

XÁC ĐỊNH MỐI TƯƠNG QUAN GIỮA HOẠT ĐỘNG HỌC TẬP VÀ KẾT QUẢ HỌC TẬP HỌC PHẦN TIẾNG ANH CƠ BẢN DỰA TRÊN PHÂN TÍCH DỮ LIỆU NHẬT KÝ TRÊN HỆ THỐNG ĐÀO TẠO TRỰC TUYẾN TẠI TRƯỜNG ĐẠI HỌC MỞ HÀ NỘI

Đặng Hải Đăng*

Email: dangdh@hou.edu.vn

Ngày tòa soạn nhận được bài báo: 03/05/2023

Ngày phản biện đánh giá: 02/11/2023

Ngày bài báo được duyệt đăng: 27/11/2023

DOI:

Tóm tắt: Việc phân tích, đánh giá kết quả học tập dựa trên quá trình đối với sinh viên đại học thường được đánh giá bởi các công cụ thủ công như điểm danh, cho điểm quá trình dựa trên các bài kiểm tra thành phần và được thực hiện trực tiếp bởi giảng viên. Việc này làm tăng khối lượng các công tác phi giảng dạy không những của giảng viên mà còn của đội ngũ hỗ trợ đào tạo do phải nhận và nhập lại dữ liệu trên hệ thống. Đối với hệ thống đào tạo trực tuyến, các công việc này không những được thu thập tự động mà còn có thể tích hợp thêm các dữ liệu khác của người học khi thao tác với hệ thống. Nghiên cứu này phân tích dữ liệu trên nhật ký hệ thống và các yếu tố nhân khẩu học bằng việc sử dụng phân tích thống kê để tìm ra các quan hệ giữa các thuộc tính của người học trên hệ thống. Học phần được lựa chọn để phân tích dữ liệu là tiếng Anh cơ bản do có dữ liệu điểm số phong phú với 4 kỹ năng. Qua đó phát hiện được các mối quan hệ giữa nhân khẩu học và hoạt động học tập trên hệ thống đối với kết quả học tập môn tiếng Anh cơ bản trên nền tảng đào tạo trực tuyến.

Từ khóa: Đào tạo trực tuyến, hoạt động học tập, thống kê mô tả, kết quả học tập, phân tích dữ liệu

I. Đặt vấn đề

Có thể nói, thời gian gần đây, đào tạo trực tuyến được áp dụng rộng rãi và được xã hội chấp nhận nhiều hơn. Đặc biệt trong giai đoạn Covid-19, đây được coi là phương thức thay thế cho đào tạo truyền thống. Một điểm đặc biệt của đào tạo trực tuyến là khả năng ghi lại toàn bộ các hoạt động dạy và học trên hệ thống, từ hình ảnh, âm thanh đến các thao tác của cá nhân lên hệ thống. Với lượng dữ liệu lớn nhật ký được lưu trữ trên các hệ thống LMS, chúng ta có thể sử dụng phân tích thống kê tùy thuộc vào đặc điểm dữ liệu để phát hiện ra các mối quan hệ giữa các biến số, sử dụng để điều chỉnh, dự đoán các biến số và theo dõi các kết quả tương ứng.

Trong khi đó, với đào tạo truyền thống thì việc thu thập, xử lý dữ liệu chủ yếu thực hiện bằng thao tác thủ công và phụ thuộc vào kỹ năng, sự cẩn thận của con người. Lượng dữ liệu được

* Trường Đại học Mở Hà Nội

xử lý bằng tay nên ngoài việc sai sót, thiếu là những nguyên nhân chủ yếu dẫn đến việc đưa ra quyết định dựa trên số liệu thường là thiếu chính xác, gây tốn kém, lãng phí nguồn lực.

II. Cơ sở lý luận

Việc phân tích các yếu tố ảnh hưởng đến kết quả học tập đã được nhiều nhóm tác giả thực hiện. Đối với nghiên cứu trên đối tượng sinh viên chính quy được nhóm nghiên cứu do Hóa Đ.T và cộng sự rút ra được các yếu tố ảnh hưởng đến kết quả học tập bao gồm tương tác lớp học, phương pháp học, kiên định học, động cơ học, bạn bè, cơ sở vật chất, ấn tượng trường học kiến thức và cách tổ chức môn học. [1]

Trong khi đó, nghiên cứu của Huyền Đ.T.T và cộng sự [2] sử dụng các yếu tố ấn tượng trường học, động cơ / phương pháp học và vai trò của cố vấn học tập cho kết quả các yếu tố giới tính, ban cán sự lớp, làm thêm không có mối tương quan với kết quả học tập. Động cơ học, phương pháp học, vai trò cố vấn học tập có mối liên quan có ý nghĩa thống kê với kết quả học tập.

Nghiên cứu của Thương P.T.T và cộng sự [3] sử dụng mô hình TAM và áp dụng thống kê Bayes trên các nhân tố ảnh hưởng như: cảm nhận dễ sử dụng / hữu ích, năng lực giảng viên, nội dung môn học, thiết kế khóa học và đặc trưng người học đối với tham gia học trực tuyến trong giai đoạn dịch bệnh Covid-19. Các nghiên cứu này chủ yếu sử dụng các công cụ là phỏng vấn sâu hoặc bảng hỏi để thu thập dữ liệu

Mansureh Kebritch và cộng sự [7] đã phân loại các vấn đề đối với đào tạo trực tuyến bao gồm: vấn đề với người học / người dạy và nội dung. Nội dung của nghiên cứu này tập trung vào các hoạt động học được lưu trữ trên hệ thống.

Trong khi nghiên cứu của Mohammed Habes và cộng sự [8] bao gồm các yếu tố: ý định hành vi, cam kết, phương tiện sử dụng, tương tác giữa học viên, tương tác giữa giáo viên thì nghiên cứu của Sandra Ricart Casadevall tập trung vào các yếu tố nhân khẩu học, sự cam kết, lịch sử quá trình học và điểm số của MOOC về phân tích địa lý của các thảm họa tự nhiên [6]. Đây chính là gợi ý cho việc thực hiện nghiên cứu này khi phân tích các hoạt động lưu lại trên hệ thống để tìm ra các mối tương quan giữa hoạt động học và kết quả học tập.

III. Đối tượng và phương pháp nghiên cứu

Thông thường các nghiên cứu xã hội học thường được triển khai bằng việc sử dụng phương pháp khảo sát bảng hỏi thu thập dữ liệu từ các đối tượng được phỏng vấn. Các dữ liệu nhân khẩu học như giới tính, bằng cấp đầu vào, độ tuổi, tình trạng hôn nhân, điều kiện sống...được đưa vào bảng hỏi để tìm hiểu xem các nhân tố này có ảnh hưởng gì đối với mục tiêu nghiên cứu. Tuy nhiên, ngoài việc dữ liệu là chủ quan do dựa trên thành kiến và kinh nghiệm của người tham thì các câu hỏi được thiết kế có thể mang tính thiên vị hoặc dẫn dắt, gây ảnh hưởng đến câu trả lời của người tham gia [10].

Thay vì sử dụng phương pháp khảo sát bảng hỏi, bài báo đề cập đến việc sử dụng chính bộ dữ liệu nhật ký được sinh ra trong quá người học tham gia hệ thống. Dữ liệu được tạo ra trong quá trình sử dụng tồn tại độc lập khách quan với nhận thức của người học do đó đảm bảo được tính khách quan và có lượng dữ liệu đủ lớn để thực hiện các mô hình phân tích trên đó. Phương pháp

nghiên cứu được triển khai ở đây là phương pháp nghiên cứu số liệu sử dụng các thống kê mô tả, nhân tố khám phá.

Đối tượng nghiên cứu là bộ dữ liệu sinh ra trong quá trình sinh viên học trực tuyến môn tiếng Anh cơ bản 1 trên LMS cùng với dữ liệu trên hệ thống thông tin tác nghiệp quản lý đào tạo. Dữ liệu trên LMS được ghi nhận trên log file (tệp nhật ký) bao gồm 635 quan sát, trên 5 bảng dữ liệu sinh viên theo lớp hành chính. Ngoài ra dữ liệu này còn được ghép thêm một số thuộc tính dữ liệu nhân khẩu học và điểm thi được lưu trên hệ thống nghiệp quản lý đào tạo. Theo quy định cụ thể của học phần Tiếng Anh trong đề cương chi tiết của chương trình đào tạo trực tuyến, điểm tổng kết học phần sẽ được tính theo công thức như sau:

Điểm TKHP (tổng kết học phần) = 0.1 x điểm QT(quá trình) + 0.2 x điểm KT (kiểm tra) + 0.7 x điểm thi KTHP (kết thúc học phần)

Trong đó: Điểm QT được tính bằng bình quân kết quả của 4 bài luyện tập trắc nghiệm, điểm KT được tính bằng bình quân của 2 bài kiểm tra trắc nghiệm, điểm thi KTHP được tính bằng bình quân 4 kỹ năng Nghe, Nói, Đọc, Viết.

Log file trên hệ thống ghi lại rất nhiều hoạt động thao tác của sinh viên, giảng viên, quản lý lớp học phần.v.v. Do đó để có thể sử dụng log file thì cần phải tiền xử lý để loại bỏ các hoạt động, thao tác của các đối tượng khác sinh viên và các thao tác của sinh viên không phục vụ cho việc phân tích. Quá trình này tiền xử lý bao gồm 2 bước là làm sạch dữ liệu [11] và biến đổi dữ liệu phù hợp với mục đích sử dụng [5]. Dưới đây là các bước cụ thể được áp dụng để làm sạch bộ dữ liệu trong nghiên cứu này [4].

Bảng 1. Các bước làm sạch và biến đổi dữ liệu

Các bước làm sạch dữ liệu	Thao tác cụ thể
Bước 1: Loại bỏ trùng lặp hoặc các quan sát không liên quan	5 sheet dữ liệu ứng với 5 lớp học hành chính, lớp học mã số 100-102 dành cho sinh viên theo học tại Đà Nẵng, số sinh viên này do phân cấp quản lý nên chỉ có điểm TP và QT mà không có điểm thi TT, do vậy loại bỏ toàn bộ số sinh viên này Ngoài ra , một số sinh viên bỏ thi, hoãn thi, được miễn môn nhưng có trong danh sách viên cũng không có điểm thi TT.
Bước 2: Chỉnh sửa các cấu trúc bị lỗi	Nhìn chung các cấu trúc dữ liệu nhật ký được định nghĩa từ hệ thống LMS nên không có cấu trúc bị lỗi, sai định dạng dữ liệu.
Bước 3: Lọc các dữ liệu ngoại vi không mong muốn	Có một số hoạt động chỉ có một số sinh viên sử dụng (các hoạt động không được tính điểm) nên số lượng thuộc tính dữ liệu sử dụng được chỉ còn 18 đặc điểm (thay vì 28 thuộc tính)
Bước 4: Xử lý dữ liệu bị khuyết, thiếu	Các hoạt động chủ yếu trên hệ thống là click tuy nhiên có một số dữ liệu da
Bước 5: Xác nhận hợp lệ dữ liệu	Các dữ liệu chuẩn hóa và được format ví dụ dữ liệu ngày, tháng, năm
Biến đổi dữ liệu	Thao tác cụ thể

Bước 1. Gộp dữ liệu	Dữ liệu từ 5 sheet được gộp thành 1 sheet để phù hợp với việc phân tích trên dữ liệu tổng thể.
Bước 2. Biến đổi dữ liệu phù hợp theo yêu cầu	Trong dữ liệu ngành học đầu vào, do khâu nhập liệu nên có rất nhiều định dạng viết tắt hoặc không cho cùng một loại dữ liệu phân loại: Ví dụ PTTH và phổ thông trung học, ngoài ra một số hạng mục dữ liệu phân loại quá chi tiết ví dụ TCCN (trung cấp cùng ngành), TCKN (trung cấp khác ngành) cũng được sửa và gộp thành TC (trung cấp)
Bước 3. Tạo sinh các trường mới	Các trường mới sinh bao gồm tổng số click cho các hoạt động tính điểm QT, tổng số click cho các hoạt động tính điểm KT, tính tuổi thực từ ngày sinh của sinh viên để phân loại đối tượng người học theo độ tuổi, phục vụ cho phân tích sau này

Nghiên cứu này cũng thực hiện các mô tả cắt lớp dữ liệu theo các yếu tố nhân khẩu học như giới tính, đầu vào, ngành học, đặc biệt đối với cắt lớp độ tuổi tác giả có sử dụng phân loại độ tuổi theo thể hệ gen X, gen Y, gen Z [9]. Bộ dữ liệu sau khi được làm sạch có tổng cộng 212 quan sát và 18 thuộc tính. Sau khi dữ liệu đã được tiền xử lý và sẵn sàng để thực hiện nghiên cứu, sau đây là một số câu hỏi nghiên cứu đối với bộ dữ liệu này:

1. Có mối tương quan nào giữa số lần click chuột và kết quả hay không? Liệu sinh viên có số lần click chuột nhiều có kết quả tốt hơn?
2. Liệu sinh viên có đầu vào thấp hơn có kết quả như thế nào so với sinh viên có đầu vào cao hơn?
3. Điểm trung bình học phần theo ngành, giới tính và độ tuổi?
4. Sinh viên thuộc thể hệ nào thì tương tác lên hệ thống nhiều nhất? Có những tương tác nào chỉ dành cho một số đối tượng nhất định?

VI. Kết quả và thảo luận

Trên hệ thống có quy chiếu điểm thang 10, điểm số theo thang 4, điểm chữ, tuy nhiên để có thể tìm hiểu các mối tương quan với điểm một cách chi tiết, bài báo sử dụng kết quả điểm tổng kết theo thang 10. Trước hết hãy xem các số liệu điểm trung bình theo bằng cấp đầu vào và giới tính và điểm tổng kết thang 10 của học phần tiếng Anh ứng với từng loại đối tượng. Nhìn chung, điểm trung bình của nam cao hơn nữ, tuy nhiên đối với đối tượng có đầu vào là trung học phổ thông thì điểm trung bình gần như tương đương và riêng đối tượng có đầu vào trung cấp thì nữ có điểm trung bình cao hơn nam. Đối tượng có bằng đại học có điểm trung bình cao nhất đạt 7.63, điều này cũng phù hợp với thực tiễn vì đây là đối tượng đã có một bằng cấp, có khối lượng tích lũy kiến thức nói chung cao hơn so với mặt bằng còn lại. Tuy nhiên đối tượng là nữ có bằng cao đẳng lại có điểm bình quân thấp nhất với 6.82 điểm.

Bảng 2. Dữ liệu điểm tổng kết theo bằng cấp đầu vào và giới tính

Bằng cấp đầu vào	Giới tính	Số lượng/giới tính	Điểm trung bình tổng kết thang 10
THPT	Nam	90	7,30
	Nữ	40	7,21

Tổng THPT		130	7,27
Trung cấp	Nam	19	7,51
	Nữ	12	7,74
Tổng Trung cấp		31	7,60
Cao đẳng	Nam	22	7,10
	Nữ	13	6,82
Tổng cao đẳng		35	7,00
Đại học	Nam	8	7,88
	Nữ	8	7,38
Tổng đại học		16	7,63
Tổng cộng		212	7,3

Trong bộ dữ liệu này chỉ có 1 sinh viên theo học ngành QTKD và có chỉ có 2 sinh viên nam theo học ngành kế toán trên tổng số 18 sinh viên của ngành nên các kết quả điểm tổng kết trung bình không được đưa vào phân tích. Trong các ngành còn lại chỉ có duy nhất ngành CNTT là nam có điểm bình quân cao hơn nữ khá đáng kể ở mức 7.89 so với 7.19 (tuy nhiên kết quả cũng chưa thực sự đáng tin cậy do tỉ lệ nam chiếm tới tổng số 85% của ngành) còn lại các ngành khác thì điểm bình quân của nữ cao hơn nam lần lượt ngành luật ở mức 6.78 / 7.08, luật kinh tế 6.91/7.21, tài chính ngân hàng 7.47/7.86.

Bảng 3. Dữ liệu điểm tổng kết theo ngành và giới tính

Ngành học	Giới tính	Số lượng / giới	TB điểm QT	TB điểm KT	TB điểm TK
CNTT	Nam	51	9,46	9,10	7,89
	Nữ	9	9,52	8,97	7,19
Tổng CNTT		60	9,47	9,08	7,79
Kế toán	Nam	2	9,75	9,65	9,05
	Nữ	16	8,63	8,86	7,16
Tổng Kế toán		18	8,76	8,95	7,37
Luật	Nam	37	8,48	8,44	6,78
	Nữ	23	8,62	8,10	7,08
Tổng Luật		60	8,53	8,31	6,90
Luật KT	Nam	35	9,31	8,68	6,91
	Nữ	15	8,92	8,97	7,21
Tổng Luật KT		50	9,19	8,77	7,00
QTKD	Nam	1	10,00	10,00	8,30
Tổng QTKD		1	10,00	10,00	8,30
TCNH	Nam	13	8,90	8,96	7,47
	Nữ	10	9,16	9,12	7,86
Tổng TCNH		23	9,01	9,03	7,64
Tổng cộng		212	9,03	8,77	7,3

Một trong những hoạt động học tập không mang tính bắt buộc và không được tính điểm thành phần là hoạt động tham gia lớp học trực tuyến. Với hoạt động này thì đối tượng đã có bằng đại học tham gia đạt mức thấp nhất với chỉ 75% trong khi đối tượng có bằng trung cấp tham gia nhiều nhất tới 90%. Trong khi đó tỉ lệ đối tượng có bằng đầu vào trung học phổ thông đạt bằng mức bình quân tất cả các đối tượng tham gia ở mức 84%.

Bảng 4. Tỉ lệ tham gia học lớp học trực tuyến điểm tổng kết theo bằng cấp đầu vào.

Đối tượng đầu vào	Số lượng tham gia lớp học TT	Tổng số đối tượng	% tham gia
Cao đẳng	30	35	86%
Đại học	12	16	75%
Trung cấp	28	31	90%
THPT	109	130	84%
Tổng cộng	179	212	84%

Một hoạt động khác không được tính điểm là bài luyện tập tổng hợp cho thấy đối tượng có bằng cấp đầu vào trung học phổ thông tham gia bình quân cao nhất là 3,9 click / người, trong khi nhóm có bằng trung cấp đạt tỉ lệ thấp nhất với 1,6 click / người.

Bảng 5. Tỉ lệ click /người các đối tượng khác nhau trong hoạt động luyện tập tổng hợp

Đối tượng đầu vào	Tổng lượt click	Tổng số đối tượng	Click /Người
THPT	511	130	3,9
Trung cấp	51	31	1,6
Cao đẳng	80	35	2,3
Đại học	47	16	2,9
Tổng cộng	689	212	3,25

Về điểm tổng kết, đối tượng có bằng đại học đạt mức bình quân cao nhất với 7.63 điểm tuy nhiên điểm trung bình kỹ năng viết của nhóm lại đạt mức thấp nhất với mức 6.84 điểm, trong khi đó đối tượng có bằng cao đẳng có điểm bình quân thấp nhất đạt 7.00 điểm với trung bình kỹ năng nghe và đọc của nhóm này cũng đạt mức thấp nhất lần lượt là 5.11 và 5.17. Đối tượng có bằng trung cấp có điểm tổng kết bình quân đứng thứ 2 với 7.60 điểm tuy nhiên điểm trung bình kỹ năng nói của nhóm đạt mức thấp nhất với chỉ 4.98 điểm.

Bảng 6. Điểm bình quân tổng kết và 4 kỹ năng thành phần theo đối tượng đầu vào

Đối tượng đầu vào	Tổng số đối tượng	TB Nghe	TB Đọc	TB Viết	TB Nói	TB điểm TK
THPT	130	5,97	7,79	7,06	5,52	7,27
Trung cấp	31	6,79	7,58	7,42	4,98	7,60
Cao đẳng	35	5,11	7,49	6,89	5,17	7,00
Đại học	16	6,28	8,06	6,84	7,32	7,63
Tổng cộng	212	5,97	7,73	7,07	5,52	7,30

Thực hiện các phân tích mô tả theo thế hệ Gen X (sinh từ 1965 - 1980), Gen Y (sinh từ 1981 - 1996) và Gen Z (sinh từ 1997 - 2012) cho thấy về mặt giới tính tham gia theo học trực tuyến trong khi tỉ lệ nam / nữ tương đối cân bằng với Gen X thì đối với Gen Y và Gen Z thì tỉ lệ nam theo học gấp 2 lần số nữ. Về trung bình kết quả học tập thì nam giới Gen X và Y cao hơn nữ

giới lần lượt với mức 7.65 / 6.81 và 7.34 / 7.19. Tuy nhiên đối với Gen Z thì trung bình kết quả học tập của nữ giới lại cao hơn với mức 7.57 / 7.27.

Bảng 7. Điểm trung bình theo giới tính và thế hệ

Thế hệ	Giới tính	Số lượng / giới tính	TB điểm TK
GenX	Nam	9	7,56
	Nữ	10	6,81
Tổng GenX		19	7,16
GenY	Nam	87	7,34
	Nữ	42	7,19
GenY Total		129	7,29
GenZ	Nam	43	7,27
	Nữ	21	7,57
Tổng GenZ		64	7,36
Tổng		212	7,30

Bảng 8. Bảng ma trận quan hệ giữa các hoạt động click và điểm số

	Tổng click	Lượt click QT	Điểm QT	Lượt click KT	Điểm KT	Điểm TK
Tổng click	1					
Lượt click QT	0,85	1				
Điểm QT	0,07	0,09	1			
Lượt click KT	0,73	0,65	0,06	1		
Điểm KT	-0,09	-0,06	0,70	-0,06	1	
Điểm TK	-0,07	-0,10	0,42	-0,04	0,50	1

Thực hiện xây dựng ma trận quan hệ (correlation matrix) trên các thuộc tổng số lượt click, số lượt click quá trình, số lượt click kiểm tra, điểm quá trình, điểm kiểm tra và điểm tổng kết ta thấy chỉ có hệ số tương quan giữa điểm quá trình và điểm kiểm tra là đồng biến mạnh ở mức 0.7.

Bảng 9. Bảng ma trận quan hệ giữa các các điểm kỹ năng thành phần và điểm tổng kết

	Nghe	Đọc	Viết	Nói	Điểm tổng kết
Nghe	1				
Đọc	0,45	1			
Viết	0,31	0,34	1		
Nói	0,45	0,28	0,22	1	
Điểm tổng kết	0,80	0,61	0,46	0,65	1

Ma trận quan hệ giữa các điểm kỹ năng thành phần Nghe, Nói, Đọc, Viết cho thấy hệ số tương quan giữa Nghe và điểm TKHP là đồng biến mạnh ở mức 0.8.

Thực hiện phân tích Anova 1 factor với điểm số quá trình, thành phần và điểm tổng kết học phần ta có các giả thuyết sau:

Giả thuyết Ho: Không có sự khác biệt đáng kể giữa giá trị trung bình của các thuộc tính

Giả thuyết Ha: Có sự khác biệt đáng kể giữa giá trị trung bình của các thuộc tính

Bảng 10. Anova 1 factor với điểm số quá trình, thành phần và tổng kết học phần

SUMMARY						
Groups	Count	Sum	Average	Variance		
Điểm QT	212	1914,6	9,031132	1,594761		
Điểm KT	212	1860,25	8,774764	2,360983		
Điểm tổng kết	212	1547,6	7,3	1,501422		
ANOVA						
Source	of					
Variation	SS	df	MS	F	P-value	F crit
Between Groups	370,1147	2	185,0573	101,7327	4,89E-39	3,009955
Within Groups	1151,462	633	1,819055			
Total	1521,577	635				

Với $p\text{-value} = 4.89 \text{ E}^{-39} < 0.05$, giả thuyết H_0 bị bác bỏ và giá trị trung bình của 3 thuộc tính này có sự khác biệt đáng kể.

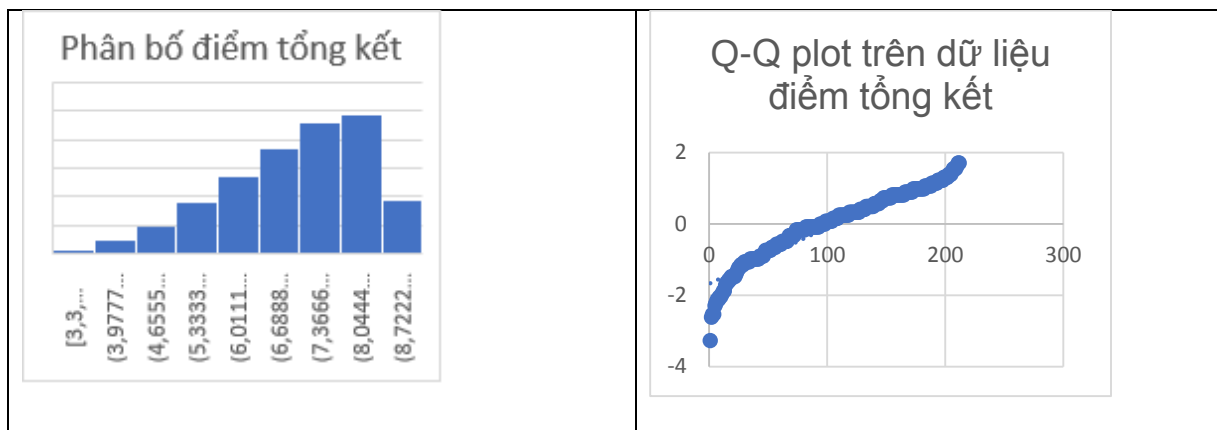
Bảng 11. Anova 1 factor với điểm số quá trình, thành phần và tổng kết học phần

SUMMARY						
Groups	Count	Sum	Average	Variance		
Nghe	212	1265,5	5,96934	9,637681		
Đọc	212	1639	7,731132	3,988979		
Viết	212	1498	7,066038	2,405571		
Nói	212	1170,2	5,519811	5,486999		
Điểm tổng kết	212	1403,3	6,61934	2,365264		
ANOVA						
Source	of					
Variation	SS	df	MS	F	P-value	F crit
Between Groups	648,674434	4	162,1686	33,94851	1,12E-26	2,380366
Within Groups	5039,628208	1055	4,776899			
Total	5688,302642	1059				

Tiếp tục thực hiện phân tích Anova 1 factor với 4 đầu điểm thành phần và điểm tổng kết có kết quả $p\text{-value} = 1.12 \text{ E}^{-26} < 0.05$ cho thấy giá trị trung bình của 5 thuộc tính này có sự khác biệt đáng kể.

Cuối cùng, để có thể kiểm định xem điểm TKHP có thực sự là ngẫu nhiên và khách quan theo trình độ người học thông qua kiểm định phân phối điểm tổng kết có phải phân phối gần chuẩn (nearly normal distribution). Sở dĩ điểm số TKHP còn bao gồm cả điểm QT và điểm KT cũng như điểm thi KTHP và những trường hợp ngoại lệ như không có các đầu điểm QT và TP thì không được tham gia thi KTHP nên phân phối điểm sẽ nằm ở dải từ 3 đến 10 điểm. Dưới đây là đồ thị phân bố điểm bằng plot chart cho thấy về mặt trực quan, điểm số có sự phân bố gần chuẩn với hình dạng left tail. Tuy nhiên để có đưa ra kết luận chắc chắn về phân phối gần chuẩn của điểm số ta phải thực hiện thêm kiểm định Q-Q plot.

Hình 1. Phân phối điểm tổng kết và kiểm định Q-Q plot để xác định phân phối chuẩn



Kiểm định Q-Q plot cho ra kết quả tiệm cận straight line (đường thẳng) nên có thể kết luận phân phối điểm là gần chuẩn. Nếu loại bỏ các đối tượng có điểm tổng kết dưới 5 (đối tượng người học trực tuyến không tích cực tham gia các hoạt động học tập) thì phân phối này càng tiệm cận với phân phối chuẩn.

Nhìn chung, điểm số được phân bố gần mức chuẩn nên điểm số đối với học phần tiếng Anh cơ bản này là thang đo tin cậy. Nếu xét điểm số theo ngành đào tạo thì nhìn chung điểm số nữ giới cao hơn nam giới (trừ trường hợp ngành công nghệ thông tin do chênh lệch nam /nữ nên không đưa vào phân tích). Sinh viên có đầu vào trung học phổ thông tham gia hoạt động nhiều hơn so với các nhóm đối tượng khác (thể hiện qua số lượt click). Sinh viên có bằng đại học có điểm tổng kết học phần cao hơn nhóm đối tượng còn lại. Sinh viên có bằng tốt nghiệp cao đẳng có kỹ năng Nghe và Nói thấp nhất trong số các nhóm đối tượng và sinh viên nữ giới của nhóm này cũng có điểm tổng kết bình quân thấp nhất. Các hoạt động như Vclass lần 1 và ghi lại có ít sinh viên tham gia.

Căn cứ các kết quả nghiên cứu nêu trên, tác giả đề xuất một số khuyến nghị đối với xây dựng và triển khai các khóa học trực tuyến nói chung và cụ thể là môn tiếng Anh nói riêng như sau:

1. Một số hoạt động có các thao tác hoạt động thấp như Vclass cần phải xem xét lại về mặt nội dung và kỹ thuật xem nguyên nhân tại sao lại có ít tương tác.
2. Cần phải có thêm các bài pre-test (kiểm tra đầu vào) và post-test (kiểm tra sau đào tạo) để có thể chẩn đoán được các kiến thức trước và sau khi tham gia khóa học.
3. Vì kĩ năng Nghe tiếng Anh có hệ số tương quan mạnh với điểm TKHP nên cần thiết phải tăng thời lượng Nghe trong học phần.
4. Căn cứ trên các mô tả cắt lớp theo nhân khẩu học để có các tư vấn phù hợp với đối tượng tuyển sinh.
5. Các thao tác hoạt động đang chủ yếu dựa vào số lượt click / hoạt động nên chưa tìm được mối quan hệ ẩn giữa các hoạt động và kết quả học tập, trong các nghiên cứu tiếp theo cần phải đưa vào thời lượng trung bình một người dùng đối ứng với từng hoạt động cụ thể; thời điểm đăng nhập trong ngày, giữa các ngày; số lượt thử bài luyện tập / kiểm tra

V. Kết luận

Các kết quả nghiên cứu được chỉ ra ở trên cho thấy một số khuynh hướng học tập theo nhân khẩu học và thao tác trên hệ thống khá rõ rệt. Đối tượng có đầu vào thấp như trung học phổ thông có xu hướng tương tác với hệ thống nhiều hơn. Đối tượng có đầu vào cao như đại học có kết quả học tập tốt hơn mặc dù thao tác trên hệ thống ít hơn. Sinh viên có điểm quá trình tốt thì đồng thời cũng có điểm kiểm tra tốt. Kỹ năng Nghe là kỹ năng quan trọng và đóng góp nhiều vào điểm số KTHP. Phân phối điểm số TKHP là phân phối gần chuẩn, thể hiện thang đo khách quan, tin cậy. Một số khuyến nghị nhằm nâng cao chất lượng học phần đào tạo trực tuyến nói chung và tiếng Anh nói riêng cũng được đề xuất. Các thông tin về nhân khẩu học cũng có thể được tham khảo và sử dụng để phục vụ công tác tuyển sinh.

Tài liệu tham khảo

- [1]. Đinh Thị Hóa và cộng sự. Phân tích các nhân tố ảnh hưởng đến kết quả học tập của sinh viên Khoa Kinh tế Trường Đại học Đồng Nai Tạp chí Khoa học - Đại học Đồng Nai số 11 - 2018 (trang 18-29) ISSN 2354-1482.
- [2]. Đinh Thị Thu Huyền và cộng sự. Thực trạng và một số yếu tố liên quan đến kết quả học tập của sinh viên đại học điều dưỡng chính quy khóa 14, Trường Đại học Điều dưỡng Nam Định năm học 2018 – 2019. Tạp chí Khoa học Điều dưỡng, Tập 5 Số 02 (2022) <https://jns.vn/index.php/journal/article/view/413>
- [3]. Phạm Thị Tình Thương và cộng sự. Những ảnh hưởng lên kết quả học tập trực tuyến của của sinh viên trong giai đoạn dịch bệnh Covid-19: Phân tích nhân tố khám phá Bayesian https://www.researchgate.net/publication/355305062_The_Factors_Affecting_Students'_Online_Learning_Outcomes_during_the_COVID-19_Pandemic_A_Bayesian_Exploratory_Factor_Analysis
- [4]. Britha Bhandari (2022). Data Cleaning a guide with Examples and Steps
- [5]. Available at: <https://www.scribbr.co.uk/research-methods/data-cleaning/>
- [6]. Cheng Fan, Meiling Chen, Xinghua Wang, Jiayuan Wang, Bufu Huang Front. Energy Res., 29 March 2021 Sec. Sustainable Energy Systems Volume 9 - 2021
- [7]. <https://doi.org/10.3389/fenrg.2021.652801>
- [8]. Goodman, Z. T., Timpano, K. R., Llabre, M. M., & Bainter, S. A. (2022). Revisiting the factor structure and construct validity of the Cognitive Failures Questionnaire. *Psychological Assessment*, 34(7), 671–683. <https://doi.org/10.1037/pas0001127>
- [9]. Mansureh Kebritch et al (2017) Issues and challenges for teaching successful online course in higher education: A literature review
- [10]. https://www.researchgate.net/publication/319013030_Issues_and_Challenges_for_Teaching_Successful_Online_Courses_in_Higher_Education_A_Literature_Review
- [11]. Mohammed Habes et al (2018) The role of modern media technology in improving collaborative learning of students in Jordanian universities. International Journal of Information Technology and Language Studies (IJITLS) <https://journals.sfu.ca/ijitls/index.php/ijitls/article/view/51/pdf>
- [12]. Prakash, K., Tiwari, P. Millennials and Post Millennials: A Systematic Literature Review. Pub Res Q 37, 99–116 (2021). <https://doi.org/10.1007/s12109-021-09794-w>
- [13]. Tempelaar D, Rienties B, Nguyen Q (2020) Subjective data, objective data and the role of bias in predictive modeling: Lessons from a dispositional learning analytics application. PLoS ONE 15(6): e0233977. <https://doi.org/10.1371/journal.pone.0233977>

[11]. Van den Broeck J, Argeseanu Cunningham S, Eeckels R, Herbst K (2005) Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. PLoS Med 2(10): e267. <https://doi.org/10.1371/journal.pmed.0020267>

Nghiên cứu này được tài trợ bởi đề tài cấp Trường của Trường Đại học Mở Hà Nội, mã số: MHN2023-03.17

DETERMINING THE CORRELATION BETWEEN LEARNING ACTIVITIES AND LEARNING OUTCOMES OF BASIC ENGLISH BASED ON ANALYZING LOG FILE OF THE ONLINE TRAINING SYSTEM AT HANOI OPEN UNIVERSITY

Dang Hai Dang[†]

Abstract: *Analyzing and evaluating process-based learning outcomes for university students is often assessed by manual tools such as attendance checking, score giving based on component tests and is performed directly by the instructor. This increases the amount of non-teaching work not only for lecturers but also for the training support team due to having to receive and re-enter data on the system. For online training systems, these tasks are not only collected automatically but can also integrate other data from learners when working with the system. This study analyzes system log data and demographic factors using statistical analysis models to find relationships between learner attributes on the system. The course selected for data analysis is basic English because it has a score data system with 4 skills. Thereby, determining the relationships between demographics and learning activities on learning outcomes of basic English subjects on the online training platform.*

Keywords: *E-learning, learning activity, descriptive statistics, learning outcome, data analysis*

[†] Hanoi Open University