

「賽馬會運算思維教育」先導計劃

評估總結報告 — 附錄



2020年5月

目錄

附錄一：「賽馬會運算思維教育」框架	1
附錄二：分析方法	2
附錄三：模型估算結果	5
附錄四：驗證分析	11

附錄一：「賽馬會運算思維教育」框架

運算思維概念： 設計員在編程時應用的概念

1. 序列	識別在解決編程任務中的一系列有序步驟
2. 事件	一件事情導致另一件事情的發生
3. 條件	基於各種不同狀況來作出決定
4. 運算子	支援數學及邏輯表達式的運算符號
5. 同步發生	讓多於一件事同時發生
6. 重複	反覆多次執行相同序列的事件
7. 命名和變數	以有意義的名字命名在程式中引用和計算的變數
8. 數據結構	數據儲存、檢索和更新的基本方式
9. 程序	創建一系列指令，從而把相關序列模組化和抽象化

運算思維實踐：解決問題技巧 設計員按照運算思維概念研發的解決問題技巧

1. 測試及除錯	確保事情能夠運作以及找出並解決所出現的問題
2. 反覆構思及漸進編程	進行一些發展，然後嘗試將發展的事情再進一步發展
3. 重用及整合	基於現有的指令、方案或想法來重複使用或重新組織來建造新的事情
4. 概念化及模組化	在整體和部份之間找出關連
5. 算法思維	以明確的規則和步驟清楚表達解決問題的方法

運算思維視野：身份與動機 設計員對身邊世界和自己的看法

1. 自我表達	以新媒體來創作和表達自己的想法
2. 提問與瞭解	自覺懂得運用科技發問及提出與科技有關的問題
3. 聯繫生活	感激別人賞識自己的作品，以及別人與自己攜手創作
4. 數碼充權	感受到自己能夠運用數碼科技來處理身邊的挑戰
5. 運算身份認同	建立主動及正面地運用運算知識及技能處理個人及身邊的人的生活事情的身份認同

附錄二：分析方法

本附錄旨在闡述是次評估所使用的數項分析方法。

局部矩陣採樣法

評核學生的運算思維表現時，研究團隊採用了矩陣採樣法，也就是將需要評核的項目分散載於不同表格/問卷，然後將這些表格隨機派發給學生回答。這方法讓團隊可以一次過收集適用於學生的多個項目的數據。矩陣採樣法常用於量度同一期學生的表現，而非個別學生的分數。舉例說，美國全國教育進度評核 (U.S. National Assessment for Educational Progress) 便應用了這個方法量度全美學生的成績。

在矩陣採樣法下，學生獲派的表格內容不一(每份表格所涵蓋的運算思維概念不盡相同)，因此不能直接比較個別學生的表現，但卻可用於比較不同組別學生的情況。我們採用的局部矩陣採樣法(矩陣採樣法的一支)，在某程度上仍可比較個別學生的表現。所有派發予學生的表格都包含一些共通的核心評核項目，其餘評核項目則分散載於不同表格。只要將局部矩陣採樣法結合「項目反應理論」分析(IRT, 詳情見下文)，即使學生受評核的項目不完全相同，也能計算出個別學生的分數，以資比較。我們按照學生受評核項目建立 IRT 模型後，便可從中估算其能力。表格內的共通評核項目是進行能力評估的基礎，確保我們使用的是同一準則，因此可作比較。我們使用局部矩陣採樣法制訂運算思維概念和運算思維實

踐的評核方案，好處是可以同一時間量度同一期學生在所有評核項目上的表現，減省個別學生用於填寫問卷/表格的時間。

在 3 年的評估期裡，我們持續精簡課程所涵蓋的概念，並相應地減少評核項目。在評核學生的運算思維概念時，我們從 2018 年起已剔除「程序」這一項，使學生在一段測試期內只需填寫一份表格/問卷，便可涵蓋全套運算思維概念。因此，在 2018 及 2019 年，局部矩陣採樣法只用於評核學生的運算思維實踐表現。

項目反應理論

我們根據「項目反應理論」(IRT) 來計算學生在運算思維概念、實踐和視野 3 方面的得分。IRT 是一種以測試對象潛在特徵 (latent variable) 為基礎的建模方式，按照評核項目的得分顯示這些項目的難度，並使用同一尺度顯示學生的能力。這種分析方法不僅可以衡量學生的整體運算思維能力，亦可監察個別學生或特定學期學生在一段時間內的進度。IRT 經過調整後，還可以處理缺失數據(missing data)的問題。這項優點對矩陣採樣法非常重要，因為在矩陣採樣法下，每名學生接受評核時只需回答部分而非全套問題；學生的能力高低，是根據他提供的答案估算。至於學生沒有回答的問題(問卷的刻意設計)，不會納入計算其能力的考慮因素。這讓我們得以在每個管理點(administration point) 估算每名學生的整體運算思維能力。

基於這種設計特色，我們可以比較不同學生在整個運算思維學習過程裡的進度。由於矩陣採樣法會隨機要求大量學生回應不同的評核項目，我們可同時比較不同學期學生在每個評核項目上的表現。這有別於追蹤個別學生在某段時期學習某項獨特技能的進度，必須注意。

我們為瞭解學生進度而進行多次評核，而為確保可以在項目難度和學生能力兩方面比較這些評核所得的結果，這些評核包含了一些用以界定衡量尺度的共通項目(IRT 一般稱之為「錨點項目」)。因此，評核學生的運算思維概念時，派發給同年級和不同年級學生的評核表格/問卷，皆包含同一套共通項目。我們可以使用這些錨點項目校正評核項目的特點(例如難度)和不同學生樣本之間的差異，並將使用不同表格/問卷和在不同年度進行的評核連繫起來。另外，我們進行的每項分析，都一併使用了(數年來透過不同表格和問卷收集的)所有數據來估算得分，確保用以估算學生能力的項目難度可作比較。如屬數年來持續出現在所有表格/問卷的共通項目，其難度是保持不變的；但我們會局部按照這些項目難度來釐訂只會出現在一份表格/問卷的項目的難度。如此，我們便可根據某個評核項目的共同連續體得知學生能力的改變/變異。

對於包含二元計分項目的運算思維概念，由於這些項目的難度有重要意義，我們使用了單參數邏輯(1PL)模型。運算思維實踐方面，由於當中的評核項目屬於二元計分或多元計分(最高為 5 分)，我們採

用了分部評分模型 (PCM)來解讀可以歸入多重評分類別的評核項目。至於使用李克特量表 (Likert-type scale) 進行評核的運算思維視野項目，我們則使用評分量表模型(RSM)來闡述這些項目的多重回應類別。有關這些模型的詳情和用處，可參閱 Embretson & Reise (2013)¹。

轉換成 NCE 分數

為使 IRT 分數較易理解，我們把其轉換成「常態曲線等值」(NCE) 分數。NCE 分數介乎 0–100，以常態分佈，平均值及標準差分別為 50 及 21.06。以 NCE 尺度顯示 IRT 分數，使人較易理解，亦有助比較不同評核項目調查結果的數量。

我們以 7 個元素/項目來衡量學生在運算思維視野上的表現：對編程的興趣；數碼自我效能感；效用動機；貢獻世界的動機；創意；參與/投入；以及歸屬感。我們使用 NCE 分數來綜合顯示學生在運算思維視野方面的整體正面表現，方法是首先將學生在每項元素/項目的 IRT 得分標準化，然後從所有經過標準化處理的分數中得出平均數，作為學生的運算思維視野綜合得分，最後將此綜合得分轉換為 NCE 分數。本報告描述學生在運算思維視野的整體表現時，只涵蓋上文提及的首 6 項元素，並不包括「歸屬感」這一項。「歸屬感」涉及學生對在編程時與他人合作的看法，而非學生本人對編程的觀感。由於「歸屬感」與其他元素的關連不大，因此其在整體評核時已被剔除。

1. Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. London: Psychology Press.

使用等級線性模型 (HLM) 進行影響分析

我們使用一個由學生和學校組成的兩層模型進行第 2 和第 3 年的影響分析。在這個模型裡，每一層都有若干可以控制的共變數(covariates)。以下是該模型的公式：

i = 學生； j = 學校； Y_{ij} = 個別學生的表現；

$CoolThink_{ij} = 1$ (先導學校)；

$CoolThink_{ij} = 0$ (比照學校)；

e_{ij} = 學生隨機效應 (student random effects)；

r_j = 學校隨機效應 (school random effects)；

β_k = 學生共變數的向量 (vector)；

β_1 = 「賽馬會運算思維教育」對學生表現的估算影響

圖 B1 列出了等級線性模型的每一級所包含的控制變數 (control variables)。

$$y_{ij} = \beta_0 + \beta_1(CoolThink_{ij}) + \beta_k(k^{th} student covariate_{ij}) + \beta_l(l^{th} school covariate_j) + e_{ij} + r_j$$

表 B1 影響模型所包含的控制變數

層次	控制變數
學校	接受經濟資助學生%
	有特殊教育需要學生%
	非華語學生%
	學校之前提供編程教育的體驗
	審批文件得分
	兩期學生綜合分析中的組別指標
學生	基線表現量度方法
	性別
	就讀年級

附錄三：模型估算結果

本附錄旨在闡述「影響分析」的基線等值 (baseline equivalence)、模型估算結果 (model estimates)、標準誤差 (standard errors) 及 p 值 (p-values)。

運算思維概念

我們分析了先導學生與比照學生的基線分數差別，以瞭解兩組學生在先導計劃開始前的表現是否等同。表

分數及樣本數量。由於兩者的標準差屬 0.25 之內，因此，對運算思維概念基線分數作出相應調整後，此分析已達致基線等值。

表 C2 顯示使用兩層 (學生及學校) HLM 模型來估算先導計劃對整體學生及不同子組別學生所產生的影響。此分析涵蓋了第一期及第二期的先導學校，有關結果已於報告正文闡述。

表 C1 進行第 2 年「影響分析」時學生的運算思維概念基線得分

	平均數	標準差	學生	學校
先導	38.8	12.9	4776	30
比照	37.1	12.1	3517	22

表 C2 進行第 2 年「影響分析」時以模型估算所得的運算思維概念得分

	估算分數	標準誤差	P 值
整體影響			
先導	5.0	2.9	0.084
對不同子組別的影響			
女生 vs 男生	-2.3*	1.0	0.020
5 年級 vs 6 年級	1.6	1.0	0.108
運算思維概念基線得分	0.0	0.0	0.209
基線數學測驗得分(經過標準化處理)	0.7	0.5	0.169
學校有%學生有特殊教育需要	0.5	0.3	0.144
學校有%學生接受經濟資助	-0.1	0.1	0.220
之前是否有編程經驗	-0.3	1.1	0.791
之前有否在課堂上進行編程練習	-0.4	1.2	0.752
可以在家上網	1.5	1.2	0.221
可以在家使用電腦	0.2	0.3	0.448
學校推行編程教育的經驗	-0.6	0.6	0.304

* $p < 0.05$.

我們沒有將「非華語學生%」列為需要分析的子組別。這是因為大部分學校表示沒有這類學生或有關學生比率低於 5%，差距不足以進行分析。

表 C3 顯示了進行第 3 年「影響分析」時，第一期先導及比照學校的運算思維概念基線分數及樣本數量。兩組學生的標準差屬 0.25 之內，因此，對運算思維概念

基線分數作出相應調整後，此分析已達致基線等值。

表 C4 顯示使用兩層 (學生及學校) HLM 模型來估算先導計劃於第 3 年對第一期學生所產生的影響。其結果與涵蓋第一期及第二期學生的第 2 年分析一致。

表 C3 進行第 3 年「影響分析」時第一期學生的運算思維概念基線分數

	平均數	標準差	學生	學校
先導	38.6	12.0	764	9
比照	36.6	11.2	535	12

表 C4 進行第 3 年「影響分析」時以模型估算所得的第一期學生運算思維概念得分

	估算分數	標準誤差	P 值
整體影響			
先導	3.8	4.6	0.412

運算思維實踐

運算思維實踐方面，我們於 2018 年 9 月首次進行評核，對象僅限於第二期學校的學生。當時，第一期先導學生已完成第一級課程，因此並沒有這批學生的運算思維實踐基線數據。

表 C5 顯示了先導與比照學校的運算思維實踐基線分數及樣本數量。兩者的標準差屬 0.25 之內，因此，對運算思維實踐分數作出相應調整後，此分析已達致基線等值。

表 C6 顯示使用兩層 (學生及學校) HLM 模型來估算先導計劃對整體學生及不同子組別學生所產生的影響。有關結果已於報告正文闡述。

我們也評估了第一期學生在完成 3 年課堂後的運算思維實踐得益。由於我們並沒對第一期學生進行運算思維實踐基線評核，分析他們的運算思維實踐表現得分時，需要對其運算思維概念基線分數作出調整。表 C7 顯示了先導與比照學校的運算思維實踐基線分數及樣本數量。兩者的標準差屬 0.25 之內，

表 C5 進行第 2 年「影響分析」時第二期學生的運算思維實踐基線分數

	平均數	標準差	學生	學校
先導	48.4	15.6	3900	20
比照	47.2	14.7	2487	20

表 C6 進行第 2 年「影響分析」時以模型估算所得的第二期學生運算思維實踐得分

	估算分數	標準誤差	P 值
整體影響			
先導	9.3***	2.8	0.001
對不同子組別的影響			
女生 vs 男生	0.6	1.2	0.605
5 年級 vs 6 年級	-1.3	1.2	0.283
運算思維概念基線得分	0.1*	0.0	0.041
基線數學測驗得分(經過標準化處理)	8.2***	0.4	0.000
學校有%學生有特殊教育需要	0.2	0.4	0.533
學校有%學生接受經濟資助	0.1	0.1	0.226
之前是否有編程經驗	-1.0	0.9	0.265
之前有否在課堂上進行編程練習	-1.4	1.0	0.179
可以在家上網	2.5**	0.9	0.007
可以在家使用電腦	0.4	0.2	0.071
學校推行編程教育的經驗	0.0	0.7	0.999

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

表 C7 進行第 3 年「影響分析」時第一期學生的運算思維實踐基線分數

	平均數	標準差	學生	學校
先導	48.0	14.2	739	10
比照	45.7	13.3	665	12

因此，對運算思維實踐分數作出相應調整後，此分析已達致基線等值。

表 C8 顯示進行第 3 年分析時，使用兩層 (學生及學校) HLM 模型來估算先導計劃對第一期學生的影響。

以影響幅度而言，這次估算結果(NCE 分數為 10.6)與對第一及第二期先導學生的第 2 年分析(NCE 分數為 9.3)所得出的估算影響相近；另外，其 p 值為 0.08，在統計學上已接近具備重要意義。

表 C8 進行第 3 年「影響分析」時以模型估算所得的第一期學生運算思維實踐得分

	估算分數	標準誤差	P 值
整體影響			
先導	10.6	6.0	0.076

運算思維視野

我們比較了先導及比照學生的運算思維視野基線分數，比較範圍涵蓋他們整體和在個別評核項目上的得分，以決定兩者在先導計劃開始前的表現是否等

同。表 C9 臚列了先導及比照學生的有關統計數字。在所有與運算思維視野有關的評核項目中，兩者在各平均數上的標準差均少於 0.25，因此，對運算思維視野基線分數作出相應調整後，我們認為先導與比照學校已達致基線等值。

表 C9 進行第 2 年「影響分析」時 30 間先導學校及 22 間比照學校的運算思維視野基線分數

		平均數	標準差	學生人數
整體	先導	51.9	18.1	4914
	比照	49.0	17.7	3652
歸屬感	先導	50.3	19.1	4114
	比照	49.4	17.8	2776
對編程的興趣	先導	51.2	21.0	4246
	比照	48.4	20.3	2922
參與/投入	先導	52.3	21.1	4243
	比照	49.6	21.0	2922
貢獻世界的動機	先導	50.9	20.0	4249
	比照	48.6	19.7	2949
數碼自我效能感	先導	50.1	20.7	4243
	比照	46.7	20.5	2944
效用動機	先導	51.4	21.1	4234
	比照	49.2	20.9	2941

表 C10 顯示使用兩層 (學生及學校) HLM 模型來估算先導計劃對學生整體運算思維視野及在個別評

核項目上所產生的影響，有關結果已於報告正文闡述。

表 C10 進行第 2 年「影響分析」時以模型估算所得的學生運算思維視野得分

	估算分數	標準誤差	P 值
整體影響			
運算思維視野	-2.3	1.5	0.131
歸屬感	-2.2	1.6	0.175
對編程的興趣	-2.3	1.8	0.188
參與/投入	-2.5	1.6	0.114
貢獻世界的動機	-0.6	1.6	0.709
創意	-2.3	1.7	0.179
數碼自我效能感*	-3.1	1.6	0.048
效用動機	-1.5	1.6	0.361

* $p < .05$

表 C11 臚列了第一期先導與比照學校在第 3 年「影響分析」中於整體和個別評核項目上的運算思維視野基線統計數字。在整體運算思維視野和對編程的興趣、參與/投入、創意、數碼自我效能感、效用動機這 5 個評核項目(全數共有 7 個評核項目)上，第一期先導學校的基線分數遠較比照學校為高(標準差高於 0.25)。

然而，「歸屬感」及「貢獻世界的動機」這 2 個評核項目卻未能達致基線等值。在這 2 個評核項目上，在

調整運算思維視野基線分數後，先導與比照學校的差距屬 0.25 的標準差之內。由於缺乏其他運算思維視野評核項目的基線等值，解讀在第 3 年對第一期學校進行運算思維視野分析所得的結果時，必須加倍小心。

表 C12 顯示進行第 3 年「影響分析」時，使用兩層 (學生及學校) HLM 模型來估算先導計劃對第一期學生的運算思維視野的影響。有關結果與第 2 年「影響分析」所得的結果一致。

表 C11 進行第 3 年「影響分析」時 10 間第一期先導學校及 12 間比照學校的運算思維視野基線分數

		平均數	標準差	學生人數
整體	先導	56.6	18.1	798
	比照	49.9	17.1	822
歸屬感	先導	45.7	16.9	408
	比照	47.8	14.7	382
對編程的興趣	先導	61.0	22.5	410
	比照	52.0	21.2	398
參與/投入	先導	59.6	21.1	410
	比照	52.3	21.6	399
貢獻世界的動機	先導	49.8	14.1	405
	比照	46.9	14.3	443
數碼自我效能感	先導	50.4	15.3	405
	比照	46.5	14.3	442
效用動機	先導	52.7	17.9	405
	比照	45.0	17.6	443

表 C12 進行第 3 年「影響分析」時以模型估算所得的第一期學生運算思維視野得分

	估算分數	標準誤差	P 值
整體影響			
運算思維視野	-5.8	3.7	0.117
歸屬感	-5.1	4.9	0.296
對編程的興趣	-7.3	5.7	0.202
參與/投入	-9.3	4.8	0.052
貢獻世界的動機	-4.1	3.6	0.254
創意	-4.1	3.8	0.277
數碼自我效能感*	-9.2	3.5	0.008
效用動機	-2.9	3.9	0.452

** $p < 0.01$.

附錄四：驗證分析

測驗效度 (Test validity) 是指有關證據和理論足以按照目標用途來闡釋測試得分的程度。驗證分析(validation analysis)現今被視為一個持續的過程：它始於設計評核方案之時，並貫穿方案的發展及執行階段。這對用以衡量先導計劃效用的運算思維工具尤其重要。由於先導計劃在不同的時間點或會對學生產生不同的潛在影響，我們期望和實質上可以怎樣使用及闡釋評核結果，亦可能因處於評估的不同階段(基線、中線及終線)而出現差異。

在基線及中線(中期)報告裡，我們扼要介紹了初步使用的各類效度證據 (validity evidence)。這些證據涵蓋評核內容及內部結構，為我們闡釋第 1 及第 2 級運算思維概念、運算思維實踐及運算思維視野基線分數提供了堅實的基礎。我們描述了這些使用「實據為本設計」(ECD) 的評核和測試如何確保運算思維全數 3 大範疇的評核工具均屬有效，因此能夠反映有關狀況。ECD 是一種讓設計者聚焦 3 條問題的評核方案設計策略：我們需要量度甚麼？需要甚麼憑證？甚麼樣的評核方案才可取得所需的證據？ECD 過程有助確保評核方案中的具體評核措施對準評核目標 (Mislevy, 2007)。2 在之前的報告裡，我們已透過檢視評核可靠性及因子結構來探討運算思維概念及運算思維視野的內部結構效度。

在本附錄裡，我們會闡述從總結(終線)評核所取得的效度證據，並將結果與基線及中線(中期)證據比較，以確保關於效度的討論是連貫和一致的。本附錄將闡述不同類型的效度證據，包括評核/測試內容、內部結構、項目效度及外部標準。

運算思維概念

從基線、中線到終線評估，我們都檢視了第 1 至 3 級運算思維概念評核表格/問卷的內部結構，並進行可靠性分析及驗證性因子分析 (CFA)，以探討表格/問卷上的評核項目能否一貫和穩定地量度該運算思維概念元素。針對運算思維概念的評核中，需要量度的子概念有 4-5 項，而針對每個子概念的評核項目不多；因此，我們相信這些子概念之間是有足夠的關連，足以讓我們認為是項評核/測試只是在量度學生在一項元素上的表現(學生的運算思維概念知識)。表 D1 顯示的是運算思維概念可靠性分析的阿爾法系數(coefficient alpha)，以及其單因子 CFA 模型的配適指標(fit indices)。綜合整個先導計劃在不同時段和學習級別的情況，運算思維概念評核表格/問卷的阿爾法系數介乎 .40 至 .66。由於學生尚未或只是剛剛開始學習運算思維，其基線可靠性相對較低，亦屬意料中事。我們不期望學生能在這個階段好好掌握有關概念，並估計他們會在學習過程中作出大量猜測，或因此令數據的可靠性降低。

1 Mislevy, R. J. (2007). Validity by design. *Educational Researcher*, 36(8), 463–469. <https://doi.org/10.3102/0013189X07311660>

另外，我們在設計評核/測試內容時，特意包含一系列難度不同的項目，而當中部分項目涵蓋的內容，是直到學生修讀較高級別課程才會學習的。雖然這有助區分和辨別學生，但由於學生並非以近似的方式與所有評核項目互動，或會令可靠性降低。然而，在這個情況下，我們認為數據的可靠性屬可接受程度。因對學生參與先導計劃前的運算思維知識水平缺乏資料而造成的下限效應(floor effect)，亦導致進行基線評核時，運算思維概念的不同評核項目之間欠缺有意義的共變數。隨著學生不斷學習，雖然評核/測試項目的數量減少，但中線和終線的阿爾法系數卻大幅上升。

也許是受到同一下限效應的影響，因子分析結果顯示中線和終線數據較基線數據更能支持以有關項目衡量一個運算思維概念元素的假設。在針對第 1、2 及 3 級課程的中線及終線評估，適配指標顯示單因子模型適用於有關數據。長期檢視共通項目的結果顯示，從基線、中線逐步邁向終線的過程中，學生有較佳表現，有關評核項目之間亦出現較大的共變數。此乃我們意料之中，這是因為較多學生可以正確回答當中部分問題。

表 D1 運算思維概念評核表格/問卷的可靠性因子分析摘要

運算思維概念 表格/問卷	學生 人數	項目 數量	阿爾法 系數*	單因子模型的適配性**			
				卡方值	df	RMSEA	CFI
2017 年 2 月(第 1 級)	289	31	.40	601.296***	434	.037	.519
2017 年 6 月/9 月(第 1 級)	439	31	.41	684.076***	434	.036	.649
2018 年 6 月(第 1 級)	16,594	16	.56	706.713***	104	.064	.890
2018 年 6 月(第 2 級)	2,355	21	.63	1627.836***	189	.057	.816
2019 年 6 月(第 1 級)	5,698	16	.50	2276.200***	104	.061	.884
2019 年 6 月(第 2 級)	9,412	21	.66	6939.003***	189	.062	.827
2019 年 6 月(第 3 級)	1,837	20	.48	1323.646***	170	.061	.809

* .70 是普遍獲認可的可靠性分界點。然而，正如上文所述，這個分界點極受一系列因素影響，包括評核/測試長度、內容域、需要量度的項目數量，以及評核/測試的預期用途。

** CFA 模型適配指數的一般分界點：CFI 等於或大於 .90 及 .95 者，分別被視為良好及卓越；等於或小於 0.01、0.05 及 RMSEA 為 0.08 者，分別被視為卓越、良好及普通/平平。

*** $p < .001$

我們透過研究項目的難度、鑑別力和數據缺失率來檢視項目效度。在每份運算思維概念評核表格/問卷裡，評核項目都是按預期目標分佈，難度各異。我們發現包含較多困難任務的項目(舉例說，屬於第 2 級課程內容的項目，難度會較屬於第 1 級課程者高；一些包含較複雜代碼的項目也是如此)顯示出較低的項目正確率。從基線(第 1 級課程為 33%及 32%)、中線(第 2 及 3 級課程分別為 50%及 44%)到終線(第 1、2、3 級課程分別為 56%, 50%及 56%)的過程中，運算思維概念項目的平均正確率持續上升，顯示評核/測試的難度恰當。

表 D2 顯示不同運算思維概念評核表格/問卷中的共通項目正確率。除了少數例外例子，從基線、中線到終線的過程中，學生的表現都有所改善或保持穩定。雖然學生在項目 q121 上的表現(基線至中線)並沒改善，但在終線卻有大幅改善。在項目 q221 方面，從基線到中線的過程中，學生的表現有所改善，但卻在終線出現輕微跌勢。這或許是因為第 2 級課程降低了對「條件」這個元素的重視程度。

表 D2 不同運算思維概念評核表格/問卷共通項目的正確率 (第 1 部分)

元素	級別	項目	2017 年 2 月	2017 年 6 月/9 月	2018 年 6 月	2018 年 6 月
			第 1 級 (N=289, J=31)	第 1 級 (N=439, J=31)	第 1 級 (N=16,594, J =16)	第 2 級 (N=2,355, J=21)
重複	1	q111	37.50%	37.27%	48.97%	51.31%
	2	q121	17.58%	14.42%	-	15.78%
	2	q122	46.64%	44.55%	48.64%	-
條件	1	q212	29.33%	28.17%	35.54%	39.91%
	1	q213	15.11%	14.82%	22.24%	-
	2	q221	26.24%	27.31%	35.32%	31.59%
同步發生及序列	1	q312a	37.94%	31.24%	40.06%	-
	1	q312b	59.71%	53.38%	54.55%	-
	2	q321	36.07%	38.55%	37.01%	38.31%
	2	q322	33.94%	31.85%	37.76%	38.02%
數據結構及算法	1	q411a	45.71%	49.88%	74.47%	62.45%
	1	q411b	44.93%	44.37%	67.68%	56.63%
	1	q411c	35.53%	36.38%	45.78%	43.30%
	3	q423a	-	-	80.60%	77.90%
	3	q423b	-	-	78.08%	76.46%
	3	q423c	-	-	65.08%	68.19%

表 D2 不同運算思維概念評核表格/問卷共通項目的正確率 (第 2 部分)

元素	級別	項目	2019 年 6 月	2019 年 6 月	2019 年 6 月
			第 1 級 (N=5,698, J=16)	第 2 級 (N=9,412, J=21)	第 3 級 (N=1,837, J=20)
重複	1	q111	56.71%	61.94%	67.21%
	2	q121	-	23.55%	-
	2	q122	-	-	-
條件	1	q212	44.67%	48.46%	-
	1	q213	29.88%	-	-
	2	q221	30.38%	25.21%	26.50%
同步發生 及 序列	1	q312a	46.25%	-	-
	1	q312b	63.64%	-	-
	2	q321	37.94%	40.56%	40.16%
	2	q322	37.87%	41.51%	-
數據結構 及 算法	1	q411a	84.68%	77.51%	76.10%
	1	q411b	78.11%	73.17%	70.87%
	1	q411c	49.92%	51.91%	52.22%
	3	q423a	83.87%	84.99%	87.19%
	3	q423b	86.53%	87.66%	90.23%
	3	q423c	73.37%	79.19%	82.52%

綜觀所有評核表格/問卷，如使用單參數 IRT 模型進行估算，其整體鑑別力從基線的 0.33 及 0.34 (第 1 級) 上升至中線的 0.62 及 0.64 (分指第 1 及第 2 級)，並於終線達到 0.57、0.71 及 0.50 的水平 (分指第 1、2 及 3 級)。上述結果顯示隨著先導學生的表現日漸進步，有關評核亦較能鑑別學生的能力高低。

運算思維實踐

我們按照 ECD 策略來設計和制訂運算思維實踐的評核方案。另外，為了檢視當中的一組項目，我們讓 8 名學生在回應/完成這些項目時口頭表達他們的想法，並

將他們的意見記錄下來，再行審視，以瞭解這些學生有否如預期般回應/完成這些項目，以及辨識他們在理解這些項目時有沒有遇上困難。我們其後根據學生的回饋作出修訂，使有關項目更加清晰。

對運算思維實踐評核方案的可靠性分析顯示，中線及終線的阿爾法系數分別為 0.61 及 0.78。我們同時使用了單因子模型及四因子模型進行數據分析，以檢視評核方案的內部結構。表 D3 顯示了這 2 個模型的因子分析結果。根據通用的適配標準，單因子模型與數據的適配度是足夠的。

表 D3 運算思維實踐終線評核表格驗證因子分析摘要 (N=1,808, J=26)

因子模型	卡方值	df	RMSEA	CFI
單因子模型	1386.751***	299	.045	.900
四因子模型	1079.168***	293	.039	.928

*** $p < .001$

四因子模型的適配度與單因子模型接近。不同評核元素之間的相關性高，在中線和終線分別介乎.62 至.84 及 .68 至 .86。從上述結果可見，將單因子模型用於運算思維實踐是恰當的，而在評核裡使用單一分數亦同樣合適。

由於運算思維實踐評核涉及大量項目，我們特意設計了 2 款可相互替代的表格/問卷，以進行驗證。學生會隨機獲派其中一款表格/問卷，即表 F 或表 G。兩者涵蓋的評核項目完全一樣，但項目的排列次序有別。中線分析結果顯示，項目正確率及項目缺失率與該項目在評核表格/問卷裡出現的位置有關：如該項目在評核的較後期出現，便會出現正確率偏低和缺失率較高的情況。由於存在上述次序效應，表 F 所顯示的學生整體表現較佳，可靠性及鑑別力亦較高。這情況與之前的發現吻合：學生進行中線評核表格/問卷時的速度過快，必然使正確率下跌。

進行終線評核時，我們修改了一些程序，以提升學生參與評核的動機。舉例說，學生獲提供額外紙張來回答/解決問題；老師則明確鼓勵學生全力以赴。

表 D4 顯示在不同的終線驗證表格/問卷裡，運算思維實踐項目的最高可能分數百分比。次序效應在某程度上仍然存在，但已不如中線結果般明顯。

運算思維視野

在之前發表的研究報告裡，我們曾經討論使用數個可以描述不同元素之間關係的替代性概念模型來評核該 7 項運算思維視野元素。這些分析的結論顯示，大部分項目能夠可靠地量度其配對的運算思維視野元素，並無一個概念框架的適配性大幅優於其他框架 (see Shear et al., 2019)。由始至終，撇開「歸屬感」這個元素，所有其他元素都是高度相關。

表 D5 顯示不同運算思維視野元素在中線的相關性。正如我們在主體報告中有關分析方法的章節裡指出，我們把 6 項元素(對編程的興趣、數碼自我效能感、效用動機、貢獻世界的動機、創意、參與/投入)綜合成一個量度運算思維視野的複合方法。這些元素有很高的相關性，複合量度結果的可靠性系數達到極高的 0.95。

表 D4 不同終線驗證表格/問卷裡的運算思維實踐項目最高可能分數百分比

元素	項目	最高可 能分數	所有表格	表格 F		表格 G	
			(N=1,808, J=26)	次序	最高 %	次序	最高 %
算法思維	q11a	1	56.64%	1	63.94%	23	48.57%
	q11b	1	50.68%	2	60.63%	24	39.79%
	q11c	1	51.10%	3	61.00%	25	40.31%
	q11d	1	40.30%	4	44.43%	26	35.79%
	q12aa	1	42.56%	16	43.30%	7	41.77%
	q12ab	1	32.29%	17	34.29%	8	30.18%
	q12ba	1	26.19%	18	26.76%	9	25.58%
	q12bb	1	26.90%	19	25.03%	10	28.89%
	q13	1	42.75%	20	44.10%	11	41.32%
		迷宮分析 (第 2 部分)					
重用及 整合	q215	1	46.65%	22	21.31%	1	24.03%
	q217	1	48.08%	23	44.80%	2	48.57%
	q21_mod	2	22.64%	24	42.34%	2	54.03%
	Q22a	1	25.31%	6	23.54%	19	27.26%
	q22b	1	19.88%	7	23.37%	20	16.05%
	q22c		16.40%	8	15.63%	21	17.23%
測試及 除錯	q32a	1	29.23%	9	34.61%	17	23.32%
	q32b	4	30.91%	10	35.76%	18	25.73%
	q34	1	29.11%	15	28.07%	12	30.21%
概念化及 模組化	q42	1	22.06%	13	67.90%	22	19.61%
	q43a	2	65.84%	14	33.02%	13	63.62%
	q43b	2	32.85%	25	23.80%	14	32.67%
	q44a	1	25.13%	26	11.96%	4	26.52%
	q44b		15.69%	21	52.01%	5	19.61%
測試及 除錯	q51	2	46.33%	21	52.01%	6	40.33%
概念化及 模組化	q52a	2	70.29%	11	68.75%	15	71.93%
算法思維	q52b	5	82.44%	12	83.35%	16	81.45%

表 D5 運算思維視野元素的相關性(終線) (N=16,341)

	歸屬感	興趣	參與/投入	貢獻世界的動機	創意	數碼自我效能感
對編程的興趣	.49***					
參與/投入	.49***	.93***				
貢獻世界的動機	.45***	.79***	.82***			
創意	.49***	.83***	.87***	.87***		
數碼自我效能感	.44***	.81***	.82***	.76***	.80***	
效用動機	.46***	.83***	.85***	.84***	.85***	.85***

*** $p < .001$