

以小學教師資格考探索人工智慧工具於師資養成輔助的 潛力－試題認知層次分析

賴碧純 文爾雅 林靜雯*

摘要

為解決複雜的教學問題，需培養教師具備高階認知層次的能力，而日新月異的人工智慧有望成為師資養成的助力。教師資格考試為師資生修習師培課程能力認證的門檻，分析試題中的認知層次分布及人工智慧工具（例：ChatGPT）的答題表現，以探索在教學輔助能力的表現有其必要性。本研究選取 2019 年到 2022 年國民小學教師資格考試中的教育專業科目，即「教育理念與實務」、「學習者發展與適性輔導」和「課程教學與評量」三科相關考科的試題，共 560 題，運用修訂後的 Bloom 認知分類六層次，並再區分為高、低層次對不同考科試題及題型進行分析。此外，研究者探查 ChatGPT 於相關考科的總得分，並以盒鬚圖比較 ChatGPT 在不同科目和題型中高、低認知層次試題得分比的分布情形。結果顯示，教師資格考試的教育專業科目中，三科考科的低認知層次試題占分比都高於高認知層次試題。符應素養導向評量要求的綜合題，高認知層次占分比較高，達 81.67%，但選擇題的高認知層次占分比僅達 6.55%，有提升高認知層次比例的必要。此外，ChatGPT 在教育專業科目平均總分為 72.73 分，成功通過教師資格考試平均 60 分的門檻，但對於高認知層次試題的表現歧異度大。這表示 ChatGPT 具有成為師資養成輔助工具的潛力，但教育工作者仍需特別注意高層次認知情境的評鑑與批判。

關鍵詞：人工智慧、師資培育、教師資格考試、聊天生成預訓練轉換器

賴碧純，國立台北教育大學 自然科學教育學系 博士生

文爾雅，國立台北教育大學 自然科學教育學系 博士生

* 林靜雯，國立台北教育大學 自然科學教育學系 教授（本文通訊作者）

電子信箱：jwlin@mail.ntue.edu.tw

來稿日期：2023 年 10 月 1 日；修稿日期：2023 年 12 月 14 日；採用日期：2024 年 1 月 15 日

Exploring the Potential of Artificial Intelligence Tools in Teacher Preparation Based on the Elementary School Teacher Qualification Examination: Analyzing the Cognitive Level of Examination Questions

Bi-Chun Lai Erh-Ya Wen * Jing-Wen Lin

Abstract

Training teachers to develop higher cognitive levels to address complex teaching challenges is necessary. Artificial intelligence (e.g., ChatGPT) holds the potential to assist in teacher education. Teacher qualification examination serves as a threshold for certifying pre-service teachers' abilities in pursuing teacher education curricula. Analyzing the distribution of cognitive levels in examination questions and assessing the performance of artificial intelligence tools in answering these questions is necessary to explore their potential in aiding teaching assistance. This study selected 560 education professional examination questions from the elementary school teacher qualification examination between 2019 and 2022. We used the revised Bloom's cognitive taxonomy to analyze and categorize questions into high and low cognitive levels. We investigated the ChatGPT scores in the relevant subject areas and compared the differences in score distribution between high and low cognitive levels across different subjects and question types using box plots. The results indicate that the percentage of low-order questions in all three subjects was higher than that of high-order questions in the teacher qualification examination of education professional subjects. The proportion of high-order comprehensive questions has increased to 81.67 % to meet the competence-oriented assessment requirements. The percentage for multiple-choice questions remains relatively low at only 6.55 %. This indicates a need to increase the proportion of high-order questions. Furthermore, ChatGPT achieved an average total score of 72.73 points in the education professional subjects, surpassing the threshold of an average score of 60 points for the teacher qualification examination. However, there was significant variability in its performance on high-order questions. This suggests that ChatGPT can potentially become an assistive tool for teacher education. However, educators must consider developing critical thinking skills in high-order cognitive contexts when using such tools.

Keywords: Artificial Intelligence, Teacher Education, Teacher Qualification Examinations, ChatGPT

Bi-Chun Lai Ph. D., Department of Science Education, National Taipei University of Education
Erh-ya Wen Ph. D., Department of Science Education, National Taipei University of Education
* Jing-Wen Lin Professor, Department of Science Education, National Taipei University of Education (Corresponding author)

E-mail: jwlin@mail.ntue.edu.tw

Manuscript received: October 1, 2023; Modified: December 14, 2023; Accepted: January 15, 2024

壹、緒論

近年來，人工智慧 (Artificial Intelligence, AI) 技術的迅速發展已引起廣泛的關注和熱潮。其中，由 OpenAI 所開發的聊天生成預訓練轉換器 (Chat Generative Pre-trained Transformer, ChatGPT) 更在學術界和大眾間引起極大的關注。ChatGPT 是一種深度學習的語言模型，能透過自然語言交互生成有意義的回答，其強大的語言理解和生成能力使其在各個領域都具有廣泛的應用潛力。目前，ChatGPT 具有 GPT-3.5 與 GPT-4 兩種引擎，其中 GPT-3.5 因免費、使用門檻低且能夠在沒有額外訓練的情況完成不同任務等優勢下，於 2022 年甫推出，短短 5 天就達一百萬名用戶。ChatGPT 可說是目前最具智慧的機器之一，但其產生內容中間的過程具有黑盒子特性，且其在原創性與隱私安全上仍是大問題，因此研究 ChatGPT 可以幫助我們探索 AI 對教育的潛在影響 (Zhai, 2022)。據此，本研究以 ChatGPT 中廣為大眾使用的免費版本 GPT-3.5 為人工智慧代表進行研究。

教育是社會進步和個人發展的關鍵，而 ChatGPT 的出現和應用無疑對教育領域帶來了新的影響和挑戰，並引發對其在教學模式和人機協作中所扮演角色的不同看法。一些學者認為，ChatGPT 能為學生提供即時的問答和解答支持，學生可以透過與 ChatGPT 對話獲得個別化的學習體驗和教育指導，這有可能取代傳統教室中教師的角色。教師可以使用 AI 工具減輕工作量，例如：開發自動評估系統 (Zhai et al., 2020)，獲

取有關學生的見解並促進教學創新 (Baker et al., 2019)。學生也應該學會使用 AI 工具完成學習任務 (Zhai, 2022)。因此這些學者致力於強化人工智慧工具的效能以取代人類。

另一些學者則強調：自 20 世紀初以來，出現許多被譽為學習和教學的創新技術，但我們常常忽略教育者如何實施這些技術以及學生如何與之互動。ChatGPT 在市場上是一個相對較新的工具，現階段缺乏足夠的實證研究來確定它對教育的影響 (Rudolph et al., 2023)，且 GPT-3.5 所使用的資料庫僅限於 2021 年 9 月以前，對於事實的準確性存在一定的限制。AI 可以協助教育工作者更即時地判斷學生的學習需求，但無法取代教育者角色 (Zhai, 2022)。當教師與學生過度依賴 ChatGPT，也可能造成創造力、批判性思考、推理和解決問題等高層次能力的降低 (Farrokhnia et al., 2023)。因此，這些學者主張人類與人工智慧各有優勢與弱勢的領域，應建立適當的人機協作模式 (Kubsch et al., 2023)。這兩種觀點對教育都有貢獻，但本研究更傾向於後者的觀點。

一名教師的養成始於師資培育教育，並透過教學實踐和持續專業發展的歷程不斷成長。檢核師資生是否能成功完成師培教育，並預測其是否有潛力擔任教師有其必要性 (Metsäpelto et al., 2021)。在臺灣，「教師資格考試」可以提供教師專業標準中專業知能面向的檢核依據 (林政逸、賴慧君，2021)，根據現行制度，修畢師資職前教育課程並取得修畢證書的候選人可以參加教

師資格考試。教師資格考試的教育專業科目包含「教育理念與實務」、「學習者發展與適性輔導」、「課程教學與評量」相關的三個科目，考題的內容與題型根據考科評量架構和學科內容的差異而有所不同。自 2021 年起，因應 12 年國民基本教育課程綱要，教師資格考試題型除原訂的選擇題與問答題外，新增綜合題，並加入了素養導向的命題（「教師資格考試辦法」2020 年，3 月 25 日），更加強調師資生以跨領域核心素養在真實情境中應用、分析等相關能力解決問題（任宗浩，2017）。Haataja 等人（2023）根據 Anderson 與 Krathwohl（2001）提出的教學目標分類法修訂版，將這些能力歸屬於高層次認知能力（亦即：應用、分析、評鑑和創作）。

師資培育教育的目標包括短期和長期兩個層面。短期目標致力於協助師資生順利通過教師資格考試，提供相應的專業支援，以確保其具備應對考試挑戰的能力。長期目標則聚焦於培養師資生的高層次認知能力，激發其應用知識、深度思考和解決問題，以期在未來教育現場中展現專業的學科教學知識。教學被視為一門高深的藝術，而人工智慧則為教育帶來新的機遇。因此，在不同的教學目標和認知層次下，我們需建立相應的人機協作模式。

在這樣的背景下，本研究旨在檢視目前教師資格考試中不同科目和題型試題下的認知分類和層次分布情形，並評估 ChatGPT 在與師資生相同條件下，亦即，不提供額外提示（prompt）及說明下，是否能通過國民小學國民小學教師資格考試？以及，其

在不同的認知層次中的表現有何差異？希冀據此為未來師資養成領域的人機協作提供建議。綜上所述，本研究針對國民小學教師資格考試科目中的專業教育考科（以下以「國民小學教師資格考試」指稱本研究之國民小學教師資格考試中的三科專業教育考科），即為「教育理念與實務」、「學習者發展與適性輔導」、「課程教學與評量」相關的考科；及國民小學教師資格考試題型，包含選擇題、問答題與綜合題三種題型，提出以下研究問題：

一、國民小學教師資格考試中

（一）不同科目下的認知分類和高層次（亦即：應用、分析、評鑑和創作）、低層次（亦即：記憶、理解）試題分數占總分比例（簡稱占分比）為何？

（二）不同題型下的認知分類和高、低層次試題占分比為何？

二、ChatGPT 在國民小學教師資格考試中，不同科目的得分情形為何？

三、ChatGPT 在國民小學教師資格考試中

（一）對於不同科目下的認知分類和高、低層次試題得分占總分比例（簡稱得分比）為何？

（二）對於不同題型下的認知分類和高、低層次試題得分比為何？

貳、文獻探討

一、人工智慧在教育領域的應用

人工智慧在教育領域的應用正逐漸引起廣泛關注和重視。從 60 年代起為適應學習者需求開始使用電腦於個性化教學上，到現今 AI 技術憑藉其強大的計算能力和演算法，為教育帶來了許多創新和轉變的可能性，並深入到許多層面，包含個性化學習、自動化管理任務、評量評估系統等，例如智慧輔導系統可模擬導師協助學習者面對問題時提供個別化幫助 (Rudolph et al., 2023)、AI 學習評測系統可以將持續性的回饋納入學習過程中，為傳統評測開闢一條新的道路 (Cope et al., 2021)、AI 輔助寫作的形成性回饋與評估可以加強學生學習外語的自主性，提高學生學習的自我效能 (Nazari et al., 2021) 等。

ChatGPT 是由 OpenAI 開發的一種基於 GPT-3.5 架構的自然語言處理語言模型。它是 GPT (Generative Pre-trained Transformer) 系列的一部分，繼承了 GPT-3 的特性並進一步進化。GPT-3 擁有強大的自然語言處理能力，能夠生成高度連貫且有邏輯的文章、回答問題、提供專業知識等，基於 GPT-3 的成功和挑戰，ChatGPT (GPT-3.5) 在 GPT-3 的基礎上進行了一系列的優化和調整，提供更具互動性和對話能力的應用體驗，能產生即時與個人化回覆，且具備自我提昇能力 (Farrokhnia et al., 2023)。於 2022 年推出至今全球使用人數已超過上億，並在多個領域引起廣泛應用與討論。以下本文未

特別說明版本時，ChatGPT 即為 GPT-3.5。

在醫學領域，ChatGPT 成功通過美國醫學執照考試 (United States Medical Licensing Examination)，它能夠提供具有高度一致性的解答和具有洞察力的見解，為醫學生提供學習資源 (Kung et al., 2023)。在法律領域，ChatGPT 在沒有額外訓練下，通過全美律師資格考試委員會 (National Conference of Bar Examiners) 制定的律師資格考試 (Bar Exam) 中的兩項類別，顯示 ChatGPT 對法律有普遍的理解 (Bommarito II & Katz, 2022)。在教育領域，ChatGPT 在美國國家教育進步評估 (National Assessment of Educational Progress) 科學項目中表現優於大部分作答學生 (Zhai et al., 2023)，在美國大學入學考試中 (Scholastic Assessment Test) ChatGPT 表現優於 70 % 以上的考生。同時，ChatGPT 能通過高等教育中初級與中級程式編碼測驗，但對於一系列推理步驟的活動則較難以勝任 (Savelka et al., 2023)。

ChatGPT 可以為各階段的學習者提供個性化學習機會並提供資訊的可近性，然而 ChatGPT 仍有許多不確定性，包含缺乏對文本的深刻理解，難以評估答案的品質、存在歧視和偏見的風險且缺乏高階思維能力 (Farrokhnia et al., 2023)，我們在使用時仍須注意 ChatGPT 缺乏可解釋性與可能存在的偏見 (Kasneji et al., 2023)。ChatGPT 已證明在協助教師上能促進更多的創新教學和學習，但在學術界仍引起許多焦慮 (Rudolph et al., 2023)。不同的 AI 工具，有著

不同的優點與侷限，作為教育工作者，我們需要瞭解人工智慧的內部侷限性，當我們與人工智慧建立協作關係時，不可忘記展現人類的思想與行動 (Cope et al., 2021)，應找出各自擅長領域，協調出最恰當的人機協作模式 (Kubsch et al., 2023)。

綜上所述，ChatGPT (GPT-3.5) 具有免費使用、強大的生成和語言理解能力等優點，但在教育領域的應用仍需進一步探索和完善，且就我們蒐集資料庫後，尚未發現與教師資格考試相關的研究。因此，本研究旨在透過評估 ChatGPT 在教師資格考試中的表現，來探索其在師資培育領域人機協作的可行性。

二、師資養成和教師資格考試的沿革

教育部高級中等以下學校及幼兒園教師資格考試是教育部在師資培育制度上的一大變革。若要擔任國民中小學及幼兒園正式教師，必須在修畢各類科師資職前教育之後，通過教師資格考試取得教師證書，以示修畢並獲得完整師資培育課程的知能（「教師資格考試辦法」2020年，3月25日）。

自2005年啟動的教師資格考試至2020年6月，課程規劃以科目和學分來計算，考試以傳統學科知識為主，包含數學能力與國語能力兩科共同科目，以及三科教育相關考科。2019年，教師資格考試實施新制，由「先實習再考試」轉為《師培法》修正後的「先考試後實習」檢定制度，避免未通過考試者在實習中浪費時間，因此2019

年有兩次考試。

而後，素養導向的教學與評量成為國際趨勢，2019年開始實施的十二年國民基本教育課程綱要，由過去知識能力進一步強化為素養，透過脈絡化情境，教師引導學生使用策略，進行思考和討論，自主地完成學習任務，以促進學生素養的發展（吳璧純，2017）。換言之，若定義素養導向之評量，其「佈題強調真實的情境與真實的問題，評量強調總綱核心素養或領域／科目核心素養、學科本質及學習重點」（任宗浩，2018，頁76）。因此，師資培育課程和教師資格考試皆隨之調整以符合這一趨勢。

教育部師資培育司於2021年6月的教師資格考試，根據《高級中等以下學校及幼兒園教師資格考試命題作業要點》（2020）的評量架構，修改了教育專業科目的名稱、內容和題型。2019年和2020年的教師資格考試對應的是九年一貫課程綱要，而2021年和2022年對應十二年國教課程綱要，由於對應的課程綱要不同，因此考試科目名稱也略有不同。「教育理念與實務」於2019和2020年時原為「教育原理與制度」（統一簡稱教育理念），「學習者發展與適性輔導」則為「兒童發展與輔導」（統一簡稱兒童發展），「課程教學與評量」則為「國民小學課程與教學」（統一簡稱教學與評量）。並新增了「綜合題型」，形式不受限制，包括選擇和問答，並加入情境內容描述，強調教學現場問題解決能力的應用，著重對應學習經驗的跨領域整合（曾建銘，2020）。其中，綜合題中的問答部分更是評量考生是否能針對問題情境提出獨立的論述和見解（「教師資格考試命題

作業要點」2020 年，11 月 13 日)。

基於 12 年國教課程基準，教育部編定教師資格檢定的素養評量指標(表 1) 要求師資生了解基本核心知識、學習重點與技能後，應用於教育現場情境。這些要求突顯素養導向評量應盡可能具有真實日常生活或學術探究的情境脈絡，並強調總綱或領域的核心素養及學習重點。亦即，特別強調於多元、真實的「情境」中運用應用、分析等「高認知層次」等相關能力，減少機械式記憶與計算練習的題目(任宗浩，2017)。

三、教師資格檢定與認知層次分類

Anderson 與 Krathwohl (2001) 考量更廣泛的教學因素，因此修訂 Bloom (1956) 提出的教學目標分類法，將認知層次從簡單到複雜依次劃分為記憶、理解、應用、分析、評鑑和創作。這一修訂後的分類法已被廣泛應用於各種課程、教育目標和評量中 (Baghaei et al., 2020; Haataja et al., 2023)。我國的基本學力測驗也曾參考 Bloom 修訂版本進行分析(李坤崇，2004)。原版與修訂版的認知層次都具有分層的概念，從基礎到高階，需要先掌握基礎認知層次才能達到高階認知能力。其中，基礎認知層次(簡稱低層次)包括記憶和理解，而高階認知層次(簡稱高

表 1
教育類專業科目評量指標

科目	評量指標
教育理念與實務	<ol style="list-style-type: none"> 1. 了解主要教育思潮和理論，並應用於各級學校教育。 2. 了解並應用教育內外在社會環境與社會脈絡，包括教育與社會公平、社會變遷等巨觀和微觀層面。 3. 了解我國主要教育體制、法規、政策、學校實務和教育改革趨勢，並應用於國民小學教育情境。 4. 了解教師專業倫理與實踐之內涵，並應用於國民小學。
學習者發展與適性輔導	<ol style="list-style-type: none"> 1. 理解兒童身心發展理論，考慮背景對發展的影響，並在教學中應用。 2. 了解學習和動機理論，以促進學生的學習，包括學習策略的應用。 3. 辨識特殊需求學生的身心特質，提供支持並進行適當的轉介。 4. 應用班級經營原理和正向支持方法，培養學生自律，促進親師生關係，並創造友善學習環境。 5. 了解輔導理論、技巧、機制、倫理和法規，並應用於協助學生適應與發展。
課程教學與評量	<ol style="list-style-type: none"> 1. 了解課程主要原理、發展與設計、評鑑原則，應用於國小課程、教學及評量。 2. 考量社會變遷、課程政策及改革趨勢，發展國小課程、教學及評量。 3. 了解教學主要理論與模式，以設計、實施、改善國小教學。 4. 了解多元教學方法、學習科技與資源，應用於國小教學規劃與實踐。 5. 了解與使用多元評量方法(包括科技)，檢視學習進展，促進學生自我成長，指引與調整教學。

資料來源：「教師資格考試命題作業要點」2020 年，11 月 13 日

層次)則包括應用、分析、評鑑和創作,值得注意的是評鑑與創作層次無法在選擇題中進行評估 (Haataja et al., 2023)。

教學極度複雜,需要透過教師培育體系和職業生涯中的正式和非正式學習,來發展專業知識,對於想要成為一名教師的人來說,掌握高層次能力是至關重要的 (Metsäpelto et al., 2021),由於教師職業的複雜性,師資生更需要擁有處理訊息的高層次能力。當評選師資生的考試中低層次試題佔較大比例時,學生有可能無需使用高層次認知能力就可以高分過關 (Haataja et al., 2023)。此外,關於教師資格檢定及甄選這方面的研究一直很匱乏 (Klassen & Kim, 2019)。因此,為了能更加瞭解教師資格考試試題的特性,本研究採用修訂後的 Bloom 認知層次分類法對試題進行分析,並進一步比較試題中高、低層次分布,瞭解教師資格考試試題

中具有高層次試題的占分比,期望能提供學術界和教育實踐領域對於教師資格考試試題設計和評估的重要參考。

參、研究方法

一、分析對象

教育部國民小學師資類科的教師資格考試共有五科,本研究扣除國語文能力與數學能力等共同考科,聚焦於國民小學教育專業考科,挑選近四年(2019年至2022年)共五次教師資格考試,教育理念、兒童發展及教學與評量相關考科為分析對象。

本研究從教育部高級中等以下學校及幼兒園教師資格考試網站 (<https://reurl.cc/b9QrEl>) 所公開的歷屆試題資料中,選取了2019年到2022年間的三個考科的考試題

表 2
各年度試題題型與配分的分布情形

科目	年度	2019-1、2019-2、2020		2021		2022	
	題型	題數	配分	題數	配分	題數	配分
教育理念	選擇	40	60	25	50	25	50
	問答	4	40	3	30	3	30
	綜合	-	-	4	20	6	20
兒童發展	選擇	30	60	25	50	25	50
	問答	4	40	3	30	3	30
	綜合	-	-	5	20	7	20
教學與評量	選擇	30	60	25	50	25	50
	問答	4	40	3	30	3	30
		-	-	5	20	5	20

目。其中，2019 年因應師培法修訂加考一次，本研究以 2019-1 與 2019-2 區分，共計五次考試的題目。這些選取的題目包括 450 題選擇題、54 題問答題和 32 題綜合題，共計 536 道題目，其中問答與綜合題可以包含多個小題，總計 560 小題。各年度試題題型與配分的分布情況如表 2 所示：

二、資料分析

(一) 試題認知分類與層次的編碼

針對研究問題一，本研究採用修訂後的 Bloom 認知分類 (Wilson, 2016) 中的六個認知層次進行試題分類。然而，許多研究指出不同認知層次之間存在著模糊的界線，因此區分時往往難以完全明確 (Baghaei et al., 2020)。為了確保分析內容的可靠性，減少語言對分析造成的影響，除參考修訂後 Bloom 認知分類的原典外，三位研究者另參考李坤崇 (2004) 使用 Bloom 認知分類修訂版解析九年一貫課綱中的命題實例，以 2019 年第一次教師資格考試中兒童發展科目的 30 題試題進行認知分類訓練。

研究團隊成員包括三位研究者，兩位研究者皆為國小專任教師，多次參加素養試題

培訓工作坊，現為教育相關系所博士生；第三位研究者 (通訊作者) 為系所教授。

在試題認知分類的訓練階段，研究者們進行了編碼一致性的評估 (認知層次判斷相同題數 / (判斷相同題數 + 判斷不同題數) x 100%)。雙方按照分類標準進行試題分類，逐題討論確保一致性達到 80% 以上。隨後，第一和第二研究者獨立進行所有試題的認知分類，再共同比對分類結果。針對分類層次不一致的試題進行討論。在討論後若仍無法達成共識，由第三位作者進行最終判斷。

在教育理念、兒童發展及教學與評量三個科目中，所有 560 題試題的認知層次分類一致性達 98.9%，只有 6 題試題無法達成共識。最後，所有研究者共同進行討論和最終的分類，以確保每道題目的認知層次分類達到高度一致性。表 3 為本研究試題認知層次分析範例，研究者分析每一道試題的評量目的，並根據六個認知層次分類進行編碼。

根據《高級中等以下學校及幼兒園教師資格考試命題作業要點》(2020) 的規定，本研究的試卷包含選擇題、問答題和綜合題，各應試科目以一百分為滿分。選擇題屬於四選一的單選題，答錯不倒扣。自 2021 年起，綜合題被納入考試內容，綜合題可能包

表 3

國民小學教師資格考試試題的認知層次編碼範例

試題範例	說明	編碼
下列何者是幫學生「搭鷹架」(scaffolding) 的做法？(A) 示範舉例 (B) 直接教學 (C) 設定目標 (D) 實施安置性測驗 (出自 2021 課程與教學評量，選擇題第 5 題)	指出搭鷹架的方法	記憶

試題範例	說明	編碼
<p>下列何者在認知領域目標中的層次最高？(A) 表達對垃圾危害海洋環境的感受 (B) 提出一個海洋垃圾減量的解決方案 (C) 說明海洋廢棄垃圾的來源與所造成的後果 (D) 用電腦軟體比較不同種類的海洋垃圾數量</p> <p>(出自 2021 課程教學與評論，選擇題第 2 題)</p>	比較選項中認知層次的差異	理解
<p>十歲的小美因父母離婚心情沮喪而接受輔導，小美的媽媽希望輔導老師能在不告知小美的情況下，讓她知道小美輔導的內容。根據輔導專業倫理原則，輔導老師該如何做最恰當？(A) 為了個案最大的權益，所以告知小美媽媽輔導內容 (B) 因為個案未成年，輔導老師有義務告知家長輔導內容 (C) 為了尊重個案的權利，應先徵求小美的同意，再跟小美媽媽說明 (D) 為了對個案談話內容保密，可避重就輕、簡略的向小美媽媽說明</p> <p>(出自 2019-2 兒童發展與輔導，選擇題第 14 題)</p>	運用已知解決情境中的問題	應用
<p>臺南市的王老師自創「王氏遊戲教學法」，一堂五年級數學的統計圖表課，他用古蹟 探查的方式進行。王老師先讓學生分組規劃古蹟探查路線，學生就交通工具、費用等 議題討論後上臺發表，再表決選出最佳路線。接著王老師帶著學生實地訪查，並畫出各區古蹟的統計圖表。除此之外，王老師讓學生自行討論規劃射飛鏢遊戲，將結果做成圖表。王老師教學設計的核心理念與哪些學習理論的觀點相符應？甲、建構論 乙、行為主義 丙、情境學習理論 丁、意義學習論 (A) 甲乙 (B) 乙丙 (C) 丙丁 (D) 甲丙</p> <p>(出自 2020 教育原理與制度，選擇題第 8 題)</p>	從文本中區別出與理論相關的線索進行判斷	分析
<p>新屋國小教師在生活課程的共備中進行同課異構，由兩位教師就同一主題進行兩種不同的教學設計。你認為以下甲、乙兩份教案大綱，何者較接近素養導向教學的精神？請說明理由。</p> <p>(出自 2019-2 國民國民小學課程與教學，問答題第 4 題)</p>	根據素養導向教學的精神檢視教學設計進行判斷	評鑑
<p>丁老師決定要設計一個體驗學習的課程，讓學生能夠欣賞、重視，進而推廣泰雅族的音樂文化。目前丁老師已經設計了體驗學習課程中第一階段「具體經驗」的學習任務：讓學生體驗不同慶典上的泰雅音樂。請接續幫他規劃「觀察與反思」、「形成抽象化的概念」以及「實踐與應用」三個階段的學習任務。</p> <p>(出自 2021 課程教學與評論，綜合題第 8 題)</p>	依據學習階段設計任務	創作

含選擇題、是非題、配合題和問答題。問答題和綜合題可以包含多個小題，每道題目的配分可能不同。

研究者按照修訂後的 Bloom 認知分類，將試卷中的每一小題進行分類，總計共 560 小題。為了呈現不同認知層次在整份試卷中的配分比例，研究者以「占分比」的方式表示。具體而言，占分比指的是某認知層次占該類別總分的比例。以 2022 年教學與評量考科的綜合題為例，該次綜合題共 5 題，配分依序為 2 分、8 分、2 分、2 分和 6 分，總共 20 分。其中，第一和第三題屬於分析層次，共占 4 分，因此，此綜合題試題的分析層次占分比為 20 %。

為了進一步進行試題分析，本研究確認所有題目的認知層次分類後，再依循 Haataja 等人(2023)提出的高、低層次框架分類所有試題，認知分類中記憶與理解屬於低層次，應用、分析、評鑑與創作屬於高層次。這樣的分類方式讓我們能夠同時觀察六大認知分類以及高、低層次之間的差異。

(二) ChatGPT 在教師資格考試科目中的得分

針對研究問題二，研究者將教育部高級中等以下學校及幼兒園教師資格考試網站公開的試題提供給 ChatGPT 作答(測驗時間：2023 年 4 月)。為了確保試題完整性，對於 ChatGPT 無法處理的圖像題(其中包括選擇題 1 題和綜合題 1 大題，共 2 題)，研究者將其轉換為文字描述，強制 ChatGPT 進行回答。例如，2021 年度課程

教學與評量第 23 題選擇題：

某國小國語期末考卷出了下面這道題目，結果某生的回答如下圖 1。

考後老師們針對部分學生誤解「相反字」的意思，提出四種修改題目的建議，哪

圖 1
某生回答情形



一項較不恰當？

因題目出現 ChatGPT 無法處理的圖像題，研究者改用文字描述如下：「某國小國語期末考卷出了寫出相反字這道題目，結果某生的回答將相反字寫成上下左右顛倒的文字，考後老師們針對部分學生誤解「相反字」的意思，提出四種修改題目的建議，哪一項較不恰當？」

此外，參考 Kung 等人(2023)檢視 ChatGPT 的表現是否能通過美國醫師執照考試的研究方法，以減少因記憶效應而產生的偏差，研究者在進行分析時會在對話中開啟新的 ChatGPT 對話框，以避免 ChatGPT

在分析不同年度和科目的試題時混淆。

在選擇題部分，研究者將原始題目和選項逐字複製並輸入到 ChatGPT 中，記錄回覆的選項，並根據公開的參考答案直接判斷 ChatGPT 生成的回答是否正確。對於問答題和綜合題部分，研究者逐字複製原始題目並要求 ChatGPT 回答文字答案。在評分方面，研究者從歷年來為本校批改教師資格考試的委員名單中邀請相關領域專家，接著提供兩份答案給領域專家，一份是由 ChatGPT 生成的答案，另一份是坊間教師資格考試試題解析的正確答案。為了減少主觀偏見，研究者不明示哪份答案是 ChatGPT 生成的，並要求領域專家對兩份答案進行獨立評分，以確保評分的客觀性和公正性。最後，分別統計 ChatGPT 在國民小學教師資格考試中三項專業教育科目總得分，以呈現 2019 年到 2022 年 15 份試卷中 ChatGPT 的表現。

(三) ChatGPT 在不同認知層次中的得分比與分布情形

針對研究問題三，根據試題的認知層次分類和高、低層次分布，統計 ChatGPT 在國民小學教師資格考試中各項得分，並依照各題目配分，計算 ChatGPT 在不同認知

層次得分占不同類別總分的比例，研究者以「得分比」表示。具體而言，得分比指的是 ChatGPT 在某認知層次的得分占該類別總分的比例。以 2022 年教學與評量考科綜合題為例，該次綜合題總分 20 分，應用層次試題配分 8 分，ChatGPT 得分 6 分，此綜合題試題的應用層次得分比為 75 %。

接著，我們使用盒鬚圖進一步了解 ChatGPT 在國民小學教師資格考試中不同科目和題型下的得分比分布情形。然而由於教師資格考試中不同題型中缺乏某些認知層次的試題，樣本無法齊全，因此我們僅針對 2019 年到 2022 年，ChatGPT 在三個科目和三種題型下高、低層次的得分比繪製盒鬚圖。

一種題型的樣本數包括 3 個科目、4 個年度以及高、低層次，共計 24 個樣本。然而，對於綜合題，由於僅在新制度實施後的 2021 和 2022 年有相關數據，所以僅有 12 個樣本可供分析。此外，如果在某一年度的某一科目或類型下完全沒有高或低層次的試題，這會導致無法比較 ChatGPT 在高、低層次之間的差異。在這種情況下，相應的年度、科目和類型的樣本數也將被扣除，因此樣本數會相應減少，如表 4。藉由資料視覺化，觀察盒鬚圖中的中位數、四分位數、

表 4
ChatGPT 分別在不同科目與題型中，高、低層次的樣本數

分類	科目			題型		
名稱	教育理念	兒童發展	教學與評量	選擇題	問答題	綜合題
樣本數	16	20	14	20	20	10

最小值、最大值和極端值等統計特徵，以更深入地了解 ChatGPT 在不同科目和題型下高、低層次試題的得分比差異。

肆、研究結果

一、國民小學教師資格考試中，不同科目和題型的認知分類和高、低層次試題占分比

(一) 不同科目下的認知分類和高、低層次試題占分比

表 5 顯示不同科目試題認知分類與高、低層次的占分比。研究結果顯示，在 2019 年至 2022 年的教師資格考試中，三考科不同認知分類試題最高的占分比皆為理解層次，分別為 56.60 %、59.20 % 和 45.20 %，平均占分比達 53.67 %。然而，在教育理念與兒童發展方面，並沒有任何創作層次的試題出現，評鑑層次的試題占分比最低，分別為 3.40 % 和 3.20 %。此外，教學與評量試題雖然涵蓋六項認知分類，但創作層次的試題僅占 8.80 %，而評鑑層次的試題占分比仍然是最低，僅為 4.00 %。這表明三個考科中高達 93.54 % 的試題占分比集中在記憶、

理解、應用與分析這四個認知層次。

另外，值得注意的是，研究結果顯示三個考科的低層次試題占分比都高於高層次試題。教育理念與兒童發展低層次試題占分比皆大於 73 %。較為特別的是教學與評量，高 (47.20 %)、低層次 (52.80 %) 占分比差距最為接近。

(二) 不同題型下的認知分類和高、低層次試題占分比

根據表 6 研究結果顯示，在 2019 年至 2022 年的教師資格考試中，三種題型的認知分類試題占分比有所不同。選擇題中占分比最高的認知層次是理解層次，占 71.31 %；問答題中占分比最高的認知層次是應用層次，占 34.81 %；而綜合題中占分比最高的認知層次是分析層次，占 36.67 %。

值得注意的是，選擇題囿限於答題型式，完全無法展現評鑑和創作的層次，而問答題與綜合題試題認知分類涵蓋六大層次。其中，問答題的記憶與創作層次兩者的占分比均為最低，皆為 5.19 %，高達 89.62 % 的問答題集中在理解、應用、分析與評鑑四大認知層次；綜合題中記憶層次占分比 (3.33 %)

表 5
不同科目試題認知分類與高、低層次的占分比 (%)

科目	記憶	理解	低層次	應用	分析	評鑑	創作	高層次
教育理念	21.60	56.60	78.20	7.40	11.40	3.40	0.00	22.20
兒童發展	14.40	59.20	73.60	18.40	4.80	3.20	0.00	26.40
教學與評量	7.60	45.20	52.80	22.40	12.00	4.00	8.80	47.20
平均	14.53	53.67	68.20	16.07	9.4	3.53	2.93	31.93

表 6
不同題型試題認知分類與高、低層次的占分比 (%)

題型	記憶	理解	低層次	應用	分析	評鑑	創作	高層次
選擇題	22.14	71.31	93.45	3.87	2.68	0.00	0.00	6.55
問答題	5.19	34.63	39.81	34.81	13.70	6.48	5.19	60.19
綜合題	3.33	15.00	18.33	16.67	36.67	15.00	13.33	81.67

最低，其次是創作層次 (13.33 %)，達 83.40 % 的綜合題同樣集中在四大認知層次，這表明兩種題型的試題認知層次多集中在理解、應用、分析與評鑑。

另外，三種題型在高、低層次的占分比情形也有所不同。選擇題的高層次試題占分比僅 6.55 %，問答題與綜合題高層次試題占分比皆高於低層次試題。其中，綜合題中高層次試題的占分比達到 81.67 %。

二、ChatGPT 在國民小學教師資格考試中，不同科目的得分情形為何？

根據表 7 研究結果顯示，ChatGPT 在 2019 年至 2022 年共五次的教師資格考試中，三科平均得分為 72.73 分，五次考試三科平均得分皆高於 60 分，其中 ChatGPT 在 2022 年的教師資格考試表現最差，三科平

均得分 69.67 分。

比較 ChatGPT 在三個考科得分情形，ChatGPT 在兒童發展的表現最好，平均得分 75.2 分，在教學與評量的表現最差，平均僅得 70.6 分。

三、ChatGPT 在國民小學教師資格考試中，對於不同科目和題型下的認知分類和高、低層次試題得分比

(一) ChatGPT 對於不同科目下的認知分類和高、低層次試題得分比

根據表 8 比較 ChatGPT 在三個考科中不同認知分類的得分比，發現 ChatGPT 在教育理念的分析層次表現最好，得分比達 87.22 %，在評鑑層次表現最差，得分比僅為 15.15 %。而在兒童發展科目上，ChatGPT

表 7
ChatGPT 在不同科目下，各年度的得分情形 (分)

科目	2019-1	2019-2	2020	2021	2022	平均
教育理念	74	68	68	79	73	72.40
兒童發展	75	77	88	66	70	75.20
教學與評量	71	78	67	71	66	70.60
平均	73.33	74.33	74.33	72.00	69.67	72.73

在應用層次取得高得分比(80.21%)，但卻在分析層次表現最差(44.45%)。而在教學與評量方面則是在應用層次取得高得分比(80.51%)，而在評鑑層次表現最差(0%)。整體趨勢而言，ChatGPT 於不同科目不同認知層次的表現並無明顯的規則。

若僅區分高、低層次，綜合 ChatGPT 在三個考科的高、低層次平均得分比差異不大，分別為 63.50% 與 67.45%，針對高層次試題的表現僅略低。然而，三個考科的表現不盡相同，大致上，ChatGPT 在教育理念與兒童發展方面的低層次表現優於高層次；但值得注意的是，在教學與評量方面則是高層次的表現(66.90%)優於低層次(57.16%)。

再將三科目的得分比分布以盒鬚圖(圖 2)呈現，可以觀察到在三個考科中，

ChatGPT 在高層次題型的得分比分布都明顯大於低層次題型，這意味著 ChatGPT 在高層次題型中的答題穩定度差、歧異度高。特別是在教學與評量中，ChatGPT 在高層次試題的表現優於低層次試題，但高層次題型的得分比也呈現最大歧異度，這表示這一考科的相關試題或試題情境特別須要教師具有高度警覺與評鑑 ChatGPT 回答的能力。

(二) ChatGPT 對於不同題型下的認知分類和高、低層次試題得分比

表 9 顯示 ChatGPT 於不同題型下六認知層次分類的表現。研究結果顯示，統整 2019 年至 2022 年所有教師資格考試，ChatGPT 的得分比介於 66.67% - 77.94% 之間，在問答題的表現最好，而綜合題的表現最差。

圖 2

ChatGPT 在不同科目中高、低層次得分比分布盒鬚圖

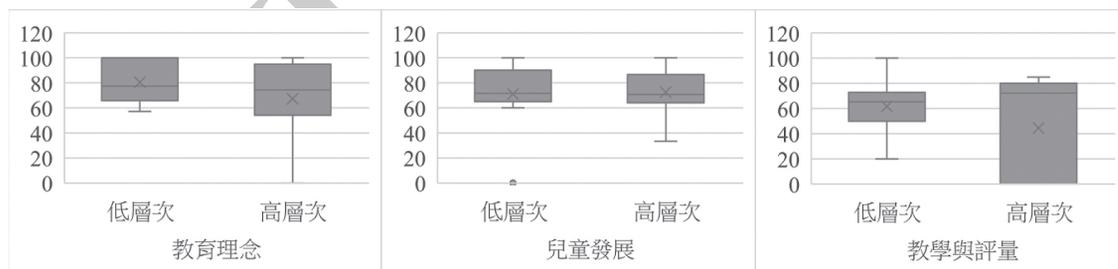


表 8

ChatGPT 在不同科目試題認知分類與高、低層次的得分比 (%)

科目	記憶	理解	低層次	應用	分析	評鑑	創作	高層次
教育理念	56.13	79.33	67.63	75.00	87.22	15.15	-	59.12
兒童發展	77.04	78.10	77.57	80.21	44.45	68.75	-	64.47
教學與評量	47.62	66.71	57.16	80.51	59.44	0.00	60.74	66.90
平均	60.26	74.71	67.45	78.57	63.70	27.96	60.74	63.50

進一步比較 ChatGPT 在三種題型中不同認知分類的表現，發現差異極大。在選擇題中，ChatGPT 在理解層次表現最好 (74.94%)；而在問答題中，六個層次中，應用、分析與創作層次的得分都高於 83%，位居高位；而在綜合題中，六個層次中，則是記憶與應用層次的表現最好，分別為 100.00% 與 93.75%。這顯示了 ChatGPT 在不同題型和認知層次上的表現存在較大的差異。

此外，ChatGPT 在三種題型的高、低層次表現也有不同。在選擇題和綜合題中，ChatGPT 的低層次表現優於高層次；然而在問答題中，高層次的表現優於低層次。

比較 ChatGPT 在三種題型中高、

低層次得分比的盒鬚圖(圖 3)，觀察到 ChatGPT 在選擇題的表現上，高層次得分比分布呈現較大的歧異度，這也是歧異度最大的題型。有些時候，ChatGPT 回答了一個推理的答案，但答案與選項不同或略有不同時，ChatGPT 也容易回答錯誤，因而造成歧異度大的情形。相反地，在問答題和綜合題中，高層次的歧異度則小於低層次。

伍、研究討論

一、國民小學教師資格考試試題中認知分類和高、低層次分布

本研究針對 2019 年到 2022 年的教師

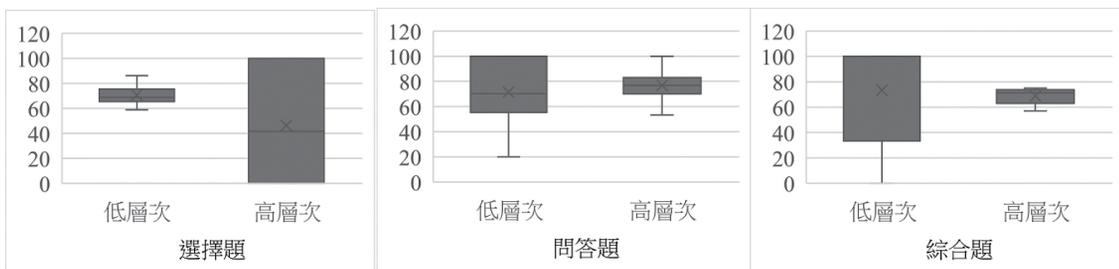
表 9

ChatGPT 在不同題型試題認知分類與高、低層次的得分比 (%)

題型	記憶	理解	低層次	應用	分析	評鑑	創作	高層次	得分比
選擇題	63.01	74.94	68.97	46.40	40.39	-	-	43.39	69.95
問答題	57.22	75.18	66.20	85.02	83.85	70.84	86.11	81.45	77.94
綜合題	100.00	58.33	79.17	93.75	71.58	25.00	55.00	61.33	66.67

圖 3

ChatGPT 在不同題型中高、低層次得分比分布盒鬚圖



資格考試進行分析。結果顯示，在這些考試的選擇題中，並沒有包含評鑑和創作認知層次的題目，這與 Haataja 等人 (2023) 的研究結果一致。2021 年起，教師資格考試為符應素養導向命題要求加入綜合題，其試題包含三種題型，涵蓋六項認知層次，高層次占分比達 81.67 % 明顯高於低層次試題 (18.33 %)，顯見素養導向評量與提高認知層次試題目標相同。

教師資格考試每份試卷選擇題占 50 到 60 分，然而，本研究觀察到選擇題低層次題目占分比極高，平均為 93.45 % ，遠高於其他研究中的結果。例如，芬蘭教師入學考試中，低層次選擇題約占 60 % (Haataja et al., 2023) ；在醫學類試卷中，測試事實回憶類的低層次選擇題超過 50 % (Palmer & Devitt, 2007) ；而在護理教育題庫中，有 47.3 % 的題目屬於知識層面 (Masters et al., 2001) 。

此外，教育理念與兒童發展無論何種題型皆沒有創作層次試題，評鑑層次占分比也是最低，且三考科低層次占分比都高於高層次占分比，表明在本國教師資格考試中，應該增加對創作和評鑑層次的試題涵蓋。根據李坤崇 (2004) 對九年一貫課綱命題的解析實例，選擇題仍然能夠評量出應用與分析高層次認知能力。此外，教師資格考試命題作業要點 (2020) 僅規定了各科目的考試題型。為了提升教師資格考試的有效性，有必要增加選擇題中高層次題目數並調整三種題型的占分比，朝向多元題型展現高層次思考的評量。藉由教師資格考試試題認知層次的提

升，促使師資培育機構更重視培養師資生的高層次認知能力，鼓勵學生進行深度思考、解決問題和應用知識，以利於培養師資生們未來在教學實踐中的能力。

二、ChatGPT 在國民小學教師資格考試中不同考科的表現情形

在教師資格考試的三個考科中，ChatGPT 平均得分 72.73 分，成功通過了平均 60 分的及格標準。這一優異的評量表現與眾多研究結果一致，包括美國醫學執照考試 (Kung et al., 2023) 、律師資格考試 (Bommarito II & Katz, 2022) 及國外大學入學考試 (Marcus, 2023) 等。

值得注意的是，在三個考科中，ChatGPT 在教學與評量方面的得分相對較低。一方面，教學與評量方面的試題中，高層次占分比遠大於教育理念與兒童發展兩個科目的高層次占分比；一方面，此科目強調師資生之情境應用、高層次思考、反思修正等表現 (教育部, 2020) ，比其他兩科目具有更多元與真實情境的變異性。此外，在 2021 年和 2022 年加入素養導向命題後的國民小學教師資格考試中，ChatGPT 的得分都比 2019 年和 2020 年低，這顯示 ChatGPT 在應對高層次試題以及應對素養導向命題方面能力相對較弱的可能性。

總之，ChatGPT 在教師資格考試中取得了優秀的整體成績，但在某些特定方面的表現有待提高，特別是在高層次試題和素養導向命題方面。

三、ChatGPT 在國民小學教師資格考試中不同認知分類和高、低層次的表現情形

ChatGPT 能通過教師資格考試標準，顯示其在教師教學知識和教學內容方面具有基本的表現，初步展現其作為師資培育教學輔助工具的可行性。然而，進一步觀察發現 ChatGPT 在不同科目和題型中表現最佳的認知層次並不相同，顯示其表現具有多樣性，且這種多樣性可能受到認知層次外其他變因，例如：考科或問題情境等其他變因之間交互作用的影響，但由於人工智慧工具的運作機制是黑盒子，我們僅能透過觀察實際表現來推斷可能的影響原因。

此外，ChatGPT 在高層次試題的得分較低層次試題低，且在高層次試題中的答題穩定性也較差。在教學與評量一科目中，ChatGPT 的高層次得分雖然平均較低層次高，但在這種情況下高層次答題的歧異度卻最大，顯示 ChatGPT 在高層次表現普遍存在不穩定性，且在應對不同問題情境時，表現不穩定的情形會更加劇烈。這進一步呼應了前述觀察，即 ChatGPT 的表現受到測試內容的多樣變因影響及交互影響。

ChatGPT 的表現受限於既有教檢試題情境中資訊的訓練，因此雖能通過教檢，達成師資生欲通過教檢的短期目標，但其在教學與評量一科目中歧異度最大的情形也可以預測其無法如資深教師般應對瞬息萬變的未知情境。因此，若使用 ChatGPT 培育師資生達成一個優良教師的長期目標，特別需要訓練師資生根據情境，運用相關知識與技

能解決問題或做出決策(葉玉珠,2002)，這也是身為教育者應該具備的能力。儘管 ChatGPT 在創意「發想」方面表現良好，但其具體成果可能相對不足，而具體成果的啟發、連結、評鑑和產出更端賴人類的參與才能取得良好的創造性結果。這進一步強調了使用人工智慧工具的目的，應該是找出人機各自擅長與弱勢的領域，建立適當的人機協作模式，而非追求訓練完美機器以取代人類(Kubsch et al., 2023)。但即使如此，在某些不容許犯錯的高風險領域，例如教育與醫療，即使在有人工智慧工具輔助下，人類仍應隨時以高認知層次能力掌握與評鑑人工智慧機器的表現，並持續於真實世界中持續研究與發展(Kung et al., 2023)，以避免類似於無人車造成意外車禍等不可逆的潛在風險。

陸、結論與建議

本文分析國民小學教師資格考試試題認知層次，以下根據三個研究問題，依序提出本研究之結論，並於每個結論之後，緊接著提出相應之建議。

一、高層次試題的比例偏低，應提高國民小學教師資格考試中高層次試題比例

師資生需要具備更高層次的認知能力，如應用、分析、評鑑和創作等能力，以勝任複雜的教學環境和滿足學生多樣化的需求，符應素養導向教學目標。根據研究結果，目前國民小學教師資格考試中綜合題的高層次試題比例高於低層次，符合素養導向評量的

要求，但整份試卷中選擇題仍然佔據主導地位，高層次試題的比例明顯偏低。

因此，為了更全面地評估師資生的能力和素養，提高教師資格考試的品質和有效性，並提供更準確的評估結果，研究者建議有必要調整三種題型的占分比，逐漸增加教師資格考試中高層次試題的比例，朝向多元題型展現高層次思考的評量。這樣的措施也將促使師資培育課程更加重視培養師資生在高層次教學方面的需求，從而提升師資生的教學能力，更好地迎接未來的教學挑戰。

二、ChatGPT 回應不同科目和類型的穩定度不同，師培課程可引入區分認知層次之人機協作模式

ChatGPT 成功通過國民小學教師資格考試教育相關科目的考試門檻，顯示其具備協助師資養成的潛力。然而，值得注意的是，ChatGPT 在國民小學教師資格考試中對於不同科目和題型的認知分類以及高、低層次試題的穩定度呈現不同。在不同科目下，ChatGPT 在兒童發展方面表現最佳，而在具有較多問題情境的教學與評量方面表現相對較差。在不同題型下，ChatGPT 在問答題中表現較佳，而在綜合題方面則表現相對較差。在不同認知分類和層次上，也呈現出明顯的多樣性與差異。

有鑑於這些差異，研究者提出區分認知層次所建立的人機協作模式之建議：在達成短期「通過教檢」的目的時，師資培育者可借助 ChatGPT 生成低層次試題的解釋性文

本，以確保文本的正確性和清晰性，進一步提高師資生對基本概念和知識的理解。而在高層次試題上，鑒於 ChatGPT 回應的穩定度不足，師資培育者可透過不同案例和問題情境，提供多個 ChatGPT 生成的答案作為參考，協助師資生評鑑選擇最佳答案，以促使師資生將教育理念與實際教學情境更密切結合。儘管 ChatGPT 具有強大的語言處理能力，師資生仍應保持批判性思考，確保能夠從多個角度評估和整合資訊，在複雜情境中進行批判和評鑑，以建立高層次思考模式。長此以往，此模式亦可能有助於協助師資培育機構實現多元化的教學，同時培養師資生未來進入教學現場時的創造力、批判思維和解決問題的高層次能力。

三、除認知分類外還有其他影響 ChatGPT 表現的變因，可設計實驗進一步探討 ChatGPT 在國民小學教師資格考試試題中輔助的影響

雖然在不同科目下，ChatGPT 在高層次試題的答題穩定度較差。然而，在三個科目與三種題型中高層次試題的得分比並非最低，顯示樣本間仍有其他因素影響。本研究僅以描述性統計觀察這五次考試中的現象，無法獲知或推論其他科目、其他年度的情形，是本研究的限制。基於這些結果，我們建議未來的研究可以進一步以其他架構，深入探討試題間的差異，以找出影響作答的關鍵因素，這將有助於提供師資培育機構在培育師資生時其他有價值的參考。值得注意的是，雖然人工智慧的進展日新月異，本研究所探

索的 ChatGPT 的能力亦僅限於 GPT-3.5 以及進行研究當下(2023 年 4 月) GPT-3.5 的能力,為本研究另一個重要限制。但本研究強調使用免費、廣泛普及使用的人工智慧工具,短時間內 GPT-3.5 仍具有此項優勢,因此仍具備研究的重要性與價值。

進一步的研究方向可以針對本研究所探索出的 ChatGPT 優勢及弱勢,引入區分認知層次的人機協作模式,設計實驗探討 ChatGPT 融入師資培育過程中的輔助影響,藉此瞭解 ChatGPT (或其他人工智慧工具)融入師資培育課程對師培生能力提升的實際效果。這樣的研究有望為人工智慧工具與師資培育過程提供因果關係的證據,與具體的示例與指引。

參考文獻

- 任宗浩(2017年12月)。素養導向評量。國家教育研究院電子報,166期。https://reurl.cc/7k1ED1
- [Jen, T. H. (2017, December). Competence-based assessment. *Electronic Newsletter of the National Academy for Educational Research*, Issue 166. https://reurl.cc/7k1ED1]
- 任宗浩(2018)。素養導向評量的界定與實踐。載於蔡清華(主編),課程協作與實踐第二輯(頁75-82)。教育部中小學師資課程教學與評量協作中心。https://reurl.cc/WveyKO
- [Jen, T. H. (2018). Definition and practice of competence-based assessment. In C. H. Tsai (Ed.), *Collaborative Curriculum and Practice, Volume 2* (pp. 75-82). Ministry of Education, Curriculum, Teaching, and Assessment Collaboration Center for Elementary and Secondary School Teacher Education.]
- 李坤崇(2004)。修訂 Bloom 認知分類及命題實例。教育研究月刊,122,98-127。
- [Li, K. C. (2004). Revised Bloom's cognitive taxonomy and examples of test item. *Journal of Education Research*, 122, 98-127.]
- 吳璧純(2017)。素養導向教學之學習評量 - 國民中小學核心素養如何評鑑。台灣教育評論月刊,6(3),30-34。
- [Wu, P. C. (2017). Assessment of competence-based teaching - How to evaluate core literacy in elementary and junior high schools. *Taiwan Educational Review Monthly*, 6(3), 30-34.]
- 林政逸、賴慧君(2021)。我國中小學教師專業標準之研究。師資培育與教師專業發展期刊,14(3),1-28。https://doi.org/10.3966/207136492021121403001
- [Lin, J. Y., & Lai, H. J. (2021). The professional standards of elementary and secondary school teachers in Taiwan. *Journal of Teacher Education and Professional Development*, 14(3), 1-28. https://doi.org/10.3966/207136492021121403001]
- 師資培育法(修正公布2017年6月14日)。https://law.moj.gov.tw/LawClass/

- LawAll.aspx?PCode=H0050001
 [Teacher Education Act (Revised on 2017 June 14). <https://law.moj.gov.tw/LawClass/LawAll.aspx?PCode=H0050001>]
- 高級中等以下學校及幼兒園教師資格考試辦法(修正公布 2020 年 3 月 25 日)。
<https://reurl.cc/N072Rn>
- [Regulations for K-12 Qualification Examination of Teachers (Revised on 2020 March 25). <https://reurl.cc/N072Rn>]
- 高級中等以下學校及幼兒園教師資格考試命題作業要點(修正公布 2020 年 11 月 13 日)。
<https://reurl.cc/4o0lXj>
- [Guidelines for K-12 Teacher Qualification Examination (Revised on 2020 November 13). <https://reurl.cc/4o0lXj>]
- 葉玉珠(2002)。高層次思考教學設計的要
 素分析。《**中山通識教育學報**·創刊號·
 75-101。
- [Yeh, Y. C. (2002). Analysis of elements
 in the design of high-level thinking
 teaching. *General Education Journal, Inaugural Issue*, 75-101.]
- 曾建銘(2020 年 11 月)。教師資格考素養導
 向試題轉型。《**國家教育研究院電子報**·
 201 期。 https://epaper.naer.edu.tw/edm.php?grp_no=1&edm_no=201&content_no=3550
- [Tseng, C. M. (2020, November). Trans-
 formation of teacher qualification
 examination into a competency-based
 items. *Electronic Newsletter of the
 National Academy for Educational Re-
 search*, Issue 201. https://epaper.naer.edu.tw/edm.php?grp_no=1&edm_no=201&content_no=3550]
- Anderson, L. W., & Krathwohl, D. R.
 (2001). *A taxonomy for learning, teach-
 ing, and assessing: A revision of Bloom's
 taxonomy of educational objectives: Complete edition*. Addison Wesley
 Longman, Inc.
- Baghaei, S., Bagheri, M. S., & Yamini, M.
 (2020). Analysis of IELTS and TOEFL
 reading and listening tests in terms of
 revised bloom's taxonomy. *Cogent Ed-
 ucation*, 7(1), 1720939. <https://doi.org/10.1080/2331186X.2020.1720939>
- Baker, T., Smith, L., & Anissa, N. (2019).
 Educ-AI-tion rebooted? Exploring the
 future of artificial intelligence in schools
 and colleges. *Nesta Foundation*. [https://
 media.nesta.org.uk/documents/Future_
 of_AI_and_education_v5_WEB.pdf](https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf)
- Bloom, B. S. (1956). *Taxonomy of education-
 al objectives: The classification of edu-
 cational goals. Handbook I, Cognitive
 Domain*. Longmans.
- Bommarito II, M., & Katz, D. M. (2022).
 GPT takes the bar exam. *arXiv pre-
 print arXiv:2212.14402*. [https://doi.
 org/10.48550/arXiv.2212.14402](https://doi.org/10.48550/arXiv.2212.14402)
- Cope, B., Kalantzis, M., & Searsmith, D.
 (2021). Artificial intelligence for edu-
 cation: Knowledge and its assessment

- in AI-enabled learning ecologies. *Educational Philosophy and Theory*, 53(12), 1229-1245. <https://doi.org/10.1080/00131857.2020.1728732>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 1-15. <https://doi.org/10.1080/14703297.2023.2195846>
- Haataja, E. S., Tolvanen, A., Vilppu, H., Kallio, M., Peltonen, J., & Metsäpelto, R. L. (2023). Measuring higher-order cognitive skills with multiple choice questions—potentials and pitfalls of Finnish teacher education entrance. *Teaching and Teacher Education*, 122, 103943. <https://doi.org/10.1016/j.tate.2022.103943>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Klassen, R. M., & Kim, L. E. (2019). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review*, 26, 32-51. <https://doi.org/10.1016/j.edurev.2018.12.003>
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS digital health*, 2(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Kubsch, M., Krist, C., & Rosenberg, J. M. (2023). Distributing epistemic functions and tasks—A framework for augmenting human analytic power with machine learning in science education research. *Journal of Research in Science Teaching*, 60(2), 423-447. <https://doi.org/10.1002/tea.21803>
- Marcus, L. (2023). How Smart is ChatGPT? *Visual Capitalist*. <https://reurl.cc/zYdKoe>
- Masters, J. C., Hulsmeyer, B. S., Pike, M. E., Leichty, K., Miller, M. T., & Verst, A. L. (2001). Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. *Journal of Nursing Education*, 40(1), 25-32. <https://doi.org/10.3928/0148-4834-20010101-07>

- Metsäpelto, R., Poikkeus, A., Heikkilä, M., Husu, J., Laine, A., Lappalainen, K., Lähteenmäki, M., Mikkilä-Erdmann, M., Warinowski, A., Iiskala, T., Hangelin, S., Harmoinen, S., Holmström, A., Kyrö-Ämmälä, O., Lehesvuori, S., Mankki, V., & Suvilehto, P. (2021). A multidimensional adapted process model of teaching. *Educational assessment, Evaluation and accountability*, 1-30. <https://doi.org/10.1007/s11092-021-09373-9>
- Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence powered digital writing assistant in higher education: randomized controlled trial. *Heliyon*, 7(5), e07014. <https://doi.org/10.1016/j.heliyon.2021.e07014>
- Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Research paper. *BMC medical education*, 7(1), 1-7. <http://www.biomedcentral.com/1472-6920/7/49>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 342-363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Savelka, J., Agarwal, A., Bogart, C., Song, Y., & Sakr, M. (2023). Can generative pre-trained transformers (GPT) pass assessments in higher education programming courses?. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education*, 1, 117-123. <https://doi.org/10.48550/arXiv.2303.09325>
- Wilson, L. O. (2016). Anderson and Krathwohl–Bloom’s taxonomy revised. *Understanding the New Version of Bloom’s Taxonomy*. <https://reurl.cc/x7rWGE>
- Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: A systematic review. *Studies in Science Education*, 56(1), 111–151. <https://doi.org/10.1080/03057267.2020.1735757>
- Zhai, X. (2022). ChatGPT user experience: Implications for education. *Available at SSRN*. <http://dx.doi.org/10.2139/ssrn.4312418>
- Zhai, X., Nyaaba, M., & Ma, W. (2023). Can generative AI and ChatGPT outperform humans on Cognitive-demanding problem-solving tasks in science? *Available at SSRN*. <http://dx.doi.org/10.2139/ssrn.4451722>

J T E P D