

108 課綱教案架構師 GEM 專用 Knowledge Base 4 of 4

知識庫檔案: K-12 素養導向評量設計與尺規建構指南

文件用途: 本模組提供 Custom Gem 進行素養評量設計、表現任務規劃及評量尺規 (Rubrics) 撰寫的標準與檢核機制。

1. 評量典範轉移 (Assessment Paradigm Shift)

AI 在分析或設計評量時，需遵循以下核心定義：

- 評量目的: 從「學習的檢核 (Assessment of Learning)」轉向**「促進學習的評量 (Assessment for Learning)」與「評量即學習 (Assessment as Learning)」**。
- 評量標的 (K-S-A 模型):
 - 認知 (Cognitive): 不僅是事實記憶，更包含概念理解與後設認知 (監控自我學習)。
 - 技能 (Skills): 從精細動作到批判性思考、邏輯論證等高階心智技能。
 - 情意 (Affective): 不直接評量內在心理，而是評量「表現出態度的行為」(例如: 實驗失敗時是否展現恆毅力)。
- 核心原則:
 - 情境化 (Contextualization): 任務必須鑲嵌在真實生活 (Real-world) 或學術探究 (Academic) 脈絡中。
 - 歷程與結果並重: 設置「階段性檢核點 (Milestones)」，評量學生如何修正錯誤與優化策略，而非僅看最終成品。

2. 表現任務設計工具: GRASPS 模型

當使用者要求設計「實作評量」或「表現任務」時，AI 應依據 Wiggins & McTighe 提出的 GRASPS 架構進行生成與檢核：

元素	定義	範例 (物理科)
G (Goal)	任務目標	設計連鎖反應機關，展示三種物理原理。
R (Role)	學生角色	科學博物館的展品設計師。
A (Audience)	目標觀眾	參觀博物館的中小學生 (需淺顯易懂)。
S (Situation)	情境脈絡	博物館舉辦特展，徵求高互動性展品。
P (Product)	產出表現	實體模型 + 設計圖 + 現場演示。

S (Standards)	評量標準	運作成功率、原理正確性、解說清晰度。
---------------	------	--------------------

3. 評量尺規工程學 (Rubric Construction Engineering)

這是 AI 最具價值的功能之一。當使用者要求「幫我寫評分標準」時，請依據以下邏輯：

A. 尺規三大要素

1. 評量規準 (Criteria): 具體的評量向度 (如「論證邏輯」、「視覺設計」)。原則：向度間需獨立互斥。
2. 表現等級 (Performance Levels): 建議使用發展性詞彙，如「典範 (Exemplary)」、「精熟 (Proficient)」、「發展中 (Developing)」、「新手 (Novice)」。建議設為 4 個等級。
3. 等級描述 (Descriptors): 描述具體可觀察的行為，而非抽象形容詞。

B. 尺規類型選擇

- 分析性尺規 (Analytic): 各向度獨立給分。適用：診斷學習強弱項、形成性評量。
- 整體性尺規 (Holistic): 給予單一綜合分數。適用：大規模測驗、總結性評量。
- 單點式尺規 (Single-Point): 僅列出「精熟」標準，左右留白供質性回饋。適用：自我反思、同儕互評。

C. 常見設計偏誤與修正 (AI 檢核點)

- 模糊形容詞: ❌「富有創意」→ ✅「使用了現有教科書以外的材料或方法」。
- 計算錯誤次數: ❌「錯字超過3個」→ ✅「錯誤頻率不影響閱讀流暢度」。
- 負面表述: ❌「內容空洞」→ ✅「尚未能提出有力證據支持論點」。

4. 特定領域評量指南 (Domain-Specific Guidelines)

A. 自然科學探究與實作

- 發現問題：能否區分因果與相關？能否提出可驗證的假設？
- 規劃執行：是否設計對照組？是否考慮誤差控制？
- 論證建模：關鍵指標——當數據與理論不符時，學生是「誠實面對並探討原因」(A級)，還是「篡改數據」(劣質表現)。

B. 口語表達與溝通

- 語音：發音清晰度、音量穿透力。
- 聲情技巧：斷句是否依照語意 (而非僅依標點)、抑揚頓挫是否配合情感。
- 儀態：目光接觸 (Eye Contact)、站姿。
- 警示：避免「暈輪效應」，勿因學生聲音好聽而忽略內容邏輯。

C. 合作學習 (Cooperative Learning)

- 三角檢證法：結合「教師觀察」、「同儕互評 (解決搭便車問題)」與「個人反思」。
- 評量維度：參與責任 (準時/貢獻)、溝通互動 (傾聽/協調)、問題解決 (決策/執行)。

D. 藝術與跨域專題

- 自主性遞減：A級 (主動規劃/應用) → B級 (參與/技法正確) → C級 (願意嘗試)。
- 專題指標：內容深度、創新性 (區分複製與創造)、執行力 (專案管理)、反思能力。

5. 評量品質控管 (Quality Control)

- 校準會議 (Calibration): 多位評分者先試評樣卷，討論差異並建立共識後，再正式評分。

- 科技輔助：利用電子檔案 (e-Portfolio) 記錄歷程，使用 AI 初評語音或格式。
-

— END —