

# ĐỀ XUẤT KHUNG PHÂN TÍCH HỌC TẬP LẤY TRI THỨC LÀM TRUNG TÂM

Vũ Xuân Hạnh<sup>1</sup>, Bùi Anh Tuấn<sup>1\*</sup>, Ngô Văn Sơn<sup>1</sup>, Nguyễn Quang Ánh<sup>1</sup>

\*Tác giả liên hệ, Email: tuanba@hou.edu.vn. ORCID: 0009-0001-3046-0951

Ngày tòa soạn nhận được bài báo: 15/06/2026

Ngày phản biện đánh giá: 22/06/2026

Ngày bài báo được duyệt đăng: 14/09/2026

DOI: 10.59266/houjs.2026.1438

**Tóm tắt:** Bài báo đề xuất khung phân tích học tập lấy tri thức làm trung tâm (Knowledge-Centered Learning Analytics Framework - KCLAF) - một khung bốn lớp nhằm giải quyết học tập phân mảnh (Fragmented Learning - FL) và quá tải nhận thức (Cognitive Overload - CL) trong học tập số. Dữ liệu từ 266 sinh viên Trường Đại học Mở Hà Nội (bảng hỏi Likert 5 mức, 57 mục, 11 cấu trúc) được phân tích bằng phân tích nhân tố khám phá (Exploratory Factor Analysis - EFA) và phân tích nhân tố khẳng định (Confirmatory Factor Analysis - CFA), hồi quy bình phương nhỏ nhất thông thường (Ordinary Least Squares - OLS) phân lớp với sai số chuẩn vững HC3 (heteroskedasticity-consistent, dạng 3) và phân tích trung gian bằng kỹ thuật lấy mẫu lặp. FL ( $M = 3,556$ ) và CL ( $M = 3,597$ ) cao hơn có ý nghĩa so với điểm giữa thang đo, với 30,5% và 36,1% sinh viên đạt mức từ 4,0 trở lên. Tổ chức tri thức (Knowledge Organizing - OR) là biến dự báo học sâu mạnh nhất ( $\beta = 0,348$ ) và trung gian khoảng 44,9% liên hệ giữa sử dụng trí tuệ nhân tạo (Artificial Intelligence - AI) và học sâu (Deep Approach Learning - DL). Trái với giả thuyết ban đầu về nghịch lý AI-OR, hai biến tương quan dương ( $r = 0,458$ ); tuy nhiên chỉ số khoảng cách chuẩn hóa giữa chúng ở cấp cá nhân có liên hệ nghịch với học sâu sau khi kiểm soát mức độ chung ( $\beta = -0,127$ ;  $p = 0,046$ ). Nghiên cứu ghi nhận hai hạn chế đo lường: năm cấu trúc kiến tạo tri thức không tách bạch được về thực nghiệm và thiên lệch đồng thuận. KCLAF được định vị là khung khái niệm cần kiểm chứng, kèm lộ trình bốn giai đoạn.

**Từ khóa:** học sâu, học tập phân mảnh, phân tích học tập, quá tải nhận thức, tổ chức tri thức

## I. Đặt vấn đề

Sự phổ biến của các nền tảng học tập số - hệ thống quản lý học tập, Youtube, trang web học thuật và mô hình ngôn ngữ lớn đã thay đổi cách sinh viên tiếp cận nguồn tri thức. Tuy nhiên theo nghiên cứu quốc tế nhận ra một nghịch lý đáng lo ngại: mặc dù nguồn tài nguyên học tập

phong phú nhưng để hiểu bản chất, nắm rõ khái niệm và đưa vào thực tiễn lại không cải thiện tương xứng (Gašević và cộng sự, 2015). Điều này chứng minh rằng vấn đề không nằm ở việc thiếu tài nguyên và nằm ở thiếu cấu trúc học và cách học.

Hai cơ chế được tài liệu học thuật xác định là nguyên nhân chính: (i) học tập

<sup>1</sup> Trường Đại học Mở Hà Nội, Hà Nội, Việt Nam

phân mảnh (Fragmented Learning - FL) - học từ các video, bài viết, nội dung ngắn, ... rời rạc trên nhiều nền tảng, dẫn đến bức tranh tri thức không liên kết; (ii) quá tải nhận thức (Cognitive Overload - CL) - tình trạng hệ thống nhận thức bị quá tải khi xử lý lượng thông tin lớn (Dabbagh và Kitsantas, 2012; Sweller và cộng sự, 1998). Cả hai hiện tượng đều được khuếch đại trong môi trường đào tạo từ xa, nơi sinh viên chủ yếu tự học với ít hướng dẫn cấu trúc trực tiếp từ giảng viên.

Trường Đại học Mở Hà Nội (HOU) - với hình thức đào tạo từ xa lớn là bối cảnh nghiên cứu đặc biệt phù hợp. Đặc thù học tập phân tán, chủ yếu qua nền tảng số, làm cho FL và CL trở thành thách thức thường trực. Câu hỏi nghiên cứu của bài báo này là: (1) FL và CL có phải là vấn đề thực sự và có quy mô đo lường được tại HOU không? (2) Tồn tại cấu trúc nhân quả nào giữa các hành vi học tập và kết quả học sâu, và cấu trúc đó có thể được mô hình hóa để thiết kế hệ thống hỗ trợ không?

Nhóm nghiên cứu đề xuất KCLAF - khung phân tích học tập lấy tri thức làm trung tâm gồm bốn lớp - và cung cấp bằng chứng thực nghiệm ban đầu từ khảo sát 266 sinh viên HOU. Đóng góp cụ thể gồm: (1) khung KCLAF bốn lớp với cơ sở lý thuyết và đối chiếu với các khung phân tích học tập hiện có; (2) bộ dữ liệu thực nghiệm HOU với kiểm định mô hình đo lường đầy đủ và phân tích trung gian bằng lấy mẫu lặp; (3) một phương pháp thao tác hóa khoảng cách AI-OR ở cấp cá nhân; (4) lộ trình kiểm chứng bốn giai đoạn kèm chỉ số hiệu năng và các yêu cầu về quyền riêng tư, công bằng và giám sát của con người.

## II. Cơ sở lý thuyết

### 2.1. *Learning analytics, tải nhận thức và học tập phân mảnh*

Phân tích học tập (Learning Analytics) được định nghĩa là quá trình đo lường, thu thập, phân tích và báo cáo dữ liệu về người học nhằm hiểu và tối ưu hóa học tập (Siemens, 2013). Từ Hội nghị LAK năm 2011, lĩnh vực này phát triển nhanh với các hướng dự đoán kết quả, phát hiện nguy cơ bỏ học và cá nhân hóa nội dung (Long và Siemens, 2011). Tuy nhiên, Gašević và cộng sự (2015) cho rằng LA còn thiên về hành vi bề mặt, thiếu tích hợp lý thuyết học tập và ít chuyển hóa dự đoán thành can thiệp giảng dạy. Trong học tập số, hạn chế này càng rõ vì dữ liệu LMS thường phản ánh tần suất truy cập hoặc thời lượng làm bài hơn là chất lượng cấu trúc tri thức bên trong người học.

Học tập phân mảnh (Fragmented Learning) là khái niệm mô tả hiện tượng tiếp thu kiến thức từ các mảnh thông tin ngắn và thiếu liên kết cấu trúc, đặc biệt phổ biến thông qua các nền tảng video ngắn như TikTok giáo dục, YouTube Shorts hoặc tương tác với chatbot AI (Ophir và cộng sự, 2009). Các nghiên cứu thực nghiệm chỉ ra rằng FL dễ tạo ra ảo giác về sự hiểu biết nhưng thực tế không xây dựng được các lược đồ nhận thức chuyên sâu, dẫn đến hệ quả là khả năng vận dụng kém trong các tình huống thực tế mới (Dunlosky và cộng sự, 2013; Muharmi và cộng sự, 2025).

### 2.2. *AI trong học tập thực trạng, tiềm năng và khoảng trống nghiên cứu*

Kasneci và cộng sự (2023) chỉ ra các lợi ích như giải thích khái niệm, phản hồi tức thời, cá nhân hóa nội dung và mở

rộng khả năng tiếp cận học liệu. Ngược lại, nghiên cứu về “Google effects on memory” cho thấy khi thông tin có thể truy xuất dễ dàng, người học có xu hướng nhớ nơi tìm thông tin hơn là tổ chức nội dung (Sparrow và cộng sự, 2011). Trong bối cảnh AI tạo sinh, rủi ro này tăng lên vì hệ thống không chỉ tìm kiếm mà còn diễn giải, tóm tắt và đề xuất câu trả lời hoàn chỉnh.

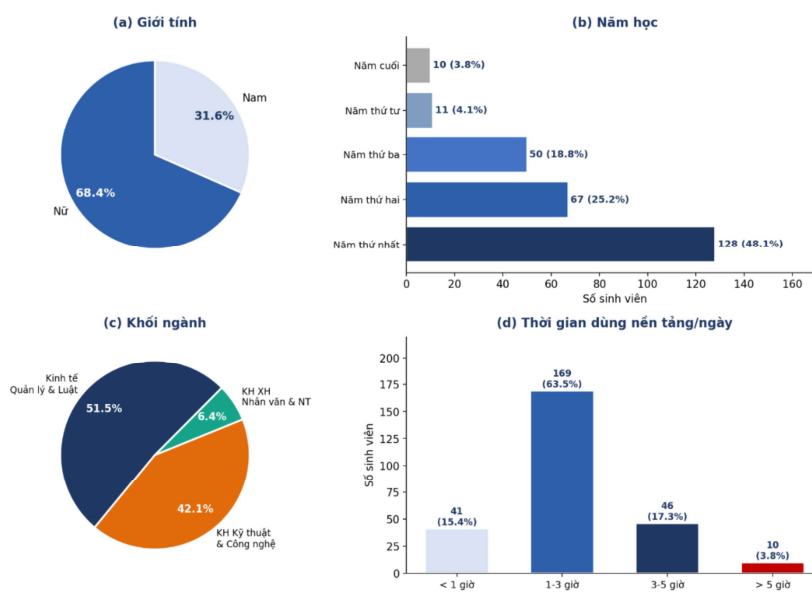
Các lý thuyết về định hướng mục tiêu, đánh giá nguồn và schema cung cấp nền tảng cho can thiệp trước khi người học đi vào tổ chức tri thức. Lý thuyết định hướng mục tiêu (*Goal Orientation Theory*) của Ames (1992) nhấn mạnh vai trò của định hướng làm chủ đối với hiểu biết sâu; mô hình của Metzger (2007) nhấn mạnh năng lực kiểm chứng và đánh giá nguồn trong môi trường web. Ở tầng sâu hơn, lý thuyết schema của Bartlett (1932) và lý thuyết học tập có ý nghĩa của Ausubel (như được thảo luận trong (Bryce và Blown, 2023) cho thấy tri thức mới chỉ bền vững khi được liên kết có ý nghĩa với cấu trúc tri thức đã có.

Dữ liệu thực nghiệm từ nghiên cứu này cho thấy sinh viên sử dụng AI nhiều nhất lại có điểm tổ chức tri thức thấp nhất, trong khi tổ chức tri thức là biến dự báo DL mạnh nhất. Đây chính là khoảng trống mà KCLAF hướng đến lấp đầy. Thay vì chỉ đo lường hành vi sử dụng nền tảng hoặc dự đoán điểm số, KCLAF đặt OR thành một lớp phân tích độc lập, nhằm theo dõi và hỗ trợ quá trình biến thông tin thành cấu trúc tri thức có thể giải thích, vận dụng và phát triển theo thời gian.

### III. Phương pháp nghiên cứu

#### 3.1. Thiết kế, mẫu và quy trình thu thập dữ liệu

Nghiên cứu sử dụng thiết kế khảo sát mặt cắt ngang nhằm kiểm chứng nhu cầu và cấu trúc quan hệ nền tảng cho khung KCLAF. Dữ liệu được thu thập bằng Google Forms trong học kỳ II năm học 2024-2025 tại HOU thông qua phương pháp chọn mẫu mục đích kết hợp thuận tiện từ danh sách lớp học phân của ba khối ngành. Sau khi loại bỏ phản hồi không hợp lệ, mẫu phân tích gồm 266 sinh viên HOU.



Hình 1. Đặc điểm của mẫu sinh viên HOU (n=266)

### 3.2. Công cụ đo lường

Bảng hỏi gồm 57 mục Likert 5 mức, đo 11 cấu trúc. Các thang đo được phát triển theo quy trình ba bước. Bước một, nhóm nghiên cứu xây dựng ngân hàng mục hỏi trên cơ sở nội hàm lý thuyết của từng khái niệm. Bước hai, đánh giá từng

mục hỏi theo hai tiêu chí: mức độ phù hợp với khái niệm và mức độ rõ ràng về diễn đạt, trên thang bốn mức. Các mục có chỉ số hiệu lực nội dung dưới 0,80 được chỉnh sửa hoặc loại bỏ. Bước ba, bảng hỏi được thử nghiệm trên một nhóm nhỏ sinh viên để hiệu chỉnh diễn đạt trước khi triển khai chính thức.

Bảng 1. Thang đo Likert 5 mức độ và độ tin cậy của các cấu trúc

| Mã | Lớp | Cấu trúc                                                  | Nội dung đo                                      | k | M (SD)           | $\alpha$ | CR    | AVE   |
|----|-----|-----------------------------------------------------------|--------------------------------------------------|---|------------------|----------|-------|-------|
| FL | L1  | Fragmented Learning (Học tập phân mảnh)                   | Học từ nội dung ngắn, rời rạc, đa nền tảng       | 6 | 3,556<br>(0,635) | 0,820    | 0,825 | 0,445 |
| CL | L1  | Cognitive Overload (Quá tải nhận thức)                    | Quá tải nhận thức khi xử lý thông tin số         | 5 | 3,597<br>(0,659) | 0,861    | 0,864 | 0,560 |
| GO | L2  | Goal Orientation (Định hướng mục tiêu)                    | Xác định mục tiêu và định hướng học tập          | 4 | 3,669<br>(0,643) | 0,871    | 0,872 | 0,629 |
| SE | L2  | Source Evaluation (Đánh giá nguồn thông tin)              | Đánh giá độ tin cậy nguồn thông tin              | 5 | 3,909<br>(0,615) | 0,892    | 0,892 | 0,624 |
| AI | L3  | AI Usage (Mức độ sử dụng trí tuệ nhân tạo)                | Mức độ và cách thức dùng AI trong học tập        | 6 | 3,863<br>(0,562) | 0,824    | 0,838 | 0,476 |
| OR | L3  | Knowledge Organizing (Tổ chức tri thức)                   | Xây dựng cấu trúc liên kết tri thức              | 5 | 3,573<br>(0,662) | 0,909    | 0,909 | 0,668 |
| EN | L3  | Encoding (Mã hóa thông tin)                               | Mã hóa và ghi chú có cấu trúc                    | 5 | 3,647<br>(0,637) | 0,877    | 0,878 | 0,590 |
| DP | L3  | Deep Processing (Xử lý sâu)                               | Tự giải thích, phản biện, mở rộng và vận dụng    | 6 | 3,633<br>(0,641) | 0,901    | 0,903 | 0,609 |
| VE | L3  | Verification & Retention (Kiểm chứng và lưu giữ tri thức) | Kiểm chứng thông tin, ôn tập và lưu trữ hệ thống | 5 | 3,765<br>(0,645) | 0,876    | 0,877 | 0,588 |
| DL | L4  | Deep Approach to Learning (Tiếp cận học sâu)              | Hiểu bản chất, vận dụng, chuyển giao             | 5 | 3,674<br>(0,577) | 0,860    | 0,866 | 0,569 |
| IN | L4  | Integration (Tích hợp tri thức)                           | Tích hợp tri thức mới với nền tảng cũ            | 5 | 3,729<br>(0,631) | 0,898    | 0,898 | 0,640 |

Ghi chú:  $k$  = số mục hỏi;  $\alpha$  = hệ số Cronbach's alpha; CR = độ tin cậy tổng hợp; AVE = phương sai trích trung bình

### 3.3. Mô hình nghiên cứu và giả thuyết

Mô hình quan hệ giả định gồm 4 lớp. Lớp L1 gồm hai vấn đề cấu trúc đầu vào (FL, CL). Lớp L2 gồm hai năng lực định hướng và lọc thông tin (GO, SE). Lớp L3

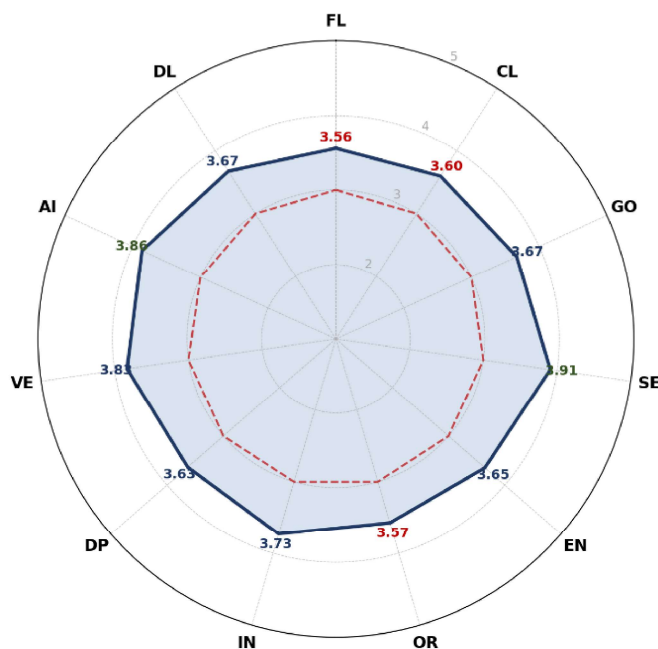
gồm các hành vi kiến tạo tri thức (AI, OR, EN, DP, VE). Lớp L4 gồm các kết quả học tập (DL, IN). Mô hình giả định rằng ảnh hưởng của L1 đến L4 được truyền qua L2 và L3 chứ không trực tiếp. Các giả thuyết được phát biểu ở Bảng 2.

Bảng 2. Giả thuyết nghiên cứu

| Mã  | Nội dung giả thuyết                                                                                                                                      | Cơ sở lý thuyết                                        |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|
| H1a | Điểm FL và CL của sinh viên cao hơn có ý nghĩa so với điểm giữa của thang đo                                                                             | Sweller và cộng sự (1998); Dabbagh và Kitsantas (2012) |
| H1b | FL và CL tương quan dương với nhau                                                                                                                       | Lý thuyết phân mảnh dữ liệu                            |
| H2a | GO và SE tương quan dương với DL                                                                                                                         | Ames (1992); Metzger (2007)                            |
| H2b | FL và CL tương quan âm với GO và SE                                                                                                                      | Kiểm soát chú ý (Ophir và cộng sự, 2009)               |
| H2c | GO và SE trung gian liên hệ giữa L1 và DL                                                                                                                | Mô hình quá trình nhận thức phân tầng                  |
| H3a | AI, OR, EN tương quan dương với DL                                                                                                                       | Bartlett (1932); Bryce và Blown (2023)                 |
| H3b | Trong nhóm L3, OR là biến dự báo DL mạnh nhất                                                                                                            | Lý thuyết lược đồ và học tập có ý nghĩa                |
| H3c | OR trung gian liên hệ giữa AI và DL                                                                                                                      | Sparrow và cộng sự (2011); Risko và Gilbert (2016)     |
| H4a | Mô hình bốn lớp giải thích phần lớn phương sai của DL                                                                                                    | Tổng hợp các cơ sở trên                                |
| H4b | Việc bổ sung L3 làm tăng $R^2$ so với mô hình chỉ có L1 và L2                                                                                            | Vai trò độc lập của lớp tổ chức tri thức               |
| H5  | Khoảng cách chuẩn hóa ở cấp cá nhân giữa mức sử dụng AI và mức tổ chức tri thức có liên hệ nghịch với DL, sau khi kiểm soát mức độ chung của hai hành vi | Sparrow và cộng sự (2011); Kasneci và cộng sự (2023)   |

#### IV. Kết quả và thảo luận

##### 4.1. Kiểm chứng mức độ hiện diện của FL và CL



Hình 2. Radar chart điểm trung bình 11 constructs theo 4 lớp KCLAF (n = 266)

FL ( $M = 3,556$ ;  $SD = 0,635$ ) và CL ( $M = 3,597$ ;  $SD = 0,659$ ) đều cao hơn có ý nghĩa so với điểm giữa 3,0:  $t(265) = 14,28$  và  $14,78$ ;  $p < 0,001$ . H1a được ủng hộ. Cần diễn giải hệ số hiệu ứng thận trọng: mốc 3,0 là điểm giữa quy ước chứ không phải ngưỡng chẩn đoán độc

lập, nên  $d = 0,876$  và  $0,906$  chỉ phản ánh khoảng cách so với mốc đó. Bằng chứng trực tiếp hơn là phân bố theo ngưỡng: 55,3% sinh viên có điểm FL từ 3,5 trở lên và 30,5% từ 4,0 trở lên; với CL là 54,1% và 36,1%. Tương quan FL-CL đạt 0,580, ủng hộ H1b.



Do có phương sai thay đổi (Breusch-Pagan LM = 21,69;  $p = 0,003$ ), sai số chuẩn ở Bảng 3 báo cáo theo phương pháp vững HC3. Mô hình đầy đủ giải thích 67,8% phương sai của DL, ủng hộ H4a; bổ sung lớp L3 làm tăng  $R^2$  thêm 5,5 điểm phần trăm, ủng hộ H4b. OR là biến

dự báo mạnh nhất ( $\beta = 0,348$ ;  $p < 0,001$ ), tiếp theo GO (0,243), SE (0,217) và AI (0,135); H3b được ủng hộ. EN không có ý nghĩa ( $\beta = -0,032$ ), phù hợp nhận định rằng phương sai chung của cặp OR-EN chủ yếu được OR hấp thụ.

Bảng 3: Hồi quy OLS phân lớp với biến phụ thuộc DL

| Mô hình       | Biến | $\beta$ | SE    | p       | KTC 95%         | VIF  | $R^2$ | $\Delta R^2$ |
|---------------|------|---------|-------|---------|-----------------|------|-------|--------------|
| M1 (L1)       | FL   | 0,394   | 0,088 | < 0,001 | [0,222; 0,566]  | 1,51 | 0,229 | —            |
|               | CL   | 0,127   | 0,081 | 0,119   | [-0,033; 0,287] | 1,51 |       |              |
| M2 (L1+L2)    | FL   | 0,154   | 0,063 | 0,014   | [0,031; 0,278]  | 1,65 | 0,623 | +0,394       |
|               | CL   | 0,005   | 0,053 | 0,924   | [-0,098; 0,108] | 1,56 |       |              |
|               | GO   | 0,446   | 0,066 | < 0,001 | [0,316; 0,576]  | 1,97 |       |              |
|               | SE   | 0,320   | 0,060 | < 0,001 | [0,203; 0,438]  | 2,04 |       |              |
| M3 (L1+L2+L3) | FL   | 0,090   | 0,056 | 0,105   | [-0,019; 0,199] | 1,76 | 0,678 | +0,055       |
|               | CL   | -0,010  | 0,052 | 0,848   | [-0,111; 0,091] | 1,70 |       |              |
|               | GO   | 0,243   | 0,081 | 0,003   | [0,085; 0,401]  | 3,07 |       |              |
|               | SE   | 0,217   | 0,069 | 0,002   | [0,082; 0,352]  | 2,58 |       |              |
|               | AI   | 0,135   | 0,065 | 0,039   | [0,007; 0,263]  | 2,13 |       |              |
|               | OR   | 0,348   | 0,081 | < 0,001 | [0,189; 0,507]  | 3,52 |       |              |
|               | EN   | -0,032  | 0,090 | 0,726   | [-0,208; 0,145] | 4,59 |       |              |

Ghi chú: Hệ số  $\beta$  là hệ số chuẩn hóa; SE là sai số chuẩn vững HC3.

Mô hình bổ sung DP và VE cho  $R^2 = 0,695$ : OR vẫn giữ hệ số cao nhất nhóm L3 ( $\beta = 0,265$ ) cùng GO ( $\beta = 0,238$ ), trong khi các biến còn lại không đạt ngưỡng 0,05 và đa cộng tuyến tăng (VIF tối đa 4,86). Hệ số của FL dương và có ý nghĩa ở M1 ( $\beta = 0,394$ ) nhưng mất ý nghĩa khi bổ sung L2 và L3.

#### 4.3. Phân tích trung gian

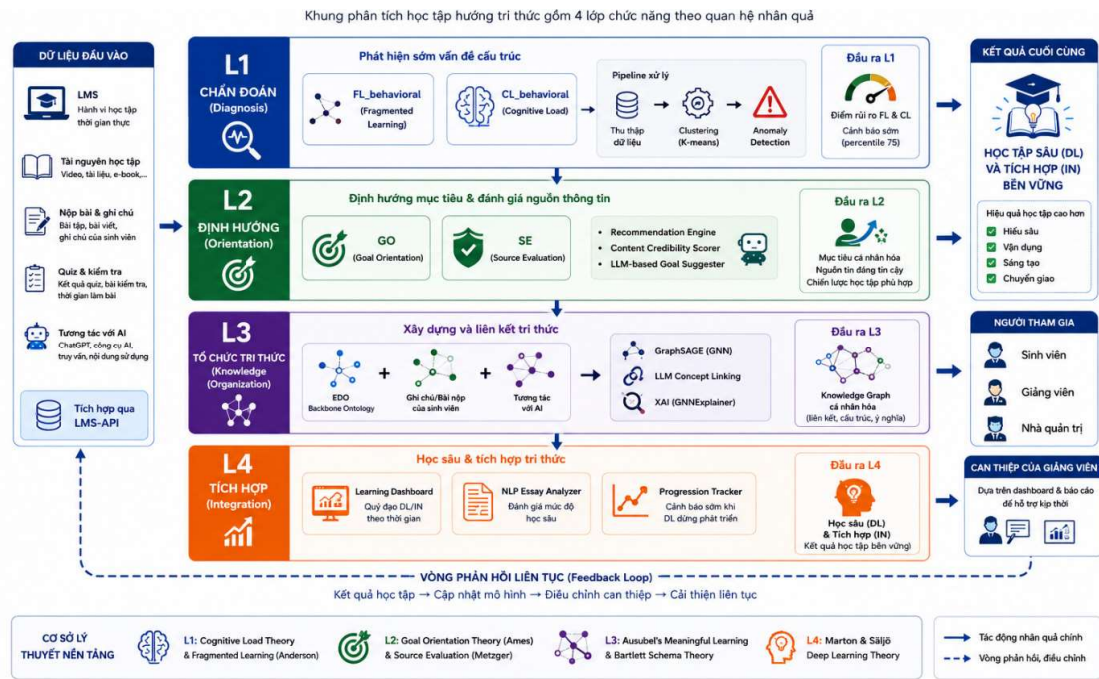
Phân tích trung gian với 5.000 mẫu lặp xác nhận H2c và H3c, vì tất cả

khoảng tin cậy 95% đều không chứa 0. Con đường AI→OR→DL có indirect effect=0,273, CI [0,170; 0,394], tương đương khoảng 46% tổng quan hệ AI-DL. Hai con đường từ FL qua L2 cũng có cường độ đáng kể: FL→SE→DL đạt 0,225 và FL→GO→DL đạt 0,222. Ngoài các giả thuyết ban đầu, con đường CL→GO→DL cũng có ý nghĩa, hàm ý rằng quá tải nhận thức có thể làm suy giảm định hướng mục tiêu, từ đó ảnh hưởng tiêu cực đến học sâu.

Bảng 4. Phân tích trung gian với 5000 mẫu lặp

| Đường dẫn    | a (X→M) | b (M→Y X) | c tổng | ab gián tiếp | KTC 95% của ab | ab/c  |
|--------------|---------|-----------|--------|--------------|----------------|-------|
| AI → OR → DL | 0,458   | 0,579     | 0,591  | 0,265        | [0,165; 0,378] | 44,9% |
| AI → EN → DL | 0,520   | 0,549     | 0,591  | 0,285        | [0,196; 0,385] | 48,3% |
| FL → GO → DL | 0,399   | 0,648     | 0,467  | 0,259        | [0,154; 0,381] | 55,3% |
| FL → SE → DL | 0,412   | 0,604     | 0,467  | 0,249        | [0,156; 0,350] | 53,3% |
| CL → GO → DL | 0,318   | 0,688     | 0,355  | 0,219        | [0,115; 0,330] | 61,6% |

#### 4.4. Khung KCLAF bốn lớp từ bằng chứng thực nghiệm



Hình 4. Khung phân tích học tập hướng tri thức

KCLAF được đề xuất như một khung phân tích học tập hướng tri thức gồm bốn lớp chức năng có quan hệ nhân quả theo thứ bậc. L1 - Chẩn đoán - đo FL và CL từ dữ liệu LMS, hoạt động học tập, quiz, bài nộp và tương tác AI. Cơ sở của L1 là CLT và nghiên cứu về học tập online (Sweller và cộng sự, 1998). L1 có thể triển khai pipeline LMS-API, chuẩn hóa dữ liệu hành vi, phân cụm mẫu học tập và cảnh báo sớm khi chỉ số FL/CL vượt ngưỡng.

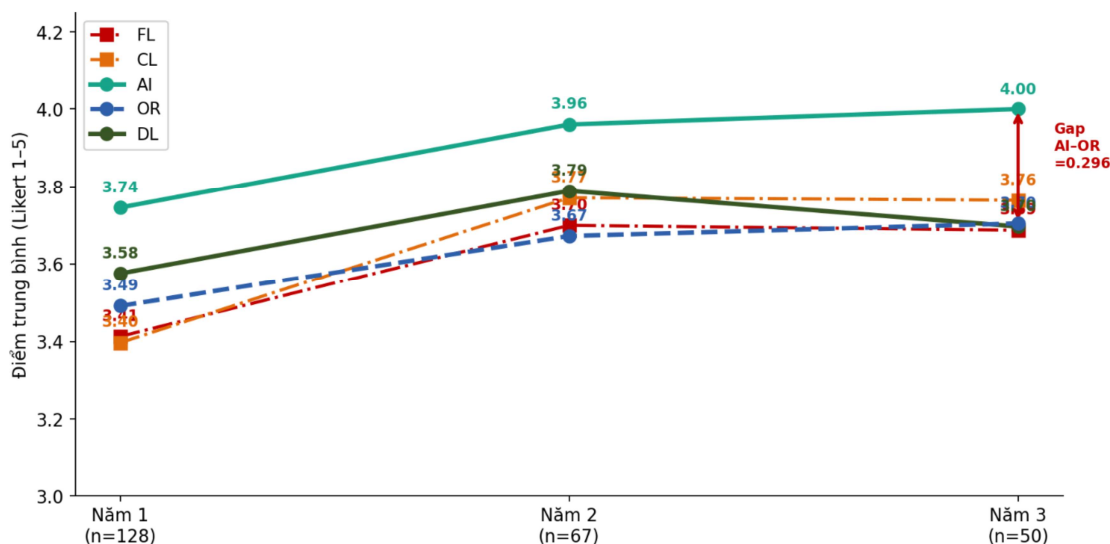
L2 - Định hướng - trả lời câu hỏi làm thế nào để người học có mục tiêu rõ ràng và đánh giá được nguồn thông tin. Lớp này dựa trên Goal Orientation Theory của Carole (1992) và mô hình đánh giá nguồn của Metzger (2007). Kết quả hồi quy và trung gian cho thấy GO và SE không chỉ dự báo DL mà còn hấp thụ một phần đáng kể ảnh hưởng của FL. L2 có thể bao gồm recommendation engine, bộ chấm độ tin cậy nội dung, gợi ý mục tiêu học tập bằng

LLM và phản hồi cá nhân hóa cho sinh viên có dấu hiệu học phân mảnh.

L3 - Tổ chức tri thức - là lớp cốt lõi của KCLAF. Lớp này đặt OR làm trung tâm, vì OR vừa là biến dự báo DL mạnh nhất vừa trung gian gần một nửa hiệu ứng AI→DL. Về mặt kỹ thuật, L3 có thể xây dựng đồ thị tri thức cá nhân từ môn học, ghi chú, bài nộp và hội thoại với AI. Các thành phần LLM không nhằm thay thế người học, mà nhằm phát hiện khái niệm thiếu, gợi ý liên kết và buộc người học kiểm tra lại cấu trúc tri thức của mình.

L4 - Tích hợp - đo mức độ học sâu và tích hợp tri thức. Lớp này dựa trên phân loại học sâu của Marton và Säljö (1976), trong đó DL được hiểu là khả năng hiểu ý nghĩa bên dưới bề mặt thông tin, kết nối với tri thức nền và vận dụng sáng tạo. L4 có thể phát hiện dấu hiệu học sâu trong bài viết, cảnh báo khi DL dừng phát triển và cung cấp báo cáo có thể hành động cho giảng viên.

#### 4.5. Phân tích khoảng cách AI-OR



Hình 5. Xu hướng điểm trung bình theo năm học

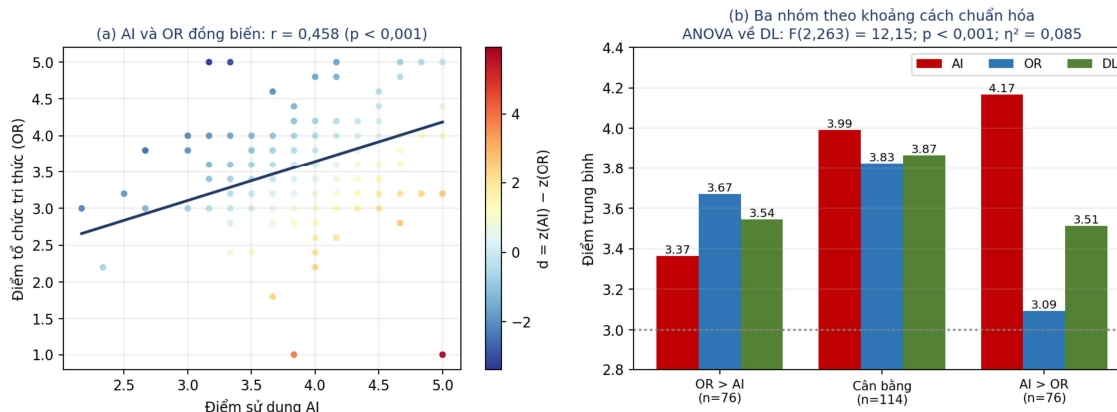
Phân tích theo năm học cho thấy FL tăng có ý nghĩa từ năm 1 ( $M=3,413$ ) đến năm 3 ( $M=3,687$ ;  $t=-2,819$ ,  $p=0,005$ ,  $d=0,496$ ), CL tăng từ 3,397 lên 3,764 ( $t=-3,609$ ,  $p<0,001$ ,  $d=0,618$ ). Tuy nhiên, DL chưa khác biệt có ý nghĩa ( $t=-1,241$ ,  $p=0,216$ ,  $d=0,211$ ), mặc dù có xu hướng tăng nhẹ. Kết quả này gợi ý môi trường học tập hiện tại làm FL và CL tăng theo thời gian, nhưng chưa chuyển hóa được áp lực học tập thành học sâu tương ứng.

Sinh viên Kinh tế - Quản lý - Luật có FL và CL cao hơn nhóm Khoa học kỹ thuật và Công nghệ ( $p<0,01$ ), nhưng DL không khác biệt ( $p=0,458$ ). Điều này gợi ý L2-L4 có thể dùng mô hình thống nhất, còn L1 cần tham số hóa ngưỡng phát hiện theo khối ngành. Nói cách khác, cường độ đọc và xử lý tài liệu đa dạng có thể khác nhau giữa các ngành, nhưng nhu cầu định hướng, tổ chức tri thức và tích hợp tri thức là nhu cầu chung.

Một phân tích ban đầu có thể so sánh  $M(AI) = 3,863$  với  $M(OR) = 3,573$  và kết luận rằng sinh viên dùng AI nhiều

nhưng tổ chức tri thức kém. Nhóm bác bỏ cách lập luận này vì AI và OR là hai thang đo khác nội dung và phân phối nên chênh lệch trung bình không có giá trị diễn giải; hơn nữa dữ liệu cho kết quả ngược lại:  $r(AI, OR) = 0,458$  ( $p < 0,001$ ) và nhóm dùng AI nhiều (từ phân vị 75,  $n = 75$ ) có điểm OR cao hơn rõ rệt nhóm dùng ít ( $n = 74$ ):  $M = 3,904$  so với 3,224;  $d = 1,04$  (Hình 6a).

Để kiểm định H5 hợp lệ, nhóm nghiên cứu xây dựng chỉ số khoảng cách chuẩn hóa ở cấp cá nhân  $d_i = z(AI_i) - z(OR_i)$ ; giá trị dương nghĩa là vị thế tương đối về mức dùng AI cao hơn vị thế về tổ chức tri thức. Trong mẫu, 36,1% sinh viên có  $d_i$  dương và 28,6% vượt nửa độ lệch chuẩn. Hình 6b cho thấy nhóm có  $d$  dương lớn đạt mức dùng AI cao nhất (4,167) nhưng có điểm tổ chức tri thức (3,095) và học sâu (3,513) thấp nhất, trong khi nhóm cân bằng đạt học sâu cao nhất (3,867);  $F(2, 263) = 12,15$ ;  $\eta^2 = 0,085$ .



Hình 6. Quan hệ AI-OR: (a) phân tán; (b) ba nhóm theo chỉ số khoảng cách  $d = z(AI) - z(OR)$

## V. Kết luận và kiến nghị

Nghiên cứu này kiểm chứng tính cần thiết và cấu trúc nền tảng của KCLAF trong bối cảnh Đại học Mở Hà Nội. Kết quả cho thấy FL và CL là hai vấn đề học tập số có quy mô đo lường được; OR là biến dự báo học sâu mạnh nhất; và khoảng cách AI-OR chỉ ra rằng việc sử dụng AI đơn thuần không bảo đảm học sâu nếu thiếu cơ chế tổ chức tri thức. Về lý luận, nghiên cứu bổ sung lớp tổ chức tri thức vào learning analytics; về thực tiễn, nghiên cứu gợi ý cách chuyển dữ liệu học tập thành can thiệp cụ thể thay vì chỉ dự đoán kết quả.

Giai đoạn tiếp theo đề xuất lộ trình xây dựng nguyên mẫu bốn pha trong 12 tháng: (1) dữ liệu hóa L1 qua LMS-API; (2) can thiệp L2 bằng module GO/SE; (3) xây dựng knowledge graph L3; (4) đánh giá tác động DL bằng quasi-experimental design, với KPI baseline từ dữ liệu  $n=266$ . Thiết kế waitlist/crossover được chọn để đảm bảo mọi sinh viên đều được tiếp cận can thiệp, đồng thời tạo dữ liệu đối chứng. Hiệu ứng carry-over sẽ được kiểm soát bằng biến đồng biến thời điểm tham gia.

Nghiên cứu còn một số giới hạn. Thiết kế cắt ngang chưa cho phép khẳng định quan hệ nhân quả theo nghĩa thực nghiệm; thang đo tự báo cáo có thể chịu thiên lệch nhận thức; mẫu khảo sát tập trung ở HOU nên cần kiểm định thêm ở các cơ sở khác. Các nghiên cứu tiếp theo nên triển khai dữ liệu log thời gian thực, đánh giá dọc nhiều học kỳ và so sánh các mô hình can thiệp để xác định hiệu quả thực tế của KCLAF đối với học sâu.

## Tài liệu tham khảo

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261-271.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.
- Bryce, T. G. K., & Blown, E. J. (2023). Ausubel's meaningful learning revisited. *Current Psychology*, 1-20. <https://doi.org/10.1007/s12144-023-04440-4>
- Dabbagh, N., & Kitsantas, A. (2012). Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and Higher Education*, 15(1), 3-8. <https://doi.org/10.1016/j.iheduc.2011.06.002>

- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58. <https://doi.org/10.1177/1529100612453266>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64-71. <https://doi.org/10.1007/s11528-014-0822-x>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & 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>
- Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 31-40.
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I-Outcome and process. *British Journal of Educational Psychology*, 46(1), 4-11. <https://doi.org/10.1111/j.2044-8279.1976.tb02980.x>
- Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58(13), 2078-2091. <https://doi.org/10.1002/asi.20672>
- Muharmi, Y., Hermanto, B., & Almadaea, V. (2025). Information system for guidance on student practical work reports in the computer science department of the web-based information management study program. *Jurnal Pepadun*, 6, 230-240. <https://doi.org/10.23960/pepadun.v6i3.301>
- Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences*, 106(37), 15583-15587. <https://doi.org/10.1073/pnas.0903620106>
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676-688. <https://doi.org/10.1016/j.tics.2016.07.002>
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400. <https://doi.org/10.1177/0002764213498851>
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043), 776-778. <https://doi.org/10.1126/science.1207745>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251-296. <https://doi.org/10.1023/A:1022193728205>

## PROPOSING A KNOWLEDGE-CENTERED LEARNING ANALYTICS FRAMEWORK

Vu Xuan Hanh<sup>1</sup>, Bui Anh Tuan<sup>1</sup>, Ngo Van Son<sup>1</sup>, Nguyen Quang Anh<sup>1</sup>

**Abstract:** *This paper proposes a Knowledge-Centered Learning Analytics Framework (KCLAF) - a four-layer framework addressing fragmented learning (FL) and cognitive overload (CL) in digital learning. Data from 266 students at Hanoi Open University (a 57-item, five-point Likert questionnaire measuring 11 constructs) were analyzed using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), hierarchical ordinary least squares (OLS) regression with HC3 heteroskedasticity-consistent robust standard errors, and mediation analysis based on resampling. FL ( $M = 3.556$ ) and CL ( $M = 3.597$ ) were significantly above the scale midpoint, with 30.5% and 36.1% of students scoring at or above 4.0. Knowledge organizing (OR) was the strongest predictor of deep learning ( $\beta = 0.348$ ) and mediated approximately 44.9% of the association between artificial intelligence (AI) use and deep approach learning (DL). Contrary to the initially hypothesized AI-OR paradox, the two constructs were positively correlated ( $r = 0.458$ ); however, an individual-level standardized discrepancy index between them was negatively associated with deep learning after controlling for their common level ( $\beta = -0.127$ ;  $p = 0.046$ ). Two measurement limitations are documented: the five knowledge-construction constructs were not empirically separable, and acquiescence bias was present. KCLAF is positioned as a conceptual framework requiring validation, accompanied by a four-stage roadmap.*

**Keywords:** *cognitive overload, deep learning, fragmented learning, knowledge analytics, knowledge organizing*

---

<sup>1</sup> Hanoi Open University, Hanoi, Vietnam