師範大學理學院科學教育研究所

碩士論文

Graduate Institute of Science Education

College of Science

National Taiwan Normal University

Master's Thesis

非理工科系大學生在科學新聞中的

統計素養之初探研究

An Exploratory Study on Non-Science-Major College

Students' Statistical Literacy for Science News

游珮詩

Yu, Pei-Shih

指導教授：顏妙璇 博士

Advisor : Yen, Miao-Hsuan, Ph.D.

中華民國 111 年 8 月

August 2022

摘要

隨著大數據時代的發展，各領域在分享資訊時也會使用大量的統計訊息，現代社會的公民能否正確判讀統計資料傳達的訊息相當重要。本研究針對非理工科系大學生了解其閱讀科學新聞文本時，統計素養相關先備知識、CARE 統計素養教學指引的學習成效、以及大學時期是否修習統計相關課程對統計素養的影響。研究工具的設計係根據 Schield (2010) 提出影響統計值需要注意 (CARE) 的四個面向：「Context 變項的脈絡」、「Assembly 變項的集合」、「Randomness 隨機性」與「Error 偏差」，改編成科學新聞中常見的 8 項統計素養測驗題型與教學概念，分析 34 位非理工科系大學生在 CARE 統計素養教學指引前後，統計素養能力的改變。

研究結果顯示：(1)非理工科系大學生的統計素養前測中，面對科學新聞中的抽樣偏差與機率的解讀較無問題，但是需要計算的基準值差異以及牽涉到統計學專業名詞的隨機性與混淆變項等三題型，受試者感受到較大的困難。而受試者是否修習統計相關課程分組中，基準值差異與比較類型的選擇兩題型，有修組的表現顯著高於沒修組。(2)教學任務的學習表現上與前測結果相同，受試者需要花較多的心力理解基準值差異、隨機性與混淆變項三個概念。是否修習統計的分組分析與前測結果相同，有修組在基準值差異與比較類型的選擇兩題型顯著高於沒修組。(3)受試者的後測表現與前兩階段表現不同，在面對科學新聞裡的題型，比例的比較與基準值差異題型較難對比教學指引，無法順利解讀語句埋藏的錯誤而使表現下降。受試者是否修習統計的分組則是皆無顯著差異；受試者學習成效在基準值差異、混淆變項、隨機性與偏差共四個題型的後測分數顯著高於前測分數，經過教學任務的指引後，受試者更能掌握科學新聞文本中的埋錯並更正。

關鍵詞：統計素養、科學新聞



Abstract

With the development of big data, a large amount of statistical information is used when sharing information. It is important for citizens in the modern society to correctly interpret the information conveyed by statistical data. This study explored statistical literacy of non-science-major college students when they read science news. Prior knowledge about statistical literacy, learning outcomes of CARE statistical literacy guidelines, and the influence of whether students take statistics-related courses in college on these measures were examined in this study. The instruments were developed based on Schield (2010), in which factors that influence the interpretation of statistical data were classified into four categories: context, assembly, randomness and error (CARE). We adapted these four factors to create 8 items which are common in science news for statistical literacy test items and concepts to be learned.

The results of this study are described as follows. In the pre-test for statistical literacy, non-science-major college students understood the concept of sampling bias and the meaning of probability in science news. But students experienced great difficulties in items concerning baseline value, randomness, and confounding variables. Whether students take statistics-related courses in college resulted in significant differences in items about baseline value and comparison choice. Comprehension levels during instruction were similar to the patterns found in the pre-test. In the post-test, students did not perform well in items about comparison of proportion and baseline value. There was no significant differences between students taking statistics-related courses in college or not. The post-test scores were significantly higher than the pre-test scores in the following four types of questions: baseline value, confounding variables, randomness and errors. After being guided by CARE statistical literacy

guidelines, students were better able to detect and correct errors in statistical information mentioned in science news.

Key words : Statistical Literacy, Science News



目 次

| | |
|-----------------------|-----|
| 摘 要..... | i |
| Abstract..... | iii |
| 目 次..... | v |
| 表 次..... | vii |
| 圖 次..... | ix |
| 第一章 緒論..... | 1 |
| 第一節 研究背景與研究動機..... | 1 |
| 第二節 研究目的與研究問題..... | 2 |
| 第三節 名詞解釋..... | 3 |
| 第二章 文獻探討..... | 5 |
| 第一節 統計素養..... | 5 |
| 第二節 統計素養測驗..... | 12 |
| 第三節 統計素養教學..... | 15 |
| 第三章 研究方法..... | 19 |
| 第一節 研究對象..... | 19 |
| 第二節 研究設計..... | 19 |
| 第三節 研究流程..... | 20 |
| 第四節 研究工具..... | 24 |
| 第五節 資料分析..... | 37 |
| 第四章 研究結果與討論..... | 41 |
| 第一節 學生先備知識之統計與分析..... | 41 |
| 第二節 統計素養教學的學習表現..... | 55 |
| 第三節 統計素養學習成效..... | 64 |

| | | |
|-----------|---------------------------------|-----|
| 第五章 | 結論與建議..... | 91 |
| 第一節 | 研究結果摘要..... | 91 |
| 第二節 | 限制與建議..... | 94 |
| 參考文獻..... | | 97 |
| 附錄一 | 科學新聞文本（前測）..... | 101 |
| 附錄二 | 科學新聞文本（後測）..... | 105 |
| 附錄三 | 統計素養教學投影片..... | 108 |
| 附錄四 | 前測試題評分規準..... | 115 |
| 附錄五 | 後測試題評分規準..... | 120 |
| 附錄六 | 單因子相依樣本 ANOVA 分析的 LSD 事後比較..... | 126 |



表 次

| | | |
|-------------|--|----|
| 表 2-1-1 | 統計素養能力層次架構 (Watson & Callingham, 2003) | 8 |
| 表 2-2-1 | 國外統計素養測驗發展 | 13 |
| 表 2-2-2 | 臺灣統計素養測驗 | 14 |
| 表 2-3-1 | 與教育部課程綱要的對照 | 17 |
| 表 3-4-1 | 研究工具簡介 | 24 |
| 表 3-4-2 | 前後測科學新聞來源範例 | 28 |
| 表 3-4-3 | 統計素養測驗題目 | 31 |
| 表 3-4-3 (續) | 統計素養測驗題目 | 32 |
| 表 3-4-4 | 前測評分規準 | 33 |
| 表 3-4-5 | 後測表現評分規準 | 33 |
| 表 3-4-6 | 教學影片範例 | 36 |
| 表 3-4-7 | 學習表現評分規準 | 37 |
| 表 3-5-1 | 研究問題與資料分析 | 38 |
| 表 4-1-1 | 前測各題型變異數分析表 | 42 |
| 表 4-1-2 | 前測各題型表現統計表 | 43 |
| 表 4-1-3 | 前測各題型得分人數分布 ($N=34$) | 44 |
| 表 4-1-4 | 大學有、無修習統計兩組在前測各題型之 t 檢定 | 52 |
| 表 4-1-5 | 大學有、無修習統計兩組在前測各題型得分人數分布 ($N=34$) | 53 |
| 表 4-2-1 | 各概念學習表現變異數分析表 | 55 |
| 表 4-2-2 | 各概念學習表現統計表 | 56 |
| 表 4-2-3 | 各概念學習表現得分人數分布 ($N=34$) | 57 |
| 表 4-2-4 | 大學有、無修習統計兩組各概念在學習表現之 t 檢定 | 61 |
| 表 4-2-5 | 大學有、無修習統計兩組在各概念學習表現得分人數分布 ($N=34$) | |

| | |
|---|----|
| | 62 |
| 表 4-3-1 後測各題型表現變異數分析表..... | 65 |
| 表 4-3-2 後測各題型表現統計表..... | 65 |
| 表 4-3-3 各題型後測表現得分人數分布 ($N=34$) | 66 |
| 表 4-3-4 前、後測相依樣本 t 檢定結果表..... | 66 |
| 表 4-3-5 「E 抽樣偏差」前後測分數變化表 | 67 |
| 表 4-3-6 「R 隨機性(信賴區間)」前後測分數變化表..... | 69 |
| 表 4-3-7 「C-3 混淆變項」前後測分數變化表..... | 70 |
| 表 4-3-8 「A-2 比例類型的選擇」前後測分數變化表..... | 72 |
| 表 4-3-9 「A-1 比較類型的選擇」前後測分數變化表..... | 73 |
| 表 4-3-10 「P 機率的解讀」前後測分數變化表..... | 75 |
| 表 4-3-11 「C-1 比例的比較」前後測分數變化表..... | 76 |
| 表 4-3-12 「C-2 基準值差異的比較」前後測分數變化表..... | 78 |
| 表 4-3-13 大學有、無修習統計兩組在後測各題型之 t 檢定..... | 81 |
| 表 4-3-14 大學有、無修習統計兩組在後測各題型得分人數分布 ($N=34$) ... | 81 |
| 表 4-3-15 組別、前後測對測驗分數的推論性統計總表..... | 83 |
| 表 4-3-16 大學有無修習統計的前後測成對樣本 t 檢定結果表..... | 84 |
| 表 4-3-17 實驗結果彙整..... | 86 |
| 表 4-3-17 (續) 實驗結果彙整 | 87 |
| 表 5-1-1 研究問題摘要表..... | 92 |

圖 次

| | |
|------------------------------|----|
| 圖 3-3-1 實驗流程圖..... | 20 |
| 圖 3-3-2 統計素養前測範例..... | 21 |
| 圖 3-3-3 統計素養後測範例..... | 23 |
| 圖 3-4-1 前測科學新聞文本..... | 27 |
| 圖 3-4-2 前測統計素養測驗..... | 29 |
| 圖 3-4-3 後測統計素養測驗..... | 30 |
| 圖 3-4-4 教學影片呈現畫面..... | 35 |
| 圖 3-4-5 統計素養教學影片格式..... | 35 |
| 圖 4-3-1 統計素養表現在三階段任務的變化..... | 88 |



第一章 緒論

本章將分成三個小節，第一節為「研究背景與研究動機」，第二節探討「研究目的與研究問題」，第三節介紹「名詞解釋」，以下將依序說明。

第一節 研究背景與研究動機

在大數據和網路時代，各領域充斥著大量的統計訊息，不論是科學文章或新聞傳播也會輔以數字、統計數據佐證可靠性。隨著政府在施政與決策需要更透明，國民具有強大的統計知識去理解技術報告變得相當重要（Ullmann, 2016）。對於資料的產製與接收兩種角度，Wild 和 Pfannkuch（1999）提出將探究脈絡與閱讀脈絡兩種方式做劃分，從實際資料進行實證研究者（例如學生，統計學家）屬於探究脈絡，對資料做統計推理使決策或行動做出最好的判斷。Gal（2002）將這些人稱為“資料生產者（data producers）”或“資料分析者（data analyzers）”，需要解釋自己的資料和結果，並報告其發現和結論，Gal 認為閱讀脈絡與探究脈絡在關注面有所不同，大多數成年人是統計訊息的“消費者（consumers）”，面對資訊滿載的社會，需要的是能對統計資訊和訊息進行解讀、批判性思考和與人交流的能力。

在現今大數據演算法都能預測與操弄行為模式並企圖做出影響的時代，民眾是否有足夠的統計素養能從媒體中理解與判讀正確的訊息是相當重要的。科學新聞應是民眾接觸科學研究的重要渠道，而且也是能同時探討科學、文化、社會、政治等等各種議題的新聞文本（黃俊儒、簡妙如，2006）。國內的科學新聞主要是來自於相關科學研究文獻的摘要或是轉譯其他國家的新聞報導，讓非科學相關背景的普羅大眾也能理解與掌握並能做出適切地判斷。然而媒體在轉寫過程中的誤報或為了吸引閱覽數而下的聳動文字時有所聞，例如鄔啓柔（2022）以「新冠肺炎」相關的科學新聞做論述分析和轉譯寫作分析，將原始文獻轉寫成科學新聞時的錯誤類型歸納為「主張型錯誤」、「證據型錯誤」、「推論型錯誤」與「無錯誤」等類型，其中的「證據型錯誤」是指描述樣本出現錯誤，以及描述數據時誤寫或誤判等等，媒體在描述研究的實驗流程以及產生的數據統計等資訊時，也會因為記者有所增添而產生同一篇原始文獻轉寫出各家媒體不同的版本。

以資料消費者的角度，學生在完成十二年國民義務教育學業後，選擇科系與未來方向並踏入專業領域的環境後，非理工背景學生是否有足夠的統計素養，還是沉浸於本科知識的深入鑽研，已經失去了解讀與判斷數字資訊的能力？文組學生在高中時期學過統計，但是到了大學較少數學相關課程的情況下，是否在閱讀科學新聞或需要統計知識的文章時，仍然有統計素養的敏感度，還是直接採用文章的結論。此外，若透過改編簡化過的統計素養要點的指引，能否幫助受試者提升批判思考的能力。

因此本研究以科學新聞為文本材料，根據 Schield (2010) 所提出影響統計值需要注意 (CARE) 的四個面向：「Context 變項的脈絡」、「Assembly 變項的集合」、「Randomness 隨機性」與「Error 偏差」，整理出科學新聞中最常出現的 8 個統計資料相關陷阱的類型，探究非理工科系學生在面對科學新聞中的統計資料，能否正確識讀並做出批判性思考，以及給予 CARE 統計素養教學指引後，能否提升受試者的科學新聞統計素養。

第二節 研究目的與研究問題

本研究目的為探討非理工科系大學生在科學新聞中的統計素養表現，以及根據 Schield (2010) 提出影響統計值的 C.A.R.E. 四個面向，改編而成科學新聞中的 CARE 統計素養指引，瞭解學生的學習成效，以及探討大學時期是否修習統計相關課程對統計素養表現的影響。故本研究以受試者在前測任務共 8 題的統計素養題型分數評估受試者的統計素養相關先備知識、統計素養教學的學習表現（以理解程度作為測量指標），以及經過統計素養概念教學後，運用 8 項 CARE 統計素養指引，感知科學新聞中的問題點並做批判性思考的能力表現（以後測表現作為測量指標），總共三階段的任務表現分數進行統計分析。

根據上述的研究目的，本研究提出三個研究問題：

1. 研究問題一：非理工科系大學生在科學新聞中 CARE 統計素養前測的表現
 - (1) 受試者在科學新聞中 CARE 統計素養前測表現，題型間是否有差異？
 - (2) 大學是否修習統計課程的學生在科學新聞中 CARE 統計素養前測表現是否有差異？

2. 研究問題二：非理工科系大學生在學習 CARE 統計素養相關概念時的學習表現
 - (1) 受試者在 CARE 統計素養相關概念的學習表現，概念間是否有差異？
 - (2) 大學是否修習統計課程對 CARE 統計素養相關概念的學習表現是否有影響？
3. 研究問題三：非理工科系大學生運用 CARE 統計素養指引檢視科學新聞中統計訊息的學習成效，以及統計素養前後測表現的差異
 - (1) 受試者的 CARE 統計素養後測表現，題型間是否有差異以及分數是否提升？
 - (2) 大學是否修習統計課程對 CARE 統計素養後測表現以及分數提升是否有影響？
 - (3) 受試者的 CARE 統計素養表現在前測、教學與後測三階段任務中的變化為何？



第三節 名詞解釋

一、統計素養 (Statistical Literacy)

Gal (2002a) 提出統計素養是一種能將資料現象、統計訊息做出解讀和批判性思考，並能有意義的與他人解釋及討論溝通觀點的能力，以及人們在評估各種媒體管道中(例如:報紙文章、出版物、廣告、廣播電視的新聞或節目)的統計訊息時資料論證 (data-based arguments) 的能力。

本研究根據 Schield (2010) 分類會對統計數據產生影響的四個面向，改編成科學新聞中常見的 8 類統計素養前後測驗題型，依此評估受試者的統計素養表現，著重於解讀與批判思考的能力。

二、CARE 統計素養指引 (CARE Statistical Literacy Guideline)

本研究的統計素養指引係根據 Schield(2010) 提出的會影響統計值的因素發展而成，總共分為四類：「Context 變項的脈絡」、「Assembly 變項的集合」、

「Randomness 隨機性」與「Error 偏差」，Schield 提出在解讀統計資料時需注意 (take CARE) 這些影響因素的重要性。Schield 認為統計素養應該滿足現代公民生活在以資料驅動的社會中的需求，幫助他們對統計資料進行批判性思考，如果設計一門課程以符合這個目標，則課程內容應包含：(1) 討論所有會影響統計值的因素；(2) 盡量根據目前盛行的媒體話題中的資料作討論；(3) 鼓勵資料消費者能從呈現的資料素材中發掘其中的價值 (Schield, 2006; Schield, 2008; Schield, 2010)。因此本研究將其改編為科學新聞中的統計素養指引，給予受試者面對科學新聞做批判性思考時的方向。

三、科學新聞 (Science News)

科學新聞 (Science News) 是以科學為報導主體，具備普及性與重要性，涵括科學事件背後的社會文化等相關因素，而非僅止於科學知識相關的論述 (黃俊儒、簡妙如，2006)。過去也有許多學者強調統計素養在媒體中的重要性 (Watson & Callingham, 2003; Reston, 2005; Merriman, 2006; Schield, 2006; Schield, 2010)。現代公民在資訊發達的這個時代，接觸及學習科學的管道與過去相比更加豐富多元，而科學新聞是民眾在走出校園步入社會後，仍能保持接觸科學新知的重要管道 (黃俊儒、簡妙如，2008)，因此本研究將媒體聚焦於科學新聞，以網路上蒐集具有統計資料以及科學相關研究結果的科學新聞，探討受試者的統計素養表現。

第二章 文獻探討

本章進行相關研究文獻探討，總共分成三節，第一節「統計素養」，回顧過去研究中對於統計素養的定義；第二節「統計素養測驗」，整理過去統計素養測驗的相關研究；第三節「統計素養教學」，討論過去統計素養教學的相關研究，並整理現今台灣高中生的統計課程範圍，以及課綱中對於統計部分的學習目標。

第一節 統計素養

在現今數據分析和網路時代，大量的統計資料主導決策方向，新聞媒體與科學傳播也會使用數字、統計數據佐證可靠性。統計素養是批判和民主素養，被認為是學校數學教育中必須追求的一種價值(Weiland, 2017)。因此，統計素養是資訊社會的國民所應具備的(Gal, 2002b)。學校教育已將統計教學扎根於數學課程(Ben-Zvi & Garfield, 2008)，但是在學校課堂環境中，數學和統計學的教學情境經常應用在虛構且中立的主題上，其實課程應該更討論社會上的實際問題，有幾篇研究主張將教學法集中在提出問題和將領域內容與社會基本問題做聯結(Frankenstein, 2009; Gutstein, 2006; Gutstein and Peterson, 2013)。因此，本研究想知道學生在面對生活上真實而非虛擬議題的科學新聞所提供的數字與統計資訊時，能夠與不能夠有效接收並處理與使用的資訊是哪些部分。

一、統計素養的意義

隨著統計素養越來越受到重視，廣泛的議題觀點與對統計素養能力的要求也有多樣性的發展，Wallman(1993)在美國統計協會演講時定義：統計素養是理解和批判性地評估滲透到我們日常生活中統計結果的能力，以及欣賞在公共和私人、專業和個人決策中以統計思維做出貢獻的能力。Wallman認為最煩惱的是人們在日常生活中無時無刻都會接觸到統計訊息，但往往沒有具備評估訊息所需的統計素養。為了提高國民的統計素養，主張進一步發展清晰、簡單以及有意義的術語和符號，讓大眾媒體可以在報告統計資訊時加以推廣和使用。

Watson(1997)認為統計素養是由三層能力與複雜性漸增的結構所組成：(1)對數學和統計術語的基本理解；(2)對社會廣泛議題中統計語言和概念的理解，和解釋統計訊息以得出結論並做出決策；(3)對未採取適當統計基礎的主張具質

疑的態度。Snell (1999) 則根據 OED (Oxford English Dictionary) 對具有素養者提供的兩種定義：(a) 有讀寫能力的人；(b) 受過教育或博學的人，延伸出 (a)，具有統計知識的人可能是了解統計語言的人；(b) 可能是一位了解統計概念和推理的人，兼備兩者的涵義將統計素養定義為具有基本層次能理解統計概念和推理能力的人。Schild (1999) 定義統計素養是能夠對統計資料做批判性思考的基本能力，並提出統計素養涉及兩種閱讀技能：理解和解讀/詮釋。

Gal (2002) 認為統計素養是資訊滿載社會中的公民期望的一項關鍵能力，除了是學校預期推動的成果，也是成人計算與讀寫的必要能力，因此提出一個關於成人而不是積極學習統計學學生的統計素養概念：(1) 面對文章脈絡中的統計訊息、資料相關的議題以及隨機現象，能夠解釋和批判的能力；(2) 對資訊含意的理解、意見與可接受性的結論，能夠討論和溝通統計訊息的反應能力。因此 Gal 提出一個模型解釋除了成年人還包含從學校或大學畢業的學生，應該運用知識與情意兩個層面，使他們能夠理解、解釋和批判性思考與對統計訊息做出反應。此模型分為知識面向與情意面向，分別包括五個知識要素（素養技能、統計知識、數學知識、脈絡知識、批判質疑）與兩個情意要素（信念和態度、批判的立場）。

Merriman (2006) 以 Gal (2000) 提出的統計素養進行定義-包括技能、知識、態度、情意、批判性思考、討論和做出判斷的能力。統計素養的內在思維能力是疊加在更廣泛的統計和數學知識庫上，在閱讀和理解文本、表格和圖表方面的讀寫能力也發揮了作用。Merriman 認為媒體報導在傳播統計訊息給民眾上扮演關鍵角色之外，也是學校用來提高學生統計素養的好方法，可以讓學生學會如何做出合理的判斷與決策，並形成在未來發展思慮周全的觀點。因此設計了一個 12 小時以媒體報導為素材的統計素養教學單元，並在課程前後皆讓受試者對統計素養概念的媒體報導做簡短的問答做為前測與後測，以此分析統計素養能力的差異，結果顯示教學實驗後學生的統計素養有顯著的提高，後測問卷也指出學生能意識到會有具誤導性的圖表陷阱，需要自己評斷所讀到資訊的正確性而不是全盤接收，發展出批判和反思的能力。

二、統計素養能力層次

Watson (1997) 認為在社會脈絡中的統計思考是統計教育重要的一部分，為

了解釋在社會媒體報導中出現的統計資訊所需的能力，提出統計素養的三層結構：
(1) 能理解基本的機率和統計術語；(2) 能將機率和統計概念的語言廣泛的應用在社會議題中；(3) 面對沒有適當統計依據的主張或結論，能運用更複雜的統計概念做批判性思考。

Watson 並在 1993、1995、1997 與 2000 年針對三年級到九年級共 3,852 名澳洲學生，使用內含 80 個問題的統計試卷進行施測調查(Watson & Callingham, 2003)，以了解學生在統計方面的理解，運用 SOLO (Structure of the Observed Learning Outcome) 分類法對學生的答案進行編碼，並藉由 Rasch 分析提出統計素養的六個能力層次，分別是：個人獨特性的 (Idiosyncratic)、非正式的 (Informal)、不一致的 (Inconsistent)、一致但無批判的 (Consistent Non-critical)、批判的 (Critical) 和數學批判的 (Critical Mathematical)。

如表 2-1-1 的六個統計素養能力層次架構所示：層次一的表現以個人經驗解讀文章，缺乏連結文章脈絡的能力；層次二雖然對複雜的術語和情境有初步概念，能夠簡單計算表格、圖形和機率解讀，但是僅限非直觀的統計概念；層次三相較前兩個層次更強調與文章脈絡的連結，主要著重在統計概念的性質而非量化的使用，能做出適當的結論但無法說明理由；層次四能深入瞭解文章內容並多方面使用統計術語，已能整合平均值、簡單機率、統計圖特性等等的數學和統計技能，但不具批判性思考的能力；層次五 (批判的) 和層次六 (數學批判的) 皆需要批判性思考的能力，差別只在於其中的數學能力，層次五能夠適當地使用統計術語，能理解統計的性質且有量化的認知，但尚無法使用複雜的比例推理；層次六則是任何文章脈絡皆能以批判和質疑的態度處理，而且在媒體的文章脈絡中使用比例推理，以及能察覺到不確定性並詮釋語言與脈絡中的細微差別，還能對那些缺乏適當統計依據的主張進行質問。後續有許多研究者參考此統計素養能力層次架構為依據，作測驗評分編碼 (翁意茹，2006；趙于甯，2011；陳展興，2015)。

表 2-1-1 統計素養能力層次架構 (Watson & Callingham, 2003)

| 層次 | 層次名稱 | 層次特色 |
|----|--------------------------------------|---|
| 一 | 個人獨特性/自我的 (Idiosyncratic) | <ul style="list-style-type: none"> • 只用自身的想法瞭解文章內容 • 重複使用專有名詞 • 逐一計算和解讀表格內容等基本數學技能 |
| 二 | 非正式的 (Informal) | <ul style="list-style-type: none"> • 僅限於口語或非正式的瞭解文章內容 • 反映出直觀的非統計概念 • 對複雜的術語和情境有初步概念 • 簡單計算表格、圖形和機會 |
| 三 | 不一致的 (Inconsistent) | <ul style="list-style-type: none"> • 通常以支持性的方式，對文章內容做選擇性瞭解 • 能接受結論但尚無法說明理由 • 只能質性地討論但沒辦法量化地使用統計概念 |
| 四 | 一致但無批判的 (Consistent Non-critical) | <ul style="list-style-type: none"> • 能適當瞭解文章內容，但不具批判性 • 能多方面使用統計術語 • 能察覺到機率的變異 • 結合平均值、簡單機率、統計圖特性等的統計技能 |
| 五 | 批判的 (Critical) | <ul style="list-style-type: none"> • 能以批判和質疑的態度面對熟悉或不熟悉的文章脈絡，而這些文章內含適當的專門術語，但不涉及比例的推理 • 能察覺機率的變異且能夠解釋其性質 |
| 六 | 數學批判的 (Critical Mathematical) | <ul style="list-style-type: none"> • 能以批判和質疑的態度面對任何文章脈絡 • 在媒體的文章脈絡中使用比例推理 • 在做預測時，能察覺到不確定性並詮釋語言的微妙層面 |

三、CARE 統計素養

Schield (2010) 針對統計資料的消費者，也就是那些接收統計資料結果而非統計資料的生產者，能夠有聚焦的重點，Schield 整理以前相關的研究與調查結果，將所有會影響到統計值的相關因素統整歸類，主要分成以下四類：「Context 變項的脈絡」、「Assembly 變項的集合」、「Randomness 隨機性」和「Error 偏差」，Schield 學者將這四個需要注意 (take CARE) 的統計數值影響合起來統稱 CARE，每個面向都對統計數據產生明顯的影響。

(一) 變項的脈絡 (Context)

變項的脈絡主要是討論各變項之間的比較：(1) 分數與比例的比較；(2) 倍數的差異；(3) 實驗設計產生的影響；以及 (4) 研究中沒考慮到的混淆變項的影響。

將「Context 變項的脈絡」置於第一項是基於人文教育中脈絡的重要性，以及統計學家將統計學從數學中區分出來的關鍵就是脈絡，因為統計學是關於收集解釋資料的一種數學學科分支，所以脈絡在其中有極高的重要性。為了瞭解脈絡對統計值的影響，一個具備統計素養的人必須了解不同類型研究設計的特徵和益處，例如：實驗研究與觀察性研究、縱貫性研究與橫斷性研究等類型的差別。具有統計素養者還必須瞭解會影響與改變統計值的初階方法，例如：比較分子分母的組成、使用平均值的概念、對不同比例做比較等等，以及更進階的方法，例如使用加權平均數改變比例和百分比的值。他們還必須有描述以及比較表格和圖表中的比例和百分比的含意。而在實驗研究的過程中，在做好控制的變項與其他被忽略的變項之間，存在著混淆變項在影響著自變項與依變項的因果關係。

本研究將「Context 變項的脈絡」結合科學新聞中常見的錯誤，改編為 CARE 統計素養指引的「C-1 比例的比較」、「C-2 基準值差異的比較」與「C-3 混淆變項」，探討科學新聞在變項之間做比較時，對統計值與結果產生的影響。

(二) 變項的集合 (Assembly)

變項的集合關注在變項的各種選擇，以下列方式做選擇時對統計值產生的影響：(1) 群組定義的選擇；(2) 平均值與中位數的選擇；(3) 比較類型的選擇，例如絕對差異與相對比例的選擇；(4) 比例類型的選擇，例如倒數的混淆或檢察

官的謬誤；(5) 選擇何種圖形、圖表或統計資料以表示統計摘要和結果。

具有統計知識的消費者需要知道，對變項做不同的選擇，例如：對群體的定義變嚴格，群體的規模會隨之減少，對於群體整體的比例則會增加；同一筆資料，取平均值或中位數有可能會產生極為不同的結果；如何使用絕對差異或是相對比例來加強資料的優勢；將比例的倒數意義混淆或是誤植，資料的涵義會完全不同；針對你想闡述的摘要與研究結果，如何選擇合適的圖表。

本研究將「Assembly 變項的集合」結合科學新聞中常見的錯誤，改編為 CARE 統計素養指引的「A-1 比較類型的選擇」與「A-2 比例類型的選擇」，探討科學新聞使用統計資料時，文章作者對數據處理所做的各種選擇，是以哪種角度在討論比例關係。

(三) 隨機性 (Randomness)

(1) 探討事件的平均和巧合對機率造成的影響，例如因為某件事發生很多次，所以誤以為會再次發生的熱手謬誤；(2) 以及統計顯著與實際顯著性之間的差異，透過樣本統計量推估信賴區間，能夠以此判定兩組之間是否具有顯著差異。

本研究將「Randomness 隨機性」結合科學新聞中常見的錯誤，改編為 CARE 統計素養指引的「R 隨機性(信賴區間)」，探討科學新聞中統計資料的抽樣樣本分佈隨機性對顯著性判斷的影響。

(四) 偏差 (Error)

探討樣本與目標人群之間會使抽樣時產生偏差的各種因素，對研究觀察的結果與現實之間產生差異的影響：(1) 抽樣偏差，抽取樣本與目標母群之間造成的差異；(2) 測量偏差，事物固有的不確定性因素造成量測與真實間產生的偏差，例如儀器失誤、設備缺陷或測量環境的變化等等；(3) 受試者偏差，實驗對象不像研究者認為的理解問題，填錯答案或是以各種意料不到的方式給出並非本意的答案。

本研究將「Error 偏差」結合科學新聞中常見的錯誤，改編為 CARE 統計素養指引的「E 抽樣偏差」，探討科學新聞中引用的研究設計，是否因為抽樣的方式造成偏差，以及對樣本代表性提出質疑。

上述四個關於影響著統計值因素的分類，共有兩篇實證研究的結果如下：

- (1) Schield 與 Schield (2007) 整理了 250 份以統計性資料為基礎的新聞報導進行分析，發現有 75% 的內容因為屬於觀察性研究而容易受到混淆變項的影響，74% 的內容涉及變項的集合，其中超過一半的內容以百分之、百分比、比例或機率等詞語來表達比率的概念。
- (2) Raymond 與 Schield (2008) 對 160 份基於數值類型的文章進行分析，發現 27% 的文章在建構類別或量測值時涉及變項的選擇，62% 的文章存在因果關係的連結。就 CARE 的四個類別而言，42% 的文章其中的統計資料受到混淆變項的影響、17% 的文章受到變項集合的影響、11% 的文章受到偏差影響以及 9% 的文章受到隨機性的影響。

這些文章或新聞大多是以健康為主的新聞報導，不包括雜誌或新聞專欄等等較長篇幅的文章，也不包含政府統計部門發布的詳細報告。不過分析還是提供了重要發現，與後兩項的「Randomness 隨機性」和「Error 偏差」相比，「Context 變項的脈絡」與「Assembly 變項的集合」在日常新聞中更為普遍，此發現對於決定評估什麼資料以及如何評估是相當重要的。

Schild(2010)發現統計素養課程在美國四年學制的大學中變得越來越普遍，這波課程數量的成長使 40% 的非主修量化專業的美國大學生有更好的機會批判性思考新聞中的數字，再加上借助媒體有關統計的數據、獨特的內容和新的評估模式，統計素養作為文組科系中的新興課程正在迅速崛起。

第二節 統計素養測驗

國內外關於研究統計素養的研究繁多，眾多學者致力於發展統計素養測驗，本節分別整理國外與臺灣現有的統計素養測驗，依照時間順序呈現於表 2-2-1 國外統計素養測驗發展以及表 2-2-2 臺灣統計素養測驗發展。

由表 2-2-1 與表 2-2-2 可知，學者在發展統計素養測驗的過程中，會依據研究對象與採用統計素養定義的不同，對於測驗內容與向度做適當的更動，如表 2-2-1 中：Watson (1997) 討論在社會脈絡下統計思考的重要性，將統計素養分為三層能力結構以解釋社會中出現的各種統計資料，並討論在課堂實踐的可能性；Reston (2005) 與 Merriman (2006) 皆主張培養學生解讀媒體常報導的統計資料，以媒體作為統計素養題目素材測驗學生的能力，幫助學生意識到統計概念的實用性並產生統計素養的學習動機，為學生提供真實的範例以因應畢業後的真實社會情境。因此不少學者開始採用媒體報導作為統計素養測驗的素材，以此評估學生的統計素養；Schield (2006) 則是將重點放在統計圖表中的比例和百分比上，認為具有統計知識的人必須能夠閱讀和解釋新聞媒體或一般工作與生活中遇到的數據，因此將對象放在學生、大學教授與數據分析師間的統計圖表解讀表現差異；Ziegler & Garfield (2018) 提出的 The Basic Literacy In Statistics (BLIS) 用來分析學生參與大學專業統計課程後具備的統計素養能力，可用於評估學生在各種統計課程主題中的學習情況並給予授課教師關於學習困難的教學回饋。

表 2-2-1 國外統計素養測驗發展

| 學者 (年代) | 測驗名稱 | 測驗內容 | 對象 |
|---------------------------------|--|---|----------------|
| Watson (1997) | Statistical Literacy Assessment | (1) 機率和統計術語的基本瞭解； (2) 對內含在社會議題內容中的統計語言和概念的瞭解； (3) 對缺乏適當統計依據的結論具質疑的態度，並應用概念作反駁。 | 高中 |
| Reston, (2005) | The Statistical Literacy Assessment Scale (SLAS) | (1) 面對真實生活中的情境問題以統計方式回答； (2) 解釋不同媒體中的圖表數據。 | 大學 |
| Merriman (2006) | Statistical Literacy Survey | 以媒體報導作為素材： (1) 統計知識； (2) 數學知識； (3) 脈絡知識； (4) 一般素養； (5) 批判技能。 | 高中 |
| Schild (2006) | Statistical Literacy Inventory | 針對表格與圖表中的比例與百分比做解讀。 | 大學師生與 數據分析師 |
| Ziegler & Garfield (2018) | The Basic Literacy In Statistics (BLIS) | (1) 樣本和母體間的差異； (2) 解讀圖表； (3) 描述性統計； (4) 經驗抽樣分布； (5) 信賴區間； (6) 隨機分布； (7) 假設檢定； (8) 迴歸與相關。 | 大學 |

國外的統計素養調查對象多針對高中、大學與成年人，學者們強調統計素養的重要性包含進入社會後的公民素養與媒體識讀。而國內對於統計素養的研究對象從國小學生至大學生皆有，其中國小至高中的主要測驗重點在統計知識的理解上，但是對於將統計概念應用於媒體素材的相關研究則仍顯不足，因此本研究希望經由對國內外相關文獻的統計素養測驗整理，以及現有統計素養測驗編製內容及對象的參考，改編成以科學新聞為文本素材的統計素養測驗，以探究我國非理工科系大學生的統計素養表現，以此作為日後統計素養教育與科學新聞做連結的參考。

表 2-2-2 臺灣統計素養測驗

| 學者 (年代) | 測驗名稱 | 測驗內容 | 對象 |
|---------------|------------------|---|------------|
| 楊靜惠 (2006) | 統計素養問卷 | (1) 數學知識/統計知識；(2) 推論； (3) 批判質疑。 | 國中 |
| 翁意茹 (2006) | 統計素養問卷 | (1) 語文知識；(2) 統計知識；(3) 數學知識；(4) 內容知識及批判質疑。 | 大學 師資生 |
| 曾瑞玫 (2009) | 統計素養測驗 題本 | (1) PISA 數學素養測驗「不確定性」；(2) 統計圖表、統計量、機率與抽樣等統計概念。 | 高中 |
| 趙于甯 (2011) | 統計素養測驗 | 以媒體報導作為試題脈絡：(1) 統計圖；(2) 統計表；(3) 統計量。 | 大學 |
| 陳展興 (2015) | 國小職前教師 統計素養問卷 | (1) 語文知識；(2) 統計知識；(3) 數學知識；(4) 內容知識；(5) 批判質疑。 | 國小職 前教師 |
| 吳思嫻 (2017) | 六年級學生的 統計素養測驗 | (1) 樣本；(2) 基本圖表概念；(3) 基本機會概念；(4) 圖表變異概念；(5) 機會變異概念。 | 國小 |

其中趙于甯(2011)綜合許多學者的論點(Gal, 2002a, 2002b; Holmes, 2003; Lehohla, 2002; Reston, 2005; Schield, 1999, 2004, 2010; Snell, 1999; Wallman, 1993; Watson, 1997; Watson & Moritz, 1994), 將統計素養的主要能力歸納為三項:能夠「理解」統計資料, 面對不正確的資訊有能力「批判質疑」, 有能力將自己的資料解讀論點與他人進行「溝通」, 以這三個能力作為統計素養測驗的評量向度, 以及參考 Watson & Callingham (2003) 的統計素養層次架構, 針對受試者在三個評量向度上的答題表現, 簡化為三個能力層次進行統計素養能力編碼, 分析大學生的統計素養與統計態度表現, 結果發現在統計素養層次表現上, 約有九成大學生達到統計素養第二層次, 具備基礎的統計溝通與推論能力, 但是只有五成具備進階的統計理解、溝通及批判能力, 能活用統計之事進行邏輯性思考, 並對缺乏適當統計依據的主張進行批判性思考。因此本研究除了討論大學生科學新聞中的統計素養現況, 以分析受試者的統計素養相關先備知識, 並探討在給予改編自 Schield(2010) 的 CARE 教學指引後, 是否能提升受試者的統計素養, 幫助受試者建立批判性思考的方向。

第三節 統計素養教學

國內外的統計素養研究主要專注於統計素養的測驗上, 以開發測驗工具評估不同研究對象較缺乏的能力與面向, 對教學提出建議。關於統計素養教學的研究, Weiland(2017)有感於許多數學教育學者主張將有意義的社會和政治問題納入學校數學課程中, 以避免學生未來沒有使用數學解決問題的能力, 以及認為統計素養是一項能夠對個人面對各種爆炸性統計資料時, 做出評估和決定的關鍵能力, Weiland 分析正規 K-12 學校數學課程中批判素養的主要思維, 並與統計素養做連結。Weiland (2019) 強調批判性思考的重要性, 認為統計學對 K-12 學生中的教育價值是使學生成為社會的公民, 方法來自於構成統計學過程的核心實踐: 提出問題、蒐集相關資料並在脈絡中分析資料, 用精確、有邏輯的論述, 將資料清楚回答別人提出的問題, 不是讓學生完全理解或解決社會問題, 而是要利用統計資料讀取文字和解讀世界議題。

Schield (2004) 發表一篇統計素養課程設計的建議, 認為完善的統計素養課程應以統計在學生日常、公民或職涯生活中的作用為基礎。由於統計的重點在於數據資料, 應該根據資料去訂定統計思考的主題: (1)研究設計-透過取得資料

的方法討論統計概念，其中有隨機性、誤差與混淆變項的影響；(2)資料類型-根據所需要的資料模型討論單變量或多變量分析的差別，並設計了一學年的統計素養課程，並蒐集學生的課堂回饋表示 43%的學生非常同意該課程有助於他們培養批判性思考的能力，18%的學生非常同意將此課程列為必修學分。Schield(2008) 探討課堂上的大學生對於百分比與比例的文句，以及以表格顯示的百分比解讀能力發現，許多學生無法正確解釋其含義，因此對統計與量化課程建議教育者應多注重如何用詞語解釋比例而不是只有使用公式代數，並加強解讀圖表的能力。以及提出對於未來不需要在大學時期學習量化專業相關科系的學生，將教學重點放在介紹以表格和圖表顯示的比率和百分比上。這幾篇研究主要是對於在課堂教學上，將統計素養連結於社會議題以及日常生活應用的建議，需要較長的教學時數將完整的統計概念教給學生，因此本研究想嘗試透過教學指引，能否幫助學生更簡潔明確的批判性思考方向。

Schield (2004) 表示統計素養是為了資料的消費者 (consumers)，而統計能力則是屬於資料的生產者 (producers)，對於主修不需要量化專業技能課程的非理工科系大學生，仍然需要統計素養課程幫助他們發展以統計學為依據評估論點的技能，並且在畢業後作為現代公民與資訊消費者，認識到統計素養在日常生活與職業領域中的價值。

為了因應世界上快速變動的教育改革，我國的國民中小學與高級中學課程也不停的在翻新以提升國家未來的競爭力，因此現行的國民中小學以及高中課程綱要經歷多次的變革，本研究以民國一零八年正式實施的最新版本課綱「十二年國民基本教育課程綱要」，所公布的學習表現為主要對照，以下簡稱「108 課綱」，將 8 個科學新聞中的統計素養概念與高中數學課程綱要內容做對照，列於表 2-3-1。

「P 機率的解讀」、「C-1 比例的比較」、「C-2 基準值差異的比較」、「A-1 比較類型的選擇」與「A-2 比例類型的選擇」題型所需要的能力皆於高中一年級課程的學習表現相符合。「A-2 比例類型的選擇」為集合單元的基礎，探討同一主體兩種不同比例的差異；「C-1 比例的比較」則是集合的中階課程，需要分辨不同主體間做比例的比較，並透過文氏圖的概念區辨母體差異對比較統計值大小產生的影響。「C-2 基準值差異的比較」與「A-1 比較類型的選擇」概念則是在數據分析單元，學習如何選用適當的統計量作為描述數據的參數。

而「C-3 混淆變項」、「R 隨機性(信賴區間)」與「E 抽樣偏差」在 108 課綱的學習表現無法對應：「信賴區間與信心水準」在 95 年開始引進高中數學課綱，此觀念的解讀難度過高，以及受限於課程時長，於 108 課綱將之移除，且本研究的實驗對象於高中課程皆以「98 年普通高級中學課程綱要」為主，因此本研究將「R 隨機性(信賴區間)」的概念與 98 課綱的學習內容作對應；而「C-3 混淆變項」與「E 抽樣偏差」兩個概念因為屬於研究設計的範疇，於高中數學領域並無做討論。

表 2-3-1 與教育部課程綱要的對照

| 主題單元 | 對應課綱 | CARE 題型/概念 |
|------|---|-----------------------------|
| 集合 | 對應「十二年國民基本教育課程綱要，數學領域」的學習表現： d-V-1 認識集合，理解並欣賞集合語言的簡潔性，能操作集合的運算，能以文氏圖作為輔助，並能用於溝通與推論。 | A-2 比例類型的選擇 C-1 比例的比較 |
| 數據分析 | 對應「十二年國民基本教育課程綱要，數學領域」的學習表現： d-V-2 能判斷分析數據的時機，能選用適當的統計量作為描述數據的參數，理解數據分析可能產生的例外，並能處理例外。 | C-2 基準值差異的比較 A-1 比較類型的選擇 |
| 不確定性 | 對應「十二年國民基本教育課程綱要，數學領域」的學習表現： d-V-3 理解事件的不確定性，並能以機率將之量化。理解機率的性質並能操作其運算，能用以溝通和推論。 | P 機率的解讀 |

| | | |
|------|---|-------------|
| 信賴區間 | 容： | R 隨機性(信賴區間) |
| | 高二數學選修教材綱要」的學習內 常態分佈、信賴區間與信心水準的解 讀。 | |

實驗設計

C-3 混淆變項

E 抽樣偏差

註 1：引自十二年國民基本教育課程綱要(頁 7~19)

註 2：引自 98 年普通高級中學課程綱要(頁 71~74)

學生在高中時學習基本的機率與統計知識，進入大學後部分非理工科系的必/選修課程規定中有統計相關課程，例如人文社會科學院或教育相關系所，因此研究者蒐集整理非理工科系的大學統計必/選修課程課程，發現課程學習共通點有推論統計的基本概念、學習各類統計方法的計算方式與正確用途、運用統計軟體進行資料分析工作與相關研究，以及應用統計學方法於教育或人文社會科學研究領域中，因此我們想探討對於高中一類組，到了大學因為科系規定修畢或正在修習的學生，與高中畢業後就不再接觸任何統計或數學相關課程的學生，在科學新聞中的統計素養是否有差異。

第三章 研究方法

本研究欲探討非理工科系大學生在科學新聞中，統計素養的能力表現，以及透過 CARE 統計素養教學的指引，瞭解學生的學習成效與在科學新聞的實際表現。因 COVID-19 疫情嚴峻，為維護受試者與主試者安全，避免接觸感染，實驗改以 Google Meet 線上會議，由受試者端分享螢幕畫面的方式進行。研究過程中記錄受試者在前後測閱讀文本後的批判性思考作答，以及教學時的理解程度，做為資料分析之來源。

本章針對研究方法共分為五節，分別為「第一節，研究對象」、「第二節，研究設計」、「第三節，研究流程」、「第四節，研究工具」與「第五節，資料分析」。

第一節 研究對象

本研究探討非理工科系大學生，在科學新聞中統計素養的能力表現，以及大學是否修習統計相關課程對統計素養的影響。因此透過網路招募高中時為一類組，並且非理工科系之大學生共計 34 名受試者進行實驗（主要學院為文學院、社會科學院與教育學院）。其中共有 17 位非理工科系受試者因科系必/選修規定或個人需求，例如教育相關或人文社會學院等相關科系，於大學時期修畢或正在修習統計相關課程，另外 17 位受試者自高中畢業後再無接觸統計相關課程。

第二節 研究設計

本研究想了解非理工科系大學生在科學新聞中的統計素養表現，以及給予 CARE 統計素養教學指引後，統計素養能力是否有所提升。因此本實驗改編網路上常見的科學新聞文本，以及根據 Schield (2010) 分類會對統計數據產生影響的四個面向，改編成科學新聞中常見的 8 類統計素養問題，作為統計素養前後測驗題型，以此探討受試者在科學新聞中的統計素養相關先備知識（以前測表現作為測量指標）、統計素養教學的學習表現（以理解程度作為測量指標），以及經過統計素養概念教學後，運用 8 項 CARE 統計素養指引，感知科學新聞中的問題點並做批判性思考的能力表現（以後測表現作為測量指標）。整個實驗根據受試者的統計素養測驗表現與學習成效進行分析，探討非理工科系大學生科學新聞的統計素養表現與 CARE 統計素養指引的成效。

第三節 研究流程

本研究為個別實驗，所有受試者單獨進行。實驗流程分為 5 個部分，依序為：實驗準備、前測、教學、後測以及實驗後的訪談，如下圖 3-3-1 所示。本節介紹將會以研究流程為主，視需要搭配研究工具截圖做輔助說明，詳細設計原則將於第四節的研究工具介紹。

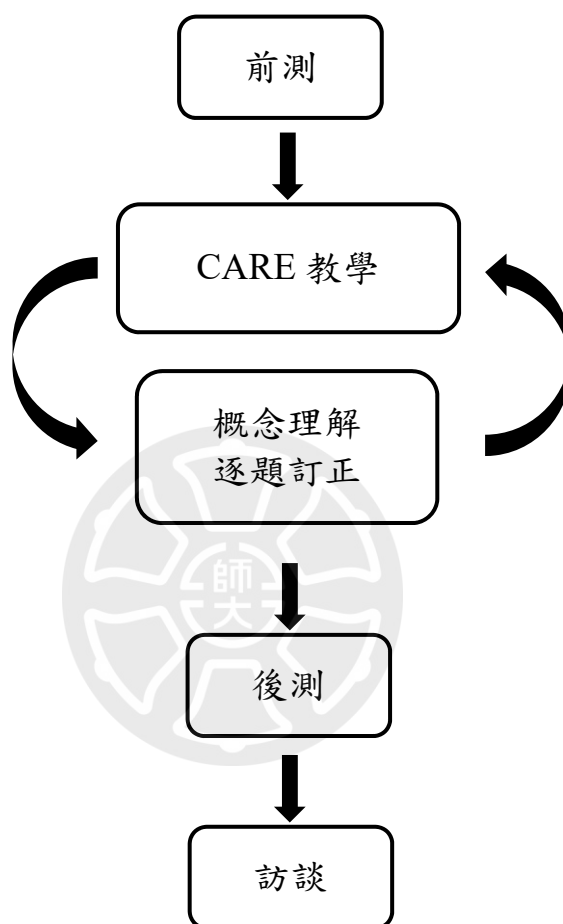


圖 3-3-1 實驗流程圖

一、實驗前

實驗以 Google Meet 線上會議進行，首先請受試者分享螢幕畫面，並詢問受試者是否同意主試者端開啟螢幕錄音錄影，以此做為研究記錄，受試者同意後閱讀實驗說明與實驗流程，了解實驗目的與流程後簽署受試者知情同意書，隨即開始正式實驗。

二、前測

受試者閱讀一份以線上版 Google 文件模擬網頁版的科學新聞，文本內容參照網路上常見的科學新聞格式，彙整多篇同一主題的科學新聞，而且為了方便計算將其中的統計資料或數據調整簡化，其餘資料盡可能符合真實情況。閱讀完畢後，針對科學新聞中的資料與相對應的題目使用線上版 Google 試算表呈現，如下圖 3-3-2。題目設計為 8 個於科學新聞中常見的陷阱或錯誤的數據解讀，每段落搭配 2~3 題的統計素養測驗題，受試者依據問題將答案以及說明填於欄位內做為書面作答，如果有語意不清的部分主試者會再口頭詢問記錄為口頭說明，以確保完整蒐集受試者的想法。依據受試者對統計素養題目解讀的作答與說明，將答案依照評分規準給予 3~0 分共四個等級的評分，作為科學新聞中的統計素養相關先備知識的表現分數。

The screenshot shows a Google spreadsheet with the following content:

| 題號/類別 | 題目 | 答案 | 理由 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|---|----|--|---------|----|----|--------|----|-------|----|-------|----|-------|----|--------|----|-------------|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|
| <p>科學新聞文本</p> <p>水中微塑膠</p> <p>近年來，塑膠垃圾污染問題越來越嚴重，塑膠微粒的污染已經滲入我們生活各方面，從河川、海洋到土壤、食物以及我們的飲用水等。美國一家調查機構提出我們平常飲用的自來水中，含有肉眼看不見的塑膠微粒，該研究對全球各地14個國家的自來水進行隨機調查，結果顯示平均83%的自來水樣本中含有塑膠微粒。</p> <p>英國新聞頻道Sky News報導指出，全球每年大約產生3億噸的塑膠垃圾，其中只有6%被回收，12%被焚燒，另外的79%被掩埋或是堆入到自然環境中，尤其是海洋，全世界每年海洋平均多出的1,270萬噸的塑膠垃圾，對海洋生物造成嚴重的威脅，例如造成海鳥甚至和鯨魚中毒。</p> <p>奧斯陸-巴黎公約(OSPAR)統計全球許多處海面上廢棄物的種類比例(圖一圓餅圖)，以此看出塑膠物品一直以來佔全球海洋廢棄物的大宗。圖二折線圖表示海面上「每100公尺有多少廢棄物」的年平均價值趨勢(圖上也顯示了各年廢棄物95%信賴區間)。</p> <p>圖一</p> <table border="1"> <tr><th>海廢廢棄物種類</th><th>比例</th></tr> <tr><td>塑膠</td><td>55.00%</td></tr> <tr><td>玻璃</td><td>1.00%</td></tr> <tr><td>紙張</td><td>1.00%</td></tr> <tr><td>金屬</td><td>1.00%</td></tr> <tr><td>其他</td><td>39.00%</td></tr> </table> <p>圖二</p> <table border="1"> <tr><th>年份</th><th>每100公尺廢棄物數量</th></tr> <tr><td>2001</td><td>100</td></tr> <tr><td>2002</td><td>120</td></tr> <tr><td>2003</td><td>140</td></tr> <tr><td>2004</td><td>160</td></tr> <tr><td>2005</td><td>180</td></tr> <tr><td>2006</td><td>200</td></tr> <tr><td>2007</td><td>220</td></tr> <tr><td>2008</td><td>240</td></tr> <tr><td>2009</td><td>260</td></tr> <tr><td>2010</td><td>280</td></tr> <tr><td>2011</td><td>300</td></tr> <tr><td>2012</td><td>320</td></tr> <tr><td>2013</td><td>340</td></tr> <tr><td>2014</td><td>360</td></tr> <tr><td>2015</td><td>380</td></tr> <tr><td>2016</td><td>400</td></tr> <tr><td>2017</td><td>420</td></tr> <tr><td>2018</td><td>440</td></tr> <tr><td>2019</td><td>460</td></tr> <tr><td>2020</td><td>480</td></tr> <tr><td>2021</td><td>500</td></tr> </table> | | | | 海廢廢棄物種類 | 比例 | 塑膠 | 55.00% | 玻璃 | 1.00% | 紙張 | 1.00% | 金屬 | 1.00% | 其他 | 39.00% | 年份 | 每100公尺廢棄物數量 | 2001 | 100 | 2002 | 120 | 2003 | 140 | 2004 | 160 | 2005 | 180 | 2006 | 200 | 2007 | 220 | 2008 | 240 | 2009 | 260 | 2010 | 280 | 2011 | 300 | 2012 | 320 | 2013 | 340 | 2014 | 360 | 2015 | 380 | 2016 | 400 | 2017 | 420 | 2018 | 440 | 2019 | 460 | 2020 | 480 | 2021 | 500 |
| 海廢廢棄物種類 | 比例 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 塑膠 | 55.00% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 玻璃 | 1.00% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 紙張 | 1.00% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 金屬 | 1.00% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 其他 | 39.00% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 年份 | 每100公尺廢棄物數量 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2001 | 100 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2002 | 120 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2003 | 140 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2004 | 160 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2005 | 180 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2006 | 200 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2007 | 220 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2008 | 240 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2009 | 260 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2010 | 280 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2011 | 300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2012 | 320 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2013 | 340 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2014 | 360 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2015 | 380 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2016 | 400 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2017 | 420 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2018 | 440 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2019 | 460 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2020 | 480 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2021 | 500 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 問題1 | 請問文中平均83%的自來水樣本中含有塑膠微粒是什麼意思? (A) 不要喝自來水，因為每一杯水的塑膠微粒濃度為83% (B) 喝自來水時，盡量避免喝到83%塑膠微粒的部分 (C) 每100人，約有83人會喝到塑膠微粒 (D) 每100處地點自來水樣品，約有83處樣品中含有塑膠微粒 (E) 隨機帶往一些地點的自來水，塑膠微粒佔整杯水83% | D | 因為相較於其他選項來說它看起來最正確，也最符合我心中的正確答案(將每一個樣本處所含之塑膠微粒百分比相加再除以樣本數而得之平均數為83%) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 問題2 | 小明看到文章中的研究對全球各地14個國家的自來水進行隨機調查，質疑隨機地點的選擇，不能代表全世界，因此認為抽樣結果有偏差(與現實有差異)，請問小明的質疑合理嗎?為什麼? | 合理 | 因文章中也確實沒提出這14個國家分別是哪些國家及其所在之區域，因此確實可以合理懷疑這14個國家可能都來自相近之處 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 問題3 | 折線圖表示海面上「每100公尺有多少廢棄物」的年平均價值趨勢(圖上也顯示了各年廢棄物95%信賴區間)，請問從折線圖能否推論出2011~2020年，海廢廢棄物逐年上升的趨勢(增加)? | 有 | 因圖表已說是「每100公尺有多少廢棄物」，因此隨著數字增加會讓人有種「海洋廢棄物之數量」也在跟著增加之感覺，再者，從文章主題來看會覺得說此處所放之資源應該是要佐證論點之素材，故廢棄物之數量之年年增加應該就是論點之基點 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

受試者書面作答

受試者口頭說明/主試者註記

圖 3-3-2 統計素養前測範例

三、統計素養教學

前測題目填寫完畢後，請受試者點擊連結至 Youtube 影音平台，播放主試者錄製的半互動式教學影片，影片中講解面對科學新聞裡埋藏的陷阱可以透過哪些統計素養的概念去批判思考，介紹 Schield (2010) 分類的四個會對統計數據產生影響的面向，以及衍伸的 8 個常見於科學新聞中產生錯誤解讀的要點，概念與前測的 8 道題目相對應。教學任務以每題為一小階段，請受試者切換 Youtube 觀看教學影片與 Google 試算表分頁進行訂正，依序完成共 8 道題目的教學任務，步驟舉例如下：(1) 受試者觀看「C-1 比例的比較」教學概念與講解前測題目；(2) 訂正「C-1 比例的比較」的前測題目。主試者依據受試者訂正題目的完整度確認學習狀況，並在受試者提出有理解困難的地方給予二次教學，依理解程度作 3~0 分的評分，作為學習表現分數。

四、後測

統計素養概念教學結束後稍作休息，接著進入後測，測驗受試者將 8 個統計素養指引應用於科學新聞中的批判思考表現。後測科學新聞以線上版 Google 文件的形式呈現，如下圖 3-3-3。與前測科學新聞設計原則相同，但改以空氣汙染為主題改編自網路媒體常見的科學新聞，將文本中設計的統計資料埋錯畫底線標註為問題點，請受試者使用大綱欄位所顯示的 CARE 統計素養指引列表，檢視問題點分別是屬於哪一種 CARE 要點的陷阱或是錯誤解讀等等，受試者使用線上版 Google 文件的註解功能加註在段落旁，書面作答問題點屬於何種 CARE 統計素養指引的代碼以及對於問題點的批判性思考說明，與前測相同，如果受試者說明文字有語意不清的部分主試者會再口頭詢問並記錄，以確保受試者將想法全部表達完整。主試者再依據受試者註解的說明，依照與前測規則一致的評分規準給予 3~0 分的評分，作為統計素養後測分數。

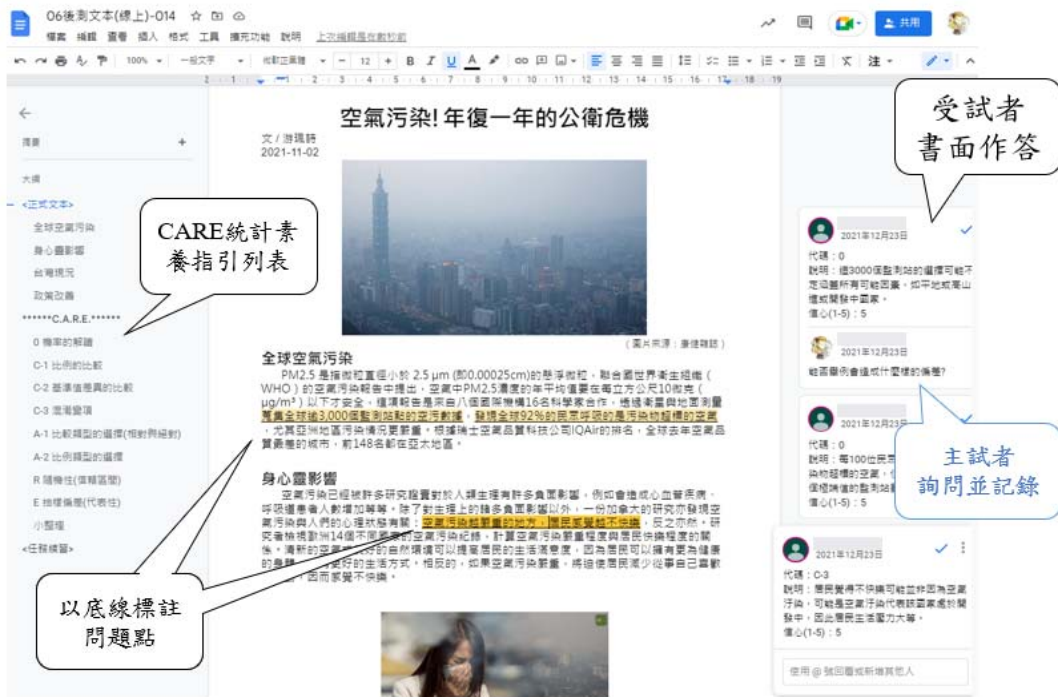


圖 3-3-3 統計素養後測範例

五、訪談

實驗最後進行訪談，了解受試者平常對於科學新聞的閱讀習慣，以及受試者在面對本實驗設計匯集許多常見統計資料錯誤的科學新聞時，進行批判性思考的困難處。

第四節 研究工具

為了解受試者在閱讀科學新聞或需要統計知識的文章時，是否有統計素養的敏感度，亦或是直接採用文章作者的推論，以及透過統計素養要點的指引，是否能幫助受試者提升批判思考的能力，本實驗設計劃分為三個階段，依序為前測、教學與後測，研究流程與研究工具的對照如表 3-4-1 所示。本節分為四個部分，第一部分先介紹根據 Schield (2010) 改編的統計素養 8 題型/概念；第二部分介紹前後測的科學新聞文本，皆取材改編自網路上常見的幾個科學新聞媒體；第三部分討論前後測的統計素養測驗，說明測驗的設計原則、評分規準、專家審查與評分者信度，呈現前後測共同的改編原則與對比其中的差異；第四部分是統計素養教學的影片內容、受試者學習表現的評分規準、專家審查與評分者信度。

表 3-4-1 研究工具簡介

| | 實驗流程 | 研究工具 |
|----|------------|-------------|
| 前測 | 閱讀科學新聞 | 科學新聞文本-塑膠汙染 |
| | 進行統計素養前測 | 統計素養前測 |
| 教學 | 觀看教學影片 | 教學影片 |
| | 前測統計素養測驗訂正 | 前測訂正區 |
| 後測 | 閱讀科學新聞 | 科學新聞文本-空氣汙染 |
| | 使用統計素養指引 | 統計素養指引表 |
| | 進行統計素養後測 | 統計素養後測 |

一、統計素養題型與概念

本研究將 Schield (2010) 提出的四個會對統計數值產生明顯影響的因素，結合科學新聞中常見的錯誤類型，並於第一項加入最常見的機率解讀，改編成 8 項科學新聞中的統計素養類別，分別在前後測驗稱為「統計素養 8 類題型」，以及教學時的「統計素養 8 個概念」，詳細前後測題目對照可參考表 3-4-3 的統計素養測驗。

Schild (2010) 總共分為四大類：「Context 變項的脈絡」、「Assembly 變項的集合」、「Randomness 隨機性」與「Error 偏差」，在解讀統計資料時需注意 (take CARE) 這些影響因素的重要性。本研究依據四大分類與其中的細項，依序統整介紹如下。

(一) P 機率的解讀

在科學新聞中常見的解讀錯誤，使用統計資料前應該先理解機率的涵義，並且在文章作者引用研究數據時，批判性思考其中是否有不同意思，例如前測使用 83% 的自來水樣品中含有塑膠微粒作為文本材料。

(二) C-1 比例的比較

根據 Schild (2010) 在「變項的脈絡」中討論各變項之間做比較的 (1) 分數與比例的比較，改編成討論科學新聞中兩筆統計資料的比例做比較，並且區辨兩比例的母體差異對比較統計值大小造成的影響，例如前測使用占比與回收率這兩類統計資料為材料做討論。

(三) C-2 基準值差異的比較

根據 Schild (2010) 在「變項的脈絡」中討論各變項之間做比較的 (2) 倍數的差異，改編成統計資料間在做比較時，比較的對象改變基準值需做相對應的調整，並計算出正確的倍率，例如前測使用塑膠容器的回收率與紙容器的回收率這兩類統計資料為材料做討論。

(四) C-3 混淆變項

根據 Schield (2010) 在「變項的脈絡」中討論各變項之間做比較的 (4) 研究中沒考慮到的混淆變項的影響，改編成討論科學新聞採用實驗研究的結論時，研究設計是否存在混淆變項的可能性，科學新聞的讀者在接受研究結論之前能夠反思有沒有其他影響因素也會影響研究結果，例如前測使用塑化劑與頸動脈硬化的相關性研究做討論。

(五) A-1 比較類型的選擇

根據 Schield (2010) 在「Assembly 變項的集合」中討論各變項之間做選擇的 (1) 比較類型的選擇，即選擇絕對差異或相對比例來比較兩個數值間的差異，改編成討論科學新聞中對統計資料採用絕對差異或是相對比例做討論的不同選擇，進而使資料消費者的下一步決策或行動方案做出不同的判斷，例如前測使用減塑總量與減塑效益 (節省 400 公斤與 80% 的塑膠原料)，討論總量與效益兩者擁有的不同優勢。

(六) A-2 比例類型的選擇

根據 Schield (2010) 在「Assembly 變項的集合」中討論各變項之間做選擇的 (2) 比例類型的選擇，例如倒數的混淆或檢察官的謬誤，改編成針對同一對象的兩種統計資料知道其分母分子的意義，並能夠區辨兩者的差異，例如前測使用塑膠容器的占比與回收率這兩類統計資料做討論。

(七) R 隨機性(信賴區間)

根據 Schield (2010) 在「Randomness 隨機性」中討論 (2) 統計顯著與實際顯著性之間的差異，透過樣本統計量推估信賴區間判定兩組間是否具有顯著差異。改編成科學新聞中使用「顯著增加」時，讀者是否理解信賴區間誤差線與顯著性的關聯，例如前測使用海灘廢棄物數量的平均線，在每年間是否有顯著增加為材料做討論。

(八) E 抽樣偏差

根據 Schield (2010) 在「Error 偏差」中討論樣本與目標母體之間使抽樣時產生偏差的各種因素，對研究觀察的結果與現實之間產生差異的各種影響：(1) 抽樣偏差，抽取樣本與目標母群之間造成的差異。改編成討論科學新聞中採用研究報告的結論時，受試者能否批判性思考研究樣本的蒐集是否因為抽樣方式的不同，造成推測母體時產生偏差，並對樣本代表性提出質疑，例如前測使用塑化劑與頸動脈硬化的相關性研究做討論。

二、科學新聞文本

為了更貼近受試者日常在網路閱讀科學新聞，版面設計以線上版 Google 文件模擬網頁版新聞如下圖 3-4-1，完整的科學新聞文本參見附錄一。



圖 3-4-1 前測科學新聞文本

內容是參照網路上常見的科學新聞，彙整多篇同主題的科學新聞，改編成一份內含 8 類統計素養相關的問題，為了方便計算將其中的統計資料或數據調整簡化，其餘資料盡可能符合真實情況，前後測的科學新聞文本引用來源範例如表 3-4-2 所示，皆挑選較有可信度的科學新聞或較常見的新聞網站。

表 3-4-2 前後測科學新聞來源範例

| 主題 | 文本段落 | 文本來源/新聞標題 | |
|----|------|---|--|
| 前測 | 塑膠汙染 | 該研究對全球各地 14 個國家的自來水進行抽樣調查，結果顯示平均 83% 的自來水樣本中含有塑膠微粒。 | BBC News 中文 塑料污染：鯨魚的痛苦 你應該懂 |
| | | 圖二折線圖表示海灘上「每 100 公尺有多少廢棄物」的年平均趨勢(圖上也顯示了各年廢棄物 95% 信賴區間)。 | 環境資訊中心 海洋生物誤食塑膠垃圾 或纏死 近十年增四成 |
| | | 台大研究團隊以多變項控制塑化劑濃度，發現塑化劑濃度與頸動脈硬化機率呈現正相關。 | ETtoday 健康雲 台大研究：塑化劑無所不在！國小教室內外都測得...也傷心血管 |
| | | 除了對生理上的諸多負面影響以外，一份加拿大的研究亦發現空氣污染與人們的心理狀態有關：空氣污染越嚴重的地方，居民感覺越不快樂。 | PanSci 泛科學 你快樂嗎？空氣污染影響我們的幸福感受 |
| 後測 | 空氣汙染 | 下圖為環保署空氣品質監測報告，呈現每年 PM2.5 濃度平均值，PM2.5 濃度自 2007 年起由 34.5 $\mu\text{g}/\text{m}^3$ 下降至 2017 年只有 20.7 $\mu\text{g}/\text{m}^3$ 。 | 台塑企業 熱門議題 台灣 PM2.5 的主要污染來源 |
| | | 全國 PM2.5 紅色警示次數由 2015 年 997 次，降至 2019 年 146 次（次數降低 851 次，降幅 85%）。 | 行政院 政策與計畫 空氣污染防制大作戰— 保護國人健康，讓臺灣 環境永續 |

三、統計素養測驗

本研究依 Schield (2010) 提出的四個會對統計數值產生明顯影響的因素，加以翻譯與改寫科學新聞中的常見錯誤類型後，設計成 8 個科學新聞中常見的統計素養題型。以下依序說明前後測共同的設計原則與對比其中的差異，介紹前後測的評分規準，以及專家審查與評分者信度。

(一) 設計原則

前測的目的為測驗非理工科系大學生的統計素養相關先備知識，因此測驗模式是請受試者在閱讀完科學新聞後，根據題目引導詢問文中作者推論的正確意思，或是思考文句是否有誤，尚未讓受試者感受到題型概念的分類，如圖 3-4-2。

The screenshot shows a web browser window with a news article titled "塑膠汙染" (Plastic Pollution) and three multiple-choice questions. The interface includes a navigation bar, a table with columns for question number, title, answer, and reason, and a footer with navigation links.

科學新聞文本 (News Text):

水中抽樣
近年來，塑膠垃圾污染問題越來越嚴重。塑膠微粒的污染已經滲入我們生活各方面，從河川、海洋到土壤、食物以及我們的飲用水管。美國一家調查機構提出我們平常飲用的自來水中，含有肉眼看不見的塑膠微粒。該研究對全球各地14個國家的自來水進行隨機調查，結果顯示平均83%的自來水樣本中含有塑膠微粒。
英國新聞頻道Sky News報導指出，全球每年大約產生3億噸的塑膠垃圾；其中只有9%被回收，12%被焚燒，另外的79%被堆埋或是流入到自然環境中，尤其是海洋。全世界每年海洋平均多出的1,270萬噸的塑膠垃圾，對海洋生物造成嚴重的威脅，例如造成海龜窒息和鯨魚中毒。
奧斯陸-巴黎公約(OSPAR)統計全球許多處海面上廢棄物的種類比例(第一圖餅圖)，以此看出塑膠物品一直以來佔全球海洋廢棄物的大宗。第二折線圖表示海面上「每100公尺有多少廢棄物」的年平均值趨勢(圖上也顯示了各年廢棄物95%信賴區間)。

圖一：海面上廢棄物種類比例餅圖

| 廢棄物種類 | 比例 |
|-------|-------|
| 塑膠 | 47.0% |
| 紙類 | 21.0% |
| 玻璃 | 12.0% |
| 金屬 | 10.0% |
| 其他 | 8.0% |

圖二：海面上廢棄物年平均值趨勢折線圖

| 年份 | 年平均值 (每100公尺) |
|------|---------------|
| 2001 | ~100 |
| 2002 | ~120 |
| 2003 | ~140 |
| 2004 | ~160 |
| 2005 | ~180 |
| 2006 | ~200 |
| 2007 | ~220 |
| 2008 | ~240 |
| 2009 | ~260 |
| 2010 | ~280 |
| 2011 | ~300 |
| 2012 | ~320 |
| 2013 | ~340 |
| 2014 | ~360 |
| 2015 | ~380 |
| 2016 | ~400 |

統計素養試題 (Statistical Literacy Questions):

問題1: 請問文中平均83%的自來水樣本中含有塑膠微粒是什麼意思?
(A) 不要喝自來水，因為每一杯水的塑膠微粒濃度為83%
(B) 喝自來水時，盡量避免喝到83%塑膠微粒的部分
(C) 每100人，約有83人會喝到塑膠微粒
(D) 每100處地點自來水樣品，約有83處樣品中含有塑膠微粒
(E) 隨機抽任一地點的自來水，塑膠微粒佔整杯水83%

問題2: 小明看到文章中的研究對全球各地14個國家的自來水進行隨機調查，質疑抽樣地點的選擇，不能代表全世界，因此認為抽樣結果有偏誤(與現實有差異)，請問小明的質疑合理嗎?為什麼?

問題3: 折線圖表示海面上「每100公尺有多少廢棄物」的年平均值趨勢(圖上也顯示了各年廢棄物95%信賴區間)，請問從折線圖是否能推論出2004~2006年，海面上廢棄物比前年的「廢棄物種類」增加?

受試者書面作答 (Written Answer):

問題1: D 因為相較於其他選項來說它看起來最正確，也最符合我心中的正確意義(將每一個樣本處所言之塑膠微粒百分比相加再除以樣本數而得之平均數為83%)

問題2: 合理 因文章中也確實沒提出這14個國家分別是哪些國家及其所在之區域，因此確實可以合理懷疑這14個國家可能都來自相近之處
因圖表已說是「每100公尺有多少廢棄物」，因此隨著數字的增加會讓人有一種「海洋廢棄物之數量」也在跟著增加之感覺。再者，從文章主題來看會覺得說此處所放之資料應該是要在論證點之棄物，故棄物之數量之增加應是廢棄物種類增加。

受試者口頭說明/主試者註記 (Oral Explanation/Note):

問題3: 有

圖 3-4-2 前測統計素養測驗

在後測的部分，因為本研究想探討經過教學，受試者是否能使用統計素養指引於科學新聞中進行批判性思考，因此後測形式如圖 3-4-3，以科學新聞中埋錯的形式，請受試者針對底線的問題點以注解的方式選擇概念代碼並作答，完整的後測科學新聞文本參見附錄二。

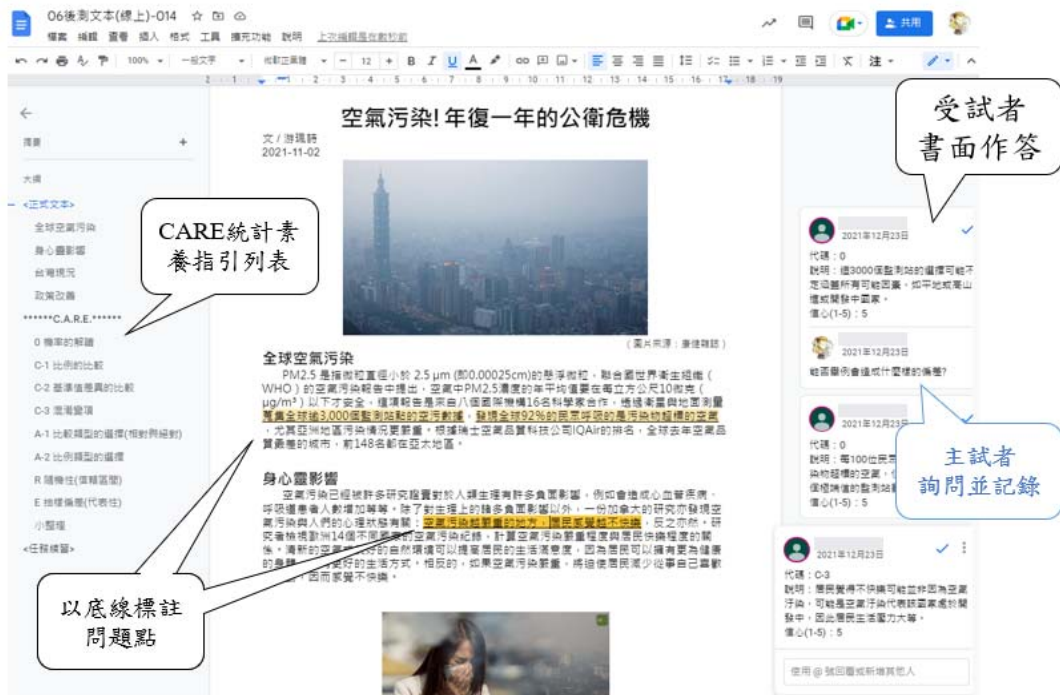


圖 3-4-3 後測統計素養測驗

統計素養測驗前後測對照如表 3-4-3，因為前後測文章所討論的議題不同，前測題目與後測問題點的語句架構不同，造成部分受試者認知資源不足因此表現分數下降，將於第四章第三節詳細討論統計素養學習成效中的前後測表現差異。

表 3-4-3 統計素養測驗題目

| 主題 | 內容 | 前測 | 後測 |
|------------------|-----------------|---|--|
| 機率的解讀 | P 機率正確意涵 | 請問文中平均 83%的自來水樣品中含有塑膠微粒是什麼意思? | 發現全球 92%的民眾呼吸的是污染物超標的空氣。 |
| | C-1 比例的比較 | 塑膠容器的回收率大約 50%，相較於紙容器的 75%高回收率仍舊不足，也就是塑膠容器回收率小於紙容器回收率；如果我們計算所有回收物中塑膠容器的占比，是否也會小於紙容器的占比? | 空氣污染中的工業污染源 18%比交通污染源 20%少，也就是工業造成的所有環境污染中空氣污染的比例，也會比交通造成的所有環境污染中空氣污染的比例少。 |
| Context 變項的比較 | C-2 基準值差異的比較 | 塑膠容器的回收率大約 50%，相較於紙容器的 75%回收率，也就是說紙容器比塑膠容器的回收率高 50%，請問是否等於塑膠容器回收率比紙容器低 50%? | 工業污染源的百分比相較於交通污染源的百分比低 1 成，且交通污染的百分比相較於其他污染的百分比也是低 1 成。 |
| | C-3 混淆變項 | 台大研究發現塑化劑濃度與頸動脈硬化機率呈現正相關，可否得到「透過生活減塑讓人攝入的塑化劑濃度降低，就可以減少心血管疾病的患病機率」的結論?是否有其他與塑化劑相關的因素同時在影響這個結論? | 一份加拿大的研究亦發現空氣污染與人們的心理狀態有關：空氣污染越嚴重的地方，居民感覺越不快樂。 |

表 3-4-3 (續) 統計素養測驗題目

| 主題 | 內容 | 前測 | 後測 |
|-------------------|---------------------|---|--|
| Assembly 變項的選擇 | A-1 絕對差異與相對比例的選擇 | 請問甲、乙的說法有何差異?你覺得誰說的比較合理? | 可見新的行動方案降幅比較多,對空氣品質改善比舊的《空氣污染防治法》成效更良好。 |
| | A-2 比例類型的選擇 | 塑膠容器的回收率為 50%,是否代表「回收物中塑膠容器的比例」也是 50%? | 空污中的交通污染源只有 20%,也就是交通產生的所有環境污染中,空氣污染也是只有佔 20%。 |
| Randomness 隨機性 | R 信賴區間與顯著性 | 折線圖表示海灘上「每 100 公尺有多少廢棄物」的年平均趨勢(圖上也顯示了各年廢棄物 95%信賴區間),請問從折線圖是否能推論出 2001~2006 年,海灘廢棄物逐年的「顯著」增加嗎? | 呈現每年 PM2.5 濃度平均值(圖上也標示各年濃度平均值的 95%信賴區間),PM2.5 濃度自 2007 年起由 $34.5 \mu\text{g}/\text{m}^3$ 下降至 2017 年只有 $20.7 \mu\text{g}/\text{m}^3$,可以看出 PM2.5 濃度逐年顯著下降,代表台灣空氣品質確實有在改善。 |
| Error 偏差 | E 樣本與母體間的抽樣偏差 | 小明看到文章中的研究對全球各地 14 個國家的自來水進行抽樣調查,質疑抽樣地點的選擇,不能代表全世界,因此認為抽樣結果有偏差(與現實有差異),請問小明的質疑合理嗎?為什麼? | 這項報告是來自八個國際機構 16 名科學家合作,透過衛星與地面測量蒐集全球逾 3,000 個監測站點的空污數據。 |

(二) 評分規準

前測表現的評分項目以受試者答題表現為主，如有詞意不清或過於簡短省略的部分，主試者會以口頭詢問與受試者確認詳細的想法。針對前測題目的答案與支持答案的理由進行評分，最低分為0分至最高分3分共四個等級的評分，分析受試者對統計素養各題型的掌握程度，如表3-4-4所示，詳細的各題評分規準及受試者答題範例如附錄四。

表 3-4-4 前測評分規準

| 分數 | 評分規準 |
|----|-----------------------------|
| 3分 | 選答正確，針對問題提出精確的判斷理由。 |
| 2分 | 選答正確，但判斷的理由只有部分正確，或說明不夠充分。 |
| 1分 | 選答正確，但無提供支持選項的判斷理由，或判斷理由錯誤。 |
| 0分 | 選答錯誤，或不會回答。 |

如同前測，後測統計素養表現的評分項目以受試者答題表現為主，如有詞意不清或過於簡短省略的部分，主試者也會以口頭詢問與受試者確認詳細的想法。受試者的後測任務是針對科學文章中的問題點做批判性思考，主試者依據受試者的概念選擇與說明進行評分，最低分為0分至最高分3分共四個等級的評分，分析受試者對統計素養各題型的掌握程度，如表3-4-5所示，詳細的各題評分規準及受試者答題範例如附錄五。

表 3-4-5 後測表現評分規準

| 分數 | 評分規準 |
|----|------------------------------------|
| 3分 | 指出正確概念，並提出精確適當的判斷理由或反思舉出其他符合概念的例子。 |
| 2分 | 指出正確概念，但判斷的理由只有部分正確或說明不夠充分。 |
| 1分 | 指出正確概念，但是完全無法解釋或不知道怎麼算。 |
| 0分 | 使用錯誤的概念解讀、只依據自身想法或經驗來判斷、無法判斷。 |

(三) 專家審查與評分者信度

統計素養前測與後測題目編寫後，經過一位統計專長科教背景的大學教授、一位科學教育研究所教授，以及一位數位學習教育所畢業並於國中任教滿 8 年的數學老師進行內容審核，確認文本內的數學、統計與科學知識正確無誤後於實驗中提供受試者閱讀與作答。

受試者前後測答題皆為開放性問答題，抽取 1/3 受試者的答案，依研究者擬定的評分規準請上述專家審查中的國中數學老師幫忙評分。研究者藉由統計軟體 SPSS 對評分者與研究者的評分結果進行相關分析，前測、後測的 Spearman's rho 係數分別為 0.986、0.985，皆達顯著水準 ($p < .001$)，代表兩位評分者具有高度的評分一致性。評分不相同處，一同討論取得共識後給予評分，後續資料以取得共識分數進行分析。

四、統計素養教學

統計素養教學分為兩個部分，第一部分受試者首先觀看教學影片講解 Schield (2010) 提出的四個會對統計數值產生明顯影響的因素，以及說明在科學新聞中常見的錯誤形式，並搭配前測題目一一舉例說明；第二部分為若受試者影片觀看完畢仍有不懂的部分，由主試者針對疑問點再次說明與解釋其觀念或算法。以下依序介紹教學影片的設計原則、教學時學習表現的評分規準，以及專家審查與評分者信度。

(一) 設計原則

本研究為個別實驗，所有受試者單獨進行，為了確保每位受試者的教學內容一致，我們將教學概念與前測講解以簡報模式錄製成影片放在 Youtube 影音平台，受試者點選連結觀看，並且能夠依據自身需要暫停影片播放進行思考，或點選右側的章節片段重複複習，如圖 3-4-4。



圖 3-4-4 教學影片呈現畫面

影片內容以統計素養概念為主軸，搭配前測的統計素養題目，解說面對科學新聞裡埋藏的陷阱時，可使用何種統計素養概念進行批判性思考。教學整體按照 CARE 的統計素養概念，但是考量到「Context 變項的脈絡」（討論變項的比較）的概念與題目皆是「Assembly 變項的集合」（討論變項的選擇）的進階內容，因此先講解「Assembly 變項的集合」的概念與解題，才進入「Context 變項的脈絡」的說明。

統計素養教學影片格式如圖 3-4-5，影片中解說 8 個題型相對應的概念時，左側區塊會出現前測題目，右側區塊為講解說明，講解完畢畫面右上方會出現紅色暫停按鈕，此時受試者需暫停影片並切換回前測的統計素養題目頁面進行訂正，主試者同時依照表 3-4-7 學習表現評分規準進行評分。



圖 3-4-5 統計素養教學影片格式

教學影片的解說大致分為兩種類型：(1)直接以前測題目為例子說明其中的統計概念，如「P 機率的解讀」、「C-1 比例的比較」與「C-2 基準值差異的比較」；(2)先講解概念再搭配前測題目說明，如「C-3 混淆變項」、「A-1 比較類型的選擇」、「A-2 比例類型的選擇」、「R 隨機性(信賴區間)」與「E 抽樣偏差」，統計素養教學影片範例如表 3-4-6，詳細教學影片可參考附錄三。

表 3-4-6 教學影片範例

| 代碼 | 影片 | 教學內容 |
|-------------------------|---|---|
| CARE 主概念 |  | 介紹 Schield (2010) 說明 CARE 對統計值的影響 |
| C-2 基準值 差異的 比較 |  | [以題目為例子說明概念] 以回收率與回收物占比為例，說明當比較的對象改變時，基準值需做相對應的調整 |
| E 抽樣 偏差 |  | [概念說明後再講解題目] 說明樣本蒐集因為抽樣方式不同，造成推測母體時產生偏差，並以自來水抽樣研究為例，討論如何對樣本代表性提出質疑 |
| CARE 指引 統整 |  | [統整為 8 項指引] 影片最後將 8 項指引列點呈現，提醒受試者依此方向對科學新聞做批判性思考 |

(二) 評分規準

學習表現評分項目以受試者訂正內容與口頭說明為主，針對前測題目補充的完整度進行評分，最低分為 0 分至最高分 3 分共四個等級的評分，分析受試者對統計素養教學各概念的掌握程度，評分規準見表 3-4-7。

表 3-4-7 學習表現評分規準

| 分數 | 評分規準 |
|-----|---------------------------------------|
| 3 分 | 前測答對（前測獲得 3 分）且理解教學中的統計概念。 |
| 2 分 | 前測有誤，經過教學理解其中的統計概念並能夠自行訂正錯誤。 |
| 1 分 | 前測有誤，教學後仍無法理解，主試者針對不理解的部分講解後方能自行訂正錯誤。 |
| 0 分 | 前測有誤，教學後無法理解且主試者講解也無法完全理解。 |

(三) 專家審查與評分者信度

統計素養教學簡報製作後，經過一位科學教育研究所教授，以及一位數位學習教育所畢業，並於國中任教滿 8 年的數學老師進行內容審核，確認教學簡報中的數學、統計與科學知識說明正確無誤，錄音錄影製作成線上影片於實驗中提供受試者學習。

學習表現分數是根據受試者在教學任務的概念理解與訂正內容的完整程度做評分，抽取 1/3 受試者的實驗影片，依研究者擬定的評分規準請上述專家審查中的國中數學老師對受試者的教學任務表現進行評分。研究者藉由統計軟體 SPSS 對評分者與研究者的評分結果進行相關分析，學習表現的 Spearman's rho 係數為 0.999，達顯著水準 ($p < .001$)，代表兩位評分者具有高度的評分一致性。評分不相同處，一同討論取得共識後給予評分，後續資料以取得共識分數進行分析。

第五節 資料分析

本研究想探討非理工科系大學生在科學新聞中的統計素養能力表現，教學任務時的學習表現，以及透過 CARE 統計素養教學的指引，瞭解學生的學習成效與

進步情形。將受試者閱讀科學新聞後的作答以及教學時的學習表現，依據對題型與概念的掌握程度分成 3 分、2 分、1 分、0 分四個等級做為表現分數。

研究問題一、二與三分別探討受試者在前測、教學與後測三階段時的表現，並在子問題一討論整體受試者的表現，子問題二討論大學是否修習統計相關課程兩組受試者的表現差異，相對應的實驗變項與分析方法如下表 3-5-1 所示。

表 3-5-1 研究問題與資料分析

| 研究問題一 | | 研究問題二 | | 研究問題三 | | 子問題 |
|--------------------|------------------------------|------------------------------|--|-------|--|---|
| 任務 | 前測 | 教學 | 後測 | | | |
| 應變項 | 前測分數 | 學習表現 | (1)後測分數 (2)前、後測分數 | | | 子問題一： 受試者在 CARE 統計 素養題型的 前測表現、 學習表現、 後測表現與 前後測表現 是否有差異？ |
| 自變項 / 統計方法 | 統計素養 8 題型 / 相依樣本 ANOVA | 統計素養 8 概念 / 相依樣本 ANOVA | (1)統計素養 8 題型 / 相依樣本 ANOVA (2)統計素養各題前後 測 / 相依樣本 <i>t</i> 檢定 | | | |
| 描述性 統計與 質性資料 | 得分比例 書面與口頭說明 | 得分比例 書面與口頭說明 | (1)得分比例 (2)前後測人數分布變 化 書面與口頭說明 | | | |
| 應變項 | 各題前測分數 | 各題學習表現 | (1)各題後測分數 (2)各題前、後測分數 | | | 子問題二： 大學是否修 習統計課程 對前測表 現、學習表 現、後測表 現與前後測 表現是否有 差異？ |
| 自變項 / 統計方法 | 是否修習統計 / 獨立樣本 <i>t</i> 檢定 | 是否修習統計 / 獨立樣本 <i>t</i> 檢定 | (1)是否修習統計 / 獨立樣本 <i>t</i> 檢定 (2-1)是否修習統計、 測驗時間(前/後測) / 二因子混合設計 ANOVA (2-2)各組測驗時間 (前/後測) / 相依樣本 <i>t</i> 檢定 | | | |
| 描述性 統計與 質性資料 | 得分比例 | 得分比例 | (1)得分比例 (2)前後測差異 | | | |
| 描述性 統計與 質性資料 | | | 三階段任務的 統計分析彙整 | | | 子問題三： 前測、教學 與後測三階 段任務的變 化為何？ |

研究問題一：非理工科系大學生在科學新聞中 CARE 統計素養前測的表現

1-1. 受試者在科學新聞中 CARE 統計素養前測表現，題型間是否有差異？

以前測的答題分數作為受試者先備知識的檢驗，分析受試者閱讀科學新聞文本後的作答，給予 3~0 分四個等級的評分，以單因子相依樣本變異數分析討論受試者 8 題得分是否有顯著差異，接著搭配受試者得分比例，與受試者的書面作答以及口頭說明，討論受試者的得分分布差異與原因。

1-2. 大學是否修習統計課程的學生在科學新聞中 CARE 統計素養前測表現是否有差異？

將受試者分為大學有無修習統計相關課程的「有」與「無」兩組，以受試者的前測分數進行獨立樣本 t 檢定，分析兩組受試者在各題得分表現是否有顯著差異，並搭配得分比例，與受試者的書面作答以及口頭說明，討論兩種組別受試者的得分分布差異與原因。

研究問題二：非理工科系大學生在學習 CARE 統計素養相關概念時的學習表現

2-1. 受試者在 CARE 統計素養相關概念的學習表現，概念間是否有差異？

以教學時受試者對前測题目的訂正與說明給予 3~0 共四個等級的評分，作為受試者的學習成效，以單因子相依樣本變異數分析，討論受試者 8 題學習表現分數是否有顯著差異，並搭配每題得分比例，與受試者的書面作答以及口頭說明，討論受試者統計素養概念的學習分數分布差異與原因。

2-2. 大學是否修習統計課程對 CARE 統計素養相關概念的學習表現是否有影響？

將受試者分為大學有無修習統計相關課程的「有」與「無」兩組，以受試者的學習分數進行獨立樣本 t 檢定，分析兩組受試者在各概念得分情形是否有顯著差異，並搭配得分比例，與受試者的書面作答以及口頭說明，討論不同組別受試者的得分分布差異與原因。

研究問題三：非理工科系大學生運用 CARE 統計素養指引檢視科學新聞中統計

訊息的學習成效，以及統計素養前後測表現的差異

- 3-1. 受試者的 CARE 統計素養後測表現，題型間是否有差異以及分數是否提升？

以後測的答題分數作為受試者的學習成效，給予 3~0 分四個等級的評分，以單因子相依樣本變異數分析討論受試者 8 題得分是否有顯著差異，以及各題前後測分數的相依樣本 t 檢定分析前後測分數是否提升，並搭配得分比例、前後測人數分布變化，與受試者的書面作答以及口頭說明，討論受試者統計素養前測與後測分數分布差異與原因。

- 3-2. 大學是否修習統計課程對 CARE 統計素養後測表現以及分數提升是否有影響？

將受試者分為大學有無修習統計相關課程的「有」與「無」兩組，對受試者的後測分數與前後測分數分別做兩項分析：(1)以後測分數做獨立樣本 t 檢定，分析兩組受試者在後測得分表現是否有顯著差異，並搭配得分比例，討論不同組別受試者的後測得分分布差異與原因；(2)分析是否修習統計課程的兩組受試者在前後測得分表現的差異，分別進行(2-1)以是否修習統計與測驗時間(前/後測)為自變項的二因子混合設計 ANOVA，與(2-2)兩組受試者分別進行測驗時間(前/後測)相依樣本 t 檢定。

- 3-3. 受試者的 CARE 統計素養表現在前測、教學與後測三階段任務中的變化為何？

彙整三階段的研究結果，討論整體受試者在 CARE 統計素養各題型與概念在三階段任務的表現與變化，以及探討統計分組對三階段任務表現與變化造成的影響。

第四章 研究結果與討論

本研究想探討非理工科系大學生在科學新聞中，統計素養相關先備知識的能力表現，以及透過 CARE 統計素養的教學指引後學生的學習表現與成效。依據受試者在統計素養前測分數、CARE 統計素養教學的學習表現分數與統計素養後測分數，三階段的任務分數做統計分析。

統計素養前測與後測是根據受試者對題型的掌握程度分成 3、2、1 和 0 分四種分數：獲得 3 分的標準為選答正確且針對題目提出精確的判斷理由；得 2 分者為選答正確但判斷的理由只有部分正確；獲得 1 分的受試者雖然選答正確但無提供支持選項的判斷理由；其餘選答錯誤或不會回答評為 0 分。而教學任務的學習表現分數則是根據受試者各題概念教學時的理解程度分成 3、2、1 和 0 分四種分數：獲得 3 分的標準為原本就理解（前測獲得 3 分）且理解教學中的統計概念；得 2 分者為前測有誤但經過教學理解其中的統計概念並能夠自行訂正錯誤；獲得 1 分的受試者則是前測有誤，而且聽完教學影片後仍無法理解，需要主試者針對不理解的部分講解才能自行訂正錯誤；教學後無法理解且主試者講解也無法完全理解者則被評為 0 分。

本章依據研究問題分成三小節，分別為：第一節「學生先備知識之統計與分析」，分析受試者前測的整體表現，並探討非理工科系學生在大學是否修習統計相關課程對前測表現的影響；第二節「統計素養教學的學習表現」，分析受試者在教學時的學習狀況，並探討非理工科系學生在大學是否修習統計相關課程對學習任務的影響；第三節「統計素養學習成效」，分析統計素養後測表現與前後測表現差異，並探討受試者是否修習統計相關課程對後測與前後測差異的影響，以及彙整受試者的 CARE 統計素養表現在前測、教學與後測三階段任務中的變化，以討論學生各概念的吸收程度。

第一節 學生先備知識之統計與分析

回答研究問題一：非理工科系大學生在科學新聞中 CARE 統計素養前測的表現

本研究收集受試者閱讀科學新聞文本後，受試者對統計素養題目的作答與說明，將答案依照評分規準給予 3~0 分共四個等級的評分，作為科學新聞中的統計

素養相關先備知識的表現分數，將研究問題做兩項分析，第一項分析非理工科系大學生 34 位受試者的整體表現，第二項分析以受試者在大學期間是否修習統計相關課程為依據，分為兩組各 17 位受試者：大學有修統計相關課程者、大學無修習任何統計相關課程者，分析兩組受試者先備知識的表現是否有差異。

一、整體前測表現

回答子研究問題 1-1：受試者在科學新聞中 CARE 統計素養前測表現，題型間是否有差異？

將研究問題分成兩個部分進行探討，第一部分探討受試者在統計素養前測的各題平均得分，並以單因子相依樣本變異數分析受試者 8 題得分是否有顯著差異，分析受試者在各題型高低分表現，第二部分則是探討受試者在每題的得分比例，搭配受試者的書面作答以及口頭說明，討論受試者的得分分布情形。

(一) 前測平均得分

分析非理工科系大學生科學新聞中的統計素養前測得分情形，以單因子相依樣本 ANOVA 分析各題型分數差異，因違反 Mauchly 球形檢定，採用 Greenhouse-Geisser 修正，結果如表 4-1-1 和表 4-1-2。結果發現：不同題型在分數表現上有顯著差異， $F(4.35, 143.38) = 42.82, p < .001, \eta_p^2 = 0.57$ 。

表 4-1-1 前測各題型變異數分析表

| 變異來源 | SS | df | MS | F | p | η_p^2 |
|------|--------|--------|-------|-------|---------|------------|
| 題型 | 173.10 | 4.35 | 39.84 | 42.82 | .000*** | 0.57 |
| 誤差 | 133.40 | 143.38 | 0.93 | | | |

註：* $p < .05$ ，** $p < .01$ ，*** $p < .001$

表 4-1-2 前測各題型表現統計表

| 主題 | 題型 | <i>M</i> | <i>SD</i> | 高低分表現 | |
|-------|-----|-----------|-----------|-------|---|
| 機率 | P | 機率的解讀 | 2.53 | 0.93 | 高 |
| | C-1 | 比例的比較 | 2.15 | 1.05 | 高 |
| 變項的比較 | C-2 | 基準值差異的比較 | 0.94 | 1.13 | 低 |
| | C-3 | 混淆變項 | 1.21 | 0.69 | 低 |
| 變項的選擇 | A-1 | 比較類型的選擇 | 2.59 | 0.70 | 高 |
| | A-2 | 比例類型的選擇 | 2.71 | 0.72 | 高 |
| 隨機性 | R | 隨機性(信賴區間) | 0.76 | 0.92 | 低 |
| 偏差 | E | 抽樣偏差 | 2.82 | 0.39 | 高 |

註：高-題型屬於高分表現群，以灰底呈現；低-題型屬於低分表現群

經過 LSD 事後比較(詳見附錄六)，發現前測 8 題的平均表現大致可以分成兩群，兩群題目彼此間有顯著的差異，因此各自命名為高分表現題共 5 題與低分表現題共 3 題，高分表現群與低分表現群間有顯著差異，而高、低分表現題群內彼此無顯著差異。

平均得分由高至低詳述如下，高分表現題：「P 機率的解讀」($M = 2.53, SD = 0.93$)、「C-1 比例的比較」($M = 2.15, SD = 1.05$)、「A-1 比較類型的選擇」($M = 2.59, SD = 0.70$)、「A-2 比例類型的選擇」($M = 2.71, SD = 0.72$)與「E 抽樣偏差」($M = 2.82, SD = 0.39$)；低分表現題：「C-2 基準值差異的比較」($M = 0.94, SD = 1.13$)、「C-3 混淆變項」($M = 1.21, SD = 0.69$)與「R 隨機性(信賴區間)」($M = 0.76, SD = 0.92$)。

(二) 前測得分比例

此部分以前測各題型得分人數分布，探討受試者對題型概念的掌握程度，依據受試者統計素養前測題目的答案給予 3~0 分共四個等級的評分，如表 4-1-3 所示。

表 4-1-3 前測各題型得分人數分布 (N=34)

| 代碼 | 題型 | 3 分 | 2 分 | 1 分 | 0 分 | 高/低 |
|-----|-----------|----------|---------|----------|----------|-----|
| P | 機率的解讀 | 26 (76%) | 2 (6%) | 4 (12%) | 2 (6%) | 高 |
| C-1 | 比例的比較 | 17 (50%) | 9 (26%) | 4 (12%) | 4 (12%) | 高 |
| C-2 | 基準值差異的比較 | 6 (18%) | 2 (6%) | 10 (29%) | 16 (47%) | 低 |
| C-3 | 混淆變項 | 1 (3%) | 9 (26%) | 20 (59%) | 4 (12%) | 低 |
| A-1 | 比較類型的選擇 | 24 (71%) | 6 (18%) | 4 (12%) | 0 (0%) | 高 |
| A-2 | 比例類型的選擇 | 28 (82%) | 3 (9%) | 2 (6%) | 1 (3%) | 高 |
| R | 隨機性(信賴區間) | 1 (3%) | 8 (24%) | 7 (21%) | 18 (53%) | 低 |
| E | 抽樣偏差 | 28 (82%) | 6 (18%) | 0 (0%) | 0 (0%) | 高 |

註：高-題型屬於高分表現群，以灰底呈現；低-題型屬於低分表現群

由表 4-1-3 的得分人數分布與表 4-1-1 的單因子相依樣本變異數分析的結果，搭配受試者的書面作答以及口頭說明依序從最高分的題型「E 抽樣偏差」到最低分的題型「R 隨機性(信賴區間)」，探討學生在各題型的前測表現。

高分表現題型

題型 E 抽樣偏差

題型「E 抽樣偏差」主要目標是測驗受試者能否注意到樣本的蒐集會因為抽樣方式的不同，造成推測母體時產生偏差，並是否能對樣本代表性提出質疑，以下為前測內容：

(問題 2) 小明看到文章中的研究對全球各地 14 個國家的自來水進行抽樣調查，質疑抽樣地點的選擇，不能代表全世界，因此認為抽樣結果有偏差(與現實有差異)，請問小明的質疑合理嗎？為什麼？

題型「E 抽樣偏差」是 5 題高分表現題型裡平均值分數最高 ($M=2.82$)，有 28 位 (82%) 受試者對於文章中的抽樣研究，都能對樣本代表性提出質疑，並提出適當的判斷理由而獲得 3 分 (例如：S33)；剩下的 6 位 (18%) 雖然也有質疑 14 個國家的抽樣是否能夠代表全世界，但說明不夠充分精確，例如只提出不知道 14 個國家有誰，但無法解釋樣本選擇對母體產生哪些可能的偏差而被評為 2 分 (例如：S11)，如以下前測回答：

S33：合理、如果抽樣國家沒有客觀地分布在已開發、開發中國家，那數據顯示一定會偏頗某類型發展國家。(書面作答，3 分)

S11：合理、因為全球有很多國家 少數不能代表全體。(書面作答，2分)

題型 A-2 比例類型的選擇

題型「A-2 比例類型的選擇」主要目的是測驗受試者能否針對單一對象的占比與回收率這兩類統計資料做說明，並能夠區辨兩者的差異，以下為前測內容：

(問題6) 請問文中提到塑膠容器的回收率為 50%，是否代表「回收物中塑膠容器的比例」也是 50%？

題型「A-2 比例類型的選擇」屬於高分表現題型($M=2.71$)，有 28 位(82%)受試者能夠針對回收率與回收物占比兩者提出精確的判斷理由，例如指出兩者分母分子的意涵，或用文氏圖解釋兩者的不同而獲得 3 分；有 3 位(9%)因為只能說明回收率或回收佔比其中之一的比例，且無法區辨兩者而被評為 2 分(例如：S01)；2 位(6%)雖然覺得兩者總量不一樣，但不知道分母各自是什麼只能獲得 1 分(例如：S11)；只有 1 位(3%)認為題目的兩個句子是指同樣意義的比例而獲得 0 分，如以下前測回答：

S01：否、如果單純就塑膠回收物中的塑膠容器，應該是 100%?感覺回收率 50%是指有一半的塑膠容器無法回收回來(書面作答，2分)

S11：不、2者總量不一樣 分母不一樣 但我不確定分母是甚麼。(書面作答，1分)

題型 A-1 比較類型的選擇

題型「A-1 比較類型的選擇」主要目的是測驗受試者是否意識到兩種方案分別是絕對差異(減塑總量)與相對比例(減塑效益)的不同，並且說明總量與效益兩者的優勢，以下為前測內容：

(問題8) 某企業希望為減塑盡分心力，因此評估販賣 10,000 個產品的情況下哪種產品經過包裝改良後更能減塑，方案如下：

1. 產品 A 經過包裝改良，每公斤塑膠原料從可製造 10 個提高到製造 20 個(也就是原本需要 1000 公斤減少到只要 500 公斤)
2. 產品 B 經過包裝改良，每公斤塑膠原料從可製造 20 個提高到製造 100 個(原本需要 500 公斤減少到只要 100 公斤)

甲、乙在討論哪一種產品經過包裝改良後更能減少塑膠使用：

甲說：產品 A 比較減塑，因為產品 A 節省 500 公斤塑膠原料，而產品 B 只節省 400 公斤。

乙說：產品 B 比較減塑，因為產品 B 節省 80% 的塑膠原料，而產品 A 只節省 50% 的塑膠原料。

請問甲、乙的說法有何差異？你覺得誰說的比較合理？

題型「A-1 比較類型的選擇」屬於高分表現題型 ($M=2.59$)，有 24 位 (71%) 受試者能夠提出精確適當的判斷理由，例如提出完整算式作為依據，說明兩者在減塑的總量與效益上皆有不同的優勢獲得 3 分；有 6 位 (18%) 因為說明不夠充分，例如只覺得都有減塑，但是無法意識到總量與效益的差異，或是對總量與效益其中一種說明有誤而被評為 2 分 (例如：S17)；4 位 (12%) 只能提出某一方覺得減塑，但是說不出明確的理由只能獲得 1 分 (例如：S02)；所有受試者皆能對題目做討論因此無人獲得 0 分，如以下前測回答：

S17：乙、甲的說法有點奇怪，不知道為什麼可以直接用最終使用的塑膠原料量去比較，這種應該會是看各別一個產品所減少的塑膠用量。乙的說法我比較認同，就一個產品使用原料多寡的面向來看，B 節省資源的效率確實比 A 要好。(書面作答，2 分)

S02：乙比較合理、通常減少或增加一些原料好像都是用百分比呈現。(書面作答，1 分)

題型 P 機率的解讀

題型「P 機率的解讀」不屬於 CARE 統計素養概念但常見於科學新聞中的錯誤解讀，因此置於第一題測驗受試者是否理解機率正確意涵，以下為前測內容：

(問題 1) 請問文中平均 83% 的自來水樣品中含有塑膠微粒是什麼意思？

- (A) 不要喝自來水，因為每一杯水的塑膠微粒濃度為 83%
- (B) 喝自來水時，盡量避免喝到 83% 塑膠微粒的部分
- (C) 每 100 人，約有 83 人會喝到塑膠微粒
- (D) 每 100 處地點抽自來水樣品，約有 83 處樣品中含有塑膠微粒
- (E) 隨機抽任一地點的自來水，塑膠微粒佔整杯水 83%

題型「P 機率的解讀」屬於高分表現題型 ($M=2.53$)，有 26 位 (76%) 受試者能夠正確說出每處或每份樣本抽樣結果為是否含塑膠微粒，以及平均 83% 是每

100 處有 83 處樣本的結果獲得 3 分；有 2 位（6%）選答正確，但是因為說明不夠充分被評為 2 分（例如：S11）；4 位（12%）無法提出判斷理由只能獲得 1 分，例如相較於其他選項來說（D）看起來比較正確；而有 2 位（6%）誤解題目獲得 0 分，例如認為題目中的百分比是質量百分濃度的平均（例如：S06），如以下前測回答：

S11：D、因為 83%是自來水樣品的形容詞。（書面作答，2 分）

憑直覺，覺得應該是很多杯，有 83 杯有塑膠微粒。（口頭說明）

S06：E、從閱讀的提題目當中我覺得因為是在不同的地方抽取自來水，所以塑膠微粒

也是從自來水當中測出來，故平均 83%就是說平均一杯自來水當中有百分之

83 是塑膠微粒。（書面作答，1 分）

題型 C-1 比例的比較

題型「C-1 比例的比較」主要目的是測驗受試者能否說明占比與回收率這兩種統計資料的意義，並能夠區辨兩者的母體差異對比較統計值大小造成的影響，以下為前測內容：

（問題 5）文中提到塑膠容器的回收率大約 50%，相較於紙容器的 75%高回收率仍舊不足，也就是塑膠容器回收率小於紙容器回收率；如果我們計算所有回收物中塑膠容器的占比，是否也會小於紙容器的占比？

題型「C-1 比例的比較」屬於高分表現題型中平均較低的題型（ $M=2.15$ ），相較於其他高分表現題的分布，只有 17 位（50%）受試者能夠針對回收率與回收物占比兩者提出精確的判斷理由，例如指出兩者分母分子的意涵，或畫出文氏圖解釋兩者的母體差異對比較統計值大小造成的影響而獲得 3 分；有 9 位（26%）說明中有部分錯誤敘述，雖然能夠說明回收率與回收物占比兩者意涵，但是認為沒有實際數量無法比較而被評為 2 分（例如：S13）；4 位（12%）因為只能說明回收率或回收佔比其中一種，且無法區辨兩者被評為 1 分（例如：S11）；而有 4 位（12%）認為能夠透過塑膠容器與紙容器的回收率大小關係，直接推估到回收物占比大小關係也相同，無仔細思考兩種比例的不同處而獲得 0 分（例如：S14），如以下前測回答：

S13：不一定、因為這兩者的數據是不一樣的，紙容器與塑膠容器的生產量不一、回

收量不一，。因此我任為沒有那個意義討論（書面作答，2分）

S11：不、因為不知道回收總量，分母不一樣，但我不確定分母是甚麼（書面作答，1分）

S14：是、因為回收率代表使用塑膠容器或紙容器後回收的比率，所以在所有回收物中，塑膠容器的比率應該也會小於紙容器。（書面作答，0分）

高分題型小結

題型「P 機率的解讀」、「A-1 比較類型的選擇」、「A-2 比例類型的選擇」與「E 抽樣偏差」70%以上的受試者能掌握題目的方向提出精確且符合統計概念的判斷理由；「C-1 比例的比較」則是 50%的受試者能夠完整解讀兩個比例的差異與大小關係，有 26%的受試者因為沒有實際數量無法比較而落在 2 分。

低分表現題型

題型 C-3 混淆變項

題型「C-3 混淆變項」主要目的是測驗受試者接受研究結論之前能否反思有沒有其他影響因素也會讓罹病風險增加，以下為前測內容：

（問題 4）根據文中的資訊，以及台大研究發現塑化劑濃度與頸動脈硬化機率呈現正相關，可否得到「透過生活減塑讓人攝入的塑化劑濃度降低，就可以減少心血管疾病患病機率」的結論？是否有其他與塑化劑相關的因素同時在影響這個結論？

題型「C-3 混淆變項」屬於低分表現題型（ $M=1.21$ ），只有 1 位（3%）受試者順著題目的方向思考，反思並提出其他與塑化劑相關的因素同時在影響結果而獲得 3 分（例如：S34）；有 9 位（26%）認為應該有其他與塑化劑相關的因素，但是舉出的例子只與塑化劑或心血管疾病機率增加其中一項有關（例如：S28），判斷理由不夠充分或理由中有部分錯誤敘述而被評為 2 分；20 位（59%）受試者雖然能夠批判思考題目的結論，但是無法舉例其他與塑化劑相關因素的例子被評為 1 分（例如：S10）；而有 4 位（12%）覺得沒有其他與塑化劑相關因素同時在影響而獲得 0 分，如以下前測回答：

S34：是、據我所知，膽固醇攝取過高，作息不正常跟常吃高鈉高油的食物像是泡麵等等，都會讓【塑化劑直接影響頸動脈硬化】這個結論有可以討論的空間。這

樣統整下來，我會比較覺得提出【現代社會的人們在講求高效率、低時間消耗下這些觀念的影響下，選擇超商的塑膠包裝微波加熱食品的機率變高，進而導致高油鹽、塑化劑等因素影響人類健康】比較正確，也比較能從更大的範圍還有視角去帮助大家提防心血管疾病。(書面作答，3分)

S28：是、文中提到有毒物質可能也跟內分泌、生殖系統、罹癌有關，所以可能是因為其他因素導致此一結論。(書面作答，2分)

S10：是、一個實驗一定會有許多變因，但舉不出例子(書面作答，1分)

題型 C-2 基準值差異的比較

題型「C-2 基準值差異的比較」主要目的是測驗受試者是否知道比較的對象改變時，基準值需做相對應的調整，並計算出正確的倍率，以下為前測內容：

(問題7)文中提到塑膠容器的回收率大約50%，相較於紙容器的75%回收率，也就是說紙容器比塑膠容器的回收率高50%，請問是否等於塑膠容器回收率比紙容器低50%？

題型「C-2 基準值差異的比較」屬於低分表現題型 ($M = 0.94$)，只有6位(18%)受試者能夠針對兩者回收率提出精確的判斷理由，例如列出完整算式得出比塑膠容器高50%為比紙容器低33.3%而獲得3分；2位(6%)因為答案有部分錯誤敘述，例如順著題目計算塑膠比紙回收率低50%，則塑膠容器回收率將僅有37.5%而非原文之50%而被評為2分(例如：S04)；10位(29%)受試者雖然有意識到兩者回收率應該不是互為50%，但是無法計算或說明被評為1分(例如：S27)；而有16位(47%)此題0分，錯誤的主要原因是直觀的認為紙容器比塑膠容器的回收率高50%，也就是塑膠容器回收率比紙容器低50%。而且在0分(完全錯誤有4位)與1分(雖然選對但是無理由有2位)受試者中共有6位表示對於此題需要計算的部分表示放棄(例如：S22)，如以下前測回答：(例如：S11)

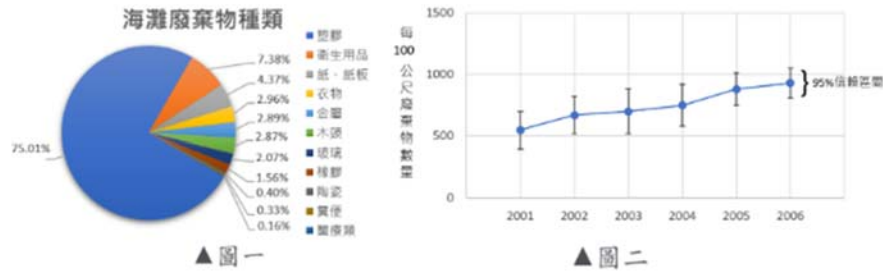
S04：否、塑膠容器比紙容器回收率低50%，則塑膠容器回收率將僅有37.5%，非原文之50%。(書面作答，3分)

S27：否、不清楚為何紙容器比塑膠容器回收率高50%，塑膠容器回收率比紙容器低25%。(書面作答，2分)

S22：不是、感覺。(書面作答，1分)

題型 R 隨機性(信賴區間)

題型「R 隨機性(信賴區間)」主要目的是測驗受試者是否知道信賴區間誤差線重疊代表的意義，因此在評估資料是否「顯著增加」時，需與信賴區間做連結，以下為前測內容：



(問題 3) 折線圖表示海灘上「每 100 公尺有多少廢棄物」的年平均趨勢(圖上也顯示了各年廢棄物 95% 信賴區間)，請問從折線圖是否能推論出 2001~2006 年，海灘廢棄物逐年的「顯著」增加嗎？

題型「R 隨機性(信賴區間)」屬於低分表現題型裡平均值分數最低($M=0.76$)，此題概念有明確的定義，受試者如果不常使用或是沒接觸過較無法提出精確的判斷。只有 1 位 (3%) 受試者能針對顯著增加提出精確的解釋，寫出信賴區間誤差線與顯著的關聯性及其意義而獲得 3 分；知道信賴區間的意思但是無法與顯著性做連結的有 8 位 (24%) 被評為 2 分 (例如：S08)。

其餘受試者沒意識到題目中的「顯著性差異」是統計學上對數據差異性的評價，而直接以圖表中的趨勢、或是誤解顯著增加是指明顯增加的意思：有 7 位 (21%) 受試者獲得 1 分，雖然認同作者所說無法推論廢棄物逐年顯著增加選答正確，但是誤解顯著的詞意是指幅度增加的大小被評為理由不正確 (例如：S04)；而有 18 位 (53%) 受試者表示不記得或不清楚信賴區間或顯著性的名詞與定義，無法對題目的顯著增加提出精確的解釋而獲得 0 分 (例如：S03)，如以下前測回答：

S08：No、因為有信賴區間，其實差距感覺沒有很大，但可能有上升趨勢。(書面作答，2分)

知道信賴區間，但是不確定顯著怎麼看 (口頭說明)

S04：否、僅些微幅度增加（書面作答，1分）

S03：無法、因為圖表上沒有給出細節的數字，因此無法得知每個區間中間上升的數值變化為多少，因此也無法判斷是否為顯著。（書面作答，0分）
區間太大（口頭說明）

低分題型小結

受試者對「C-2 基準值差異的比較」、「C-3 混淆變項」與「R 隨機性(信賴區間)」較無法提供選項的判斷理由，因為題型 C-2 的基準值概念需要敏感的數學解讀，而 C-3 的混淆變項與 R 的顯著性則是如果沒學習過相關概念無法做出正確的推論，因此有 70%受試者落在 1~0 分。

從前測整體表現來看，受試者對於高分表現題型平常多少會接觸到相關討論而有基本的概念，例如使用不同意涵的百分比做比較，也能夠提出支持選填答案的理由，得分集中在 2 分以上，而需要計算的「C-2 基準值差異的比較」或有定義概念的「C-3 混淆變項」與「R 隨機性(信賴區間)」3 題低分表現題型則集中在 0 或 1 分，無法說明或判斷理由錯誤。

二、修習統計相關課程對前測表現的影響

回答子研究問題 1-2：大學是否修習統計課程的學生在科學新聞中 CARE 統計素養前測表現是否有差異？

以非理工科系大學生在大學期間是否修習統計相關課程為依據，分為兩組：大學有修統計相關課程者、大學無修習任何統計相關課程者。本小節分成兩個部分進行探討，第一部分探討兩組受試者在統計素養前測的各題平均得分，並以獨立樣本 t 檢定分析兩組受試者在 8 題型得分是否有顯著差異，討論不同組別受試者在各題型的表現，第二部分則是探討兩組受試者在每題的得分比例，討論得分分布情形是否受到大學修習統計課程的影響。

(一) 大學有、無修習統計兩組的前測平均

獨立樣本 t 檢定分析結果如表 4-1-4 所示，結果發現：大學時期有、無修習統計課程兩組受試者在「C-2 基準值差異的比較」與「A-1 比較類型的選擇」兩類題型有顯著差異。題型「C-2 基準值差異的比較」： $t(26.94) = 2.26, p = .03, d = 0.78$ ，有修統計課程者 ($M = 1.35, SD = 1.27$) 的平均得分高於沒修統計課程者 ($M = 0.53, SD = 0.80$)。題型「A-1 比較類型的選擇」： $t(20.79) = 2.66, p = .02, d = 0.91$ ，有修統計課程者 ($M = 2.88, SD = 0.33$) 的平均得分高於沒修統計課程者 ($M = 2.29, SD = 0.85$)。其餘六題無顯著差異 ($ps > .28$)。

表 4-1-4 大學有、無修習統計兩組在前測各題型之 t 檢定

| 題型 | 平均值 (標準差) | | 自由度 | t 值 | p | 效果量 (d) |
|--------------|----------------|----------------|-------|-------|------|-------------|
| | 有 ($N = 17$) | 無 ($N = 17$) | | | | |
| P 機率的解讀 | 2.71 (0.85) | 2.35 (1.00) | 32 | 1.11 | .28 | 0.38 |
| C-1 比例的比較 | 2.12 (1.17) | 2.18 (0.95) | 32 | -0.16 | .87 | -0.06 |
| C-2 基準值差異的比較 | 1.35 (1.27) | 0.53 (0.80) | 26.94 | 2.26 | .03* | 0.78 |
| C-3 混淆變項 | 1.29 (0.69) | 1.12 (0.70) | 32 | 0.74 | .46 | 0.26 |
| A-1 比較類型的選擇 | 2.88 (0.33) | 2.29 (0.85) | 20.79 | 2.66 | .02* | 0.91 |
| A-2 比例類型的選擇 | 2.82 (0.53) | 2.59 (0.87) | 32 | 0.95 | .35 | 0.33 |
| R 隨機性(信賴區間) | 0.82 (1.01) | 0.71 (0.85) | 32 | 0.37 | .72 | 0.13 |
| E 抽樣偏差 | 2.88 (0.33) | 2.76 (0.44) | 32 | 0.88 | .38 | 0.30 |

註 1：* $p < .05$

註 2：有-大學有修統計相關課程；無-大學無修習任何統計相關課程

(二) 大學有、無修習統計兩組的得分比例

此部分將受試者分為在大學期間有修習與無修習統計相關課程，分析兩組受試者在前測各題型得分的人數分布，探討兩組受試者對題型概念掌握程度的差異，依據受試者統計素養前測題目的答案給予 3~0 分共四個等級的評分，如表 4-1-5 所示。

表 4-1-5 大學有、無修習統計兩組在前測各題型得分人數分布 (N=34)

| 代碼 | 題型 | 修習 統計 | 3 分 | 2 分 | 1 分 | 0 分 |
|------|-----------|----------|----------|---------|----------|----------|
| P | 機率的解讀 | 有 | 15 (88%) | 0 (0%) | 1 (6%) | 1 (6%) |
| | | 無 | 11 (65%) | 2 (12%) | 3 (18%) | 1 (6%) |
| C-1 | 比例的比較 | 有 | 9 (53%) | 4 (24%) | 1 (6%) | 3 (18%) |
| | | 無 | 8 (47%) | 5 (29%) | 3 (18%) | 1 (6%) |
| C-2* | 基準值差異的比較 | 有 | 5 (29%) | 2 (12%) | 4 (24%) | 6 (35%) |
| | | 無 | 1 (6%) | 0 (0%) | 6 (35%) | 10 (59%) |
| C-3 | 混淆變項 | 有 | 0 (0%) | 7 (41%) | 8 (47%) | 2 (12%) |
| | | 無 | 1 (6%) | 2 (12%) | 12 (71%) | 2 (12%) |
| A-1* | 比較類型的選擇 | 有 | 15 (88%) | 2 (12%) | 0 (0%) | 0 (0%) |
| | | 無 | 9 (53%) | 4 (24%) | 4 (24%) | 0 (0%) |
| A-2 | 比例類型的選擇 | 有 | 15 (88%) | 1 (6%) | 1 (6%) | 0 (0%) |
| | | 無 | 13 (76%) | 2 (12%) | 1 (6%) | 1 (6%) |
| R | 隨機性(信賴區間) | 有 | 1 (6%) | 4 (24%) | 3 (18%) | 9 (53%) |
| | | 無 | 0 (0%) | 4 (24%) | 4 (24%) | 9 (53%) |
| E | 抽樣偏差 | 有 | 15 (88%) | 2 (12%) | 0 (0%) | 0 (0%) |
| | | 無 | 13 (76%) | 4 (24%) | 0 (0%) | 0 (0%) |

註 1：*表示獨立樣本 t 檢定中 $p < .05$ (見表 4-1-4)

註 2：有-大學有修統計相關課程，以灰底呈現；無-大學無修習任何統計相關課程

由表 4-1-5 的得分人數分布搭配獨立樣本 t 檢定分析結果，針對「C-2 基準值差異的比較」與「A-1 比較類型的選擇」兩題有顯著差異的題目進行討論如下。

「C-2 基準值差異的比較」此題在討論「紙容器比塑膠容器的回收率高 50%，請問是否等於塑膠容器回收率比紙容器低 50%」，因為需要知道比較的對象改變時，對倍數產生的影響，大學仍持續接觸統計的有修統計課程者比沒修統計課程者對數據推理的意願較高，所以有 5 位 (29%) 是完全計算正確得 3 分，2 位 (12%) 部分錯誤得 2 分，4 位 (24%) 錯誤理由但是答案選對得 1 分，以及 6 位 (36%) 完全錯誤而 0 分；而沒修統計組只有 1 位 (6%) 針對兩者回收率列出完整的算式，得出正確的回收率獲得 3 分，其餘受試者落在 1 分的有 6 位 (35%)，與獲得 0 分的 10 位 (59%) 受試者接受題目指稱的紙容器比塑膠容器的回收率高 50%，等於塑膠容器回收率比紙容器低 50% 或放棄計算。

「A-1 比較類型的選擇」因為需理解兩個方案分別是絕對差異與相對比例的不同，或是減塑總量與減塑比例 (效益) 的差異，並且意識到總量與效益兩者有不同優勢，有修統計課程者共有 15 位 (88%) 獲得 3 分，幾乎都能完整說明兩

者在減塑的總量與效益上不同的優勢，剩下的 2 位（12%）只是總量與效益其中一個說明有誤而被評為 2 分；但是沒修統計課程者較容易將兩個方案皆以同一種方式解讀題目的數據，例如皆以減塑結果的總量做比較，不清楚怎麼看減塑比例，或是認為只要討論減少或增加都應該用百分比呈現，總量較沒有參考價值，有 9 位（53%）能夠回答正確獲得 3 分，比起有修統計組 15 位少，4 位（24%）部分錯誤評為 2 分，4 位（24%）只覺得都有減塑，但是說不出明確的理由只能獲得 1 分。

剩下的六題組別沒顯著差異的題型中，將探討得分人數分布較不同的「P 機率的解讀」與「C-3 混淆變項」兩題，其餘的四題因為得分人數兩組幾乎一致不進行討論。

「P 機率的解讀」兩組的表現，有修統計課程者（ $M=2.71, SD=0.85$ ）的平均得分略高於沒修統計課程者（ $M=2.35, SD=1.00$ ），有修統計者共有 15 位（88%）獲得 3 分，幾乎都能正確解讀機率的意涵，剩下的 1 位（6%）雖然選項正確但是無法說明落在 1 分，另 1 位（6%）表示看錯題目得 0 分；然而沒修統計者相對有修統計者較高比例因為憑感覺選答無法明確說明理由而落在 2 分以下，獲得 3 分的 11 位（65%），部分錯誤評為 2 分的 2 位（12%），憑感覺選對但是無法說明獲得 1 分的有 3 位（18%），1 位（6%）誤解百分比是質量百分濃度而 0 分。

「C-3 混淆變項」中，雖然大部分受試者都能根據題目的引導，試著批判性思考文本中的實驗設計是否有潛藏其他因素的影響，有修統計課程者（ $M=1.29, SD=0.69$ ）較高比例嘗試思考是否有其他因素也會增加罹病風險，因此主要得分落在 2 分有 7 位（41%）與 1 分有 8 位（47%），2 位（12%）認為應該沒有其他因素影響得 0 分；而沒修統計課程者（ $M=1.12, SD=0.70$ ）有 1 位（6%）成功反思並提出其他與塑化劑相關的因素得到 3 分，以及 2 位（12%）舉出的因素只能與心血管疾病或塑化劑其中之一有關聯獲得 2 分，大部分受試者（12 位，71%）因為無法舉例子做說明而只獲得 1 分，以及 2 位（12%）完全沒有想法而 0 分。

綜合以上統計結果得知，大學有修習統計課程者在科學新聞中的統計素養表現比無修習統計課程者較佳，更容易理解題目並做出批判性思考，尤其是在題型「C-2 基準值差異的比較」與「A-1 比較類型的選擇」兩組受試者有顯著差異。

第二節 統計素養教學的學習表現

回答研究問題二：非理工科系大學生在學習 CARE 統計素養相關概念時的學習表現

本節聚焦在進行統計素養概念的教學時，受試者的理解程度，將受試者經過教學任務的概念學習後，以前測作答的訂正內容與口頭說明為依據給予 3~0 分共四個等級的評分，作為教學任務的學習表現，分析受試者對統計素養教學各概念的掌握程度。因此在此節將研究問題做兩項分析，第一項分析 34 位非理工科系大學生在教學時的受試者整體學習表現，第二項以受試者在大學期間是否修習統計相關課程為自變項，學習表現為應變項進行分析，討論兩組受試者學習表現是否有差異。

一、學習表現

回答子研究問題 2-1：受試者在 CARE 統計素養相關概念的學習表現，概念間是否有差異？

將研究問題分成兩個部分進行探討，第一部分探討受試者在統計素養教學任務中各概念的學習表現平均得分，並以單因子相依樣本變異數分析討論受試者 8 題得分是否有顯著差異，分析受試者在各概念的高低分表現，第二部分則是分析受試者在每個概念的得分比例，搭配受試者的書面作答以及口頭說明，討論受試者的得分分布情形。

(一) 學習表現平均得分

分析非理工科系大學生科學新聞中的統計素養教學任務學習表現得分，以單因子相依樣本 ANOVA 分析各概念的分數差異，因違反 Mauchly 球形檢定，採用 Greenhouse-Geisser 修正，結果如表 4-2-1 和表 4-2-1。結果發現：不同概念在學習表現的分數上有顯著差異， $F(4.47, 147.42) = 32.63, p < .001, \eta_p^2 = 0.50$ 。

表 4-2-1 各概念學習表現變異數分析表

| 變異來源 | SS | df | MS | F | p | η_p^2 |
|------|-------|--------|-------|-------|---------|------------|
| 概念 | 61.09 | 4.47 | 13.67 | 32.62 | .000*** | 0.50 |
| 誤差 | 61.79 | 147.42 | 0.42 | | | |

註：* $p < .05$ ，** $p < .01$ ，*** $p < .001$

表 4-2-2 各概念學習表現統計表

| 主題 | | 概念 | <i>M</i> | <i>SD</i> | 高低分表現 |
|-------|-----|-----------|----------|-----------|-------|
| 機率 | P | 機率的解讀 | 2.76 | 0.43 | 高 |
| | C-1 | 比例的比較 | 2.38 | 0.74 | 高 |
| 變項的比較 | C-2 | 基準值差異的比較 | 1.85 | 0.78 | 低 |
| | C-3 | 混淆變項 | 1.56 | 0.56 | 低 |
| 變項的選擇 | A-1 | 比較類型的選擇 | 2.53 | 0.79 | 高 |
| | A-2 | 比例類型的選擇 | 2.82 | 0.39 | 高 |
| 隨機性 | R | 隨機性(信賴區間) | 1.82 | 0.46 | 低 |
| 偏差 | E | 抽樣偏差 | 2.82 | 0.39 | 高 |

註：高-題型屬於高分表現群，以灰底呈現；低-題型屬於低分表現群

經過 LSD 事後比較(詳見附錄六)，發現如同前測表現的分析結果，學習表現 8 題的平均大致可以分成彼此間有顯著差異的兩群，命名為高分表現題共 5 題與低分表現題共 3 題，高、低分表現題群間有顯著差異，而兩題群內彼此無顯著差異。

平均得分由高至低詳述如下，高分表現題：「P 機率的解讀」($M=2.76, SD=0.43$)、「C-1 比例的比較」($M=2.38, SD=0.74$)、「A-1 比較類型的選擇」($M=2.53, SD=0.79$)、「A-2 比例類型的選擇」($M=2.82, SD=0.39$)與「E 抽樣偏差」($M=2.82, SD=0.39$)；低分表現題：「C-2 基準值差異的比較」($M=1.85, SD=0.78$)、「C-3 混淆變項」($M=1.56, SD=0.56$)與概念「R 隨機性(信賴區間)」($M=1.82, SD=0.46$)，與前測各題型理解分數的統計結果相符。

(二) 學習表現得分比例

此部分以各概念在教學任務的學習表現得分人數分布，探討受試者對統計素養教學各概念的掌握程度，依據受試者統計素養學習表現給予 3~0 分共四個等級的評分，如表 4-2-3 所示。

表 4-2-3 各概念學習表現得分人數分布 (N=34)

| 代碼 | 概念 | 3 分 | 2 分 | 1 分 | 0 分 | 高/低 |
|-----|-----------|----------|----------|----------|--------|-----|
| P | 機率的解讀 | 26 (76%) | 8 (24%) | 0 (0%) | 0 (0%) | 高 |
| C-1 | 比例的比較 | 17 (50%) | 14 (41%) | 2 (6%) | 1 (3%) | 高 |
| C-2 | 基準值差異的比較 | 6 (18%) | 19 (56%) | 7 (21%) | 2 (6%) | 低 |
| C-3 | 混淆變項 | 1 (3%) | 17 (50%) | 16 (47%) | 0 (0%) | 低 |
| A-1 | 比較類型的選擇 | 24 (71%) | 4 (12%) | 6 (18%) | 0 (0%) | 高 |
| A-2 | 比例類型的選擇 | 28 (82%) | 6 (18%) | 0 (0%) | 0 (0%) | 高 |
| R | 隨機性(信賴區間) | 1 (3%) | 26 (76%) | 7 (21%) | 0 (0%) | 低 |
| E | 抽樣偏差 | 28 (82%) | 6 (18%) | 0 (0%) | 0 (0%) | 高 |

註：高-題型屬於高分表現群，以灰底呈現；低-題型屬於低分表現群

由表 4-2-3 的得分人數分布搭配單因子相依樣本變異數分析的結果，搭配受試者的書面作答以及口頭說明，依序從最高分的概念「E 抽樣偏差」到最低分的概念「C-3 混淆變項」，探討學生對各概念的掌握程度。

高分表現概念

概念 E 抽樣偏差

概念「E 抽樣偏差」主要介紹統計資料中的抽樣，在使用樣本去推估母體時，可能因為抽樣方式不同所造成的偏差，讓受試者有對樣本代表性作批判性思考的方向，是 5 題高分表現類型裡平均值分數最高 ($M = 2.82$)，有 28 位 (82%) 受試者在前測原本就理解，並於訂正時再對原本的想法做補充而獲得 3 分，例如原本批判性思考 14 個國家的背景是否夠具代表性，在訂正時補充水源抽樣的方式不同所造成的樣本偏差；剩下的 6 位 (18%) 是前測答案有誤，在經過教學理解其中的統計概念後能夠自行訂正錯誤而獲得 2 分；此概念無人需要再次講解被評為 1 分或是無法完全理解而被評為 0 分。

概念 A-2 比例類型的選擇

概念「A-2 比例類型的選擇」主要討論單一對象的兩種統計資料，例如占比與回收率這兩種比例的選擇並區辨兩者的差異，與「E 抽樣偏差」平均數同分 ($M = 2.82$) 且分數分布也相同，有 28 位 (82%) 受試者前測原本就理解並且針對回收率與回收物占比說明兩種比例的不同獲得 3 分；剩下的 6 位 (18%) 在前測只能說明回收率或回收佔比其中之一且無法區辨兩者，但是經過教學說明後即可自行訂正獲得 2 分；此概念無人需要再次講解被評為 1 分或是無法完全理解而被評

為 0 分。

概念 P 機率的解讀

概念「P 機率的解讀」不屬於 CARE 統計素養概念但常見於科學新聞中的錯誤解讀，因此置於前測第一題，並於教學任務放置於影片開頭解說機率正確意涵，屬於高分表現類型 ($M = 2.76$)，有 26 位 (76%) 受試者前測原本就理解，知道樣本抽樣結果為是否含塑膠微粒獲得 3 分；剩下的 8 位 (24%) 在經過教學說明後皆理解並訂正獲得 2 分；此概念無人需要再次講解被評為 1 分或是無法完全理解而被評為 0 分。

概念 A-1 比較類型的選擇

概念「A-1 比較類型的選擇」主要討論針對同一份資料，使用絕對差異與相對比例產生不同的評估結果，假設甲、乙因為兩種結果有不同優勢而做出的選擇，例如前測以減少的塑膠材料總量(絕對差異)與減少的塑膠材料比例(相對比例)為例，此概念屬於高分表現類型 ($M = 2.53$)，有 24 位 (71%) 受試者前測原本就理解且能將原本的說明以教學概念做補充，例如以絕對差異與相對比例的角度，說明減塑的總量與效益兩者不同的優勢獲得 3 分；有 4 位 (12%) 在前測雖然無法意識到總量與效益的差異，或是對總量與效益其中一種說明有誤，但是經過教學說明後能夠理解獲得 2 分；6 位 (18%) 在聽完教學後仍有部分疑惑，需要主試者再次說明被評為 1 分，例如對於兩種方案是以甲、乙兩種不同面向去解讀感到困難，需要主試者將題目一句句轉成數字做解析；無受試者無法完全理解而被評為 0 分。

概念 C-1 比例的比較

概念「C-1 比例的比較」主要經由塑膠容器與紙容器兩種母體為例子，討論其各自的占比與回收率兩種統計資料的大小比較，影片中以文氏圖解釋兩者的母體大小如何影響占比與回收率的大小值，屬於高分表現類型中平均較低分題 ($M = 2.38$)，相較於其他高分表現題的分布，只有 17 位 (50%) 受試者在前測即能夠針對回收率與回收物占比兩者提出精確的說明而獲得 3 分；14 位 (41%) 雖然在前測答案有部分錯誤，例如雖然能夠說明回收率與回收物占比兩者意涵，無法假設塑膠容器與紙容器總量大小，在兩種情形的占比與回收率做比較，但是經過文氏圖的解說能夠理解而被評為 2 分；2 位 (6%) 表示難以消化，需要主試者再

次說明才能夠理解評為 1 分；而有 1 位（3%）聽完教學影片，並再針對不理解之處做說明，受試者仍然無法自行完整解說此題被評為 0 分。

高分概念小結

概念「E 抽樣偏差」、「A-2 比例類型的選擇」與「P 機率的解讀」75%以上的受試者原本就理解得 3 分（見表 4-2-X），剩下的受試者也都能在教學影片後即可理解其中的統計概念並自行訂正錯誤獲得 2 分；而「A-1 比較類型的選擇」雖然在教學後能夠理解的有 80%以上的受試者（2 分以上者），但是 18%的受試者需要逐句解析評為 1 分；「C-1 比例的比較」則是 50%的受試者原本即能夠解讀兩個比例的差異與大小關係，有 41%的受試者教學後也能夠理解，但是 6%受試者需要再次講解且有 3%受試者無法完全理解。

低分表現概念

概念 C-2 基準值差異的比較

概念「C-2 基準值差異的比較」主要解釋百分比的計算放在不同的基準值上，計算得出的值也會有所不同，例如前測题目的「紙容器比塑膠容器的回收率高 50%，是否等於塑膠容器回收率比紙容器低 50%？」，屬於低分表現類型（ $M = 1.85$ ），只有 6 位（18%）受試者能在前測即可列出完整算式得出比塑膠容器高 50%為比紙容器低 33.3%而獲得 3 分；19 位（56%）雖然前測有部分錯誤，但是經過教學影片的說明後能夠自行完整計算評為 2 分；7 位（21%）受試者需要主試者再次逐句列出相對應的算式才能夠理解而被評為 1 分；而有 2 位（6%）在二次解說後仍表示無法理解，無法分辨兩句話分別是以何者為基準值，且表示自己在後測應該也不會算被評為 0 分。

概念 R 隨機性(信賴區間)

概念「R 隨機性(信賴區間)」主要介紹科學新聞裡，如果統計圖表解讀有使用顯著性增加等對數據差異性的評價，如何從信賴區間做判斷，屬於低分表現類型（ $M = 1.82$ ），因為此題概念有明確的統計定義，只有 1 位（3%）受試者在前測即能解釋信賴區間誤差線與顯著的關聯性獲得 3 分；有 26 位（76%）經過教學說明後即能重新解讀兩者的關聯獲得 2 分；剩下的 7 位（21%）受試者也能夠經由二次說明，排除各自的問題點評為 1 分；無受試者無法完全理解而 0 分。

概念 C-3 混淆變項

概念「C-3 混淆變項」主要內容是介紹如何思考實驗研究中的混淆變項，屬於低分表現類型裡平均值分數最低 ($M = 1.56$)，只有 1 位 (3%) 受試者在前測反思並舉例與塑化劑相關的因素，並經過教學理解混淆變項的定義而獲得 3 分；17 位 (50%) 雖然前測答案有錯誤，但是教學後理解，而且能夠重新解釋並舉出符合混淆變項的因素而被評為 2 分；剩下的 16 位 (47%) 受試者雖然無法在影片說明後就能完全理解，但是經過主試者針對各自的疑問之處說明後也能夠理解混淆變項的意涵獲得 1 分；無受試者完全無法理解而 0 分。

低分概念小結

低分表現類型整體來說，較高比例落在 2 分與 1 分，學習困難的主要原因是對答錯的受試者而言，基準值、信賴區間與顯著性的定義與混淆變項都是專有名詞，需要花時間理解，尤其是「C-3 混淆變項」概念牽涉到研究設計與變項之間的關係，常見的科學新聞較少呈現完整的實驗設計，礙於篇幅的限制選擇直接呈現實驗結果，閱聽者如果沒有查詢原始文獻無法做批判性思考，以及大部分受試者表示沒聽過混淆變項，因此 47% 受試者落在 1 分，需要主試者引導才能完全理解。

總結來說，受試者對於高分表現類型前測時已有基本概念，經過教學任務將題目與背後的統計概念連結後幾乎能夠自行訂正，並透過教學增強與 CARE 統計概念的連結，得分集中在 2 分以上。而計算類型的「C-2 基準值差異的比較」或有定義概念的「R 隨機性(信賴區間)」與「C-3 混淆變項」低分表現類型，相較於高分表現類型，有較多比例受試者前測有誤但經過教學後理解落在 2 分、以及受試者需要二次講解概念而落在 1 分。

二、修習統計相關課程對學習狀況的影響

回答子研究問題 2-2：大學是否修習統計課程對 CARE 統計素養相關概念的學習表現是否有影響？

本小節一樣以非理工科系大學生在大學期間是否修習統計相關課程為依據，

分為兩組：大學有修統計相關課程者、大學無修習任何統計相關課程者。本小節分成兩個部分進行探討，第一部分探討兩組受試者在統計素養概念的學習表現平均得分，並以獨立樣本 t 檢定分析兩組受試者在 8 個統計素養概念的學習表現是否有顯著差異，討論不同組別受試者在各概念的表現，第二部分則是根據兩組受試者在各概念的得分比例，討論受試者的得分分布情形。

(一) 大學有、無修習統計兩組的學習表現平均

獨立樣本 t 檢定分析結果如表 4-2-4 所示，結果發現：大學時期有、無修習統計課程兩組受試者在「C-2 基準值差異的比較」與「A-1 比較類型的選擇」兩類題型有顯著差異。概念「C-2 基準值差異的比較」： $t(32) = 2.61, p = .01, d = 0.90$ ，有修統計課程者 ($M = 2.18, SD = 0.64$) 的平均得分高於沒修統計課程者 ($M = 1.53, SD = 0.8$)。概念「A-1 比較類型的選擇」： $t(25.81) = 2.32, p = .03, d = 0.80$ ，有修統計課程者 ($M = 2.82, SD = 0.53$) 的平均得分高於沒修統計課程者 ($M = 2.24, SD = 0.9$)。其餘六題無顯著差異 ($ps > .11$)。

表 4-2-4 大學有、無修習統計兩組各概念在學習表現之 t 檢定

| 概念 | 平均值 (標準差) | | 自由度 | t 值 | p | 效果量 (d) |
|--------------|----------------|----------------|-------|-------|------|-------------|
| | 有 ($N = 17$) | 無 ($N = 17$) | | | | |
| P 機率的解讀 | 2.88 (0.33) | 2.65 (0.49) | 28.06 | 1.63 | .11 | 0.56 |
| C-1 比例的比較 | 2.47 (0.62) | 2.29 (0.85) | 32 | 0.69 | .50 | 0.24 |
| C-2 基準值差異的比較 | 2.18 (0.64) | 1.53 (0.80) | 32 | 2.61 | .01* | 0.90 |
| C-3 混淆變項 | 1.71 (0.47) | 1.41 (0.62) | 32 | 1.56 | .13 | 0.54 |
| A-1 比較類型的選擇 | 2.82 (0.53) | 2.24 (0.90) | 25.81 | 2.32 | .03* | 0.80 |
| A-2 比例類型的選擇 | 2.88 (0.33) | 2.76 (0.44) | 32 | 0.88 | .38 | 0.30 |
| R 隨機性(信賴區間) | 1.94 (0.43) | 1.71 (0.47) | 32 | 1.53 | .14 | 0.52 |
| E 抽樣偏差 | 2.88 (0.33) | 2.76 (0.44) | 32 | 0.88 | .38 | 0.30 |

註 1：* $p < .05$

註 2：有-大學有修統計相關課程；無-大學無修習任何統計相關課程

(二) 大學有、無修習統計兩組的得分比例

此部分將受試者分為在大學期間有修習與無修習統計相關課程，分析兩組受試者各概念學習表現得分的人數分布，探討兩組受試者對統計素養概念教學的掌握程度差異，依據受試者統計素養的學習表現給予 3~0 分共四個等級的評分，如表 4-2-5 所示。

表 4-2-5 大學有、無修習統計兩組在各概念學習表現得分人數分布 ($N=34$)

| 代碼 | 題型 | 修習 統計 | 3 分 | 2 分 | 1 分 | 0 分 |
|------|-----------|----------|----------|----------|----------|---------|
| P | 機率的解讀 | 有 | 15 (88%) | 2 (12%) | 0 (0%) | 0 (0%) |
| | | 無 | 11 (65%) | 6 (35%) | 0 (0%) | 0 (0%) |
| C-1 | 比例的比較 | 有 | 9 (53%) | 7 (41%) | 1 (6%) | 0 (0%) |
| | | 無 | 8 (47%) | 7 (41%) | 1 (6%) | 1 (6%) |
| C-2* | 基準值差異的比較 | 有 | 5 (29%) | 10 (59%) | 2 (12%) | 0 (0%) |
| | | 無 | 1 (6%) | 9 (53%) | 5 (29%) | 2 (12%) |
| C-3 | 混淆變項 | 有 | 0 (0%) | 12 (71%) | 5 (29%) | 0 (0%) |
| | | 無 | 1 (6%) | 5 (29%) | 11 (65%) | 0 (0%) |
| A-1* | 比較類型的選擇 | 有 | 15 (88%) | 1 (6%) | 1 (6%) | 0 (0%) |
| | | 無 | 9 (53%) | 3 (18%) | 5 (29%) | 0 (0%) |
| A-2 | 比例類型的選擇 | 有 | 15 (88%) | 2 (12%) | 0 (0%) | 0 (0%) |
| | | 無 | 13 (76%) | 4 (24%) | 0 (0%) | 0 (0%) |
| R | 隨機性(信賴區間) | 有 | 1 (6%) | 14 (82%) | 2 (12%) | 0 (0%) |
| | | 無 | 0 (0%) | 12 (71%) | 5 (29%) | 0 (0%) |
| E | 抽樣偏差 | 有 | 15 (88%) | 2 (12%) | 0 (0%) | 0 (0%) |
| | | 無 | 13 (76%) | 4 (24%) | 0 (0%) | 0 (0%) |

註 1：*表示獨立樣本 t 檢定中 $p < .05$ (見表 4-2-4)

註 2：有-大學有修統計相關課程；無-大學無修習任何統計相關課程

由表 4-2-4 的得分人數分布搭配獨立樣本 t 檢定分析結果，針對「C-2 基準值差異的比較」與「A-1 比較類型的選擇」兩題有顯著差異的題目進行討論如下。

「C-2 基準值差異的比較」此題因為需要知道基準值（也就是比較的對象）改變時，對加成結果產生的影響，大學有修統計課程者有 5 位（29%）受試者原本就會得到 3 分，經過教學任務即可理解其中的計算概念並自行訂正錯誤的也有 10 位（59%），只有 2 位（12%）受試者需要主試者再次列出詳細的算式說明後才能理解並訂正而評為 1 分，無受試者無法完全理解而被評為 0 分；但是沒修統計課程者只有 1 位（6%）原本就會得到 3 分，9 位（53%）教學後理解得到 2 分，有 5 位（29%）經過教學後仍需要主試者再次說明獲得 1 分，2 位（12%）受試者在二次解說後仍然無法理解被評為 0 分，有修統計課程者持續接觸數據與理論的推導，而大學沒修統計組有許多受試者表示幾乎沒有需要接觸數據的課程，因此對教學後根據概念自行訂正感到困難。

「A-1 比較類型的選擇」因為需理解絕對差異與相對比例在不同的角度解讀有不同優勢，例如前測例子是減塑總量與減塑效益的差異，有修統計課程者共有

15 位 (88%) 獲得 3 分，幾乎在前測即可說明兩者的減塑總量與效益上不同的優勢，1 位 (6%) 經過教學後也能理解並使用此概念訂正前測評為 2 分，剩下的 1 位 (6%) 經過二次解說而得到 1 分，無受試者無法完全理解而被評為 0 分；但是沒修統計課程者較多位原本只用絕對差異與相對比例其中一種方式解讀兩種方案，因此教學任務時需要較長時間思考另一種方式的邏輯，有 9 位 (53%) 能夠回答正確獲得 3 分。相較有修統計組大部分受試者 (15 位，88%) 原本就理解獲得 3 分，沒修統計組只有一半受試者 (9 位，53%) 原本就理解獲得 3 分，有 3 位 (18%) 經過教學影片後理解為 2 分，而且有 5 位 (29%) 需要主試者再次針對各自的困難點做說明而被評為 1 分，無受試者無法完全理解而被評為 0 分。

剩下的六題組別沒顯著差異的題型中，將探討得分人數分布較不同的「C-3 混淆變項」與「R 隨機性(信賴區間)」兩題，其餘的四題因為得分人數兩組幾乎一致，在此不特別做討論：概念「C-3 混淆變項」大部分受試者在教學任務中的研究設計與變項間的關係都能理解，有修統計課程者 ($M=1.71, SD=0.47$) 較高比例在前測即嘗試思考其他因素的可能性，因此教學時的表現主要落在 2 分 (12 位，71%) 與 1 分 (5 位，29%)，而沒修統計課程者 ($M=1.41, SD=0.62$) 在教學任務時除了理解混淆變項的概念，還需訂正其他增加罹病風險的因素並說明，使用較多的認知資源，因此只有 5 位 (29%) 在聽完影片就了解，較多人還是需要主試者二次說明而只獲得 1 分 (11 位，65%)；「R 隨機性(信賴區間)」概念中，有修統計課程者 ($M=1.94, SD=0.43$) 在教學任務後皆表示對於信賴區間的概念在大學時期的統計課程有學到，聽完後即可理解並自行訂正因此主要分布在 2 分 (14 位，82%)，而沒修統計課程者 ($M=1.71, SD=0.47$) 因為是在高中時期接觸信賴區間與顯著性，甚至有部分受試者表示申請入學階段已錄取大學，高三最後一章節沒有學習到，大部分受試者表示重新學習而感到困難，因此主要得分落在 2 分 (12 位，71%) 與 1 分 (5 位，29%)。

綜合以上統計結果得知，大學有修習統計課程者在教學任務的表現比無修習統計課程者較佳，更容易理解統計概念並自行訂正前測的錯誤，尤其是在概念「C-2 基準值差異的比較」與「A-1 比較類型的選擇」兩組受試者有顯著差異。

第三節 統計素養學習成效

回答研究問題三：非理工科系大學生運用 CARE 統計素養指引檢視科學新聞中統計訊息的學習成效，以及統計素養前後測表現的差異

本節聚焦在受試者是否能運用 CARE 統計素養指引，針對後測科學文章中的統計素養問題點做批判性思考，依據受試者的概念選擇與問題說明程度給予 3~0 分共四個等級的分數，以此評估受試者對統計素養概念的掌握程度。因此在此節將研究問題三做三項分析：第一項分析非理工科系大學生 34 位受試者的後測整體表現與前後測表現的差異，討論受試者的學習成效；第二項分析將受試者劃分為在大學期間是否修習統計相關課程兩組，對後測表現以及分數提升的影響；第三項分析 CARE 統計素養各題型與概念在三階段任務的表現變化情形。

一、後測表現與前後測差異

回答子研究問題 3-1：受試者的 CARE 統計素養後測表現，題型間是否有差異以及分數是否提升？

將研究問題 3-1 分成三個部分進行探討，第一部分探討整體受試者在統計素養後測的各題平均得分，並以單因子相依樣本變異數分析討論受試者 8 題得分是否有顯著差異，討論受試者在各題型高低分表現，第二部分則是分析受試者在每題的得分比例，搭配第三部分以相依樣本 t 檢定，分析受試者各題前後測分數是否有顯著差異，並根據受試者的後測高低分表現、前後測題目差異、得分分布情形，搭配受試者的書面作答以及口頭說明，討論受試者的答題表現差異。

(一) 後測表現平均得分

分析非理工科系大學生科學新聞中的統計素養後測表現得分，以單因子相依樣本 ANOVA 分析各題型分數差異，因違反 Mauchly 球形檢定，採用 Greenhouse-Geisser 修正，結果如表 4-3-1。結果發現：不同題型在後測表現的分數上有顯著差異， $F(4.10, 135.43) = 9.52, p < .001, \eta_p^2 = 0.22$ 。

表 4-3-1 後測各題型表現變異數分析表

| 變異來源 | SS | df | MS | F | p | η_p^2 |
|------|--------|--------|-------|------|---------|------------|
| 題型 | 41.41 | 4.10 | 10.09 | 9.52 | .000*** | 0.22 |
| 誤差 | 143.59 | 135.43 | 1.06 | | | |

註：* $p < .05$ ，** $p < .01$ ，*** $p < .001$

表 4-3-2 後測各題型表現統計表

| 主題 | 題型 | M | SD | 高/低分表現 | |
|-------|--------------|------|------|--------|----|
| | | | | 前測 | 後測 |
| 機率 | P 機率的解讀 | 2.15 | 0.99 | 高 | 低 |
| 變項的比較 | C-1 比例的比較 | 2.03 | 1.22 | 高 | 低 |
| | C-2 基準值差異的比較 | 1.91 | 1.24 | 低 | 低 |
| | C-3 混淆變項 | 2.71 | 0.52 | 低 | 高 |
| 變項的選擇 | A-1 比較類型的選擇 | 2.59 | 0.99 | 高 | 高 |
| | A-2 比例類型的選擇 | 2.68 | 0.84 | 高 | 高 |
| 隨機性 | R 隨機性(信賴區間) | 2.94 | 0.24 | 低 | 高 |
| 偏差 | E 抽樣偏差 | 3.00 | 0.00 | 高 | 高 |

註：高-題型屬於高分表現群，以灰底呈現；低-題型屬於低分表現群

經過 LSD 事後比較(詳見附錄六)，也是分為高分表現題共 5 題與低分表現題共 3 題，兩群題目彼此間有顯著的差異，但是與前測及學習的高/低分表現題型分布略有不同，高分表現題平均值由高至低為：「C-3 混淆變項」($M=2.71, SD=0.52$)、「A-1 比較類型的選擇」($M=2.59, SD=0.99$)、「A-2 比例類型的選擇」($M=2.68, SD=0.84$)、「R 隨機性(信賴區間)」($M=2.94, SD=0.24$)與「E 抽樣偏差」($M=3.00, SD=0.00$)；低分表現題平均值由高至低為：「P 機率的解讀」($M=2.51, SD=0.99$)、「C-1 比例的比較」($M=2.03, SD=1.22$)與「C-2 基準值差異的比較」($M=1.91, SD=1.24$)。

(二) 後測表現得分比例

此部分以各題型在後測的得分人數分布，探討受試者在統計素養教學任務後對各概念的掌握程度，依據受試者統計素養後測題目的答案給予 3~0 分共四個等級的評分，如表 4-3-3 所示，將與第三部分的前後測表現差異分析結果共同討論。

表 4-3-3 各題型後測表現得分人數分布 (N=34)

| 代碼 | 題型 | 3 分 | 2 分 | 1 分 | 0 分 | 高/低 |
|-----|-----------|-----------|---------|----------|---------|-----|
| P | 機率的解讀 | 19 (56%) | 1 (3%) | 14 (41%) | 0 (0%) | 低 |
| C-1 | 比例的比較 | 18 (53%) | 6 (18%) | 3 (9%) | 7 (21%) | 低 |
| C-2 | 基準值差異的比較 | 17 (50%) | 4 (12%) | 6 (18%) | 7 (21%) | 低 |
| C-3 | 混淆變項 | 25 (74%) | 8 (24%) | 1 (3%) | 0 (0%) | 高 |
| A-1 | 比較類型的選擇 | 28 (82%) | 2 (6%) | 0 (0%) | 4 (12%) | 高 |
| A-2 | 比例類型的選擇 | 29 (85%) | 1 (3%) | 2 (6%) | 2 (6%) | 高 |
| R | 隨機性(信賴區間) | 32 (94%) | 2 (6%) | 0 (0%) | 0 (0%) | 高 |
| E | 抽樣偏差 | 34 (100%) | 0 (0%) | 0 (0%) | 0 (0%) | 高 |

註：高-題型屬於高分表現群，以灰底呈現；低-題型屬於低分表現群

(三) 前後測表現差異

以相依樣本 t 檢定分析所有受試者之前後測表現之結果如表 4-3-4，結果發現：「C-2 基準值差異的比較」、「C-3 混淆變項」、「R 隨機性(信賴區間)」與「E 抽樣偏差」共四種題型的前測和後測平均值有顯著差異。題型「C-2 基準值差異的比較」： $t(33) = -3.41, p = .002, d = 0.79$ ，後測 ($M = 1.91, SD = 1.31$) 的平均得分高於前測 ($M = 0.94, SD = 1.13$)。題型「C-3 混淆變項」： $t(33) = -11.68, p < .001, d = 2.46$ ，後測 ($M = 2.71, SD = 0.52$) 的平均得分高於前測 ($M = 1.21, SD = 0.69$)。題型「R 隨機性(信賴區間)」： $t(33) = -13.11, p < .001, d = 3.23$ ，後測 ($M = 2.94, SD = 0.24$) 的平均得分高於前測 ($M = 0.76, SD = 0.92$)。題型「E 抽樣偏差」： $t(33) = -2.66, p = .01, d = 0.65$ ，後測 ($M = 3.00, SD = 0.00$) 的平均得分高於前測 ($M = 2.82, SD = 0.39$)。

表 4-3-4 前、後測相依樣本 t 檢定結果表

| 題型 | 平均值 (標準差) | | t 值 | p | 效果量 (d) |
|--------------|----------------|----------------|--------|---------|-------------|
| | 前 ($N = 34$) | 後 ($N = 34$) | | | |
| P 機率的解讀 | 2.53 (0.93) | 2.15 (0.99) | 1.65 | .11 | -0.40 |
| C-1 比例的比較 | 2.15 (1.05) | 2.03 (1.22) | 0.44 | .67 | -0.10 |
| C-2 基準值差異的比較 | 0.94 (1.13) | 1.91 (1.31) | -3.41 | .002** | 0.79 |
| C-3 混淆變項 | 1.21 (0.69) | 2.71 (0.52) | -11.68 | .000*** | 2.46 |
| A-1 比較類型的選擇 | 2.59 (0.70) | 2.59 (0.99) | 0.00 | 1.00 | 0.00 |
| A-2 比例類型的選擇 | 2.71 (0.72) | 2.68 (0.84) | 0.18 | .86 | -0.04 |
| R 隨機性(信賴區間) | 0.76 (0.92) | 2.94 (0.24) | -13.11 | .000*** | 3.23 |
| E 抽樣偏差 | 2.82 (0.39) | 3.00 (0.00) | -2.66 | .01* | 0.65 |

註 1：* $p < .05$ ，** $p < .01$ ，*** $p < .001$

註 2：前-統計素養前測；後-統計素養後測

綜合表 4-3-1 的單因子相依樣本變異數分析後測的高低分表現題型、表 4-3-2 的後測得分人數分布與表 4-3-3 的成對樣本 t 檢定分析前後測表現差異，搭配受試者的書面作答以及口頭說明，依序從後測平均最高分的題型「E 抽樣偏差」到最低分的題型「C-2 基準值差異的比較」，探討學生在各題型的後測表現與前後測差異情形。

高分表現題型

題型 E 抽樣偏差

題型「E 抽樣偏差」主要目標是測驗受試者能否對科學新聞擷取的研究結果，批判性思考其中的樣本代表性，以下為後測內容：

(問題點 1) 這項報告是來自八個國際機構 16 名科學家合作，透過衛星與地面測量蒐集全球逾 3,000 個監測站點的空污數據，發現全球 92% 的民眾呼吸的是污染物超標的空氣，尤其亞洲地區污染情況更嚴重。

題型「E 抽樣偏差」與前測一樣屬於高分表現題型，是 5 題高分表現題型裡平均值分數最高 ($M=3.00, SD=0.00$)，顯著高於前測平均 ($M=2.82, SD=0.39$)，搭配前後測分數變化表一起討論，如表 4-3-5。

表 4-3-5 「E 抽樣偏差」前後測分數變化表

| | | 後測 | | | | 合計 |
|----|-----|-----|-----|-----|-----|----|
| | | 3 分 | 2 分 | 1 分 | 0 分 | |
| 前測 | 3 分 | 28 | 0 | 0 | 0 | 28 |
| | 2 分 | 6 | 0 | 0 | 0 | 6 |
| | 1 分 | 0 | 0 | 0 | 0 | 0 |
| | 0 分 | 0 | 0 | 0 | 0 | 0 |
| | 合計 | 34 | 0 | 0 | 0 | 34 |

全部受試者 34 位中，有 28 位 (82%) 在前測的掌握度就已經獲得 3 分，在後測能以正確概念解讀，對樣本代表性提出質疑 (例如：S10、S24)；有 6 位 (12%) 前測獲得 2 分的受試者，在經過教學介紹抽樣偏差的統計概念與各種常見的例子後，後測也成功說明問題點的抽樣代表性獲得 3 分 (例如：S22)，如以下後測回答：

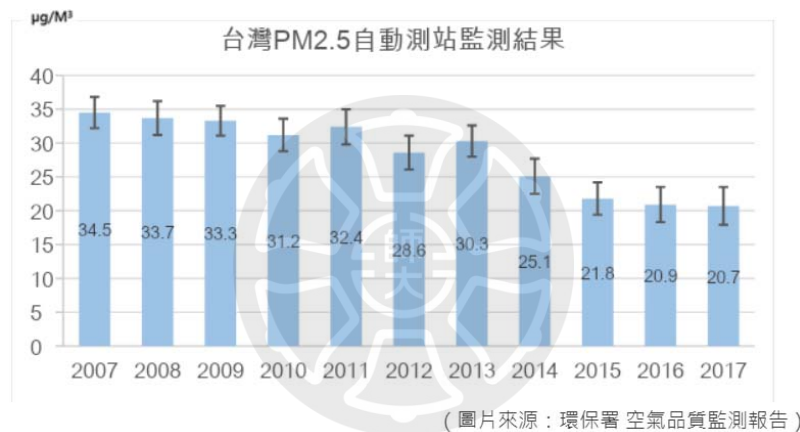
S10：E，3000 個監測站點並不一定代表全世界，若樣本大部分分佈在工業區，這樣會大大影響到實驗的準確性。屬於樣本偏差。(書面作答，3 分)

S24：E，3000 個監測站不一定具代表性，全球還有很多監測站。在開發中國家的監測站和已開發或是未開發國家的監測站所測量到的空污數據不會相同(書面作答，3 分)

S22：E，從 3000 個監測站點在哪邊?監測點是在城市還是鄉村，會影響數據，城市就會偏高，鄉村就會過低。(書面作答，3 分)

題型 R 隨機性(信賴區間)

題型「R 隨機性(信賴區間)」主要目標是測驗受試者在科學新聞評估資料是否「顯著增加」時，能否與信賴區間做連結，以下為後測內容：



(問題點 7) 圖為環保署空氣品質監測報告，呈現每年 PM2.5 濃度平均值(圖上也標示各年濃度平均值的 95%信賴區間)，PM2.5 濃度自 2007 年起由 34.5µg/m³ 下降至 2017 年只有 20.7µg/m³，可以看出 PM2.5 濃度逐年顯著下降，代表台灣空氣品質確實有在改善。

受試者在題型「R 隨機性(信賴區間)」後測表現提升至高分表現題型 ($M = 2.94, SD = 0.24$)，顯著高於前測平均 ($M = 0.76, SD = 0.92$)，經過教學任務聯結題型與其中富含的統計素養概念，使受試者有明確的分析方向與統計概念，較容易在科學新聞中發現文本的埋錯並更正，使得後測平均顯著的提升，搭配前後測分數變化表一起討論，如表 4-3-6。

表 4-3-6 「R 隨機性(信賴區間)」前後測分數變化表

| | | 後測 | | | | 合計 |
|----|-----|-----|-----|-----|-----|----|
| | | 3 分 | 2 分 | 1 分 | 0 分 | |
| 前測 | 3 分 | 1 | 0 | 0 | 0 | 1 |
| | 2 分 | 7 | 1 | 0 | 0 | 8 |
| | 1 分 | 7 | 0 | 0 | 0 | 7 |
| | 0 分 | 17 | 1 | 0 | 0 | 18 |
| | 合計 | 32 | 2 | 0 | 0 | 34 |

後測獲得 3 分的 32 位 (94%) 受試者中，除了 1 位是原本就會以外，部分錯誤得到 2 分與 1 分的受試者共 14 位經過教學理解概念並能夠正確作答，以及 17 位前測完全不會的學生在後測也能夠寫出信賴區間誤差線與顯著的關聯與其意義提升至 3 分；後測獲得 2 分的受試者中，1 位在前測知道信賴區間但是不清楚顯著的意義，在後測仍然忘記顯著的定義被評為 2 分 (例如：S08)，另一位前測完全不會的受試者，在後測知道如何解讀信賴區間與顯著下降的關聯，但是因為粗心將題目誤解成比較 2007 與 2017 兩年的顯著差異，而只獲得 2 分 (例如：S32)，如以下後測回答：

S08：R，要注意信賴區間，有可能抽出的值在平均之上或之下，造成的所謂改善可能就有差別。(書面作答，2 分)

其實我不知道顯著的意思是什麼，但是我確定他有下降的原因是因為 2007 年的信賴區間最低點跟 2017 年的最高點還是有下降 (口頭說明)

S32：R，有超出 95 的信賴區間，代表有顯著，這個說法可信！(書面作答，2 分)

我將題目誤看成比較 07、17 年顯著，忽略他寫的是逐年顯著。(口頭說明)

題型 C-3 混淆變項

題型「C-3 混淆變項」主要目標是測驗受試者面對科學新聞中引用的研究，在根據研究結論討論環境議題之前能否反思有沒有其他影響因素也會讓空氣污染指數增加，以下為後測內容：

(問題點 3) 除了對生理上的諸多負面影響以外，一份加拿大的研究亦發現空氣污染與人們的心理狀態有關：空氣污染越嚴重的地方，居民感覺越不快樂，反之亦然。

題型「C-3 混淆變項」在後測提升至高分表現題型 ($M = 2.71, SD = 0.52$)，顯著高於前測平均 ($M = 1.21, SD = 0.69$)，透過教學任務介紹實驗設計中會出現的混淆變項，讓受試者有明確的思考方向，了解如何批判性思考科學新聞中的實驗研究設計，讓受試者在後測平均有顯著的提升，搭配前後測分數變化表一起討論，如表 4-3-7。

表 4-3-7 「C-3 混淆變項」前後測分數變化表

| | | 後測 | | | | 合計 |
|----|-----|-----|-----|-----|-----|----|
| | | 3 分 | 2 分 | 1 分 | 0 分 | |
| 前測 | 3 分 | 1 | 0 | 0 | 0 | 1 |
| | 2 分 | 7 | 2 | 0 | 0 | 9 |
| | 1 分 | 15 | 5 | 0 | 0 | 20 |
| | 0 分 | 2 | 1 | 1 | 0 | 4 |
| | 合計 | 25 | 8 | 1 | 0 | 34 |

後測有 25 位受試者 (74%) 獲得 3 分，其中除了 1 位原本就會以外，前測因為舉出的塑化劑相關因素只與塑化劑或心血管疾病機率增加其中一項有關，例如認為攝入塑化劑的時間長短也會影響，但是說不出與心血管疾病機率的關聯 (S31)，屬於部分錯誤得到 2 分的 7 位，無法舉出其他與塑化劑相關因素的例子只有 1 分的 15 位受試者，以及 2 位前測完全不會的學生在後測也能夠根據統計素養教學的指引，批判性思考空氣汙染程度與居民快樂程度之間的混淆變項而提升至 3 分；後測獲得 2 分的 8 位 (24%) 受試者中，2 位在前測部分錯誤得到 2 分 (例如：S24)，在後測舉的混淆變項仍有部分錯誤被評為 2 分 (例如：S16)，5 位在前測覺得應該會有因素影響但是無法舉例只有 1 分，以及 1 位前測覺得不會有其他因素影響實驗結果得 0 分的，在後測能夠列舉出部分因素，提升至 2 分；1 位 (3%) 受試者在前測認為沒有其他與塑化劑相關的因素在影響，經過教學理解混淆變項的存在後，在後測雖然已經能察覺其中應該有混淆變項的可能，但仍然無法舉出例子而只有提升到 1 分 (例如：S04)；沒有人完全無法作答被評為 0 分。如以下後測回答：

S24：C-3，還有很多會讓人覺得不快樂，像是考試、睡眠、憂鬱指數之類的 (書面作答，2 分)

空氣的品質跟人感到快樂，可能還有很多影響快樂的原因，例如：睡不好。或

是過敏(空氣造成)所以睡不好(口頭說明)

S16：C-3，相關不等於因果，造成民眾不快，可能空氣污染較嚴重的都市，生活壓力也較大(書面作答，2分)

S04：C-3，無說明(書面作答，1分)

不快樂可能會有其他原因，有很多變項，但是一時想不到有什麼。(口頭說明)

題型 A-2 比例類型的選擇

題型「A-2 比例類型的選擇」主要目標是測驗受試者對單一對象的兩種統計資料，分辨空氣汙染中的交通汙染占比與交通產生的所有各類環境汙染中空氣汙染的比例這兩類統計資料的意義，以下為後測內容：

(問題點 5)

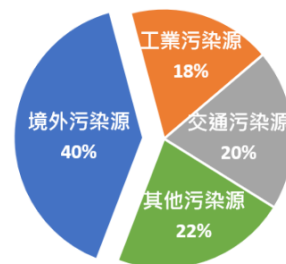
台灣現況

臺灣的境外空氣汙染主要源於中國，但境外汙染只佔臺灣空污成因的四成，另外六成汙染源都在我們的環境生成。境內汙染源包括交通汙染源約佔20%、工業汙染源18%以及其他零碎的汙染源加總約22%；交通汙染源主要為交通運輸排放之廢氣，工業汙染源包括火力發電廠、鋼鐵與石化等高汙染產業工廠排放。

由此可知三件事：

2. 空污中的交通汙染源只有 20%，也就是交通產生的所有環境汙染中，空氣汙染也是只有佔 20%；

我國細懸浮微粒(PM2.5)濃度貢獻源



(圖片來源：[行政院環保署106年度空氣汙染防制總檢討](#))

題型「A-2 比例類型的選擇」在後測雖然仍屬於高分表現題型 ($M=2.68, SD=0.84$)，但是低於前測平均 ($M=2.71, SD=0.72$)，前後測無顯著差異。教學任務中理解如何解讀單一對象的兩種比例，但是部分受試者因為題目形式的改變，將問題點埋藏在科學新聞的文章段落中，與題型「C-1 比例的比較」和「C-2 基準值差異的比較」的語句混淆，所以後測分數略微下降，搭配前後測分數變化表一起討論，如表 4-3-8。

表 4-3-8 「A-2 比例類型的選擇」前後測分數變化表

| | | 後測 | | | | 合計 |
|----|-----|-----|-----|-----|-----|----|
| | | 3 分 | 2 分 | 1 分 | 0 分 | |
| 前測 | 3 分 | 26 | 0 | 0 | 2 | 28 |
| | 2 分 | 1 | 1 | 1 | 0 | 3 |
| | 1 分 | 2 | 0 | 0 | 0 | 2 |
| | 0 分 | 0 | 0 | 1 | 0 | 1 |
| | 合計 | 29 | 1 | 2 | 2 | 34 |

後測有 29 位 (85%) 受試者獲得 3 分，其中有 26 位在前測即可對回收率與回收物占比兩者提出精確的說明得 3 分，剩下 2 分的 1 位與 1 分的 2 位在經過教學後理解在後測提升至 3 分；而獲得 2 分的只有 1 位 (3%) 受試者，在前測只能就回收率做說明，評為部分錯誤得到 2 分，在後測只能指出兩種比例分母不一樣，仍然因為部分錯誤獲得 2 分 (例如：S01)；2 位 (6%) 受試者的後測只得到 1 分，其中 1 位在前測僅能夠說明回收率得 2 分，在後測誤用「C-2 基準值差異的比較」做解讀，但是因為說明中有指出兩種比例的分母不同而給予 1 分 (例如：S04)，另 1 位受試者則是前測的比例都不能理解，經過教學後能夠指出應該是「A-2 比例類型的選擇」的問題，但是無法做說明只得 1 分；後測被評為 0 分的受試者有 2 位 (6%)，都是原本前測獲得 3 分的受試者，在後測因為題目形式的改變無法辨別此句屬於哪個概念的問題點而無法回答得到 0 分，但是主試者指引使用「A-2 比例類型的選擇」做解讀後，都能正確說明到 3 分的答題標準。如以下後測回答：

S01：A-2，分母不一樣出來的數值也會不一樣。(書面作答，2 分)

S04：C-2，因為圖表指說空汙當中的各項比例，所以畫線的地方母體變成不是空汙，變成是交通的所有汙染，奠基的東西是不一樣的 (書面作答，1 分)

題型 A-1 比較類型的選擇

題型「A-1 比較類型的選擇」主要目的是測驗受試者能否意識兩種政策分別屬於絕對差異與相對比例兩種角度，發現問題點 8 只以其中一種角度做解讀，以下為後測內容：

(問題點 8) 政府為了全面改善國內空氣汙染，針對上述 3 種可控制的境內汙染源(工業、交通與其他)提出因應對策，提出新的「空氣汙染防制行動方案」，建構更

完整的空氣品質管理制度。比較舊有的《空氣污染防治法》實行四年(2015~2019年)與新的行動方案(2020年)的成果：

全國 PM2.5 紅色警示次數由 2015 年 997 次，降至 2019 年 146 次（次數降低 851 次，降幅 85%）；但是新的行動方案推出即在 2020 年讓警示次數降至只有 15 次（次數降低 131 次，降幅 90%）；可見新的行動方案降幅比較多，對空氣品質改善比舊的《空氣污染防治法》成效更良好。

題型「A-1 比較類型的選擇」在後測也屬於高分表現題型（ $M = 2.59, SD = 0.99$ ），與前測平均（ $M = 2.59, SD = 0.70$ ）相同無顯著差異，透過教學任務說明同一份資料，如何用絕對差異與相對比例做評估，會得到不同的結果，搭配前後測分數變化表一起討論，如表 4-3-9。

表 4-3-9 「A-1 比較類型的選擇」前後測分數變化表

| | | 後測 | | | | 合計 |
|----|-----|-----|-----|-----|-----|----|
| | | 3 分 | 2 分 | 1 分 | 0 分 | |
| 前測 | 3 分 | 22 | 0 | 0 | 2 | 24 |
| | 2 分 | 4 | 1 | 0 | 1 | 6 |
| | 1 分 | 2 | 1 | 0 | 1 | 4 |
| | 0 分 | 0 | 0 | 0 | 0 | 0 |
| | 合計 | 28 | 2 | 0 | 4 | 34 |

後測有 28 位（82%）受試者獲得 3 分，其中有 22 位在前測即可辨別絕對差異與相對比例的不同得 3 分，剩下 2 分的 4 位與 1 分的 2 位在經過教學後理解在後測提升至 3 分；後測獲得 2 分的 2 位（6%）受試者中，1 位在前測部分錯誤得到 2 分，在後測有指出政策分別是次數與降幅，但是不知道怎麼說明絕對差異跟相對比例分別各有其優勢被評為 2 分，另一位前測只能說明減塑百分比，但是不會比較減塑總量與效益只獲得 1 分的受試者，在後測已經能夠說明降幅的意義，可惜仍不知道如何比較絕對差異與相對比例的不同，只有提升至 2 分；無人在後測獲得 1 分；後測被評為 0 分的受試者有 4 位（12%），都是因為無法辨別此句屬於哪個概念的問題點，其中 2 位是原本前測獲得 3 分的受試者，因此經過主試者指引使用「A-1 比較類型的選擇」做解讀，在後測經過提示後也能正確說明到 3 分的答題標準（例如：S09、S18），剩下 2 位原本在前測即部分答錯，經過提示仍然無法正確回答因此被評為 0 分（例如：S13、S22）。如以下後測回答：

S09：C-3，新方案雖使降幅變多，但也有其他因素能使降幅變多，像是現在大家對環境議題比過去幾年更加重視，也有可能影響，所以不能由此來判定新法絕對比舊方法好，還須考量其他因素。(書面作答，0分)

因為找不到，所以用混淆變項去寫，重想過有更改，但是沒覺得會是 A-1 (口頭說明)

S18：C-2，以舊行動方案為基準看新行動方案 (書面作答，0分)

S13：C-3，新的行動方案與舊的內容未必一樣，加上可能以前真的品質太爛了哈哈，所以若新的很容易達成，降幅多就會造成誤會 (書面作答，0分)

S22：C-2，比較的年份不同，所以降幅不是在同一個基準值 (書面作答，0分)

高分題型小結

受試者在「A-1 比較類型的選擇」、「A-2 比例類型的選擇」與「E 抽樣偏差」這三題的前測、教學與後測皆屬於高分表現群，受試者擁有較充足的先備知識，能夠掌握題目的概念，而「C-3 混淆變項」與「R 隨機性(信賴區間)」因為不理解專有名詞在前測與教學屬於低分表現群，經過教學理解如何使用此概念因此在後測提升為高分表現。整體來說，這五個高分表現題型較容易從語句線索與關鍵字分辨後測的問題點屬於何種統計概念，並依此分析其中的陷阱做批判性思考。

低分表現題型

題型 P 機率的解讀

題型「P 機率的解讀」在後測除了基本的機率意涵，需解讀 92% 的民眾是什麼意思，還多考驗受試者思考抽樣的對象是民眾，還是以 3000 個檢測站點有 92% 的檢測站點空氣污染超標，以下為後測內容：

(問題點 2) 這項報告是來自八個國際機構 16 名科學家合作，透過衛星與地面測量蒐集全球逾 3,000 個監測站點的空污數據，發現全球 92% 的民眾呼吸的是污染超標的空氣，尤其亞洲地區污染情況更嚴重。

受試者在題型「P 機率的解讀」後測表現下降至低分表現題型 ($M=2.15, SD=0.99$)，比前測平均略低 ($M=2.53, SD=0.93$) 但無顯著差異，透過教學任務解說機率正確意涵，更正部分受試者的錯誤觀念，但是部分受試者因為沒有細讀題

型的變化，因此後測分數些微下降，搭配前後測分數變化表一起討論，如表 4-3-10。

表 4-3-10 「P 機率的解讀」前後測分數變化表

| | | 後測 | | | | 合計 |
|----|-----|-----|-----|-----|-----|----|
| | | 3 分 | 2 分 | 1 分 | 0 分 | |
| 前測 | 3 分 | 15 | 1 | 10 | 0 | 26 |
| | 2 分 | 1 | 0 | 1 | 0 | 2 |
| | 1 分 | 1 | 0 | 3 | 0 | 4 |
| | 0 分 | 2 | 0 | 0 | 0 | 2 |
| | 合計 | 19 | 1 | 14 | 0 | 34 |

後測有 19 位受試者（56%）獲得 3 分，除了能理解 92% 是每 100 人會有 92 人呼吸超標空氣，還須能往前延伸至此研究的資料蒐集來自監測站點，數據資料可能來自 3000 個站點有 92% 的空氣污染物超標，其中有 15 位在前測對機率的意涵很清楚獲得 3 分，前測答案屬於部分錯誤得 2 分的 1 位、只有選對選項無法說明只有 1 分的 1 位，以及完全誤解百分比而 0 分的 2 位受試者，在經過教學理解概念後在後測提升至 3 分；而獲得 2 分的只有 1 位（3%）受試者，在前測能說明機率的意思，但是後測只有以人口抽樣說明 92% 獲得 2 分（例如：S10）；14 位（41%）受試者的後測只得到 1 分，其中有 10 位在前測能夠正確解讀機率得 3 分、1 位部分錯誤得 2 分以及 3 位選對但是無法說明只有 1 分的受試者，在後測都因為沒提出抽樣的概念，只提到平均 92% 是每 100 人有 92 人吸到超標空氣只有 1 分（例如：S24、S32）；沒有人完全無法作答被評為 0 分。如以下後測回答：

S10：P，指的是所有樣本數中，有還是沒有呼吸到污染物超標的空氣。假設抽 100 人，92% 指的是其中 92 人有吸到污染物超標的空氣。（書面作答，2 分）

S24：P，全球每 100 個人就有 92 個人呼吸道污染物超標的空氣，但各地人類呼吸道的空氣品質不一樣（書面作答，1 分）

S32：P，92% 指的是，每 100 個人裡面有 92 人符合後面敘述的概念。（書面作答，1 分）

題型 C-1 比例的比較

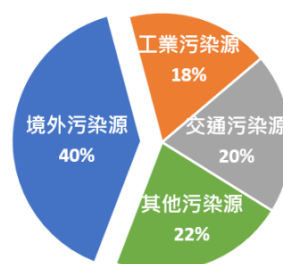
題型「C-1 比例的比較」主要目的是測驗受試者能否說明空汙中的汙染占比與所有環境汙染中空氣汙染的比例這兩類統計資料的意義，並能夠區辨兩者的母體差異對比較統計值大小造成的影響，以下為後測內容：

(問題點 6)

台灣現況

臺灣的境外空氣汙染主要源於中國，但境外汙染只佔臺灣空污成因的四成，另外六成汙染源都在我們的環境生成。境內汙染源包括交通汙染源約佔20%、工業汙染源18%以及其他零碎的汙染源加總約22%；交通汙染源主要為交通運輸排放之廢氣，工業汙染源包括火力發電廠、鋼鐵與石化等高汙染產業工廠排放。

我國細懸浮微粒(PM2.5)濃度貢獻源



(圖片來源：[行政院環保署106年度空氣汙染防制總檢討](#))

由此可知三件事：

3. 空氣汙染中的工業汙染源 18%比交通汙染源 20%少，也就是工業造成的所有環境汙染中空氣汙染的比例，也會比交通造成的所有環境汙染中空氣汙染的比例少。

題型「C-1 比例的比較」在後測表現下降至低分表現題型 ($M = 2.03$, $SD = 1.22$)，比前測平均略低 ($M = 2.15$, $SD = 1.05$) 但無顯著差異，教學任務時講解如何區分不同對象的兩種比例，並以文氏圖解釋兩者母體大小影響占比與回收率的大小關係，但是部分受試者因為題目形式的改變，問題點埋藏在科學新聞文章段落中，與題型「A-2 比例類型的選擇」和「C-2 基準值差異的比較」的語句混淆，所以後測分數略微下降，搭配前後測分數變化表一起討論，如表 4-3-11。

表 4-3-11 「C-1 比例的比較」前後測分數變化表

| | 後測 | | | | 合計 | |
|----|----|----|----|----|----|----|
| | 3分 | 2分 | 1分 | 0分 | | |
| 前測 | 3分 | 9 | 4 | 1 | 3 | 17 |
| | 2分 | 5 | 1 | 1 | 2 | 9 |
| | 1分 | 2 | 0 | 0 | 2 | 4 |
| | 0分 | 2 | 1 | 1 | 0 | 4 |
| | 合計 | 18 | 6 | 3 | 7 | 34 |

後測有 18 位 (53%) 受試者獲得 3 分，其中有 9 位在前測即可解釋兩者的母體差異對比較統計值大小造成的影響得 3 分，其餘是前測部分錯誤得 2 分的 5

位與得 1 分的 2 位，以及完全誤解百分比 0 分的 2 位，以上總共 9 位在經過教學理解概念後，在後測成功提升至 3 分；而獲得 2 分的有 6 位（18%）受試者，有 4 位前測能夠完整說明，但是後測受到題目變化的影響，在辨別後測關於空氣污染的兩種比例並比較母體大小造成的差異上感到困難，評為部分錯誤得到 2 分（例如：S15），1 位前測部分錯誤得到 2 分，後測仍然無法完整說明，以及 1 位前測完全不會，經過教學後的後測表現提升至 2 分；3 位（9%）受試者的後測只得到 1 分，1 位在前測 3 分以及 1 位前測部分錯誤得 2 分的受試者，後測都只能說明空汙來源的佔比降為 1 分，剩下 1 位則是前測完全不理解比例，但是在後測已經能夠說明其中一種比例：母體大小對空汙來源佔比的關係獲得 1 分；後測被評為 0 分的受試者有 7 位（21%），其中原本前測獲得 3 分的 3 位（例如：S09、S18）、前測部分錯誤 2 分的 2 位，以及前測只有 1 分的其中 1 位，共 6 位受試者，都是因為後測的題目形式改變，無法辨別此句屬於哪個概念的問題點，使用錯誤的概念解讀得到 0 分，但是主試者指引使用「C-1 比例的比較」做解讀後，有 4 位能正確說明到 3 分的答題標準，有 2 位仍然部分錯誤只達到 2 分的答題標準，剩下 1 位前測 1 分的受試者，在後測給予提示仍然回答無法說明被評為 0 分。如以下後測回答：

S15：C-1，此比例是空氣污染中的工業與交通佔比，並無法對應到各自佔比中空氣污染的比例（書面作答，2 分）

分母比例不一樣，沒辦法判斷，資料不夠完整，無法斷定。（口頭說明）

S09：A-1，假設空氣污染佔所有環境污染中的 20%，工業所造成的所有環境污染中空氣污染比例為 $18\%*20\%$ ，交通則是 $20\%*20\%$ ，為相對比較的選擇。（書面作答，0 分）

S18：A-2，內含工業和交通兩種比例（書面作答，0 分）

題型 C-2 基準值差異的比較

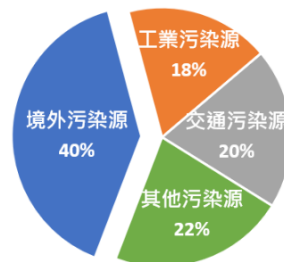
題型「C-2 基準值差異的比較」主要目的是測驗受試者是否知道比較的對象改變時，基準值需做相對應的調整，並計算出正確的倍率，以下為後測內容：

(問題點 4)

台灣現況

臺灣的境外空氣污染主要源於中國，但境外污染只佔臺灣空污成因的四成，另外六成污染源都在我們的環境生成。境內污染源包括交通污染源約佔20%、工業污染源18%以及其他零碎的污染源加總約22%；交通污染源主要為交通運輸排放之廢氣，工業污染源包括火力發電廠、鋼鐵與石化等高污染產業工廠排放。

我國細懸浮微粒(PM2.5)濃度貢獻源



(圖片來源：[行政院環保署106年度空氣污染防制總檢討](#))

由此可知三件事：

1. 工業污染源的百分比相較於交通污染源的百分比低 1 成，且交通污染的百分比相較於其他污染的百分比也是低 1 成；

受試者在題型「C-2 基準值差異的比較」後測表現雖然仍屬於低分表現題型 ($M = 1.91, SD = 1.31$)，但是顯著高於前測平均 ($M = 0.94, SD = 1.13$)，雖然有小部分受試者因為題目形式的改變，將問題點埋藏在科學新聞的文章段落中，與題型「C-1 比例的比較」和「C-2 基準值差異的比較」的語句混淆，後測分數反而下降，但是大部分的受試者能夠分辨出問題點 4 屬於「C-2 基準值差異的比較」，經過教學任務學習到如何分辨百分比的計算在不同的基準值上，此題有顯著的進步，搭配前後測分數變化表一起討論，如表 4-3-12。

表 4-3-12 「C-2 基準值差異的比較」前後測分數變化表

| | | 後測 | | | | 合計 |
|----|-----|-----|-----|-----|-----|----|
| | | 3 分 | 2 分 | 1 分 | 0 分 | |
| 前測 | 3 分 | 5 | 0 | 0 | 1 | 6 |
| | 2 分 | 1 | 0 | 0 | 1 | 2 |
| | 1 分 | 2 | 1 | 3 | 4 | 10 |
| | 0 分 | 9 | 3 | 3 | 1 | 16 |
| | 合計 | 17 | 4 | 6 | 7 | 34 |

後測有 17 位 (50%) 受試者獲得 3 分，其中有 5 位在前測即可對兩者回收率之間的倍率關係做計算得 3 分，剩下 2 分的 1 位、1 分的 2 位以及前測完全不會的 0 分受試者，經過教學後理解，在後測能發現問題點 4 的百分比低 1 成是錯置在不同的基準值上，正確計算出交通污染的百分比相較於其他污染的百分比應是低 9.1% 因此提升至 3 分；而獲得 2 分的有 4 位 (12%) 受試者，分別是 1 位前

測答案正確但無法說明的 1 分，以及 3 位完全不會的 0 分受試者提升至 2 分；6 位（18%）受試者的後測只得到 1 分，在前測分別有 3 位答案正確但無法說明的 1 分以及 3 位完全不會的 0 分，在後測皆是能夠指出是「C-2 基準值差異的比較」的問題，但是無法說明獲得 1 分（例如：S13）；其中 3 位在前測能夠僅能夠說明回收率得 2 分，在後測誤用「C-2 基準值差異的比較」做解讀，但是因為說明中有指出兩種比例的分母不同給予 1 分，另 1 位受試者則是前測的比例都不能理解，經過教學後能夠指出應該是「A-2 比例類型的選擇」的問題，但是無法做說明只得 1 分；後測被評為 0 分的受試者有 7 位（21%），皆是無法辨別此句屬於哪個概念的問題點所以錯誤解讀或是無法回答得到 0 分（例如：S08），但是主試者指引使用「C-2 基準值差異的比較」做解讀後，有 3 位能正確說明到 3 分的答題標準，有 2 位仍然部分錯誤只達到 2 分的答題標準，2 位受試者給予提示後仍然無法計算被評為 0 分。如以下後測回答：

S13：C-2，我覺得可能有點我知道是 C-2，可是不能確定是如何計算（書面作答，1 分）
之前的複習就覺得應該不會算（口頭說明）

S08：C-1，這是比較，可能交通本身的量更大，母體大，使得交通污染的百分比會較其他污染大。（書面作答，1 分）

低分題型小結

受試者的「P 機率的解讀」與「C-1 比例的比較」前測與教學表現是屬於高分表現群的，後測分別受到題目難度進階以及測驗模式的改變，在後測落入低分表現群，「C-2 基準值差異的比較」的三階段任務皆屬於低分表現，因為此概念除了理解基準值的意涵還需做計算，要在後測馬上運用較有難度，因此後測平均雖有提升仍屬於低分表現。整體來說，3 個低分表現題型因為測驗模式改變造成新聞脈絡的不同，部分受試者表示較難對比教學指引，無法順利解讀語句所埋藏的錯誤推論。

總結來看，受試者對於高分表現題型較容易應用 CARE 指引從語句與關鍵字分辨屬於何種統計概念並分析其中的陷阱，得分集中在 3 分與 2 分。因為後測文本的語句埋藏在文章中，低分表現題型較難對比教學指引，受試者無法順利解讀語句所埋藏的錯誤推論，因此有一半比例的受試者散佈在 2~0 分。受試者前後測的差異表現中，除了題型「E 抽樣偏差」在後測全部受試者皆答對，與前測有顯著差異，「C-2 基準值差異的比較」、「C-3 混淆變項」與「R 隨機性(信賴區間)」三題皆屬於前測低分表現題型，受試者經過教學任務後了解其中的統計概念，更能掌握科學新聞文本的埋錯並更正，與前測相比有顯著提升，而「P 機率的解讀」、「C-1 比例的比較」、「A-1 比較類型的選擇」與「A-2 比例類型的選擇」的前測和後測平均值均無顯著差異，且後測分數平均值略為下降，推測是因為受試者對這四題前測的掌握度就很高，雖然透過教學能夠幫助受試者確立思考方向，但是後測的統計素養測驗回歸於科學新聞中的語句做批判性思考，受試者需多方思考語句中的脈絡。

二、修習統計相關課程對後測表現的影響

回答子研究問題 3-2：大學是否修習統計課程對 CARE 統計素養後測表現以及分數提升是否有影響？

以非理工科系大學生在大學期間是否修習統計相關課程為依據，分為兩組：大學有修統計相關課程者、大學無修習任何統計相關課程者。本小節分成兩個部分進行探討，第一部分探討兩組受試者在統計素養後測的各題平均得分，並以獨立樣本 t 檢定分析兩組受試者在 8 題型得分是否有顯著差異，討論不同組別受試者在各題型的表現，第二部分則是探討兩組受試者在每題的得分比例，討論受試者的得分分布情形。

(一) 大學有、無修習統計兩組的後測平均

以獨立樣本 t 檢定分析結果如表 4-3-13 所示，結果發現：兩組人在 8 種題型皆無顯著差異 ($ps > .15$)。

表 4-3-13 大學有、無修習統計兩組在後測各題型之 t 檢定

| 題型 | 平均值 (標準差) | | 自由度 | t 值 | p | 效果量 (d) |
|--------------|--------------|--------------|-------|-------|-----|----------------|
| | 有 ($N=17$) | 無 ($N=17$) | | | | |
| P 機率的解讀 | 2.29 (0.99) | 2.00 (1.00) | 31.99 | 0.86 | .40 | 0.23 |
| C-1 比例的比較 | 2.12 (1.05) | 1.94 (1.39) | 32.00 | 0.42 | .68 | 0.14 |
| C-2 基準值差異的比較 | 2.06 (1.30) | 1.65 (1.32) | 32.00 | 0.91 | .37 | 0.32 |
| C-3 混淆變項 | 2.59 (0.62) | 2.82 (0.39) | 32.00 | -1.32 | .20 | -0.46 |
| A-1 比較類型的選擇 | 2.76 (0.75) | 2.41 (1.18) | 27.22 | 1.04 | .31 | 0.36 |
| A-2 比例類型的選擇 | 2.71 (0.85) | 2.65 (0.86) | 32.00 | 0.20 | .84 | 0.07 |
| R 隨機性(信賴區間) | 2.88 (0.33) | 3.00 (0.00) | 32.00 | -1.46 | .15 | -0.50 |
| E 抽樣偏差 | 3.00 (0.00) | 3.00 (0.00) | | | | |

註 1：*表示 $p < .05$

註 2：有-大學有修統計相關課程；無-大學無修習任何統計相關課程

(二) 大學有、無修習統計兩組的得分比例

此部分將受試者分為在大學期間有修習與無修習統計相關課程，分析兩組受試者在後測各題型得分的人數分布，探討兩組受試者對題型概念掌握程度的差異，依據受試者統計素養前測題目的答案給予 3~0 分共四個等級的評分，如表 4-3-14 所示。

表 4-3-14 大學有、無修習統計兩組在後測各題型得分人數分布 ($N=34$)

| 代碼 | 題型 | 修習 統計 | 3 分 | 2 分 | 1 分 | 0 分 |
|-----|-----------|----------|-----------|---------|---------|---------|
| | | | | | | |
| P | 機率的解讀 | 有 | 11 (65%) | 0 (0%) | 6 (35%) | 0 (0%) |
| | | 無 | 8 (47%) | 1 (6%) | 8 (47%) | 0 (0%) |
| C-1 | 比例的比較 | 有 | 8 (47%) | 5 (29%) | 2 (12%) | 2 (12%) |
| | | 無 | 10 (59%) | 1 (6%) | 1 (6%) | 5 (29%) |
| C-2 | 基準值差異的比較 | 有 | 10 (59%) | 2 (12%) | 2 (12%) | 3 (18%) |
| | | 無 | 7 (41%) | 2 (12%) | 4 (24%) | 4 (24%) |
| C-3 | 混淆變項 | 有 | 11 (65%) | 5 (29%) | 1 (6%) | 0 (0%) |
| | | 無 | 14 (82%) | 3 (18%) | 0 (0%) | 0 (0%) |
| A-1 | 比較類型的選擇 | 有 | 15 (88%) | 1 (6%) | 0 (0%) | 1 (6%) |
| | | 無 | 13 (76%) | 1 (6%) | 0 (0%) | 3 (18%) |
| A-2 | 比例類型的選擇 | 有 | 15 (88%) | 0 (0%) | 1 (6%) | 1 (6%) |
| | | 無 | 14 (82%) | 1 (6%) | 1 (6%) | 1 (6%) |
| R | 隨機性(信賴區間) | 有 | 15 (88%) | 2 (12%) | 0 (0%) | 0 (0%) |
| | | 無 | 17 (100%) | 0 (0%) | 0 (0%) | 0 (0%) |
| E | 抽樣偏差 | 有 | 17 (100%) | 0 (0%) | 0 (0%) | 0 (0%) |
| | | 無 | 17 (100%) | 0 (0%) | 0 (0%) | 0 (0%) |

註：有-大學有修統計相關課程；無-大學無修習任何統計相關課程

由表 4-3-14 的得分人數分布搭配獨立樣本 t 檢定分析結果，有修統計與沒修統計兩組的 8 題型皆無顯著差異，大部分皆為有修統計組略高於沒修統計的後測平均，因此我們針對「C-3 混淆變項」與「R 隨機性(信賴區間)」兩題沒修統計組的後測平均分數略高於有修統計組進行討論如下。

「C-3 混淆變項」有修統計課程者共有 11 位 (88%) 獲得 3 分，能夠批判性思考空氣汙染程度與居民快樂程度之間的混淆變項，有 5 位 (29%) 在後測舉的混淆變項仍有部分錯誤而被評為 2 分，1 位 (6%) 在後測雖然已經能察覺其中應該有混淆變項的可能，但仍然無法舉出例子只獲得 1 分；但是沒修統計課程者依照教學影片的指引，較多人成功對後測的問題點舉例出混淆變項，有 14 位 (82%) 能夠回答正確獲得 3 分，比起有修統計組 11 位多一些，3 位 (18%) 部分錯誤評為 2 分，而且沒有受試者落在 1 分。

「R 隨機性(信賴區間)」題型中兩組人唯一的人數分布差異就是在後測獲得 2 分的 2 位皆是大學有修統計的學生，1 位忘記顯著定義，另 1 位則是粗心將題目誤解成比較 2007 與 2017 兩年的顯著差異，而只獲得 2 分；沒修統計課程的學生皆在後測正確使用概念獲得 3 分。

綜合以上統計結果得知，經過教學任務後，前測與教學階段在「C-2 基準值差異的比較」與「A-1 比較類型的選擇」兩題皆有顯著差異的兩組受試者，在後測已無顯著差異。

(三) 大學有、無修習統計兩組的前後測表現差異

以二因子混合設計變異數分析大學有無修習統計的組別、統計素養前後測對測驗分數之效果，其中組別為獨立樣本變項，測驗時間(統計素養前/後測)為相依樣本變項，結果如表 4-3-15 所示，結果發現：8 組題型的組別與前後測題對測驗分數皆沒有顯著交互作用效果。

表 4-3-15 組別、前後測對測驗分數的推論性統計總表

| 題型 | 組別 | 前後測 | 交互作用 |
|--------------|---|--|--|
| P 機率的解讀 | $F(1,32) = 1.97,$ $p = .17,$ $\eta^2 = .06$ | $F(1,32) = 2.65,$ $p = .11,$ $\eta^2 = .08$ | $F(1,32) = 0.02,$ $p = .90,$ $\eta^2 < .001$ |
| C-1 比例的比較 | $F(1,32) = 0.04,$ $p = .84,$ $\eta^2 = .001$ | $F(1,32) = 0.19,$ $p = .67,$ $\eta^2 = .006$ | $F(1,32) = 0.19,$ $p = .67,$ $\eta^2 = .01$ |
| C-2 基準值差異的比較 | $F(1,32) = 5.11,$ $p = .03, *$ $\eta^2 = .14$ | $F(1,32) = 11.45,$ $p = .002, **$ $\eta^2 = .26$ | $F(1,32) = 0.52,$ $p = .48,$ $\eta^2 = .02$ |
| C-3 混淆變項 | $F(1,32) = 0.03,$ $p = .86,$ $\eta^2 = .001$ | $F(1,32) = 143.50,$ $p < .001, ***$ $\eta^2 = .82$ | $F(1,32) = 2.70,$ $p = .11,$ $\eta^2 = .08$ |
| A-1 比較類型的選擇 | $F(1,32) = 4.51,$ $p = .04, *$ $\eta^2 = .12$ | $F(1,32) = 0.000,$ $p = 1.00,$ $\eta^2 < .001$ | $F(1,32) = 0.42,$ $p = .52,$ $\eta^2 = .01$ |
| A-2 比例類型的選擇 | $F(1,32) = 0.46,$ $p = .50,$ $\eta^2 = .01$ | $F(1,32) = 0.03,$ $p = .86,$ $\eta^2 = .001$ | $F(1,32) = 0.30,$ $p = .59,$ $\eta^2 = .01$ |
| R 隨機性(信賴區間) | $F(1,32) = 0.000,$ $p = 1.000,$ $\eta^2 < .001$ | $F(1,32) = 169.14,$ $p < .001, ***$ $\eta^2 = .84$ | $F(1,32) = 0.49,$ $p = .49,$ $\eta^2 = .02$ |
| E 抽樣偏差 | $F(1,32) = 0.78,$ $p = .38,$ $\eta^2 = .02$ | $F(1,32) = 7.02,$ $p = .01, *$ $\eta^2 = .18$ | $F(1,32) = 0.78,$ $p = .38,$ $\eta^2 = .02$ |

註 1：* $p < .05$ ，** $p < .01$ ，*** $p < .001$

註 2：組別-大學有無修習統計的組別；前後測-統計素養前測與後測

因此接著以相依樣本 t 檢定，將受試者劃分為大學有無修習統計課程共兩組別，分別分析兩組前後測表現差異，如表 4-3-16，結果發現：對照整體受試者的相依樣本 t 檢定前後測表現，「C-2 基準值差異的比較」、「C-3 混淆變項」、「R 隨機性(信賴區間)」與「E 抽樣偏差」共四種題型的前後測平均值有顯著差異，其餘題型無顯著差異，而在受試者劃分兩組別的前後測表現上與整體表現大致上一致，惟有「C-2 基準值差異的比較」與「E 抽樣偏差」的分組檢定有所不同，兩題型有修統計組受試者的前後測表現分數無顯著差異。

表 4-3-16 大學有無修習統計的前後測成對樣本 t 檢定結果表

| 題型 | 修習 統計 | 平均值 (標準差) | | t 值 | p | 效果量 (d) |
|----------------------------------|----------|-------------|-------------|--------|---------|----------------|
| | | 前測 | 後測 | | | |
| P 機率的解讀 | 有 | 2.71 (0.85) | 2.29 (0.99) | 1.16 | .26 | -0.45 |
| | 無 | 2.35 (1.00) | 2.00 (1.00) | 1.14 | .27 | -0.35 |
| C-1 比例的比較 | 有 | 2.12 (1.17) | 2.12 (1.05) | 0.00 | 1.00 | 0.00 |
| | 無 | 2.18 (0.95) | 1.94 (1.39) | 0.64 | .53 | -0.20 |
| C-2 ⁺ 基準值差異 的比較 | 有 | 1.35 (1.27) | 2.06 (1.30) | -2.07 | .06 | 0.61 |
| | 無 | 0.53 (0.80) | 1.65 (1.32) | -2.68 | .02* | 1.11 |
| C-3 ⁺ 混淆變項 | 有 | 1.29 (0.69) | 2.59 (0.62) | -6.91 | .000*** | 1.98 |
| | 無 | 1.12 (0.70) | 2.82 (0.39) | -10.25 | .000*** | 3.02 |
| A-1 比較類型 的選擇 | 有 | 2.88 (0.33) | 2.76 (0.75) | 0.62 | .54 | -0.20 |
| | 無 | 2.29 (0.85) | 2.41 (1.18) | -0.38 | .71 | 0.12 |
| A-2 比例類型 的選擇 | 有 | 2.82 (0.53) | 2.71 (0.85) | 0.52 | .61 | -0.17 |
| | 無 | 2.59 (0.87) | 2.65 (0.86) | -0.25 | .81 | 0.07 |
| R ⁺ 隨機性 (信賴區間) | 有 | 0.82 (1.01) | 2.88 (0.33) | -7.80 | .000*** | 2.73 |
| | 無 | 0.71 (0.85) | 3.00 (0.00) | -11.14 | .000*** | 3.82 |
| E ⁺ 抽樣偏差 | 有 | 2.88 (0.33) | 3.00 (0.00) | -1.46 | .16 | 0.50 |
| | 無 | 2.76 (0.44) | 3.00 (0.00) | -2.22 | .04* | 0.76 |

註 1：* $p < .05$ ，** $p < .01$ ，*** $p < .001$

註 2：前-統計素養前測；後-統計素養後測

註 3：題型中的⁺表示整體受試者的前後測有顯著差異

不論是分組或整體做前後測成對樣本 t 檢定，結果皆為有顯著差異的兩題型：題型「C-3 混淆變項」的有修統計組： $t(33) = -6.91$ ， $p < .001$ ， $d = 1.98$ ，後測 ($M = 2.59$, $SD = 0.62$) 的平均得分高於前測 ($M = 1.29$, $SD = 0.69$)；沒修統計組： $t(33) = -10.25$ ， $p < .001$ ， $d = 3.02$ ，後測 ($M = 2.82$, $SD = 0.39$) 的平均得分也高於前測 ($M = 1.12$, $SD = 0.70$)。題型「R 隨機性(信賴區間)」的有修統計組： $t(33) = -7.80$ ， $p < .001$ ， $d = 2.73$ ，後測 ($M = 2.88$, $SD = 0.33$) 的平均得分高於前測 ($M = 0.82$, $SD = 1.01$)；沒修統計組： $t(33) = -11.14$ ， $p < .001$ ， $d = 3.82$ ，後測 ($M = 3.00$, $SD = 0.00$) 的平均得分也高於前測 ($M = 0.71$, $SD = 0.85$)。

而在整體受試者的前後測有顯著差異，但是分組後有所變化的兩題型：題型「C-2 基準值差異的比較」的沒修統計組： $t(33)=-2.68$ ， $p=.02$ ， $d=1.11$ ，後測（ $M=1.65$ ， $SD=1.32$ ）的平均得分高於前測（ $M=0.53$ ， $SD=0.80$ ），而有修統計組的前後測表現並沒有顯著差異： $t(33)=-2.07$ ， $p=.06$ ， $d=0.61$ 。題型除了需要理解基準值的概念還需做運算，因為有修統計組對數據推理的意願較高，兩組別在前測的表現就有顯著差異，但是此題型較有難度，有修統計組的受試者進步幅度較小因此前後測無顯著差異；題型「E 抽樣偏差」的沒修統計組： $t(33)=-2.22$ ， $p=.04$ ， $d=0.76$ ，後測（ $M=3.00$ ， $SD=0.00$ ）的平均得分高於前測（ $M=2.76$ ， $SD=0.44$ ），而有修統計組的前後測表現並沒有顯著差異： $t(33)=-1.46$ ， $p=.16$ ， $d=0.50$ 。因為受試者在前測的普遍表現良好，有修統計組的受試者更是高達 15 位（88%）在前測即獲得 3 分，因此後測整體進步的空間不大因此無顯著差異。

三、CARE 統計素養表現在三階段任務的變化

回答子研究問題 3-3：受試者的 CARE 統計素養表現在前測、教學與後測三階段任務中的變化為何？

本節統整非理工科系大學生在科學新聞中的先備知識（前測分數）、教學任務的學習表現（理解程度）與學習成效（後測分數）表現的統計分析結果，結果如表 4-3-17。總共分成兩個部分進行探討，第一部分討論整體受試者在 CARE 統計素養各題型與概念在三階段任務的表現與變化，第二部分則是探討統計分組對三階段任務表現與變化造成的影響。

表 4-3-17 實驗結果彙整

| 主題 | 題型 | 修習統計 | 前測分數 | | | | 學習表現 | | | | 後測分數 | | | | 前後測比較 |
|-------------------|------------------|------|------|------|-----|-----|------|------|-----|-----|------|------|-----|-----|-------|
| | | | M | SE | 高/低 | 有/無 | M | SE | 高/低 | 有/無 | M | SE | 高/低 | 有/無 | |
| Probability 機率 | P 機率的解讀 | 有 | 2.71 | 0.21 | | | 2.88 | 0.08 | | | 2.29 | 0.24 | | | |
| | | 無 | 2.35 | 0.24 | 高 | | 2.65 | 0.12 | 高 | | 2.00 | 0.24 | 低 | | |
| | | 總計 | 2.53 | 0.16 | | | 2.76 | 0.07 | | | 2.15 | 0.17 | | | |
| Context 變項的比較 | C-1 比例的比較 | 有 | 2.12 | 0.28 | | | 2.47 | 0.15 | | | 2.12 | 0.26 | | | |
| | | 無 | 2.18 | 0.23 | 高 | | 2.29 | 0.21 | 高 | | 1.94 | 0.34 | 低 | | |
| | | 總計 | 2.15 | 0.18 | | | 2.38 | 0.13 | | | 2.03 | 0.21 | | | |
| Context 變項的比較 | C-2 基準值 差異的比較 | 有 | 1.35 | 0.31 | | | 2.18 | 0.15 | | | 2.06 | 0.32 | | | |
| | | 無 | 0.53 | 0.19 | 低 | V | 1.53 | 0.19 | 低 | V | 1.65 | 0.32 | 低 | | 後>前 |
| | | 總計 | 0.94 | 0.19 | | | 1.85 | 0.13 | | | 1.85 | 0.22 | | | |
| Context 變項的比較 | C-3 混淆變項 | 有 | 1.29 | 0.17 | | | 1.71 | 0.11 | | | 2.59 | 0.15 | | | |
| | | 無 | 1.12 | 0.17 | 低 | | 1.41 | 0.15 | 低 | | 2.82 | 0.10 | 高 | | 後>前 |
| | | 總計 | 1.21 | 0.12 | | | 1.56 | 0.10 | | | 2.71 | 0.09 | | | |

註 1：有-大學有修統計相關課程；無-大學無修習任何統計相關課程。如兩組有顯著差異，以 V 表示 $p < .05$ 。

註 2：高-高分表現題型/概念；低-高分表現題型/概念。

表 4-3-17 (續) 實驗結果彙整

| 主題 | 題型 | 修習統計 | 前測分數 | | | | 學習表現 | | | | 後測分數 | | | | 前後測比較 |
|-------------------|-----------------|------|------|------|-----|-----|------|------|-----|-----|------|------|-----|-----|-------|
| | | | M | SE | 高/低 | 有/無 | M | SE | 高/低 | 有/無 | M | SE | 高/低 | 有/無 | |
| Assembly 變項的選擇 | A-1 比較類 型的選擇 | 有 | 2.88 | 0.81 | | | 2.82 | 0.13 | | | 2.76 | 0.18 | | | |
| | | 無 | 2.29 | 0.21 | 高 | V | 2.24 | 0.22 | 高 | V | 2.41 | 0.29 | 高 | | |
| | | 總計 | 2.59 | 0.12 | | | 2.53 | 0.14 | | | 2.59 | 0.17 | | | |
| | A-2 比例類 型的選擇 | 有 | 2.82 | 0.13 | | | 2.88 | 0.08 | | | 2.71 | 0.21 | | | |
| | | 無 | 2.59 | 0.21 | 高 | | 2.76 | 0.11 | 高 | | 2.65 | 0.21 | 高 | | |
| | | 總計 | 2.71 | 0.12 | | | 2.82 | 0.07 | | | 2.68 | 0.15 | | | |
| Randomness 隨機性 | R 隨機性 (信賴區間) | 有 | 0.82 | 0.25 | | | 1.94 | 0.10 | | | 2.88 | 0.08 | | | |
| | | 無 | 0.71 | 0.21 | 低 | | 1.71 | 0.11 | 低 | | 3.00 | 0.00 | 高 | | 後>前 |
| | | 總計 | 0.76 | 0.16 | | | 1.82 | 0.08 | | | 2.94 | 0.04 | | | |
| Error 偏差 | E 抽樣偏差 | 有 | 2.88 | 0.08 | | | 2.88 | 0.08 | | | 3.00 | 0.00 | | | |
| | | 無 | 2.76 | 0.11 | 高 | | 2.76 | 0.11 | 高 | | 3.00 | 0.00 | 高 | | 後>前 |
| | | 總計 | 2.82 | 0.07 | | | 2.82 | 0.07 | | | 3.00 | 0.00 | | | |

註 1：有-大學有修統計相關課程；無-大學無修習任何統計相關課程。如兩組有顯著差異，以 V 表示 $p < .05$ 。

註 2：高-高分表現題型/概念；低-高分表現題型/概念

(一) 三階段任務的題型與概念表現與變化

經過表 4-3-17 彙整，本研究的統計素養 8 類題型的平均表現大致可以分成四種變化：三個任務階段都是高分表現群的「A-1 比較類型的選擇」、「A-2 比例類型的選擇」與「E 抽樣偏差」；前測與教學屬於高分表現群，在後測落入低分表現群的「P 機率的解讀」與「C-1 比例的比較」；前測與教學屬於低分表現群，在後測提升至高分表現群的「C-3 混淆變項」與「R 隨機性(信賴區間)」；以及三個任務階段都在低分表現群的「C-2 基準值差異的比較」，如圖 4-3-1。

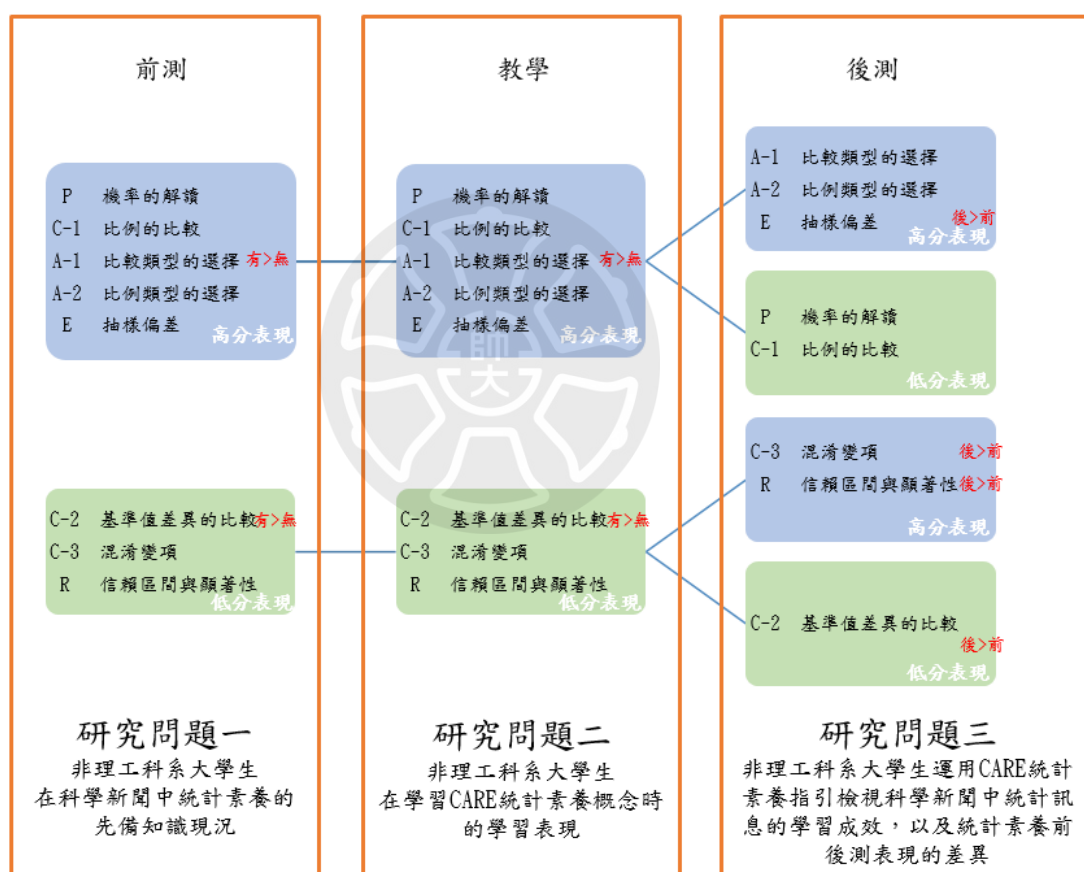


圖 4-3-1 統計素養表現在三階段任務的變化

三階段任務皆為高分表現群的題型

題型「A-1 比較類型的選擇」與「A-2 比例類型的選擇」同為變項的選擇，受試者較常見以及常使用於數學單元、各領域以及科學新聞中，而且此概念屬於高中一年級的集合與數據分析單元，此時學生尚未依照類組做分班，所有受試者

皆有完整學習到，因此前測表現分數較高，以及受試者對前測題目的「E 抽樣偏差」較能夠反思其中的樣本代表性，因此三題的前測表現幾乎(70%)都落在 3 分，剩下的受試者在教學任務也都能理解其中的統計概念並自行訂正錯誤，因此就算後測的題型進階為從整篇文章中找出 8 個問題點，後測表現仍提升至有 80% 以上的受試者發揮 3 分的能力指出正確概念，對問題點做批判性思考，尤其是「E 抽樣偏差」在後測全部受試者皆能正確回答獲得 3 分者，與前測表現相比有顯著提升。

前測與教學屬於高分表現群，在後測落入低分表現群

受試者對「P 機率的解讀」與「C-1 比例的比較」兩題型在前測能夠了解題目的問題方向，在教學也能夠理解題目與統計概念的關聯，但是受到後測題型改變的影響落入低分表現群。「P 機率的解讀」此題為了避免前後測只有數字改變，題型在後測除了基本的機率解讀，還需思考資料的來源屬於抽樣人口還是抽樣監測站點，部分受試者因為忽略此問題造成後測分數下降。而「C-1 比例的比較」則是因為前測為單一問答題型，而後測是隱藏在文章中的問題點，部分受試者較難對比教學指引，無法順利解讀語句所埋藏的錯誤推論。

前測與教學屬於低分表現群，在後測提升至高分表現群

受試者面對前測的「C-3 混淆變項」與「R 隨機性(信賴區間)」兩題型，因為屬於實驗設計與統計學上的專有名詞，較少受試者有此先備知識而屬於低分表現題型，因此教學時也需要花較多時間理解概念仍屬於低分表現群，但是在經過教學任務後了解其中的統計概念，而且題型中的專有名次概念相較其他題型更容易區分，使受試者在後測更能掌握科學新聞中的問題點做批判性思考，因此與前測相比皆有顯著提升。

三階段任務皆為低分表現群的題型

「C-2 基準值差異的比較」中的概念除了區辨比較的對象為何也就是應該以誰做為基準值，還須對基準值做計算，對主修非理工科系的大學生負擔較大，因此受試者在前測的題型表現較低，主要分布在 1 分與 0 分，經過教學說明其中的

概念以及計算方式後，50%的受試者能夠完全正確獲得3分，其中有少部分受試者對於需要計算的題目表示放棄，雖然仍屬於低分表現群的題型，但是與前測相比有顯著提升。

(二) 三階段任務統計分組造成的影響

彙整三階段任務中依據大學是否修習統計做分組，各題型的表現差異結果：題型「A-1 比較類型的選擇」與「C-2 基準值差異的比較」皆為前測與教學任務時，有修統計組的表現顯著高於沒修統計組；其餘6題無顯著差異。

前測與教學的「A-1 比較類型的選擇」是因為需要討論絕對差異與相對比例兩者的優勢，但是沒修習統計課程者較容易將兩個方案皆以同一種方式解讀題目的數據，以及「C-2 基準值差異的比較」在討論「紙容器比塑膠容器的回收率高50%，請問是否等於塑膠容器回收率比紙容器低50%」，此概念對於大學仍持續接觸統計的有修統計課程者，比沒修統計課程者對數據推理的意願較高，前測能力表現評分較高，且教學概念的吸收也較好。因此兩題型的前測與教學兩階段皆有顯著差異，有修統計課程者的平均得分顯著高於沒修統計課程者；而在後測表現差異上，透過教學概念給予受試者思考的方向，沒修統計組較多人因為統計素養指引給予思考後測問題點的方向，分數獲得提升，因此「A-1 比較類型的選擇」與「C-2 基準值差異的比較」兩個後測題型的有修統計課程者的平均得分與沒修統計課程者無顯著差異。

第五章 結論與建議

本研究想探討非理工科系大學生在科學新聞中，統計素養相關先備知識的能力表現，以及經過 CARE 統計素養教學指引後學生的學習成效。因此蒐集受試者在統計素養前測分數、CARE 統計素養教學的學習表現分數與統計素養後測分數，總共三階段的任務分數做為資料分析來源，了解受試者在三階段的任務表現，並做前後測的比較。也希望能透過本實驗，對未來有興趣做這方面研究的研究者，提供一些參考資料。

故本章第一節整理研究結果與回應研究問題，第二節根據研究設計與結果提出限制與未來研究上的建議，以及對於科學新聞中的統計素養教學提供教學建議。

第一節 研究結果摘要

本節以表 5-1-1 做研究問題與研究結果的彙整，根據研究問題逐一討論並提出結論。

研究問題一：非理工科系大學生在科學新聞中 CARE 統計素養前測的表現

研究結果顯示，統計素養前測的不同題型在分數表現上有顯著差異，分為高分表現題共 5 題與低分表現題共 3 題，代表受試者在不同題型的先備知識有顯著的差別：高分表現題型的「P 機率的解讀」、「C-1 比例的比較」、「A-1 比較類型的選擇」、「A-2 比例類型的選擇」與「E 抽樣偏差」，非理工科系大學生平常多少會接觸到相關討論而有基本的概念，而且在新聞媒體中也較常使用相關的詞彙做討論 (Raymond & Schield, 2008)，例如使用不同意涵的百分比做比較，也較能夠提出支持選填答案的理由；在低分表現題型表現中，非理工科系大學生對於需要將統計資料做交叉比較的「C-2 基準值差異的比較」感到計算困難，而在牽涉到統計學專業名詞的「R 隨機性(信賴區間)」以及研究設計的「C-3 混淆變項」，兩種用語在新聞媒體中較少出現 (Raymond & Schield, 2008)，因此受試者先備知識較不足。

在非理工科系大學生分為大學期間有修習與無修習統計相關課程兩組的結果顯示，在統計素養前測的不同題型中，「C-2 基準值差異的比較」與「A-1 比較類型的選擇」是有修組顯著高於沒修組的，可以說大學期間無修習統計相關課程受試者的「C-2 基準值差異的比較」與「A-1 比較類型的選擇」先備知識比之有修統計課程者較不足。

表 5-1-1 研究問題摘要表

| 研究問題 | | | 研究結果 | | | | | | | |
|---|---|---------------|--------|--------|--------|--------|--------|--------|--------|--------|
| | | | P | C-1 | C-2 | C-3 | A-1 | A-2 | R | E |
| 研究問題一： 非理工科系大學生在科學新聞中 CARE 統計素養前測的表現 | 1-1 受試者在科學新聞中 CARE 統計素養前測表現，題型間是否有差異？ | 高低分群 | 高 | 高 | 低 | 低 | 高 | 高 | 低 | 高 |
| | 1-2 大學是否修習統計課程的學生在科學新聞中 CARE 統計素養前測表現是否有差異？ | 修習統計差異 | | | V | | V | | | |
| 研究問題二： 非理工科系大學生在學習 CARE 統計素養相關概念時的學習表現 | 2-1 受試者在 CARE 統計素養相關概念的學習表現，概念間是否有差異？ | 高低分群 | 高 | 高 | 低 | 低 | 高 | 高 | 低 | 高 |
| | 2-2 大學是否修習統計課程對 CARE 統計素養相關概念的學習表現是否有影響？ | 修習統計差異 | | | V | | V | | | |
| 研究問題三： 非理工科系大學生運用 CARE 統計素養指引檢視科學新聞中統計訊息的學習成效，以及統計素養前後測表現的差異 | 3-1 受試者的 CARE 統計素養後測表現，題型間是否有差異以及分數是否提升？ | 高低分群 前後測差異 | 低 | 低 | 低 | 高 | 高 | 高 | 高 | 高 |
| | 3-2 大學是否修習統計課程對 CARE 統計素養後測表現以及分數提升是否有影響？ | 後測統計 前後測差異 | | | V | V | | | V | V |
| | 3-3 受試者的 CARE 統計素養表現在前測、教學與後測三階段任務中的變化為何？ | 高低變化 | 高 低 | 高 低 | 低 低 | 低 高 | 高 高 | 高 高 | 低 高 | 高 高 |

註 1：高-題型屬於高分表現群，以灰底呈現；低-題型屬於低分表現群。註 2：●表示只有沒修統計組前後測有差異，●表示兩組前後測皆有差異。

研究問題二：非理工科系大學生在學習 CARE 統計素養相關概念時的學習表現

學習表現結果顯示，統計素養學習表現的不同概念在分數表現上有顯著差異，代表受試者在不同概念的學習表現有顯著的差別，與前測的高低分表現群相同，高分表現題共 5 題與低分表現題共 3 題：高分表現題型的「P 機率的解讀」、「C-1 比例的比較」、「A-1 比較類型的選擇」、「A-2 比例類型的選擇」與「E 抽樣偏差」，受試者在前測時已有充足先備知識的情況下，經過教學任務學習如何理解題目與背後的統計概念並做連結後，幾乎都能夠自行訂正，因此得分集中在 2 分以上；在低分表現題型表現中，比較資料關係的計算類型「C-2 基準值差異的比較」或有定義概念的「R 隨機性(信賴區間)」與「C-3 混淆變項」，受試者需要花較多的心力，因此有較多比例受試者前測有誤但經過教學後理解落在 2 分、以及受試者需要二次講解概念而落在 1 分。

將受試者分為有修習與無修習統計相關課程兩組的學習表現結果，在學習統計素養的不同概念中，也是在「C-2 基準值差異的比較」與「A-1 比較類型的選擇」這兩個概念上，有修統計的受試者表現分數顯著高於沒修統計組，可以說大學期間有修習統計相關課程受試者的「C-2 基準值差異的比較」與「A-1 比較類型的選擇」兩概念比之無修統計課程者較能夠理解吸收，。

研究問題三：非理工科系大學生運用 CARE 統計素養指引檢視科學新聞中統計訊息的學習成效，以及統計素養前後測表現的差異

在後測表現結果顯示，受試者在不同統計素養題型的分數表現上有顯著差異，代表受試者面對科學新聞中的問題點埋錯，使用統計素養指引批判性思考的表現有顯著的差別，高分表現題共 5 題與低分表現題共 3 題，但是與前測以及學習表現的高低分群不同：高分表現題型的「C-3 混淆變項」、「A-1 比較類型的選擇」、「A-2 比例類型的選擇」、「R 隨機性(信賴區間)」與「E 抽樣偏差」，受試者較容易應用 CARE 指引，從語句與關鍵字分辨屬於何種統計概念並分析其中的陷阱，得分集中在 3 分與 2 分；在低分表現題型表現中，「C-1 比例的比較」與「C-2 基準值差異的比較」因為相較於前測是以一個問題一個概念的測驗模式，受試者在面對後測文章的語句解讀需要分辨各是哪種概念的錯誤時，部分受試者表示較難對比教學指引，無法順利解讀語句所埋藏的錯誤因而表現下降，在「P 機率的解讀」則是因為新聞脈絡的不同，需多考慮抽樣的推估對象，造成後測平均下降。

在前後測表現差異結果顯示，統計素養前後測的「C-2 基準值差異的比較」、「C-3 混淆變項」、「R 隨機性(信賴區間)」與「E 抽樣偏差」共四題，受試者後測顯著高於前測表現，代表這四題經過教學任務的指引後，受試者更能掌握科學新聞文本中的埋錯並更正，與前測相比有顯著提升。

而在受試者分為有修習與無修習統計相關課程兩組的後測表現上，經過教學任務後，前測與教學階段在「C-2 基準值差異的比較」與「A-1 比較類型的選擇」兩題皆有顯著差異，以及前兩階段任務皆無顯著差異的其餘六題型，兩組受試者在後測皆已無顯著差異。

第二節 限制與建議

一、研究限制與建議

本研究參考 Schield(2010)提出的四大影響統計值的面向，挑選日常科學新聞中較常見的幾個問題點，整理成 CARE 統計素養指引，因考量實驗總長對受試者的負擔，題目只能精選 8 類日常科學新聞中最常出現的錯誤類型做討論，每個概念只以一題進行前後測驗，整個實驗平均進行 2.5 小時，對於一些思考較緩慢的同學甚至進行至 4 小時，故往後的研究可以考慮分批實驗，以同一概念重複多題的測驗模式，提升結果的可信度。

關於實驗對象的部分，本研究受試者組成皆為北部非理工科系的大學學生，每個區域的環境、學生程度與特性皆有所不同，是否時常閱讀網路文章的習慣以及關注的議題也可能有所不同，建議未來研究對象能夠擴展至更多不同學校的大學生做為樣本，並比較其結果差異，建議收集更完整的數據，讓科學新聞中的統計素養研究能更加完善。

本研究為個別實驗，所有受試者單獨與主試者進行一次性的實驗流程，為了確保所有受試者接收到一致的教學內容，教學任務採取錄製影片的方式，疫情考量透過線上 Google Meet 會議以受試者分享螢幕畫面的狀態進行實驗，將教學影片上傳至 Youtube 影音平台，由受試者端播放影片進行學習與訂正的教學任務，因此研究者較難根據受試者的動作反應及時評估受試者的吸收程度，對於給予的教學幫助也有所桎梏，後續研究如果能以面對面的現場教學方式，對於評估受試

者的學習表現並及時給予幫助會更有效果。

二、教學上的建議

接下來針對研究中發現的受試者困難點，分為兩部分提出教學上的建議。

首先，根據教學任務中學習表現的人數分布分析結果：「A-1 比較類型的選擇」因為是針對同一筆資料，使用絕對差異與相對比例產生不同結果去判斷其優勢，較多受試者（6位，18%）需要主試者將題目一句句轉成數字做解析，之後教學建議能夠將絕對差異與相對比例以並列的方式呈現，輔助學生理解其中的不同之處；「C-2 基準值差異的比較」學習困難的主要原因是除了需要辨別語句的敘述是以何者為基準值，還需列出相對應的算式並做計算，有7位（21%）受試者需由主試者逐句列出相對應的算式才能夠理解，建議之後教學除了字句的拆解，還能夠同時列式呈現對學生會有更好的學習效果；而「R 隨機性(信賴區間)」與「C-3 混淆變項」兩類低分表現概念，因為都屬於專有名詞的討論，對於許多受試者來說像是重新學習一個新概念，分別有7位（21%）與16位（47%）受試者需要主試者針對各自的疑問之處再次說明，建議之後針對這兩類概念的教學如果能有更充裕的時間理解，並提供更多例子讓學生類推，以獲得更好的吸收效果。

不論之後繼續發展線上課程，或是改編發展為課室教學，建議針對本研究結果中受試者較低分表現的題型或概念，給予更詳細的說明以及例題練習，並期望將本研究發展成較長時間的課程設計，評估受試者是否能將統計素養應用在不同領域、不同議題的科學新聞中。

其次，子問題二將受試者分為大學有修習與無修習統計相關課程兩組，結果發現只有在題型「A-1 比較類型的選擇」與「C-2 基準值差異的比較」兩類概念上有顯著差異，也就是即使受試者在大學仍持續進修統計相關課程，與高中畢業後即沒接觸統計相關課程這兩類受試者的差異並不大，對於非理工科系大學生解讀科學新聞以及對數據判斷的正確性都是需要引導的。因為應用在生活中解讀科學新聞的統計素養與課室教學的統計概念不同，對屬於資料消費者的非理工科系大學生而言，教授他們如何解讀生活中的統計資訊更為重要。建議高中或大學統計相關課程，除了幫助學生了解科系所需的統計方法，也能加入統計知識的應用，以提升在科學新聞中的統計素養。



參考文獻

- Ben-Zvi, D., & Garfield, J. (2008). Introducing the emerging discipline of statistics education. *School Science and Mathematics, 108*(8), 355-361.
- Dieckmann, N. F., Slovic, P., & Peters, E. M. (2009). The use of narrative evidence and explicit likelihood by decisionmakers varying in numeracy. *Risk Analysis: An International Journal, 29*(10), 1473-1488.
- Frankenstein, M. (2009). Developing a criticalmathematical numeracy through real real-life word problems. In *Words and worlds* (pp. 111-130). Brill Sense.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives, 19*(4), 25-42.
- Gal, I. (2002a). Statistical literacy: Conceptual and instructional issues. *Perspectives on adults learning mathematics: Research and practice, 135-150*.
- Gal, I. (2002b). Adults' statistical literacy: Meanings, components, responsibilities. *International statistical review, 70*(1), 1-25.
- Gutstein, E. (2006). Reading and writing the world with mathematics: Toward a pedagogy for social justice. Taylor & Francis.
- Gutstein, E., & Peterson, B. (Eds.). (2005). *Rethinking mathematics: Teaching social justice by the numbers*. Rethinking Schools.
- Holmes, P. (2003). Statistical Literacy, Numeracy and the Future. *Augsburg College, Minneapolis, MI, 31*.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review, 87*(4), 329.
- Knouse, S. B. (1983). The letter of recommendation: Specificity and favorability of information. *Personnel Psychology, 36*(2), 331-341.
- Lehohla, P. (2002). Promoting statistical literacy: A South African perspective.
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical decision making, 21*(1), 37-44.
- Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science, 21*(1), 31-35.
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological science, 17*(5), 407-413.
- Raymond, R., & Schield, M. (2008). Numbers in the news: A survey. In *Proceedings of the ASA Section on Statistical Education* (Vol. 9).

- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372.
- Reston, E. D. (2005). Assessing Statistical Literacy in Graduate Level Statistics Education. Paper presented at the 55. In *th Session of the International Statistical Institute*.
- Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological bulletin*, 135(6), 943.
- Schild, M. (1999). Statistical literacy: Thinking critically about statistics. *Of Significance*, 1(1), 15-20.
- Schild, M. (2004). Statistical literacy curriculum design. *IASE Curriculum Design Roundtable*. See [www. StatLit. org/pdf/2004SchildIASE. pdf](http://www.StatLit.org/pdf/2004SchildIASE.pdf).
- Shields, M. (2005). Information literacy, statistical literacy, data literacy. *IASSIST quarterly*, 28(2-3), 6-6.
- Schild, M. (2006). *Statistical literacy survey analysis: Reading graphs and tables of rates and percentages*. Paper presented at the seventh International Conference on Teaching Statistics, Brazil.
- Schild, M., & Schild, C. (2007). Numbers in the news: A survey. In *Proceedings of the ASA Section on Statistical Education* (pp. 2323-2328).
- Schild, M. (2008). Quantitative literacy and school mathematics: Percentages and fractions. *Calculation vs. context: Quantitative literacy and its implications for teacher education*, 87-107.
- Schild, M. (2010). Assessing statistical literacy: Take CARE. *Assessment methods in statistical education: An international perspective*, 133-152.
- Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of internal medicine*, 127(11), 966-972.
- Snell, L. (1999). Using chance media to promote statistical literacy. In *Joint Statistical Meetings, Dallas, TX*.
- Ullmann, P. (2016). Communicating evidence? On the interaction of politics, data and the public. In *Promoting Understanding of Statistics about Society: Proceedings of the Roundtable Conference of the International Association of Statistics Education*.
- Viswanathan, M., & Childers, T. L. (1996). Processing of numerical and verbal product

- information. *Journal of Consumer Psychology*, 5(4), 359-385.
- Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1-8.
- Watson, J. M. (1997). Assessing statistical thinking using the media. *The assessment challenge in statistics education*, 12, 107-121.
- Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3-46.
- Watson, J. M., Collis, K., & Moritz, J. (1994). Authentic assessment in statistics using the media. *Report prepared for the National Center for Research in Mathematical Sciences Education-Models of Authentic Assessment Working Group (University of Wisconsin)*. Hobart, Australia: University of Tasmania, School of Education.
- Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics*, 96(1), 33-47.
- Wild, C., Triggs, C., & Pfannkuch, M. (1997). Assessment on a budget: Using traditional methods imaginatively. In *The Assessment Challenge in Statistics Education*, Eds. I. Gal and J.B. Garfield. Amsterdam : IOS Press.
- Yalch, R. F., & Elmore-Yalch, R. (1984). The effect of numbers on the route to persuasion. *Journal of Consumer Research*, 11(1), 522-527.
- 陳展興 (2015)。國小職前教師統計素養—以南部某國立大學為例。國立臺南大學應用數學系數學科教學碩士班碩士論文(未出版)。
- 翁意茹 (2006)。依據國小統計教材探討職前教師之統計素養-以國立臺北教育大學為例。國立臺北教育大學數學教育研究所碩士論文(未出版)。
- 黃俊儒、簡妙如 (2006)。科學新聞文本的論述層次及結構分佈：構思另個科學傳播的起點。 *新聞學研究*，(86)，135-170。
- 黃俊儒、簡妙如 (2008)。「科學家發明了什麼?!」--解析學生對於科學新聞中的科技產物意象。 *科學教育學刊*，(16:4)，415-438。
- 曾瑞玫 (2009)。高中學生統計素養表現之探討。國立花蓮教育大學數學系碩士論文(未出版)。
- 楊靜惠 (2006)。九年一貫數學課程實施下北部地區中學生統計素養之探究-以九年級為例。國立臺北教育大學數學教育研究所碩士論文(未出版)。
- 趙于甯 (2011)。大學生統計素養測驗與統計態度量表之發展及其相關研究。國

立臺南大學測驗統計研究所碩士論文(未出版)。

塑料污染：鯨魚的痛苦你應該懂（2018年2月7日）。**BBC News 中文**。取自：

<https://www.bbc.com/zhongwen/trad/science-42946353>

Richard Thompson（2018年2月7日）。海洋生物誤食塑膠垃圾或纏死近十年增
四成。**環境資訊中心**。取自：

<https://e-info.org.tw/node/94154>

洪巧藍（2019年5月24日）。台大研究：塑化劑無所不在！國小教室內外都測
得...也傷心血管。**ETtoday 健康雲**。取自：

<https://health.ettoday.net/news/1451728>

李柏昱（2013年5月14日）。你快樂嗎？空氣污染影響我們的幸福感受。

PanSci 泛科學。取自：

<https://pansci.asia/archives/54514>

台灣 PM2.5 的主要污染來源（2018年9月14日）。**台塑企業 熱門議題**。取
自：

<https://www.fpg.com.tw/tw/issue/1/115>

空氣污染防治大作戰—保護國人健康，讓臺灣環境永續（2020年10月14
日）。**行政院 政策與計畫**。取自：

<https://www.ey.gov.tw/Page/5A8A0CB5B41DA11E/23c411a0-1b20-42e7-9843-daf6cdedc61b>

正視塑膠汙染! 便利背後的環境悲歌

文 / 游珮詩
2021-09-12



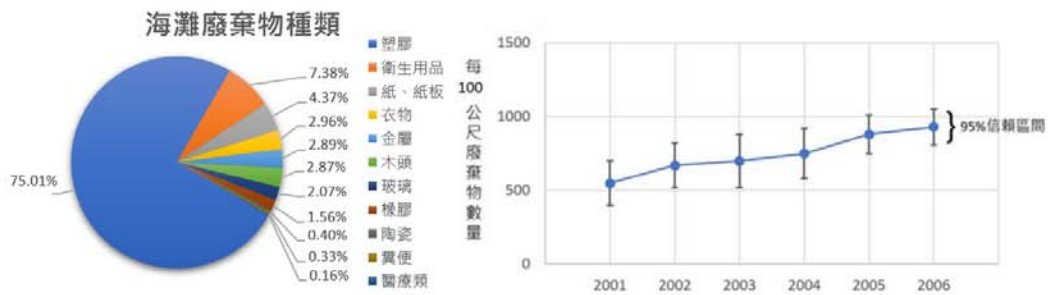
(圖片來源：[plasticsoupnews](https://plasticsoupnews.com))

水中抽樣

近年來，塑膠垃圾污染問題越來越嚴重，塑膠微粒的污染已經滲入我們生活各方面，從河川、海洋到土壤、食物以及我們的飲用水等。美國一家調查機構提出我們平常飲用的自來水中，含有肉眼看不見的塑膠微粒。該研究對全球各地 14 個國家的自來水進行抽樣調查，結果顯示平均 83% 的自來水樣本中含有塑膠微粒。

英國新聞頻道 Sky News 報導指出，全球每年大約產生 3 億噸的塑膠垃圾；其中只有 9% 被回收，12% 被焚燒，另外的 79% 被掩埋或是流入到自然環境中，尤其是海洋。全世界每年海洋平均多出約 1,270 萬噸的塑膠垃圾，對海洋生物造成嚴重的威脅，例如造成海龜窒息和鯨魚中毒。

奧斯陸-巴黎公約(OSPAR)統計全球許多處海灘上廢棄物的種類比例(圖一圓餅圖)，以此看出塑膠物品一直以來佔全球海洋廢棄物的大宗。圖二折線圖表示海灘上「每 100 公尺有多少廢棄物」的年平均值趨勢(圖上也顯示了各年廢棄物 95% 信賴區間)。



▲圖一

▲圖二

健康影響

塑膠垃圾會在降解的過程中，成為無法分解的塑膠微粒，不僅對海洋生物造成危害，也可能透過食物鏈進入人體中。塑膠微粒常被海洋浮游生物或貝類誤食，經由食物鏈，累積在中高階掠食者的體內，衍生出更多嚴重的生態和環境污染問題，也間接影響人類的健康。

雖然接觸塑膠製品並不會對人體造成立即危害，長期大量攝入才會對人體健康造成慢性影響，而透過食物鏈的累積，塑膠微粒在體內的含量已不可小覷。這些塑膠含有的有毒物質可能會干擾人類的內分泌、生殖系統，甚至可能致癌。台大研究團隊以多變項控制塑化劑濃度，發現塑化劑濃度與頸動脈硬化機率呈現正相關，身體中塑化劑濃度越高，罹病風險越高。因此生活減塑，不再是一句口號，而是你我不得不面對的燃眉健康危機了！



(圖片來源：翻攝自荒野保護協會)

回收現況

塑膠袋與塑膠包材便宜又方便，環保署統計台灣人平均一年塑膠容器的回收率大約 50%，相較於紙容器的 75% 高回收率仍舊不足，代表還有一半的塑膠容器成為垃圾，更何況還有許多的塑膠製品未計入調查。沒回收的塑膠不是被亂丟，就是焚燒或掩埋，低回收率的原因是塑膠的材質多種又難以分類，材料價值極低，非常不利於後端回收再利用，因此多數業者不收。但是無法妥善回收造成地球的沉重負擔，最後仍然是你我要買單。



(圖片來源：[greenpeace](#))

源頭減塑

氾濫的塑膠使用，源於「用了就丟」的習慣，要靠我們從生活中改變。儘管臺灣人在生活中力行減塑，養成自帶環保杯、環保袋的習慣，但當我們去超市、量販店等採購時，仍然帶回很多不必要的塑膠包裝產品。近年來，政府鼓勵產業推動事業廢棄物源頭減量，除要求產品輕量包裝，在產品設計及製造過程導入重新設計，落實綠色設計思維。



生產者如能對產品以及包裝成為廢棄物的處理負起責任，也就是說在生產的階段，加入最終處理成本的考量，進而從一開始就使用可回收物質，並使用便於回收再利用的設計。因此督促企業一起加入減塑的行列，包括與日常生活密切相關的零售通路，非常關鍵！

閱讀完畢請告知，謝謝



空氣汙染! 年復一年的公衛危機

文 / 游珮詩
2021-11-02



(圖片來源：康健雜誌)

全球空氣汙染

PM2.5 是指微粒直徑小於 2.5 μm (即 0.00025cm)的懸浮微粒，聯合國世界衛生組織 (WHO) 的空氣汙染報告中提出，空氣中 PM2.5 濃度的年平均值要在每立方公尺 10 微克 ($\mu\text{g}/\text{m}^3$) 以下才安全，這項報告是來自八個國際機構 16 名科學家合作，透過衛星與地面測量蒐集全球逾 3,000 個監測站點的空污數據，發現全球 92% 的民眾呼吸的是污染物超標的空氣，尤其亞洲地區汙染情況更嚴重。根據瑞士空氣品質科技公司 IQAir 的排名，全球去年空氣品質最差的城市，前 148 名都在亞太地區。

身心靈影響

空氣汙染已經被許多研究證實對於人類生理有許多負面影響，例如會造成心血管疾病、呼吸道患者人數增加等等。除了對生理上的諸多負面影響以外，一份加拿大的研究亦發現空氣汙染與人們的心理狀態有關：空氣汙染越嚴重的地方，居民感覺越不快樂，反之亦然。研究者檢視歐洲 14 個不同國家的空氣汙染紀錄，計算空氣汙染嚴重程度與居民快樂程度的關係。清新的空氣或良好的自然環境可以提高居民的生活滿意度，因為居民可以擁有更為健康的身體，或有更好的生活方式。相反的，如果空氣汙染嚴重，將迫使居民減少從事自己喜歡的活動，因而感覺不快樂。

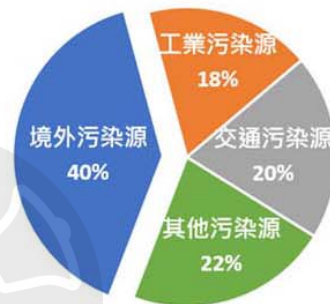


(圖片來源 : Heho 健康)

台灣現況

臺灣的境外空氣污染主要源於中國，但境外污染只佔臺灣空污成因的四成，另外六成污染源都在我們的環境生成。境內污染源包括交通污染源約佔 20%、工業污染源 18% 以及其他零碎的污染源加總約 22%；交通污染源主要為交通運輸排放之廢氣，工業污染源包括火力發電廠、鋼鐵與石化等高污染產業工廠排放。

我國細懸浮微粒(PM2.5)濃度貢獻源



(圖片來源 : [行政院環保署 106 年度空氣汙染防制總檢討](#))

由此可知兩件事：

1. 工業污染源的百分比相較於交通污染源的百分比低 1 成，且交通污染的百分比相較於其他污染的百分比也是低 1 成；

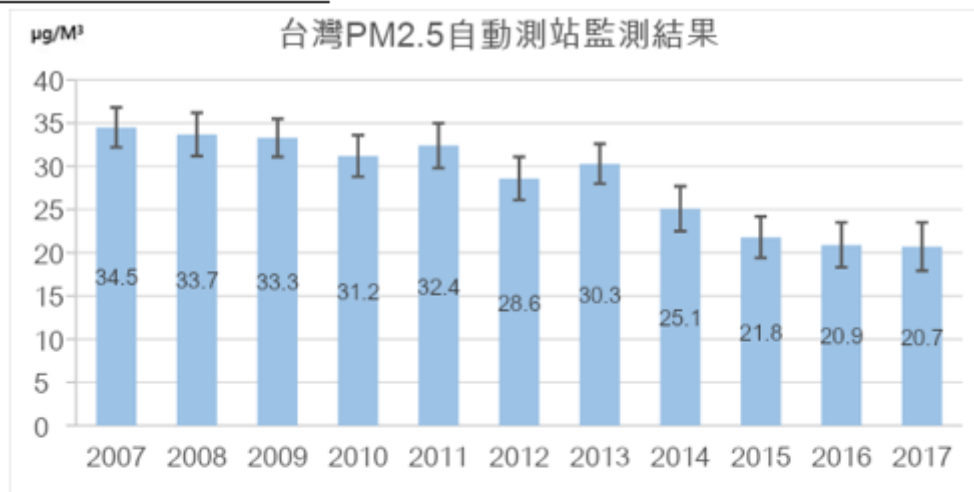
2. 空污中的交通污染源只有 20%，也就是交通產生的所有環境污染中，空氣污染也是只有佔 20%；

此外，空氣汙染中的工業污染源 18% 是所有汙染源比例最少，就是工業造成的各種環境污染，其中空汙比例也是最少，所以工業背負環境汙染禍首這個罵名其實是有待商榷的。

政策改善

下圖為環保署空氣品質監測報告，呈現每年 PM2.5 濃度平均值(圖上也標示各年濃度平均值的 95% 信賴區間)，PM2.5 濃度自 2007 年起由 34.5 $\mu\text{g}/\text{m}^3$

下降至 2017 年只有 $20.7\mu\text{g}/\text{m}^3$ ，可以看出 PM2.5 濃度逐年顯著下降，代表台灣空氣品質確實有在改善。



(圖片來源：環保署 空氣品質監測報告)

政府為了全面改善國內空氣污染，針對上述 3 種可控制的境內污染源(工業、交通與其他)提出因應對策，提出新的「空氣污染防制行動方案」，建構更完整的空氣品質管理制度。比較舊有的《空氣污染防制法》實行四年(2015~2019 年)與新的行動方案(2020 年)的成果：

全國 PM2.5 紅色警示次數由 2015 年 997 次，降至 2019 年 146 次 (次數降低 851 次，降幅 85%)；但是新的行動方案推出即在 2020 年讓警示次數降至只有 15 次 (次數降低 131 次，降幅 90%)；可見新的行動方案降幅比較多，對空氣品質改善比舊的《空氣污染防制法》成效更良好。

空氣汙染是一個慢性問題，但是許多國家及部分政治人物卻把它當成是追求發展的必要代價，決定與空污共存。若各國沒有更徹底的環境政策，這場沒有疫苗的公衛危機，恐怕還會年復一年地發生。



附錄三 統計素養教學投影片

1. P-機率的解讀

正視塑膠汙染! 便利背後的环境悲歌

請問文中平均83%的自來水樣品中含有塑膠微粒是什麼意思?

(A) 不要喝自來水，因為每一升水的塑膠微粒濃度為83%

(B) 喝自來水時，盡量避免喝到83%塑膠微粒的部分

(C) 每100人，約有83人會喝到塑膠微粒

(D) 每100處地點自來水樣品，約有83處樣品中含有塑膠微粒

(E) 隨機抽任一地點的自來水，塑膠微粒佔整杯水83%

(D) 單一地點自來水抽樣結果為含塑膠微粒與否，因此平均每100處有83處有含是正確的解讀。

2. 新聞中的統計素養

新聞中的統計素養

正視塑膠汙染! 便利背後的环境悲歌

近年來，塑膠垃圾污染問題越來越嚴重，塑膠微粒的污染已經進入我們生活各個角落。食物以及我們的飲用水等。美國一家調查機構提出我們平常飲用的自來水含有塑膠微粒。該研究對全球各地14個國家的自來水進行抽樣，平均83%的自來水樣品中含有塑膠微粒。

根據美國Sky News報導指出，全球每年大約產生3億噸的塑膠垃圾，其中約有9%被回收，12%被焚燒，其餘的7%則被埋進垃圾堆。這些塑膠垃圾在自然環境中分解成微小的塑膠微粒，這些微粒會隨風飄散，隨水流動，甚至會進入我們的食物鏈。

3. 介紹 CARE

對統計值的四大影響 (Schield, 2010)

操縱變因 → 應變變因

C Context
變項的脈絡(比較)

A Assembly
變項的集合(選擇)

R Randomness
隨機性

E Error
偏差

4. Assembly 變項的集合介紹

Assembly 變項的集合

以下方式做選擇時對統計值的影響：

- (1) 群組的選擇
- (2) 平均值與中位數的選擇
- (3) 比較類型的選擇
絕對和相對-簡單差異與倍率
- (4) 比例類型的選擇

群組大小
失業定義不同得出相異的失業率

平均值與中位數
平均薪資 vs 薪資中位數

「对不起，是我拉低了平均薪資！」

CARE

5. A-1 比較類型的選擇

Assembly 變項的集合

以下方式做選擇時對統計值的影響：

- (3) 比較類型的選擇
絕對和相對-簡單差異與倍率

某企業希望為減塑盡分心力，因此評估販賣10,000個產品的情況下哪種產品經過包裝改良後更能減塑，方案如下：

1. 產品A經過包裝改良，每公斤塑膠原料從可製造10個提高至製造20個
(也就是原本需要1000公斤減少到只要500公斤)
2. 產品B經過包裝改良，每公斤塑膠原料從可製造20個提高至製造100個(原本需要500公斤減少到只要100公斤)

甲、乙在討論哪一種產品經過包裝改良後更能減少塑膠使用：

甲說：產品A比較減塑，因為產品A節省500公斤塑膠原料，而產品B只節省400公斤。

乙說：產品B比較減塑，因為產品B節省80%的塑膠原料，而產品A只節省50%的塑膠原料。

請問甲、乙的說法有何差異？你覺得誰說的比較合理？

甲的觀點
產品A改良所減少的塑膠原料總重量大於產品B改良所減少的塑膠原料，500kg>400kg，A減少較多的塑膠原料重量。

乙的觀點
產品B節省的比例大於產品A節省的比例，80%>50%，B減少使用塑膠的效益較佳，更有效率地使用原料。

兩者在減塑的總量與效益上皆有不同的優勢

CARE

6. A-2 比例類型的選擇

Assembly 變項的集合

以下方式做選擇時對統計值的影響：

- (4) 比例類型的選擇

塑膠回收率
 $\frac{\text{塑膠容器回收}}{\text{塑膠容器}}$

回收物佔比
 $\frac{\text{塑膠容器}}{\text{所有回收物}}$

塑膠容器製品
塑膠容器回收物
所有回收物

問題6 /A-2 請問文中提到塑膠容器的回收率為50%，是否代表「回收物中塑膠容器的比例」也是50%？

兩者需要的資料不一樣，值也會不一樣。

CARE

7. Assembly 變項的集合總結

Assembly
變項的集合

以下方式做選擇時對統計值的影響：

- (1) 群組的選擇
- (2) 平均值與中位數的選擇
- (3) 比較類型的選擇
絕對和相對-簡單差異與倍率
- (4) 比例類型的選擇

CARE

8. C-1 比例的比較

Context
變項的脈絡

回收率 VS 回收物占比

回收物總量

塑膠製品 50%
紙容器回收物 75%

塑膠製品 50%
紙容器回收物 75%

玻璃回收物 金屬

CARE

9. C-2 基準值差異的比較

Context
變項的脈絡

變項對統計值的影響：

(2) 基準值差異的比較

問題7 /C-2 文中提到塑膠容器的回收率大約50%，相較於紙容器的75%回收率，也就是說紙容器比塑膠容器的回收率高50%，請問是否等於塑膠容器回收率比紙容器低50%?

紙容器比塑膠容器的回收率高50% 塑膠容器低50%
紙容器75%

塑膠容器回收率比紙容器低50%? 塑膠容器低33.3%(1/3)
塑膠容器50%
紙容器75%

CARE

10. Context-研究設計

Context 變項的脈絡

變項對統計值的影響：
(3) 研究設計
實驗研究 vs. 觀察性研究
controlled vs. uncontrolled

宣稱：多吃蔬菜，減重瘦身更容易

操縱變項 → 應變變項

蔬菜份量
大量蔬菜
少量蔬菜

控制變項
生活習慣
蔬菜種類、
烹調方式

體重變化

CARE

11. Context-混淆變項常見例子

Context 變項的脈絡

變項對統計值的影響：
(4) 討論在研究中沒考慮到的混淆變項的影響

老師?
學生素質?

混淆變項

操縱變項x → 應變變項y

宣稱

教學法
新教學法
傳統教學

學習成效

CARE

12. C-3 混淆變項

Context 變項的脈絡

變項對統計值的影響：
(4) 討論在研究中沒考慮到的混淆變項的影響

職業
家庭
飲食習慣

混淆變項

操縱變項x → 應變變項y

宣稱

生活減塑
自備環保餐具餐盒
正常使用塑膠耗材

心血管疾病

問題4
根據文中的資訊，以及台大研究發現塑化劑與頸動脈硬化呈現正相關，可否得到「透過生活減塑讓人攝入的塑化劑濃度降低，就可以減少心血管疾病的患病機率」的結論？
混淆變項
是否有其他與塑化劑相關的因素同時在影響這個結論？

研究結果相關不代表因果關係，研究只證明體內塑化劑濃度與頸動脈硬化呈現正相關，無法推論生活減塑可以減少罹患心血管疾病機率。

體內塑化劑濃度高的人或許有可能是：
1. 職業、家庭經濟影響(高壓、廉價塑膠製品)
2. 飲食習慣(常外食、常吃高熱量、油膩食物等等造成心血管疾病)

CARE

13. C-3 混淆變項反向影響舉例

Context 變項的脈絡

變項對統計值的影響：
(4) 討論在研究中沒考慮到的混淆變項的影響

都市醫院V.S.地方醫院
根據患者病情嚴重程度有不同的死亡率
狀況良好的患者可以自行走進去；狀況不佳的患者被抬進去。

| 醫院/病情 | 良好 | 不佳 |
|-------|----|----|
| 都市 | 1% | 6% |
| 地方 | 2% | 7% |

→ 都市醫院平均死亡率4%
→ 地方醫院平均死亡率5%

標準化重症比例後，醫院皆為60%的患者狀況不佳。
→ 都市醫院平均死亡率4%比地方醫院平均死亡率5%低。

混淆變項=

操縱變項x 宣稱 應變變項y

醫院醫術? 都市醫院 5.5% 地方醫院 3.5%

都市醫院有90%患者病情不佳
地方醫院有30%患者病情不佳
重症比例

CARE

14. Context 變項的脈絡總結

Context 變項的脈絡

變項對統計值的影響：
(1) 比例的比較
(2) 基準值差異的比較
(3) 研究設計

研究中沒考慮到變項的影響：
(4) 混淆變項

CARE

15. 統計資料抽樣介紹與隨機性

統計資料的抽樣

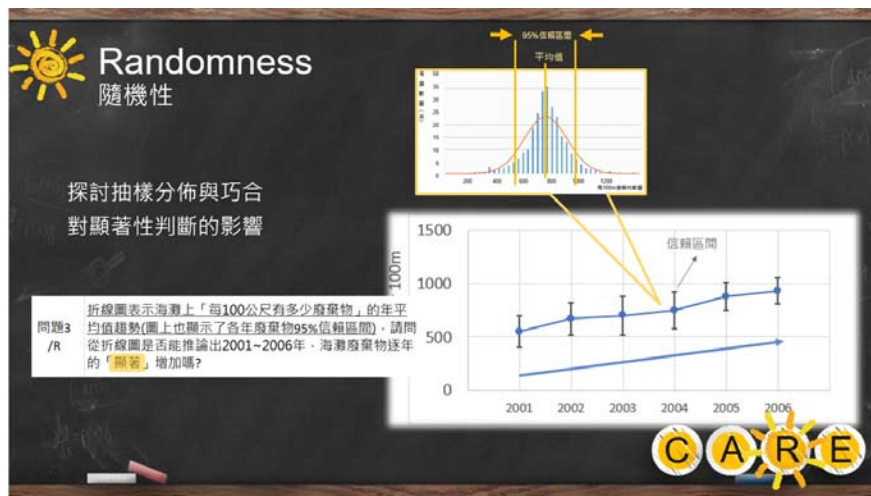
母體 母體平均數

樣本 樣本平均數

隨機性 Randomness

CARE

16. R 隨機性(信賴區間)



17. 統計資料抽樣與偏差



18. E 抽樣偏差

Error
偏差

會造成研究觀察的結果與現實之間
產生差異的影響:

(1) 抽樣偏差
(2) 測量偏差

問題2 /E
小明看到文章中的研究對全球各地14個國家的自來水進
行抽樣調查，質疑抽樣地點的選擇，不能代表全世界，
因此認為抽樣結果有偏差(與現實有差異)，請問小明的
質疑合理嗎?為什麼?

畢業生流向調查
對師大有感情想反饋的校友

1. 14個國家的背景環境是否夠具代表性
2. 各地的自來水抽樣數量(一國或一城市?)
3. 是否針對工業區或水源加強抽樣
4. 研究中的自來水抽樣地點定義(例如:各國的首
都或是自來水處理場?有無處理過?)

CARE

19. 其他偏差介紹

Error
偏差

會造成研究觀察的結果與現實之間產生差異的影響:

- (1) 抽樣偏差
- (2) 測量偏差
- (3) 受試者偏差

☀️ 台灣青少年男生平均身高174.5公分
使用身高測量儀或是家裡貼在牆壁上的身高尺

☀️ 台灣人資源回收率高達55%
調查流程差異、人們在回答時指的是全部還是偶爾

CARE

20. CARE 指引回顧與統整

Context
變項的脈絡(比較) **C** 比例的比較
基準值差異的比較
混淆變項

Assembly
變項的集合(選擇) **A** 比較類型的選擇-相對與絕對(減塑效率與減塑量)
比例類型的選擇-回收率與佔比

Randomness
隨機性 **R** 隨機性-信賴區間與顯著

Error
偏差 **E** 抽樣偏差-抽樣代表性

idea

附錄四 前測試題評分規準

| 代碼 | 評分規準 |
|-----|--|
| P | <p>【題目】 (問題1) 請問文中平均 83% 的自來水樣品中含有塑膠微粒是什麼意思？</p> <p>(A) 不要喝自來水，因為每一杯水的塑膠微粒濃度為 83%</p> <p>(B) 喝自來水時，盡量避免喝到 83% 塑膠微粒的部分</p> <p>(C) 每 100 人，約有 83 人會喝到塑膠微粒</p> <p>(D) 每 100 處地點抽自來水樣品，約有 83 處樣品中含有塑膠微粒</p> <p>(E) 隨機抽任一地點的自來水，塑膠微粒佔整杯水 83%</p> |
| | <p>【評分規準】</p> <p>3 分：選答 (D)，能夠正確說出每處/每份抽樣為含塑膠微粒與否，以及平均 83% 是每 100 處有 83 處。</p> <p>2 分：選答 (D)，但判斷的理由只有部分正確。</p> <p>1 分：選答 (D)，但無提供支持選項的判斷理由，或判斷理由錯誤敘述。</p> <p>0 分：錯誤選項 (A/B/C/E)，或不會回答、回答無意義。</p> |
| | <p>【受試者答題情況範例】</p> <p>3 分：D、平均 83% 代表為數據分析下的綜合平均數，所以意思就是百分之 83，100 個自來水樣品有 83 個含有塑膠微粒。(S33)</p> <p>2 分：D、因為 83% 是自來水樣品的形容詞(口頭說明：憑直覺，覺得應該是很多杯，有 83 杯有塑膠微粒) (S11)</p> <p>1 分：D、因為其他選項看起來是錯的，就覺得 D 是對的 (S22)</p> <p>0 分：E、從閱讀的提題目當中我覺得因為是在不同的地方抽取自來水，所以塑膠微粒也是從自來水當中測出來，故平均 83% 就是說平均一杯自來水當中有百分之 83 是塑膠微粒。(S06)</p> |
| C-1 | <p>【題目】 (問題5) 文中提到<u>塑膠容器的回收率大約 50%，相較於紙容器的 75% 高回收率仍舊不足</u>，也就是塑膠容器回收率小於紙容器回收率；如果我們計算所有回收物中塑膠容器的占比，是否也會小於紙容器的占比？</p> |
| | <p>【評分規準】</p> <p>3 分：判斷正確 (否)，並針對回收率與回收物占比兩者提出精確的判斷理由，例如指出兩者分母分子的意涵，或畫出文氏圖解釋兩者的母體差異對比較統計值大小造成的影響</p> <p>2 分：答案正確 (否)，但理由中有部分錯誤敘述，雖然知道兩者意涵，但是認為沒有實際數量無法比較。</p> <p>1 分：答案正確 (否)，但理由中有部分錯誤敘述，雖然知道兩者意涵，但是認為沒有實際數量無法比較。</p> <p>0 分：答案錯誤 (是) 只依據自身想法或經驗來回答 / 不會回答或回答無意義。</p> |

| | |
|-----|--|
| | <p>【受試者答題情況範例】</p> <p>3分：不一定、如果塑膠容器本身的量較紙容器多出一定程度，則就算回收率較低，在所有回收物中塑膠容器的占比可能還是會高於紙容器的占比(較大量的 50%可能大於較小量的 75%) (S27)</p> <p>2分：否、只是所有塑膠容器中可回收的占比約 50%，只能反映塑膠可回收率不高，不能反映塑膠容器回收的量總共有多少 (S29)</p> <p>1分：不、因為不知道回收總量 分母不一樣 但我不確定分母是甚麼。(S11)</p> <p>0分：Yes、依照文中敘述應該可以這樣推論。雖然沒有辦法百分之百保證。(S08)</p> |
| C-2 | <p>【題目】</p> <p>(問題7)文中提到<u>塑膠容器的回收率大約 50%，相較於紙容器的 75%回收率</u>，也就是說紙容器比塑膠容器的回收率高 50%，請問是否等於塑膠容器回收率比紙容器低 50%？</p> |
| | <p>【評分規準】</p> <p>3分：答案正確(否)，並針對兩者回收率提出精確的判斷理由，例如列出完整算式得出比塑膠容器高 50%為比紙容器低 33.3%</p> <p>2分：答案正確(否)，但判斷理由不夠精確充分，或理由中有部分錯誤敘述，例如順著題目計算塑膠比紙回收率低 50%，則塑膠容器回收率將僅有 37.5%而非原文之 50%。或是知道紙比塑膠高 50%，但不知道塑膠比紙高多少百分比，或是知道對象不同所以不會是高低都是 50%，但是不會列式無法知道正確數字。</p> <p>1分：答案正確(否)，但未寫出判斷理由或判斷理由不正確(例如覺得應該不會相反但說不出原因、或是覺得應該就是差 25%) / 回答正確(否)但只依據自身想法或經驗來回答。</p> <p>0分：答案錯誤(是) / 不會回答或回答無意義。</p> |
| | <p>【受試者答題情況範例】</p> <p>3分：否、$50\%(1+50\%)=75\%$, but $75\%(1-50\%) \neq 50\%$, 塑膠容器回收比紙類低 33% (S23)</p> <p>2分：否、塑膠容器比紙容器回收率低 50%，則塑膠容器回收率將僅有 37.5%，非原文之 50% (S04)</p> <p>1分：否、這兩個回收率為什麼可以直接比較?(口頭補充：否，不能等於也是 50%，但不知道為什麼要這樣比) (S17)</p> <p>0分：Yes、感覺這個敘述沒有問題。(S08)</p> |
| C-3 | <p>【題目】</p> <p>(問題4)根據文中的資訊，以及台大研究發現塑化劑濃度與頸動脈硬化機率呈現正相關，可否得到「透過生活減塑讓人攝入的塑化劑濃度降低，就可以減少心血管疾病的患病機率」的結論？是否有其他與塑化劑相關的因素同時在影響這個結論？</p> |
| | <p>【評分規準】</p> <p>3分：判斷正確(是)，生活減塑不一定就會降低體內塑化劑濃度，並且能夠反思其他與塑化劑相關的因素同時在影響結論。</p> <p>2分：答案正確(是)，但判斷理由不夠充分，或理由中有部分錯</p> |

| | |
|-----|--|
| | <p>誤敘述，因素只與其中一項有關，例如只與塑化劑或心血管有關。</p> <p>1分：答案正確（是），但理由中有一半部分錯誤敘述，或是覺得應該有其他可能但說不出原因。</p> <p>0分：答案錯誤（否）只依據自身想法或經驗來回答 / 不會回答或回答無意義。</p> <p>【受試者答題情況範例】</p> <p>3分：是、據我所知，膽固醇攝取過高，作息不正常跟常吃高鈉高油的食物像是泡麵等等，都會讓【塑化劑直接影響頸動脈硬化】這個結論有可以討論的空間。這樣統整下來，我會比較覺得提出【現代社會的人們在講求高效率、低時間消耗下這些觀念的影響下，選擇超商的塑膠包裝微波加熱食品的機率變高，進而導致高油鹽、塑化劑等因素影響人類健康】比較正確，也比較能從更大的範圍還有視角去帮助大家提防心血管疾病。(S34)</p> <p>2分：是、文中提到有毒物質可能也跟內分泌、生殖系統、罹癌有關，所以可能是因為其他因素導致此一結論 (S28)</p> <p>1分：是、文中提到有毒物質可能也跟內分泌、生殖系統、罹癌有關，所以可能是因為其他因素導致此一結論 (S15)</p> <p>0分：否、我其實不確定，因為沒有研究這個部分 (S23)</p> |
| A-1 | <p>【題目】</p> <p>(問題8) 某企業希望為減塑盡分心力，因此評估販賣 10,000 個產品的情況下哪種產品經過包裝改良後更能減塑，方案如下：</p> <ol style="list-style-type: none"> 1. 產品 A 經過包裝改良，每公斤塑膠原料從可製造 10 個提高到製造 20 個（也就是原本需要 1000 公斤減少到只要 500 公斤） 2. 產品 B 經過包裝改良，每公斤塑膠原料從可製造 20 個提高到製造 100 個（原本需要 500 公斤減少到只要 100 公斤） <p>甲、乙在討論哪一種產品經過包裝改良後更能減少塑膠使用：</p> <p>甲說：產品 A 比較減塑，因為產品 A 節省 500 公斤塑膠原料，而產品 B 只節省 400 公斤。</p> <p>乙說：產品 B 比較減塑，因為產品 B 節省 80% 的塑膠原料，而產品 A 只節省 50% 的塑膠原料。</p> <p>請問甲、乙的說法有何差異？你覺得誰說的比較合理？</p> |
| | <p>【評分規準】</p> <p>3分：答案（都合理），並提出精確適當的判斷理由，例如提出完整算式作為判斷證明兩者在減塑的總量與效率上皆有不同的優勢；答案（甲/乙），有提出精確適當的判斷理由，例如提出完整算式作為判斷證明兩者在減塑的總量與效率上皆有不同的優勢。但是因為其他理由選擇總量或效益。</p> <p>2分：答案（有選擇），但說明不夠充分精確，例如只覺得都有減塑，但是無法意識到總量與效率的差異。/對總量與效率對一半，錯一半。</p> <p>1分：答案正確（都合理），但無法說明不夠充分精確或太過含糊，例如只覺得都有減塑，但是說不出原因。</p> <p>0分：無作答/有作答但未寫出判斷理由/不會回答。</p> |
| | <p>【受試者答題情況範例】</p> |

| | |
|------------|--|
| | <p>3分：都合理，如果以絕對值去算減量，那A比較減塑。但是以比值去看，產品B減塑的效果會大於產品A。我認為AB都合理，因為如果兩個產品本身消耗量就不同，都用同一個標準去看的話會不公平。(S33)</p> <p>2分：乙，甲的說法有點奇怪，不知道為什麼可以直接用最終使用的塑膠原料量去比較，這種應該會是看個別一個產品所減少的塑膠用量。乙的說法我比較認同，就一個產品使用原料多寡的面向來看，B節省資源的效率確實比A要好。"(S17)</p> <p>1分：乙比較合理，通常減少或增加一些原料好像都是用百分比呈現(S22)</p> <p>0分：無</p> |
| <p>A-2</p> | <p>【題目】 (問題6)請問文中提到<u>塑膠容器的回收率為50%</u>，是否代表「回收物中塑膠容器的比例」也是50%?</p> <p>【評分規準】</p> <p>3分：判斷正確(否)，並針對回收率與回收物占比兩者提出精確的判斷理由，例如指出兩者分母分子的意涵，或用文氏圖解釋兩者的不同。</p> <p>2分：答案正確(否)，只知道回收率或回收佔比其一，且無法區辨兩者。</p> <p>1分：答案正確(否)，但無提供支持選項的判斷理由，或判斷理由錯誤敘述，或只根據自身的想法或經驗來判斷。</p> <p>0分：答案錯誤(是)，只依據自身想法或經驗來回答/不會回答或回答無意義。</p> <p>【受試者答題情況範例】</p> <p>3分：否、兩者意義不同，前者代表所有塑膠容器中有50%被回收，後者代表所有回收物中有50%為塑膠容器。(S27)</p> <p>2分：否、塑膠容器回收率僅表示所有塑膠容器取得回收的比率，與所有回收物中各項材質占比無關(S04)</p> <p>1分：否、塑膠容器回收率僅表示所有塑膠容器取得回收的比率，與所有回收物中各項材質占比無關(S11)</p> <p>0分：是、兩句話是換句話說(S02)</p> |
| <p>R</p> | <p>【題目】</p> <div style="display: flex; justify-content: space-around;"> <div data-bbox="406 1590 774 1859"> <p>▲圖一</p> </div> <div data-bbox="790 1590 1292 1859"> <p>▲圖二</p> </div> </div> <p>(問題3)折線圖表示<u>海灘上「每100公尺有多少廢棄物」的年平均趨勢(圖上也顯示了各年廢棄物95%信賴區間)</u>，請問從折線圖是否能推論出2001~2006年，<u>海灘廢棄物逐年的「顯著」增加嗎?</u></p> |

| | |
|----------|---|
| | <p>【評分規準】</p> <p>3分：答案正確（否），並針對顯著增加提出精確的解釋，例如寫出信賴區間誤差線與顯著的關聯性及其意義。</p> <p>2分：答案正確（否），但判斷理由不夠精確充分，或理由中有部分錯誤敘述，例如覺得誤差線有包含到前一年結果，但是不確定與顯著的關聯性。</p> <p>1分：回答正確（否），但未寫出判斷理由或判斷理由不正確。</p> <p>0分：答案錯誤（是），只依據自身想法或經驗來回答 / 不會回答或回答無意義。</p> |
| | <p>【受試者答題情況範例】</p> <p>3分：無法、因為只有平均值的趨勢是在增加的，但95%的信賴區間範圍有重疊、也可能有持平甚至降低可能，沒辦法確認是顯著增加。(S12)</p> <p>2分：否、雖然有點忘記之前高中學過的信賴區間的意義，但2001~2005年之間，他們的信賴區間都有重疊的部分，我認為這代表這些年中的垃圾量其實有可能也是差不多的，2001~2006年間整體雖然垃圾量的確是有上升趨勢，但應該不能算是「顯著」增加。(S19)</p> <p>1分：否、僅些微幅度增加。(S04)</p> <p>0分：可以、因為短短五年，長了快兩倍，所以算顯著。(S32)</p> |
| E | <p>【題目】</p> <p>(問題2) 小明看到文章中的<u>研究對全球各地14個國家的自來水進行抽樣調查</u>，質疑抽樣地點的選擇，不能代表全世界，因此認為抽樣結果有偏差(與現實有差異)，請問小明的質疑合理嗎？為什麼？</p> <p>【評分規準】</p> <p>3分：提出支持論點的適當判斷理由，例如製造出塑膠微粒偏差可能的背景或原因、提出樣本中國家/自來水源選擇。</p> <p>2分：提出支持論點的理由，但說明不夠充分精確，例如只提出不知道14個國家有誰，但無法解釋樣本選擇對母體產生什麼可能的偏差。</p> <p>1分：提出論點，但無提供支持選項的判斷理由，或判斷理由錯誤敘述，或只根據自身的想法或經驗來判斷。</p> <p>0分：不會回答、回答無意義。</p> <p>【受試者答題情況範例】</p> <p>3分：合理、如果抽樣國家沒有客觀地分布在已開發、開發中國家，那數據顯示一定會偏頗某類型發展國家。(S33)</p> <p>2分：合理、因為只有14個國家，樣本數太少。地點跟...但是不知道會有什麼偏差。(S22)</p> <p>1分：無</p> <p>0分：無</p> |

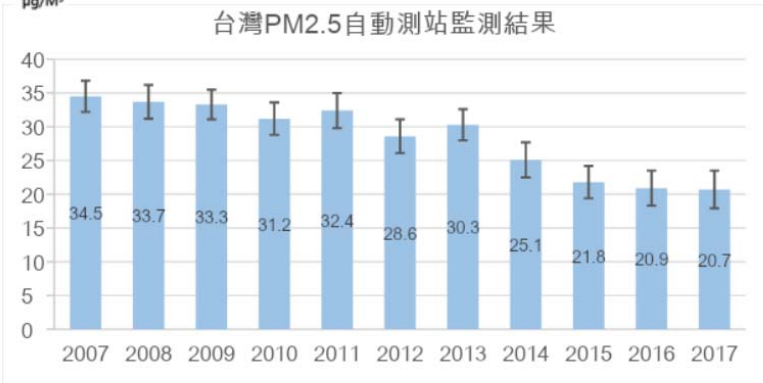
附錄五 後測試題評分規準

| 代碼 | 評分規準 | | | | | | | | | | |
|-------|---|-------|-----|-------|-----|-------|-----|-------|-----|-------|-----|
| P | <p>【題目】 (問題點 2) 這項報告是來自八個國際機構 16 名科學家合作，透過衛星與地面測量蒐集全球逾 3,000 個監測站點的空污數據，發現全球 92% 的民眾呼吸的是污染物超標的空氣，尤其亞洲地區污染情況更嚴重。</p> <p>【評分規準】 3 分：能夠正確說出每處抽樣為超標與否，以及平均 92% 是每 100 處有 92 處的地點，且發現作者推估到人口百分比。或是認為 92% 空汙站點 x 當地人口比例計算出來的平均數。 2 分：提出抽樣概念，利用抽樣概念解釋 92% 的概念，例如隨機抽 100 個地方的民眾，有無呼吸污染物超標得空氣得來的結果。 1 分：沒提出抽樣的概念，只提到平均 92% 是每 100 人有 92 人吸到超標空氣。 0 分：使用錯誤的概念解讀/無法判斷、只依據自身想法或經驗來判斷 / 不會回答或回答無意義。</p> <p>【受試者答題情況範例】 3 分：P，透過前述數據，推出全球 92% 的人呼吸的空氣污染物有超標(口頭說明：從這 3000 個監測點中，分為超標與沒有超標，將超標處與未超標處居民人數各自算出來，推出 92% 的人) (S09) 2 分：P，指的是所有樣本數中，有還是沒有呼吸道污染物超標的空氣。假設抽 100 人，92% 指的是其中 92 人有吸到污染物超標的空氣。(S10) 1 分：P，92% 的民眾是指約 100 人裡就有 92 人呼吸汙染超標空氣 (S11) 0 分：無</p> | | | | | | | | | | |
| C-1 | <p>【題目】 (問題點 6)</p> <div style="display: flex; justify-content: space-between;"> <div data-bbox="475 1447 587 1473" style="width: 45%;"> <p>台灣現況</p> <p>臺灣的境外空氣污染主要源於中國，但境外污染只佔臺灣空污成因的四成，另外六成污染源都在我們的環境生成。境內污染源包括交通污染源約佔 20%、工業污染源 18% 以及其他零碎的污染源加總約 22%；交通污染源主要為交通運輸排放之廢氣，工業污染源包括火力發電廠、鋼鐵與石化等高污染產業工廠排放。</p> </div> <div data-bbox="919 1447 1321 1473" style="width: 45%;"> <p>我國細懸浮微粒(PM2.5)濃度貢獻源</p> <table border="1" style="margin-top: 10px;"> <caption>我國細懸浮微粒(PM2.5)濃度貢獻源</caption> <thead> <tr> <th>污染源類別</th> <th>百分比</th> </tr> </thead> <tbody> <tr> <td>境外污染源</td> <td>40%</td> </tr> <tr> <td>工業污染源</td> <td>18%</td> </tr> <tr> <td>交通污染源</td> <td>20%</td> </tr> <tr> <td>其他污染源</td> <td>22%</td> </tr> </tbody> </table> <p style="font-size: small; text-align: center;">(圖片來源：行政院環保署106年度空氣污染防制總檢討)</p> </div> </div> <p>由此可知三件事：</p> <p>3. 空氣污染中的工業污染源 18% 比交通污染源 20% 少，也就是工業造成的所有環境污染中空氣污染的比例，也會比交通造成的所有環境污染中空氣污染的比例少。</p> <p>【評分規準】 3 分：針對空汙來源與汙染類別占比兩者提出精確的判斷理由，例如指出兩者分母分子的意涵，或畫出文氏圖解釋兩者的母體差異對</p> | 污染源類別 | 百分比 | 境外污染源 | 40% | 工業污染源 | 18% | 交通污染源 | 20% | 其他污染源 | 22% |
| 污染源類別 | 百分比 | | | | | | | | | | |
| 境外污染源 | 40% | | | | | | | | | | |
| 工業污染源 | 18% | | | | | | | | | | |
| 交通污染源 | 20% | | | | | | | | | | |
| 其他污染源 | 22% | | | | | | | | | | |

| | |
|-----|--|
| | <p>比較統計值大小造成的影響</p> <p>2分：雖然知道兩者意涵，但是不知道怎麼比較，或是認為沒有實際數量無法比較。</p> <p>1分：只知道空汙來源或汙染類別的佔比其一，且無法區辨兩者。</p> <p>0分：使用錯誤的概念解讀/無法判斷、只依據自身想法或經驗來判斷 /不會回答或回答無意義。</p> |
| | <p>【受試者答題情況範例】</p> <p>3分：C-1，兩者分母不同，一個是工業造成的所有環境污染總量，一個是交通所造成的所有環境污染總量，前者若比後者的比值小得多，有可能最終結果是工業的環境污染比例比較少。主要要看總量差異。(S33)</p> <p>2分：C-1，空氣污染裡面的18%跟20%，只是指這兩個污染佔空氣污染的比例，並不知道他們兩個污染的基數是多少，所以會無法判斷對這句話的正確性。(S23)</p> <p>1分：C-1，若工業污染源的總量比較大的話，18%可能就會比較多，而工業污染源的總量若比較少的話，18%可能就會比較少。(S02)</p> <p>0分：A-2，他們的分母不一樣。(S22)</p> |
| C-2 | <p>【題目】 (問題點4)</p> <p>台灣現況</p> <p>臺灣的境外空氣污染主要源於中國，但境外污染只佔臺灣空污成因的四成，另外六成污染源都在我們的環境生成。境內污染源包括交通污染源約佔20%、工業污染源18%以及其他零碎的污染源加總約22%；交通污染源主要為交通運輸排放之廢氣，工業污染源包括火力發電廠、鋼鐵與石化等高污染產業工廠排放。</p> <p>我國細懸浮微粒(PM2.5)濃度貢獻源</p> <p>(圖片來源：行政院環保署106年度空氣污染防制總檢討)</p> <p>由此可知三件事：</p> <p>1. 工業污染源的百分比相較於交通污染源的百分比低1成，且交通污染的百分比相較於其他污染的百分比也是低1成；</p> |
| | <p>【評分規準】</p> <p>3分：針對兩者百分比提出精確的理由，例如列出完整算式，或是寫出比交通低10%，則交通比其他低9.1%或1/11；或是點出各自的基準值是工業18%與其他22%。</p> <p>2分：判斷理由不夠精確充分，或理由中有部分錯誤敘述，例如知道工業比交通低10%，覺得不可能交通比其他也是10%，但是不會列式無法知道正確數字/或是計算錯誤。</p> <p>1分：知道是C-2但是完全無法解釋或不知道怎麼算。</p> <p>0分：使用錯誤的概念解讀/無法判斷、只依據自身想法或經驗來判斷 /不會回答或回答無意義。</p> |
| | <p>【受試者答題情況範例】</p> <p>3分：C-2，交通與其他相差2%，相較於其他污染的百分比指的是$2\%/22\%=1/11$，不是低一成。(S10)</p> <p>2分：C-2，基準值的差異不同(說：這我不太會算，我知道是剛才</p> |

| | |
|-----|--|
| | <p>C-2，我知道他是以交通汙染當基準，和後面是其他汙染當基準)。(S25)</p> <p>1分：C-2。(口頭說明：知道應該是這個但是無法說明)(S17)</p> <p>0分：C-1，這是比較，可能交通本身的量更大，母體大，使得交通汙染的百分比會較其他汙染大。(S08)</p> |
| C-3 | <p>【題目】 (問題點3)除了對生理上的諸多負面影響以外，一份加拿大的研究亦發現空氣汙染與人們的心理狀態有關：<u>空氣汙染越嚴重的地方，居民感覺越不快樂</u>，反之亦然。</p> <p>【評分規準】 3分：反思舉出其他與空氣汙染相關的因素同時在影響快樂程度。 2分：判斷理由不夠充分，或理由中有部分錯誤敘述，例如因素只與其中一項有關，例如只與空氣汙染有關或只與快樂指數有關。 1分：知道是C-3但是完全無法解釋或不知道怎麼算。 0分：使用錯誤的概念解讀/無法判斷、只依據自身想法或經驗來判斷 /不會回答或回答無意義。</p> <p>【受試者答題情況範例】 3分：C-3，混淆變因來說，它可能沒注意到空汙高代表發展也比較高，可能在大都會，居民便不只是因汙染而不快樂，也因都市的競爭與快節奏而壓力大、不快樂。(S307) 2分：C-3，影響情緒的因素有很多，空汙可能是其中一個(口頭說明：不開心是因為減少喜歡的活動，所以工作事情變多，減少喜歡活動影響情緒；工作量變多，使用電子產品或交通會造成空氣汙染)(S28) 1分：C-3。(口頭說明：不快樂可能會有其他原因，有很多變項，但是一時想不到有什麼。)(S04) 0分：無</p> |
| A-1 | <p>【題目】 (問題點8)政府為了全面改善國內空氣汙染，針對上述3種可控制的境內污染源(工業、交通與其他)提出因應對策，提出新的「空氣汙染防制行動方案」，建構更完整的空氣品質管理制度。比較舊有的《空氣汙染防制法》實行四年(2015~2019年)與新的行動方案(2020年)的成果： 全國PM2.5紅色警示次數由2015年997次，降至2019年146次(次數降低851次，降幅85%)；但是新的行動方案推出即在2020年讓警示次數降至只有15次(次數降低131次，降幅90%)；<u>可見新的行動方案降幅比較多，對空氣品質改善比舊的《空氣汙染防制法》成效更良好。</u></p> <p>【評分規準】 3分：提出精確適當的判斷理由，例如提出完整算式作為判斷證明兩者在次數與降幅上皆有不同的優勢。 2分：說明不夠充分精確，例如只覺得都有降低，但是無法意識到次數與降幅的差異。 1分：說明太過含糊，例如只覺得都有優點，但是說不出原因。</p> |

| | |
|-----|--|
| | <p>0分：使用錯誤的概念解讀/無法判斷、只依據自身想法或經驗來判斷 /不會回答或回答無意義。</p> <p>【受試者答題情況範例】</p> <p>3分：A-1，舊有的法制是使得降低的次數大幅下降，新有法治則是提高了降幅比率，兩者各有其優點與效益，沒有絕對的誰比較好與不好。(S29)</p> <p>2分：A-1，前面施行舊有的實行了四年，有括號次數與降幅，次數是絕對的，降幅是相對的，但是不知道怎麼說明相對跟絕對分別各有其優勢 (S04)</p> <p>1分：無</p> <p>0分：無</p> |
| | <p>【題目】 (問題點 5)</p> <p>台灣現況</p> <p>臺灣的境外空氣污染主要源於中國，但境外污染只佔臺灣空污成因的四成，另外六成污染源都在我們的環境生成。境內污染源包括交通污染源約佔20%、工業污染源18%以及其他零碎的污染源加總約22%；交通污染源主要為交通運輸排放之廢氣，工業污染源包括火力發電廠、鋼鐵與石化等高污染產業工廠排放。</p> <p style="text-align: right;">我國細懸浮微粒(PM2.5)濃度貢獻源</p> <p style="text-align: right;">(圖片來源：行政院環保署106年度空氣污染防制總檢討)</p> <p>由此可知三件事：</p> <p style="padding-left: 40px;">2. 空污中的交通污染源只有 20%，<u>也就是交通產生的所有環境污染中，空氣污染也是只有佔 20%</u>；</p> |
| A-2 | <p>【評分規準】</p> <p>3分：能準確說出兩者的比例組成，例如指出兩者分母分子的意涵，或用文氏圖解釋兩者的不同。</p> <p>2分：雖然知道兩者意涵，但是無法具體說明。</p> <p>1分：只知道其中一個比例，且無法區辨兩者。</p> <p>0分：使用錯誤的概念解讀/無法判斷、只依據自身想法或經驗來判斷 /不會回答或回答無意義。</p> <p>【受試者答題情況範例】</p> <p>3分：A-2，空污的交通污染為交通空污/全部空污成因，但不代表交通環污中空氣污染就只佔了 20%，所需調查資料不一樣，數值也會不一樣。(S29)</p> <p>2分：A-2，分母不一樣出來的數值也會不一樣。(S01)</p> <p>1分：A-2，不能如此類推 (口頭說明：有點混亂，覺得可能) (S02)</p> <p>0分：X，空白不會。(S18)</p> |

| | |
|---|--|
| | <p>【題目】</p>  <p>(圖片來源：環保署 空氣品質監測報告)</p> <p>(問題點 7) 圖為環保署空氣品質監測報告，呈現每年 PM2.5 濃度平均值(圖上也標示各年濃度平均值的 95%信賴區間)，PM2.5 濃度自 2007 年起由 34.5µg/m³ 下降至 2017 年只有 20.7µg/m³，<u>可以看出 PM2.5 濃度逐年顯著下降，代表台灣空氣品質確實有在改善。</u></p> |
| R | <p>【評分規準】</p> <p>3 分：指出問題點，並針對顯著增加提出精確的解釋，例如寫出信賴區間誤差線與顯著的關聯性及其意義。</p> <p>2 分：指出問題點，但判斷理由不夠精確充分，或理由中有部分錯誤敘述，例如覺得誤差線有包含到前一年結果，但是不確定與顯著的關聯性。</p> <p>1 分：指出問題點，但無法提出判斷理由或判斷理由不正確</p> <p>0 分：使用錯誤的概念解讀/無法判斷、只依據自身想法或經驗來判斷 /不會回答或回答無意義。</p> |
| | <p>【受試者答題情況範例】</p> <p>3 分：R，不符合逐年，也不符合顯著，以 2007 及 2008 來比較，2008 的平均值仍落在 2007 的信賴區間中，之後幾年亦是如此，顯示其嚴格來說並不能稱為顯著下降。(S07)</p> <p>2 分：R，要注意信賴區間，有可能抽出的值在平均之上或之下，造成的所謂改善可能就有差別。(說:其實我不知道顯著的意義是什麼，但是我確定他有下降的原因是因為 2007 年的信賴區間最低點跟 2017 年的最高點還是有下降) (S08)</p> <p>1 分：無</p> <p>0 分：無</p> |
| E | <p>【題目】</p> <p>(問題點 1) 這項報告是來自八個國際機構 16 名科學家合作，透過衛星與地面測量<u>蒐集全球逾 3,000 個監測站點的空污數據</u>，發現全球 92% 的民眾呼吸的是污染物超標的空氣，尤其亞洲地區污染情況更嚴重。</p> <p>【評分規準】</p> <p>3 分：提出精確適當的判斷理由，例如提出樣本中測站/地點選擇。</p> <p>2 分：說明不夠精確，例如只提出不知道 3000 個站點有誰，但無法解釋樣本選擇對母體產生什麼可能的偏差。</p> <p>1 分：知道是 E 概念，但是完全無法解釋</p> |

| |
|---|
| 0分：使用錯誤的概念解讀/無法判斷、只依據自身想法或經驗來判斷 /不會回答或回答無意義。 |
| 【受試者答題情況範例】 3分：合理、如果抽樣國家沒有客觀地分布在已開發、開發中國家，那數據顯示一定會偏頗某類型發展國家。(S33) 2分：無 1分：無 0分：無 |



附錄六 單因子相依樣本 ANOVA 分析的 LSD 事後比較

附錄六為第四章第 1-1 節「整體前測表現」、第 2-1 節「學習表現」與第 3-1 節「後測表現」中，對「各題型/概念的表現分數差異」做單因子相依樣本 ANOVA 分析的 LSD 事後比較表。

1. 第四章第 1-1 節「整體前測表現」LSD 事後比較的 p 值

| 代碼 | P | C-1 | C-2 | C-3 | A-1 | A-2 | R | E |
|-----|-------|-------|-------|-------|-------|-------|-------|---|
| P | - | | | | | | | |
| C-1 | .135 | - | | | | | | |
| C-2 | .000* | .000* | - | | | | | |
| C-3 | .000* | .000* | .239 | - | | | | |
| A-1 | .763 | .020* | .000* | .000* | - | | | |
| A-2 | .414 | .000* | .000* | .000* | .292 | - | | |
| R | .000* | .000* | .439 | .026 | .000* | .000* | - | |
| E | .096 | .000* | .000* | .000* | .019 | .160 | .000* | - |

註 1：*表示 $p < .05$

註 2：P 機率的解讀；C-1 比例的比較；C-2 基準值差異的比較；C-3 混淆變項；

A-1 比較類型的選擇；A-2 比例類型的選擇；R 隨機性(信賴區間)；E 抽樣偏差

2. 第四章第 2-1 節「學習表現」LSD 事後比較的 p 值

| 代碼 | P | C-1 | C-2 | C-3 | A-1 | A-2 | R | E |
|-----|-------|-------|-------|-------|-------|-------|-------|---|
| P | - | | | | | | | |
| C-1 | .010* | - | | | | | | |
| C-2 | .000* | .002* | - | | | | | |
| C-3 | .000* | .000* | .106 | - | | | | |
| A-1 | .118 | .304 | .000* | .000* | - | | | |
| A-2 | .571 | .000* | .000* | .000* | .006* | - | | |
| R | .000* | .000* | .831 | .059 | .000* | .000* | - | |
| E | .535 | .000* | .000* | .000* | .006* | 1.000 | .000* | - |

註 1：*表示 $p < .05$

註 2：P 機率的解讀；C-1 比例的比較；C-2 基準值差異的比較；C-3 混淆變項；
A-1 比較類型的選擇；A-2 比例類型的選擇；R 隨機性(信賴區間)；E 抽樣偏差

3. 第四章第 3-1 節「後測表現」LSD 事後比較的 p 值

| 代碼 | P | C-1 | C-2 | C-3 | A-1 | A-2 | R | E |
|-----|-------|-------|-------|-------|-------|-------|------|---|
| P | - | | | | | | | |
| C-1 | .640 | - | | | | | | |
| C-2 | .367 | .613 | - | | | | | |
| C-3 | .011* | .007* | .002* | - | | | | |
| A-1 | .066 | .001* | .005* | .563 | - | | | |
| A-2 | .022* | .001* | .000* | .856 | .521 | - | | |
| R | .000* | .000* | .000* | .030* | .056 | .095 | - | |
| E | .000* | .000* | .000* | .002* | .021* | .032* | .160 | - |

註 1：*表示 $p < .05$

註 2：P 機率的解讀；C-1 比例的比較；C-2 基準值差異的比較；C-3 混淆變項；
A-1 比較類型的選擇；A-2 比例類型的選擇；R 隨機性(信賴區間)；E 抽樣偏差