

SỬ DỤNG XÁC XUẤT THỐNG KÊ TRONG NGHIÊN CỨU GIÁO DỤC: MỘT VÍ DỤ VỀ CÁCH DÙNG MÔ HÌNH PHƯƠNG TRÌNH CẤU TRÚC (SEM)

TS. Tô Thị Thu Hương

Khoa Ngôn ngữ & Văn hoá Anh-Mỹ

Đại học Ngoại Ngữ, ĐHQG Hà Nội

Mô hình phương trình cấu trúc (SEM) là một phương pháp thống kê toàn diện để kiểm định các giả thuyết về các mối quan hệ giữa các biến quan sát được và không quan sát được. Phương pháp này được xây dựng trên cơ sở lý thuyết thống kê của đầu thập kỷ 70. Tuy vậy mãi tới gần đây phương pháp này mới được coi là phương pháp chuẩn dùng để kiểm định các giả thuyết nghiên cứu. Thuật ngữ SEM thường được nhiều tác giả và nhà nghiên cứu dùng hoán đổi với các thuật ngữ phân tích đường dẫn, các mô hình LISREL/kiểu LISREL, mô hình nguyên nhân, phân tích cấu trúc hợp biến, các mô hình với biến ẩn và mô hình cấu trúc [path analysis, LISREL/LISREL-type models, causal modelling, analysis of covariance structures, latent variable models, and structural modelling (theo Pedhazur & Schmelkin 1991: 695; Kunnan 1995: 19)].

Bước đầu tiên trong việc sử dụng SEM là xác định các mô hình hay còn gọi là ra các tuyên bố có tính chất thống kê (nêu các tham số) về các mối quan hệ giữa các biến. Đây là bước cơ bản trong SEM. Theo Hoyle (1995: 2), "không một phân tích nào có thể thực hiện được cho đến khi nhà nghiên cứu xác định được một mô hình về các mối quan hệ giữa các biến cần phân tích". Trong nghiên cứu về giá trị tiên đoán của đề thi Hệ thống kiểm tra tiếng Anh quốc tế (IELTS) với các sinh viên Việt Nam được Cơ quan Hợp tác Phát triển Ôxtrâyliya (AusAID) tài trợ học bổng du học tại Ôxtrâyliya, SEM được sử dụng để kiểm định giả thiết rằng:

Năng lực tiếng Anh được đo bằng bài thi IELTS (viết tắt là English Ability) có tác động đến hoạt động học tập sau này tại các trường giảng dạy bằng tiếng Anh được đo lường bởi điểm trung bình chung của học kỳ một và học kỳ hai tại các trường đại học tại Ôxtrâyliia (viết tắt là Academic Performance)

Việc xác định mô hình có thể nêu cụ thể độ lớn và các dấu của các tham số. Tuy nhiên, các tham số thường là cố định hoặc tự do. Các tham số cố định không được ước đoán từ các số liệu nghiên cứu và giá trị của chúng thường được đặt cố định ở mức không. Các tham số tự do được ước đoán từ các số liệu và là các con số mà nhà nghiên cứu tin là khác không. Các chỉ số khác nhau về sự thích hợp của mô hình, đặc biệt là cách thử Khi bình phương (χ^2 goodness-of-fit test) cho thấy độ nhất quán giữa mẫu hình của các tham số cố định và tự do được xác định trong một mô hình và mẫu hình của các phương sai cũng như hiệp biến lấy từ một bộ số liệu quan sát được (Hoyle 1995: 3).

Trên cơ sở mẫu hình của các tham số cố định và tự do, hai thành phần của mô hình cấu trúc chung gọi là *mô hình đo lường* và *mô hình cấu trúc* được xác định. Mô hình đo lường xác định các biến ẩn. Các biến ẩn là các biến không quan sát được nhưng được thể hiện bởi các hiệp biến giữa hai hoặc nhiều chỉ số. Các biến này hoàn toàn không mang các sai số ngẫu nhiên và các tính chất riêng thường gắn với các chỉ số của chúng. Mô hình cấu trúc thể hiện các mối quan hệ giữa các biến ẩn và các biến quan sát được nhưng không phải là các chỉ số của các biến ẩn.

Trong nghiên cứu về giá trị tiên đoán của IELTS, mô hình đo lường của biến ẩn *Năng lực tiếng Anh (English Ability)* được xác định bởi điểm số quan sát được qua bốn bài thi IELTS Nghe, Đọc, Viết, Nói và Điểm toàn bộ (LIST, READ, WRITE, SPEAK and OVERALL). Mô hình đo lường của biến ẩn *Hoạt động học (Academic Performance)* được xác định bởi các biến quan sát được là điểm trung bình chung học kỳ 1 và học kỳ 2 (SEM1, SEM2).

Mô hình cấu trúc thể hiện mối quan hệ giữa hai biến ẩn *English Ability* và *Academic Performance*.

Trong SEM, việc nhận dạng là điều cơ bản khi xác định mô hình. Việc này liên quan đến sự tương ứng giữa các tham số tự do và các

phương sai và hiệp biến quan sát được. Mô hình vừa được nhận dạng nếu với một tham số tự do có thể có được một giá trị qua một lần thao tác duy nhất các số liệu quan sát được. Trong trường hợp này, mô hình có không (0) bậc tự do. Mô hình được nhận dạng quá mức khi một giá trị cho một hoặc nhiều tham số tự do có thể có được bằng nhiều cách thao tác các số liệu quan sát được. Lấy số các hiệp biến và phương sai quan sát được trừ đi số các tham số tự do ta có bậc tự do của mô hình. Nếu với một hoặc nhiều tham số tự do không thể tính được một giá trị duy nhất từ các số liệu quan sát được, mô hình sẽ không thể ước đoán được và gọi là nhận dạng thiếu. Vì vậy một hạn định cho xác định mô hình là mô hình chỉ được nhận dạng quá mức hoặc vừa được nhận dạng.

Bước tiếp theo trong SEM là ước lượng các tham số tự do từ bộ số liệu quan sát được. Các ước lượng về những tham số tự do thể hiện một ma trận hiệp biến giống ma trận quan sát được gọi là các phương pháp lặp. Ma trận hiệp biến được thể hiện sau mỗi phép lặp được so sánh với ma trận quan sát được để hình thành một ma trận số dư với các phần tử là hiệu số của các giá trị tương ứng giữa ma trận thể hiện và ma trận quan sát được. Khi giá trị của các phần tử trong ma trận số dư không thể biến đổi nhỏ hơn nữa, ước tính kết thúc và cho một con số duy nhất tóm tắt độ tương ứng giữa ma trận hiệp biến thể hiện và quan sát được. Giá trị này được dùng để kiến thiết các chỉ số thích hợp của mô hình.

Bước tiếp theo là đánh giá sự thích hợp của mô hình. Nếu các phần tử của ma trận gần tới không, hoặc ma trận hiệp biến thể hiện tương tự ma trận hiệp biến quan sát được, mô hình được coi là thích hợp với số liệu quan sát được. Cách thử Khi bình phương (χ^2 goodness-of-fit test) là chỉ số thích hợp phổ biến nhất. Đánh giá sự thích hợp của mô hình bao gồm việc so sánh các mô hình, tức là so sánh hai hoặc nhiều mô hình dựa trên lý thuyết trên cùng một bộ số liệu để quyết định mô hình nào giải thích các số liệu quan sát được một cách tốt hơn.

Bảng 1 sau đây được trích từ hai tác giả Schumacker and Lomax (1996:121) cho thấy các chỉ số khác nhau về độ thích hợp của mô hình. So với các phương thức đa biến như phân tích phương sai, hồi quy bội, phân tích đường dẫn, việc quyết định độ thích hợp của mô hình trong SEM không trực tiếp bằng. Ngoại trừ cách thử Khi bình phương (χ^2

goodness-of-fit test), các chỉ số về độ thích hợp của mô hình trong SEM không có phép thử thống kê về ý nghĩa đi kèm. Hầu hết các chỉ số đều được tính toán nằm trong giới hạn từ 0 (thể hiện sự không thích hợp) đến 1 (thể hiện sự thích hợp hoàn toàn). Baldwin (1989), Bentler và Bonnett (1980) gợi ý rằng với một mô hình SEM, giá trị Độ thích hợp (GOF) bằng 0.90 hoặc cao hơn là có thể chấp nhận được.

Bảng 1. Các tiêu chí độ thích hợp (GOF) và cách hiểu các độ thích hợp chấp nhận được

Các tiêu chí GOF	Mức chấp nhận được	Cách hiểu
Chi-square	Tabled χ^2 value	Compares obtained χ^2 value with tabled value for given degrees of freedom (df).
Normed chi-square	1.0 to 2.0	Less than 1.0 is a poor model fit. Higher than 2 reflects a need for improvement.
Root-mean-square error of approximation (RMSEA)	< 0.05	Value less than 0.05 indicates a good model fit.
Goodness of fit (GFI)	0 (no fit) to 1 (perfect fit)	Value close to 0.95 reflects a good fit.
Adjusted goodness of fit (AGFI)	0 (no fit) to 1 (perfect fit)	Value adjusted for df, with 0.95 being a good model fit.
Tucker-Lewis or Non-Normed Fit Index (NNFI)	0 (no fit) to 1 (perfect fit)	Value close to 0.95 reflects a good fit.
Comparative Fit Index (CFI)	0 (no fit) to 1 (perfect fit)	Value close to 0.95 reflects a good fit.

Vì SEM được coi là một trong những tiến bộ mới nhất trong nghiên cứu định tính nhằm xác định giá trị của các đề thi (Bachman & Eignor 1997: 227), tác giả bài viết đã sử dụng SEM để kiểm định các giả thuyết nhằm thử nghiệm một phương thức mới trong nghiên cứu xác định giá trị tiên đoán của đề thi IELTS. Phạm vi bài viết này không cho phép mô tả toàn bộ nghiên cứu nên tác giả chỉ nêu một phần có sử dụng SEM trong bài viết để minh họa.

Các mô hình đo lường và cấu trúc sau đã được dùng để kiểm định giả thuyết "Năng lực tiếng Anh được đo bằng bài thi IELTS (viết tắt là English Ability) có tác động đến hoạt động học tập sau này tại các trường giảng dạy bằng tiếng Anh được đo lường bởi điểm trung bình chung của học kỳ một và học kỳ hai tại các trường đại học tại Ôxtrâyli (viết tắt là Academic Performance)":

Mô hình đo lường:

1) Academic Performance (APerform) = Semester 1 GPA (SEM1) + Semester 2 GPA (SEM2)

2) English Ability (EAbility) = IELTS Listening Score (LIST) + Reading Score (READ) + Writing Score (WRITE) + Speaking Score (SPEAK) + Overall Score (OVERALL)

Mô hình cấu trúc:

Academic Performance (APerform) = Hằng số x English Ability (EAbility)

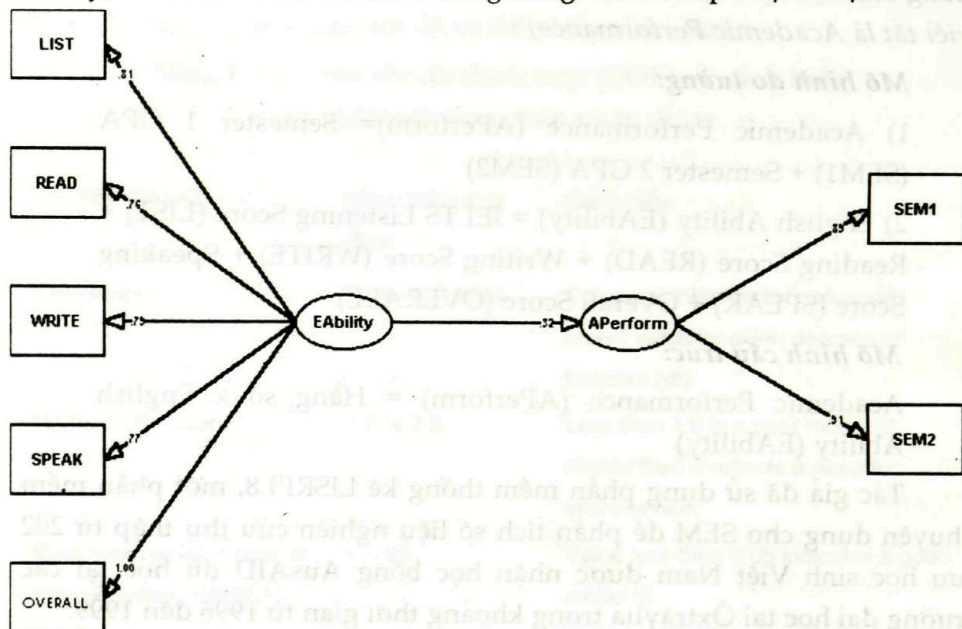
Tác giả đã sử dụng phần mềm thống kê LISREL8, một phần mềm chuyên dụng cho SEM để phân tích số liệu nghiên cứu thu thập từ 202 lưu học sinh Việt Nam được nhận học bổng AusAID du học tại các trường đại học tại Ôxtrâyli trong khoảng thời gian từ 1996 đến 1999.

Bảng 2. Các tính toán của LISREL

Hệ số hồi quy	Tính toán	Sai số chuẩn	Giá trị t
APerform → SEM1	0.89	0.15	6.03
APerform → SEM2	0.91	0.19	4.84
EAbility → LIST	0.81	0.061	13.28
EAbility → READ	0.76	0.092	8.29
EAbility → WRITE	0.79	0.057	13.69
EAbility → SPEAK	0.77	0.064	12.15
EAbility → APerform	0.32	0.041	7.82

Bảng 2 sau đây mô tả kết quả thu được. Cột 1 cho thấy đường hồi quy từ biến ẩn đến các biến quan sát được và từ biến ẩn EAbility đến biến ẩn APerform. Các đường này được thể hiện bằng các mũi tên một chiều. Cột 2 thể hiện các tính toán về mối quan hệ giữa các đường tương

ứng. Cột 3 và 4 cho thấy các sai số chuẩn (SE) và giá trị t của các tính toán. SE là số đo độ chính xác của các tham số tính được. Giá trị t là tỉ số giữa tính toán và sai số chuẩn của nó. Một giá trị t lớn hơn 1.96 được coi là dấu hiệu tính toán tốt (Joreskog & Sorbom 1993: 107). Theo tiêu chí này, tất cả các tính toán nêu trong Bảng 2 đều chấp nhận được.



Hình 1. Đồ thị đồ họa của mô hình nghiên cứu

Hình 1 trình bày biểu đồ đồ họa mô hình được nghiên cứu. Trong mô hình này, *Năng lực tiếng Anh* (EAbility) được thể hiện bằng 5 chỉ số quan sát được là LIST, READ, WRITE, SPEAK, và OVERALL. Hoạt động học tập (APerform) được biểu hiện qua SEM1 và SEM2. Có một đường dẫn từ EAbility đến APerform thể hiện tác động trực tiếp của EAbility đến APerform. Hệ số của đường dẫn này là 0.32. Sai số chuẩn của tính toán là 0.041. Giá trị t khá lớn ở mức 7.82 nhấn mạnh độ tin cậy của tính toán. Hệ số này có thể hiểu là hệ số tương quan thực giữa EAbility và APerform. Nói một cách khác, *Năng lực tiếng Anh* được đo bằng bộ đề thi IELTS có thể giải thích khoảng 10% cho sự khác biệt trong

Hoạt động học tập được thể hiện bằng điểm trung bình chung học kỳ 1 và học kỳ 2 cho nhóm đối tượng nghiên cứu.

Trên đây là một ví dụ về ứng dụng xác suất thống kê trong nghiên cứu giáo dục, cụ thể là dùng Mô hình phương trình cấu trúc (SEM), một phương pháp thống kê toàn diện để kiểm định giả thuyết về mối quan hệ giữa các biến quan sát được (điểm IELTS và điểm trung bình chung học kỳ 1 và 2) và các biến không quan sát được (năng lực tiếng Anh, hoạt động học tập tại các trường đại học giảng dạy bằng tiếng Anh), trong nghiên cứu giá trị tiên đoán của Bộ đề thi Hệ thống Kiểm tra tiếng Anh Quốc tế (IELTS). Hy vọng bài viết có thể góp phần cho việc tăng cường hơn nữa các ứng dụng của xác suất thống kê trong nghiên cứu giáo dục.

TÀI LIỆU THAM KHẢO

1. **Bachman, L. F. & Eignor, D. R.** (1997) Recent advances in quantitative test analysis. In Clapham, C. & Corson, D.(eds) *Encyclopaedia of language & education, Volume 7: Language Testing and Assessment*. Dordrecht ; Boston, Kluwer. 1997: 227-242.
2. **Baldwin, B.** (1989) A primer in the use and interpretation of structural equation models. *Measurement and Evaluation in Counseling and Development*, 22: 100-112.
3. **Bentler, P. M. & Bonnett, D. G.** (1980) Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88: 588-606.
4. **Hoyle, R.H.** (ed) (1995) *Structural equation modelling-Concepts, issues, and applications*. Thousand Oaks, London, Sage.
5. **Joreskog, K. & Sorbom, D.** (1993) *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ, SSI.
6. **Kunnan, A.J.** (1995) *Test taker characteristics and test performance; a structural modeling approach*. Cambridge, Cambridge University Press.

7. **Pedhazur, E.J. & Schmelkin, L.P.** (1991) *Measurement, design, and analysis, An integrated approach*. Hillsdale, New Jersey, Lawrence Erlbaum.
8. **Schumacker, R. E. & Lomax, R. G.** (1996) *A beginner's guide to structural equation modeling*. Mahwah, New Jersey, Lawrence Erlbaum.
9. **Tô Thị Thu Hương** (2000) *Nghiên cứu giá trị tiên đoán của đề thi Hệ thống kiểm tra tiếng Anh quốc tế (IELTS) với các sinh viên Việt Nam*. Luận văn Tiến Sĩ, Khoa Giáo dục, Đại học tổng hợp Melbourne, Ôxtrâyliia.

TÀI LIỆU THAM KHẢO

Bachman, L. F. & Eignar, D. R. (1997) *Key to success in quantitative test analysis*. In Chapman, C. & Chapman, D. (eds) *Psychometrics of language & education*. London: Lawrence Erlbaum, 137-157.

Baldwin, B. (1989) A primer in the use and interpretation of structural equation models. *Measurement and Evaluation in Counseling and Development* 22, 100-117.

Bentler, P. M. & Bonett, D. C. (1980) Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin* 88, 588-606.

Hoyle, R.H. (ed) (1995) *Structural equation modeling: concepts and practice*. Thousand Oaks, London, Sage.

Joreskog, K. & Sorbom, D. (1993) *Likert & structural equation modeling with the SIMPLIS computer package*. Hillsdale, NJ: SI.

Kunnan, A.J. (1995) *Test item characteristics and test performance: a structural modeling approach*. Cambridge: Cambridge University Press.