

AI 輔助數學教科書分析之可行性研究

陳冠丞、林勇吉

國立清華大學數理教育研究所

摘要

隨著生成式大型語言模型（LLM）的突破性發展，AI 已成為推動數位教育轉型的新興動能。在當前教育環境下，AI 不僅是輔助學習的媒介，更具備轉化為「教學研究工具」的潛力教科書作為課程實施的核心，其內容設計與問題編排直接影響教與學的品質。然而，傳統內容分析法面臨巨大的人力與時間成本挑戰，導致研究範疇往往受限於特定向度。若能導入 AI 輔助分析，有望突破人力瓶頸，實現更全面、多維度的教材結構觀測。本研究問題在比較「人工分析」與「AI 分析」在教科書評鑑上的一致性。以 Gemini 為 AI 分析工具，針對龍騰版高中數學第一冊「實數」單元進行內容分析。研究採用教科書常見理論架構作為編碼準則：(1) Stein 等人 (2000) 提出的「數學任務認知需求架構」(2)Zhu 與 Fan (2006) 在中美數學教科書比較研究中之分類，針對例行性與非例行性、開放性與封閉性、以及表徵形式等七大維度(3) Lesh 與 Lemon (1992) 之理論架構，將數學問題依其情境化程度區分為「真實情境」、「半情境」與「抽象情境」三種範疇為分析架構。本研究針對實數單元共 104 道題目進行評測，結果顯示 AI 在各向度的分析正確率普遍達 95% 以上，具體在「傳統/非傳統」、「例行/非例行」及「情境分析」三個向度中，AI 與人工分析結果完全一致；在「應用性」、「條件充足性」與「開放/封閉性」僅出現 1 題偏差。研究發現 AI 易忽略「承上題」等語境導致誤判「應用性」，在開放性以及條件充足性判斷上，AI 傾向依據課本後附之解釋進行解讀，而非單純從題目判斷；在「單步驟與多步驟」判別中，AI 與學生分析有兩題相異，推斷原因與判別任務認知需求架構類別有關；在數學表徵型態 AI 的判斷與學生有 5 題的差

異，其主要原因亦為 AI 易忽略「承上題」此字眼，而造成判斷誤差；在「數學任務認知需求架構」中表現最不理想，一致率約為 88%（共 12 題歧異），主因為該架構定義複雜，從不同教學觀點出發易產生詮釋差異。以「 $\sqrt{2}$ 到底是多少呢？」為例，此題 AI 在「條件充足性」、「開放/封閉性」以及「數學任務認知需求架構」皆出現判斷誤差，其原因為 AI 將問題後接續提到的課本內容：

「可以以十分逼近法來判斷」以及後續解法納入考量，因此 AI 認定此題為具連結程序型、封閉型以及條件充足型，然而在本研究的分析架構中，應以學生看到此題時的難度作為判斷基準，在題目中沒有提到要使用何種方法來判斷 $\sqrt{2}$ 的值也沒有提到應該逼近至小數點後幾位，因此此題應認定為作數學型、條件不充足型以及開放型的題目。此外研究發現 AI 容易將「記憶型」題目誤判為「具連結程序型」或「作數學型」，顯示 AI 過於專注於題目本身的文本解讀，

而忽略了課本脈絡及其既有定義。舉例來說，在課後練習觀念澄清「 $\frac{21}{13}$ 與 $\frac{35}{21}$ 之間有無有理數」中，此題為檢驗學生對於「有理數稠密性」的觀念，因此應為記憶型，然而 AI 因未將課本內容全部閱讀完畢，將此題判斷為具連結程序型的題目。綜上所述，AI 在多數分析維度中展現出極高的精準度。本研究建議，未來若能透過「提示工程」優化引導，並提供更多編碼範例進行校準，AI 確實具備成為數學教科書分析與課程研發輔助工具之巨大潛力。

關鍵字:人工智慧、教科書分析、提示詞工程