

## Bài báo nghiên cứu

# ĐỀ XUẤT QUY TRÌNH XÂY DỰNG THANG ĐO NĂNG LỰC CỦA HỌC SINH DỰA TRÊN MÔ HÌNH CẤU TRÚC TUYẾN TÍNH

*Tạ Thanh Trung*

*Công ty Cổ phần Khoa học và Giáo dục Sài Gòn, Việt Nam*

*Tác giả liên hệ: Tạ Thanh Trung – Email: [trungtphysics@gmail.com](mailto:trungtphysics@gmail.com)*

*Ngày nhận bài: 26-4-2023; ngày nhận bài sửa: 10-8-2023; ngày duyệt đăng: 24-8-2023*

### TÓM TẮT

*Trước xu thế chuyển từ dạy học cung cấp nội dung sang dạy học định hướng phát triển năng lực của học sinh, xây dựng và chuẩn hóa các công cụ đánh giá năng lực của người học là nhiệm vụ quan trọng nhằm đảm bảo chất lượng giáo dục. Mặc dù, gần đây nhiều nghiên cứu tập trung xây dựng các thang đo năng lực của học sinh làm tiền đề để hình thành các công cụ đánh giá năng lực với nhiều phương pháp tiếp cận khác nhau nhưng chưa nhiều nghiên cứu chỉ ra được độ tin cậy và độ giá trị của các thang đo năng lực được đề xuất. Vì vậy, bài báo này giới thiệu một quy trình để chuẩn hóa thang đo năng lực của học sinh gồm các giai đoạn: (1) Xây dựng thang đo lý thuyết; (2) Nghiên cứu sơ bộ định tính với phương pháp Delphi; (3) Nghiên cứu sơ bộ định lượng với phương pháp EFA; (4) Nghiên cứu chính thức định lượng với phương pháp CFA. Bên cạnh đó, nghiên cứu cũng chỉ ra sự đáp ứng của quy trình này với các tiêu chuẩn nghiêm ngặt của độ tin cậy và độ giá trị của thang đo năng lực của học sinh. Quy trình này có thể giúp cho các nhà giáo dục chuẩn hóa các thang đo năng lực phức hợp và giúp giáo viên có định hướng tốt hơn trong lựa chọn chỉ số hành vi để đánh giá năng lực học sinh.*

**Từ khóa:** phân tích nhân tố khám phá; phân tích nhân tố khẳng định; đánh giá năng lực; mô hình cấu trúc tuyến tính; thuyết hệ thống

### 1. Giới thiệu

Từ đầu thế kỉ XXI đến nay, xu thế chung của nhiều chương trình giáo dục phổ thông trên thế giới là chuyển từ dạy học cung cấp nội dung sang dạy học định hướng phát triển năng lực của người học (Colby, 2019). Tại nước ta, nghị quyết số 29-NQ/TW ngày 4/11/2013 Hội nghị Trung ương 8 khóa XI về đổi mới căn bản, toàn diện giáo dục và đào tạo của Ban Chấp hành Trung ương Đảng cũng đã nhấn mạnh cần nhanh chóng có sự chuyển đổi từ chủ yếu dạy học truyền thụ kiến thức sang phát triển toàn diện phẩm chất, năng lực học sinh. Vì thế, các nghiên cứu đánh giá trong giáo dục theo hướng hỗ trợ sự phát triển phẩm chất và năng lực học sinh đang được các nhà khoa học giáo dục đặc biệt quan tâm

---

*Cite this article as:* Tạ Thanh Trung (2023). Proposing the process of developing a scale for students' competency: A structural equation modeling approach. *Ho Chi Minh City University of Education Journal of Science*, 20(8), 1337-1352.

(Pham & Le, 2022). Điều này thể hiện qua sự gia tăng số lượng các công bố về phát triển công cụ đánh giá năng lực người học với đa dạng hình thức và phương pháp khác nhau (Hoang, 2015). Tuy nhiên, đa số các công cụ đánh giá năng lực này được xây dựng định tính dựa trên việc kế thừa các khung năng lực lí thuyết, áp dụng vào đặc thù của hoạt động dạy học mà chưa được chuẩn hóa bằng dữ liệu thực tiễn (Ta, Tran, Nguyen, & Nguyen, 2022). Điều này có thể dẫn đến kết quả là những công cụ đo lường thiếu độ giá trị và độ tin cậy, làm cho các hoạt động đánh giá và tự đánh giá chưa phản ánh đầy đủ, chính xác năng lực người học. Để khắc phục vấn đề này, một số nhà khoa học đã thay đổi cách tiếp cận để phát triển và xác nhận thang đo (scale development and validation) về hành vi của con người nói chung và năng lực của học sinh nói riêng bằng cách xây dựng các thiết kế nghiên cứu kết hợp định tính với định lượng thông qua việc vận dụng mô hình cấu trúc tuyến tính (structural equation modeling – SEM) (Nguyen, Nguyen, Dang, & Nguyen, 2020; Ta, Tran, Nguyen, & Nguyen, 2022). Tất nhiên, việc phát triển và chuẩn hóa một thang đo năng lực của người học theo cách tiếp cận này được đánh giá là không đơn giản đối với người mới tiếp cận vì những khó khăn như nghiên cứu sẽ cần tuân thủ nhiều bước phát triển mô hình để đảm bảo độ giá trị và độ tin cậy với nhiều thuật ngữ chuyên ngành, sự tốn kém về chi phí và thời gian và thường phải sử dụng nhiều kĩ thuật phân tích thống kê phức tạp (Boateng et al., 2018). Chính vì vậy, mục tiêu chính của bài báo này là mô tả quy trình hình thành, phát triển và chuẩn hóa một thang đo năng lực của học sinh theo cách đơn giản nhất có thể, cũng như cung cấp các tiêu chuẩn đánh giá kết quả phân tích thang đo năng lực tương ứng từng giai đoạn. Bài viết có ý nghĩa như một sự hướng dẫn cơ bản để tạo điều kiện thuận lợi cho các nghiên cứu phát triển các thang đo năng lực, công cụ đánh giá năng lực mới hợp lệ và đáng tin cậy, vừa giúp cải thiện các thang đo năng lực hiện có.

## 2. Nội dung

### 2.1. Thang đo năng lực của học sinh

#### 2.1.1. Thang đo năng lực

Năng lực (competency) được hiểu là “khả năng cá nhân đáp ứng các yêu cầu phức hợp và thực hiện thành công nhiệm vụ trong một bối cảnh cụ thể” (Rychen & Salganik, 2002, p. 8). Đó là sự huy động phẩm chất tâm lí và sinh lí như: kiến thức, kĩ năng và các thuộc tính tâm lí cá nhân khác như hứng thú, niềm tin, ý chí... tạo cho con người khả năng hoàn thành một loại hoạt động với chất lượng cao trong một bối cảnh nhất định. Năng lực giúp học sinh tổng hợp các khả năng và kĩ năng nhận thức sẵn có hoặc có thể học được ở cá nhân nhằm giải quyết những vấn đề nảy sinh trong cuộc sống và cũng hàm chứa động cơ, ý chí để cá nhân hành động một cách có trách nhiệm, có sự phê phán để đi đến giải pháp (Weinert et al., 2011). Trong bối cảnh giáo dục Việt Nam cũng dần tiếp cận dạy học định hướng phát triển năng lực của học sinh, chương trình giáo dục phổ thông của nước ta đưa ra một định nghĩa tương đối trọn vẹn cho khái niệm này: “năng lực của học sinh được hiểu là thuộc tính cá nhân được hình thành, phát triển nhờ tố chất sẵn có và quá trình học tập, rèn luyện, cho phép học sinh huy động tổng hợp các kiến thức, kĩ năng và các thuộc tính cá nhân khác như hứng thú,

niềm tin, ý chí, ... thực hiện thành công một loại hoạt động nhất định, đạt kết quả mong muốn trong những điều kiện cụ thể” (MOET, 2018, p.37). Qua đó có thể hiểu, thang đo năng lực học sinh là một tập hợp các biểu hiện hành vi có ý nghĩa, thể hiện được các kiến thức, kỹ năng và đặc điểm cá nhân của người học cần để hoàn thành tốt một nhiệm vụ học tập hoặc giải quyết một vấn đề nào đó trong thực tiễn đời sống. Thang đo năng lực có thể sử dụng làm cơ sở để đề ra các chuẩn đầu ra cho các chương trình giáo dục, xây dựng các công cụ đánh giá và tự đánh giá năng lực của học sinh.

### 2.1.2. Tiêu chuẩn về độ tin cậy và độ giá trị của thang đo năng lực học sinh

Theo trường phái thực chứng, độ giá trị (validity) và độ tin cậy (reliability) là hai công cụ nhận thức cơ bản và là tiêu chuẩn căn bản để chuẩn hóa một thang đo nói chung và thang đo năng lực học sinh nói riêng (Winter, 2000).

#### ❖ Tiêu chuẩn về độ giá trị của thang đo năng lực học sinh

Độ giá trị của công cụ đo lường thể hiện tính đo lường chính xác đối tượng cần đo của công cụ. Cụ thể, để một công cụ đánh giá có độ giá trị thì nó phải thỏa mãn các tiêu chí: tính giá trị về nội dung (content validity); tính giá trị đo lường (measurement validity) và tính giá trị cấu trúc (construct validity). Về **giá trị về nội dung**, tiêu chí này phản ánh mức độ chính xác nội dung của lý thuyết về năng lực đang được nghiên cứu. Cụ thể, ở đây là các mô tả về chỉ số hành vi trong thang đo năng lực phải đảm bảo người đọc dễ dàng tiếp cận chính xác nội dung mà nhà nghiên cứu muốn thu thập dữ liệu thông qua các mô tả đơn nghĩa, hàm súc và trọng tâm. Căn cứ theo lý thuyết về độ giá trị của Guion (1977) để đảm bảo tiêu chí này, thang đo năng lực của học sinh cần được xem xét các khía cạnh như: nội dung mô tả chỉ số hành vi có ý nghĩa, các định nghĩa đạt được chấp nhận chung và có sự đồng thuận giữa các chuyên gia; tên thành tố năng lực hoặc được xác định rõ ràng; tên thành tố năng lực có liên quan đến mục đích đo lường năng lực học sinh; người cung cấp thông tin (học sinh, giáo viên) hiểu chính xác các nội dung được trình bày và phản hồi một cách khách quan. Về **giá trị đo lường**, tiêu chí này được xác định thông qua giá trị hội tụ (convergent validity) và giá trị phân biệt (discriminant validity). Giá trị hội tụ cho biết mức độ tương quan giữa các chỉ số hành vi trong cùng một thành tố. Giá trị phân biệt thể hiện sự phân biệt giữa thành tố năng lực này so với nhân tố khác trong cùng cấu trúc. Về **giá trị cấu trúc**, theo lý thuyết hệ thống, một năng lực nào đó được coi như một chỉnh thể “hệ thống” toàn vẹn một tổ hợp các yếu tố cấu trúc liên quan chặt chẽ với nhau trong một chỉnh thể, chứa đựng trong nó các năng lực thành tố và cấu trúc sắp xếp của các thành tố đó (Bradley & Keating, 2014). Các thành phần này có mối quan hệ chặt chẽ với nhau và có thể được xác định dựa vào quá trình thực hiện nó. Trong đó, hợp phần của năng lực liên quan đến lĩnh vực chuyên môn tạo nên năng lực. Từ đó, các năng lực thành tố được bóc tách từ năng lực hợp thành mang những tính chất đặc trưng từ các lĩnh vực chuyên môn. Cả hai thành phần hợp phần năng lực và thành tố năng lực đều là các yếu tố tiềm ẩn, không thể được đánh giá trực tiếp mà cần được đúc kết thông qua các biểu hiện của tập hợp các chỉ số hành vi tương ứng, trong một khoảng thời gian nhất định (Ta, Tran, Nguyen, & Nguyen, 2022). Chỉ số hành vi là dữ liệu cung cấp thông tin về

năng lực phục vụ kiểm tra đánh giá xuyên suốt quá trình học tập. Các hành vi được thể hiện ra bên ngoài qua các hoạt động: viết được, phát biểu được, tạo ra được... là cơ sở để đo lường nhằm xác định mức độ đạt được của năng lực. Vì vậy, để đánh giá một năng lực nào đó thì người đo lường cần xác định các cấu trúc hay các năng lực thành phần của năng lực cùng các tiêu chí tương ứng với các chỉ số hành vi cùng các mức độ đạt được của học sinh.

❖ **Tiêu chuẩn về độ tin cậy của thang đo năng lực học sinh**

Mặt khác, độ tin cậy là mức độ mà tại đó kết quả không thay đổi theo thời gian và thể hiện tính đại diện chính xác cho tổng thể được nghiên cứu (Joppe, The Research Process, 2000). Nếu kết quả nghiên cứu có thể được nhân rộng theo phương pháp luận tương tự thì khi đó, công cụ đo lường được xem là đáng tin cậy. Độ tin cậy của thang đo được đánh giá qua các tiêu chuẩn về tính ổn định (stability), tính tương đương (equivalence) và tính nhất quán (internal consistency). Về **tính ổn định**, thang đo phải đảm bảo sự ổn định theo thời gian qua các mẫu tương tự. Nghĩa là, nếu có một thang đo được kiểm tra trên các nhóm mẫu có cùng đặc tính trong khoảng thời gian phù hợp thì thang đo sẽ cho kết quả tương tự. Ngoài ra, tính ổn định có thể kiểm tra khi thực hiện trên một nhóm mẫu trong một thời gian nhất định. Giữa các lần thực hiện đo cho cùng một mẫu, kết quả thu được cũng phải tương tự nhau. Về **tính tương đương**, tiêu chí này được đảm bảo khi kiểm chứng thang đo bằng cách thực hiện sử dụng thang đo khác nhau để đo lường cùng các mẫu, kết quả thu được tương đương nhau. Ngoài ra tính tương đương còn được kiểm tra với các nhóm chuyên gia. Việc thu thập dữ liệu phỏng vấn, mỗi thành viên trong nhóm chuyên gia có cùng quan điểm về thang đo, khi đó thang đo đảm bảo tính tương đương. Công cụ đánh giá năng lực được kiểm chứng tính tương đương bằng cách xin ý kiến chuyên gia, những biến được sự đồng thuận cao sẽ được giữ nguyên, các biến có sự khác biệt về quan điểm sẽ được xem xét hiệu chỉnh. Về **tính nhất quán**, để đảm bảo tiêu chí này, trước hết chia mẫu thành các nhóm nhỏ chứa đầy đủ dữ liệu của mẫu, thực hiện đo lường trên từng nhóm mẫu. Các nhóm đối tượng trong nội tại của mẫu khi đo lường bằng thang đo phải cho kết quả nhất quán. Nếu có sự không phù hợp phải xem xét và có sự hiệu chỉnh. Trong thống kê, phương pháp thường được sử dụng để kiểm tra tính nhất quán là phương pháp phân tích đa nhóm và kiểm định Cronbach' Alpha.

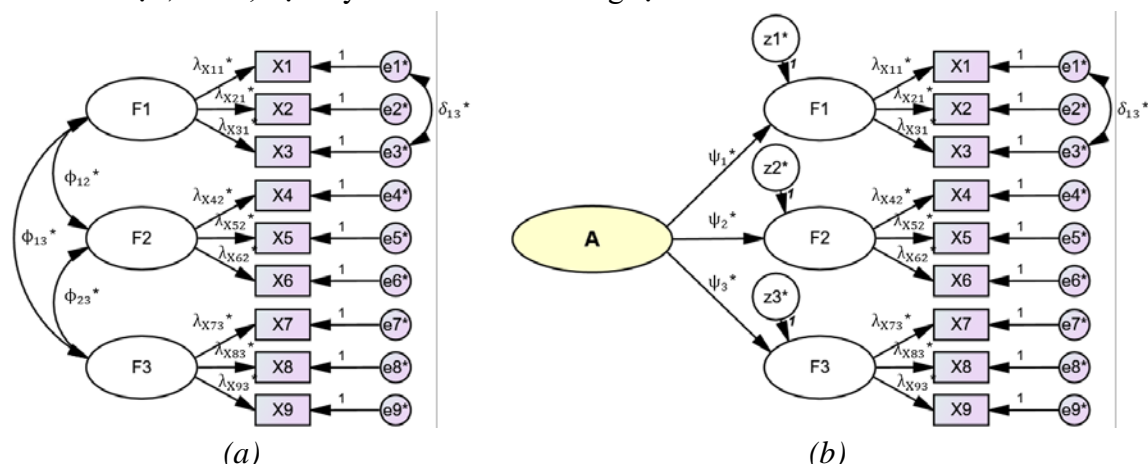
## 2.2. Mô hình hóa năng lực của học sinh bằng mô hình cấu trúc tuyến tính

### 2.2.1. Sự tương đồng giữa cấu trúc năng lực của học sinh với mô hình cấu trúc tuyến tính

Về cấu trúc, có thể mô hình hóa năng lực như một hệ thống được cấu thành từ các thành phần gồm: năng lực hợp phần (các lĩnh vực chuyên môn tạo nên năng lực), thành tố (các năng lực hoặc kỹ năng bộ phận tạo nên mỗi hợp phần), chỉ số hành vi (bộ phận được chia tách từ mỗi thành tố) được biểu hiện thông qua quá trình học tập (Hoang, 2015). Thông qua sự quy giản có chủ đích, kết quả của quá trình mô hình hóa này là một mô hình ý niệm phản ánh các khía cạnh nào đó của năng lực đang xét đến của học sinh, cho biết những thuộc tính bản chất nhất, những nguyên lý cơ bản nhất, những đặc điểm nổi bật nhất hiện có của nó (Dang, 2017).

Mô hình cấu trúc tuyến tính là một phương pháp phân tích thống kê được phát triển để phân tích mối quan hệ đa chiều giữa nhiều biến trong một mô hình (Ta & Nguyen, 2022).

Các yếu tố trong cấu trúc năng lực của học sinh được mô hình hóa trong mô hình cấu trúc tuyến tính như sau: chỉ số hành vi năng lực của học sinh được mô hình hóa thành các biến quan sát (observed variable); thành tố và năng lực hợp phần được mô hình hóa thành các biến tiềm ẩn (latent variable). Trong mô hình cấu trúc tuyến tính, các biến quan sát được dùng phản ánh một cấu trúc tiềm ẩn (khái niệm) không thể đo lường trực tiếp. Vì vậy, người nghiên cứu cần thao tác các khái niệm thành các biến mà nó đại diện. Qua đó, có thể đo lường được các biến không quan sát được khi kết nối chúng với các biến quan sát. Trong nghiên cứu khoa học hành vi, năng lực được định nghĩa là “khả năng có thể làm” và thực hiện năng lực được định nghĩa là “khả năng hoàn thành tốt một nhiệm vụ”. Việc đánh giá năng lực (competency assessment) của người học không thể được thực hiện trực tiếp và chỉ có thể đánh giá gián tiếp thông qua hiệu quả của việc thực hiện năng lực (performance assessment) với minh chứng là các biểu hiện về hành vi mà người đánh giá có thể quan sát, ghi nhận. Hay có thể hiểu, việc đánh giá năng lực của người học bản chất là quá trình ghi nhận các biểu hiện hệ quả của năng lực được người học bộc lộ trong quá trình tương tác với bối cảnh thực, từ đó, nội suy ra các thành tố năng lực tiềm ẩn.

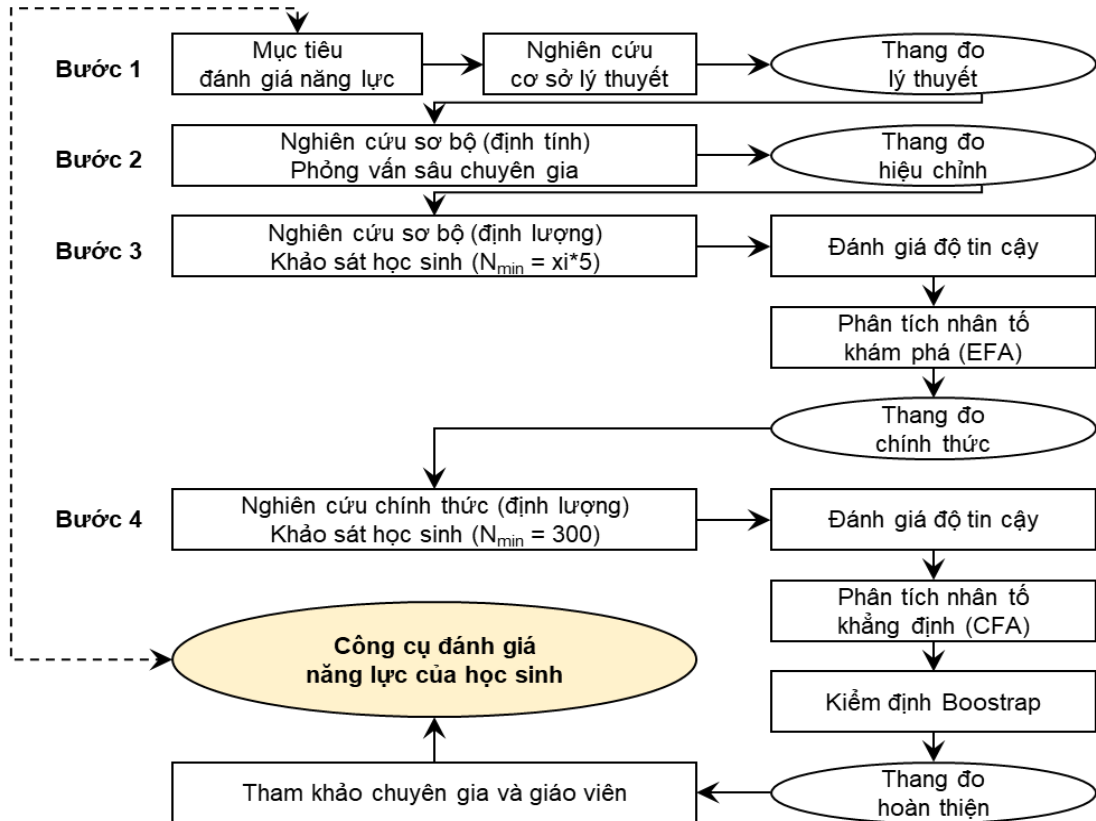


**Hình 1.** Dạng cấu trúc năng lực của học sinh được mô hình hóa bằng phần mềm AMOS: (a) mô hình cấu trúc tuyến tính bậc 1, (b) mô hình cấu trúc tuyến tính bậc 2

Hình 1 minh họa mô hình cấu trúc năng lực hợp phần A được mô hình hóa bằng mô hình cấu trúc tuyến tính với cấu trúc gồm ba thành tố năng lực  $F_1, F_2, F_3$  được biểu diễn bằng hình ellipse. Trong đó, mô hình cấu trúc tuyến tính bậc 1 phản ánh mối liên hệ giữa các chỉ số hành vi được biểu diễn bằng hình chữ nhật với các thành tố năng lực (Ví dụ:  $X_1, X_2, X_3$  là các chỉ số hành vi của thành tố  $F_1$ ). Do thành tố năng lực là nguyên nhân sinh ra các chỉ số hành vi nên để mô tả mối quan hệ này cần dùng mũi tên một chiều nối từ thành tố đến các chỉ số hành vi của thành tố ấy. Các thành tố năng lực trong cấu trúc không độc lập, mà tồn tại sự tương quan được biểu diễn bởi mũi tên hai chiều. Trong mô hình cấu trúc bậc 2, tồn tại hai mức độ của các cấu trúc. Ở cấp độ đầu tiên, cấu trúc của các thành tố vẫn được xác định thông qua các chỉ số hành vi. Tuy nhiên, sự tương quan giữa các thành tố năng lực này được giả định bởi cơ sở lí thuyết phản ánh cấu trúc không quan sát bậc 2 cho một cấp độ

khái niệm cao hơn - năng lực hợp phần A. Ngoài ra, mô hình này là chưa đạt được tính đơn nguyên do tồn tại sự tương quan giữa các sai số của hai biến chỉ số hành vi, nên để khắc phục cần nối hai phần sai số này bằng mũi tên hai chiều. Dấu \* trong mô hình thể hiện cho các tham số chưa biết (unknown model parameters) và cần được ước lượng như: hệ số tải nhân tố ( $\lambda, \psi$ ), hiệp phương sai (tương quan) giữa các thành tố năng lực ( $\phi$ ), sai số đo lường ( $e, z$ ), hiệp phương sai của các sai số chỉ số hành vi ( $\delta$ ). Trong đó, hệ số tải nhân tố đặc trưng cho mối quan hệ giữa chỉ số hành vi với thành tố năng lực hoặc mối quan hệ giữa thành tố bậc năng lực với năng lực hợp phần. Hiệp phương sai là độ đo mối quan hệ giữa của hai thành tố năng lực. Trong mô hình đo lường, mỗi chỉ số hành vi có một số hạng sai số  $e_i$  và mỗi thành tố năng lực có số hạng sai số  $z_i$ , chúng thể hiện tính không chắc chắn và không chính xác của việc đo lường, đồng thời chúng còn thể hiện tính chất này cho cả các chỉ số hành vi hay thành tố năng lực chưa được phát hiện và không được đo lường trong mô hình. Ngoài ra, nếu hiệp phương sai  $\delta_i$  của cặp sai số giữa hai chỉ số hành vi trong cùng một thành tố năng lực quá cao sẽ phản ánh hai chỉ số hành vi này có nội dung mô tả tương tự nhau dẫn đến sự chồng chéo và trùng lặp dữ liệu. Điều này sẽ được mô hình hóa bằng mũi tên hai chiều nối giữa cặp sai số của hai chỉ số hành vi bị đánh giá là có sự trùng lặp.

2.2.2. Quy trình để xây dựng và chuẩn hóa thang đo năng lực của học sinh



Hình 2. Quy trình để xây dựng và chuẩn hóa thang đo năng lực của học sinh

Để đảm bảo độ giá trị và độ tin cậy của nghiên cứu, Hình 2 mô tả quy trình xây dựng và chuẩn hóa thang đo năng lực của học sinh thông qua mô hình cấu trúc tuyến tính.

#### ❖ Bước 1. Xây dựng thang đo lí thuyết

Trên cơ sở lí thuyết về cấu trúc năng lực của học sinh đã được vận dụng để đo lường năng lực này hay một số thành tố của năng lực này ở các công trình nghiên cứu trước đó, một tập những mô tả biểu hiện hành vi được liệt kê ứng với năng lực của học sinh được bố cục trong các hoạt động dạy học cụ thể. Vận dụng cách xây dựng thang đo lí thuyết (McCoach, Gable, & Madura, 2013), nghiên cứu trong bước này sẽ gồm các công đoạn như: (1) xác định mục đích nghiên cứu (năng lực hợp phần và cấu trúc năng lực cần nghiên cứu và phát triển thành khung năng lực); (2) xác nhận rằng chưa có thang đo hiện có nào đáp ứng được mục đích nghiên cứu (nếu đã tồn tại một thang đo tương tự thì cần giải thích lí do tại sao việc phát triển một công cụ mới là cần thiết và nó sẽ khác với các thang đo hiện có như thế nào); (3) mô tả sơ bộ các khái niệm về năng lực hợp phần và thành tố năng lực; (4) giới hạn số lượng chỉ số hành vi của từng thành tố năng lực.

#### ❖ Bước 2. Nghiên cứu sơ bộ định tính

Do sự khác biệt về mức độ phát triển giáo dục, các yếu tố về văn hóa và xã hội, có thể các thang đo năng lực đã được thiết lập và vận dụng chưa thật sự phù hợp với đối tượng đang nghiên cứu. Vì vậy, quá trình nghiên cứu định tính nhằm mục tiêu điều chỉnh, bổ sung và hoàn thiện các thang đo nháp sao cho phù hợp để đánh giá năng lực của học sinh (Pham & Le, 2022). Việc nghiên cứu định tính này thường được triển khai qua phương pháp xin ý kiến chuyên gia. Trong đó, thành phần của chuyên gia gồm những người có kiến thức sâu rộng về năng lực đang được quan tâm ở học sinh (ví dụ: xây dựng năng lực vật lí của học sinh thì cần sự đánh giá của các chuyên gia về giáo dục và phương pháp dạy học Vật lí) và đội ngũ giáo viên với vai trò là đối tượng mục tiêu (target population judges) trực tiếp sử dụng công cụ đánh giá năng lực và có sự am hiểu nhất định về các biểu hiện năng lực của học sinh trong thực tế (Haynes, Richard, & Kubany, 1995; Devellis & Thorpe, 2021). Các kết quả đánh giá của chuyên gia có thể được thu thập bằng thang Likert và phân tích định lượng bằng cách sử dụng các phép thống kê, chẳng hạn như hệ số Cohen's Kappa để đo lường sự đồng thuận giữa các chuyên gia với nhau (Landis & Koch, 1977; Boateng, Neilands, Frongillo, Melgar-Quinonez, & Young, 2018).

#### ❖ Bước 3. Nghiên cứu sơ bộ định lượng

Sau khi điều chỉnh và hoàn thiện các thang đo nháp trên cơ sở nghiên cứu định tính, các thang đo tiếp tục được sử dụng cho quá trình nghiên cứu định lượng sơ bộ, thường dưới hình thức khảo sát ý kiến tự nhận thức về năng lực của người học thông qua bảng khảo sát. Tự nhận thức về năng lực là niềm tin của cá nhân về khả năng tạo ra những hiệu quả công việc nhất định (Bandura, 1977; In'am & Sutrisno, 2021). Cách thu thập thông tin này tồn tại hạn chế nhất định khi một số nghiên cứu chỉ ra rằng người học thường có xu hướng đánh giá năng lực mình cao hơn so với thực tế (Kruger & Dunning, 1999), tuy nhiên đây vẫn là cách

thu thập thông tin về năng lực người học chủ yếu khi chuẩn hóa thang đo năng lực vì một phần thông tin này dễ thu thập và một phần cũng do có rất ít lựa chọn thay thế khả thi và không tốn nhiều nhiều nguồn, đặc biệt là đối với các phương pháp đòi hỏi mẫu nghiên cứu lớn (Cohen, Manion, & Morrison, 2007). Ngoài ra, kết quả nghiên cứu của Gehringer (2017) chỉ ra rằng người học có cơ hội trải nghiệm và tiếp xúc nhiều hơn với việc tự đánh giá trong quá trình học thì sẽ nâng cao khả năng tự đánh giá tốt hơn. Do đó, tự đánh giá có thể là một công cụ đáng tin cậy để đo lường trong các nghiên cứu nếu có sự kết hợp giữa kinh nghiệm học tập (tức là học sinh của các năm học khác nhau) trong mẫu.

Trong bảng khảo sát, ngoài các câu hỏi về thông tin cá nhân, các câu hỏi chính tập trung vào các khái niệm nghiên cứu được đo lường qua các phát biểu về biến quan sát. Các phát biểu được hỏi với thang đo Likert. Quy mô mẫu phải đáp ứng tiêu chuẩn về cỡ mẫu tối thiểu. Theo (Hair et al., 2019), để phân tích thang đo trong phân tích nhân tố khám phá - EFA (Exploratory Factor Analysis) đạt đủ độ tin cậy thì kích cỡ mẫu phải gấp tối thiểu 5 lần tổng số biến quan sát trong mô hình đo lường và cỡ mẫu lớn hơn 100 là đủ độ tin cậy để phân tích các nhân tố. Từ kết quả khảo sát, các thang đo sẽ tiếp tục được kiểm định mức độ tin cậy và sự phù hợp với dữ liệu thực tế thông qua các bước phân tích đánh giá như sau:

- **Phân tích Cronbach's Alpha**

Phân tích Cronbach's Alpha nhằm để kiểm tra độ tin cậy của các chỉ số hành vi trong tập dữ liệu theo từng nhóm thành tố năng lực trong mô hình bằng hệ số Cronbach's Alpha và hệ số tương quan biến tổng (item-total correlation). Phân tích Cronbach's Alpha nằm trong nhóm phương pháp đánh giá tương quan trong hay còn gọi là đánh giá độ tin cậy nhất quán nội tại. Mục đích của việc kiểm định này là tìm hiểu xem các biến quan sát có cùng đo lường cho một khái niệm cần đo hay không (có cùng ý nghĩa trong một nhóm) hay nói cách khác, tư tưởng chung của phương pháp này là tìm kiếm sự vô lí nếu có trong các câu trả lời. Ví dụ, nếu một học sinh nào đó đã tự đánh giá rằng “biểu hiện hành vi năng lực  $X_1$  thuộc năng lực thành tố  $F_1$  của mình là rất thấp” và cho điểm thấp nhất đối với biểu hiện này ở câu hỏi đầu tiên; nhưng ở câu hỏi tiếp theo, khi được hỏi về biểu hiện hành vi  $X_2$  của cùng nhóm năng lực thành tố  $F_1$  thì HS này lại cho rằng “ $X_2$  của bản thân rất tốt” thì tương quan dữ liệu không phù hợp với suy luận logic. Điều đó dẫn đến các sai lệch có thể khi khai thác dữ liệu. Do vậy, để thang đo đạt độ tin cậy, kết quả phân tích cần phải thỏa mãn các tiêu chí sau:

+ **Hệ số tin cậy Cronbach's Alpha ( $\alpha$ ):** Hệ số này dựa trên tính nhất quán nội tại, thể hiện mức độ tương quan hay sự liên kết giữa các biến quan sát có trong thang đo (Tavakol & Dennick, 2011). Hệ số  $\alpha$  càng lớn thì độ tin cậy nhất quán nội tại càng cao (Devellis & Thorpe, 2021). Trong các nghiên cứu về giáo dục, hệ số  $\alpha$  lớn hơn 0,7 là một tiêu chuẩn bắt buộc (Taber, 2018).

+ **Hệ số tương quan biến tổng của các biến quan sát:** Giá trị này là hệ số tương quan của một chỉ số hành vi với điểm trung bình của các chỉ số hành vi khác trong cùng một thang đo thành tố. Chỉ số hành vi nào có hệ số tương quan biến tổng càng cao thì sự tương quan



của biến đó với các biến khác trong nhóm càng cao và cần đạt giá trị tối thiểu là 0,3 (Nunnally & Bernstein, The Assessment of Reliability, 1994).

- **Phân tích nhân tố khám phá (EFA)**

Phân tích EFA là một phương pháp thống kê được sử dụng khi mối quan hệ giữa các biến quan sát và thang đo hay biến tiềm ẩn là không rõ ràng hoặc không chắc chắn. Phân tích EFA được tiến hành theo kiểu khám phá để xác định xem phạm vi, mức độ quan hệ giữa các biến quan sát và các nhân tố cơ sở như thế nào, làm nền tảng cho một tập hợp các phép đo để rút gọn hay giảm bớt số biến quan sát tải lên các nhân tố cơ sở (Hair, Black, Babin, & Anderson, 2019). Hay nói cách khác, phân tích EFA giúp rút gọn một tập gồm nhiều biến quan sát chỉ số hành vi thành một nhóm thành tố năng lực có ý nghĩa hơn nhưng vẫn thể hiện hầu hết nội dung thông tin của các biến ban đầu và đồng thời còn xác định được giá trị hội tụ, độ giá trị phân biệt của nhóm biến. Các tiêu chí đánh giá thang đo khi thực hiện phân tích EFA như sau:

+ *Tiêu chí để thang đo đạt giá trị hội tụ*: hệ số tương quan đơn giữa các chỉ số hành vi và các thành tố (factor loading) phải lớn hơn hoặc bằng 0,5 trong cùng một thành tố. Những biến quan sát chỉ số hành vi nào có hệ số tải này nhỏ hơn 0,5 trong phân tích EFA sẽ tiếp tục bị loại bỏ khỏi mô hình (Hair, Black, Babin, & Anderson, 2019).

+ *Tiêu chí để thang đo đạt giá trị phân biệt*: độ chênh lệch giữa hệ số tải tối đa và các hệ số tải khác phải đạt từ 0,3 trở lên đối với mỗi biến quan sát chỉ số hành vi.

+ *Xác định số lượng nhân tố*: Eigenvalue là chỉ số xác định số lượng thành tố. Eigenvalue không nhỏ hơn 1 mới được giữ lại trong mô hình phân tích.

+ *Tiêu chí về phương sai trích (variance explained criteria)*: tổng phương sai trích - TVE (total variance explained) phải lớn hơn 50% (Gerbing & Anderson, 1988).

+ *Hệ số KMO (Kaiser-Meyer-Olkin measure of sampling adequacy)*: chỉ số dùng để so sánh độ lớn của hệ số tương quan giữa hai chỉ số hành vi với độ lớn của hệ số tương quan riêng phần của chúng, thể hiện mức độ phù hợp của phương pháp phân tích EFA, nếu hệ số KMO lớn hơn 0,5 và nhỏ hơn 1 thì phân tích nhân tố được coi là phù hợp (Hair, Black, Babin, & Anderson, 2019).

+ *Kiểm định Bartlett*: xem xét giả thiết mức độ tương quan giữa các biến quan sát trong tổng thể bằng 0. Nếu như kiểm định này có ý nghĩa thống kê ( $p < 0,05$ ) thì các biến quan sát chỉ số hành vi có tương quan nhau trong tổng thể.

- **Bước 4. Nghiên cứu chính thức định lượng**

Nghiên cứu chính thức sẽ được thực hiện với quy mô mẫu lớn hơn sau khi đã có những hiệu chỉnh thang đo từ bước nghiên cứu định lượng sơ bộ. Một số nhà khoa học đề xuất tiêu chuẩn để xác định cỡ mẫu trong nghiên cứu phát triển một thang đo như: Guadagnoli và Velicer (1988) dựa trên nghiên cứu mô phỏng bằng cách sử dụng các cỡ mẫu khác nhau, đề xuất rằng cần tối thiểu khoảng 300 - 450 mẫu để quan sát khả năng so sánh có thể chấp nhận được của các mẫu và cần thiết lặp lại nghiên cứu nếu cỡ mẫu nhỏ hơn 300; Comrey và Lee

(2013) đề xuất một thang đánh giá phân loại cho cỡ mẫu để phát triển thang đo: khoảng 100 là kém, 200 là khá, 300 là tốt, 500 là rất tốt, lớn hơn 1000 là xuất sắc; (Hair et al., 2019) cho rằng mô hình SEM có từ 7 khái niệm nghiên cứu (trong trường hợp này là các thành tố năng lực) trở xuống thì 300 là cỡ mẫu tối thiểu. Bên cạnh đó, nếu thang đo năng lực được áp dụng riêng cho một nhóm đối tượng học sinh mà đã biết kích thước của tổng thể thì cần áp dụng thêm tiêu chuẩn cỡ mẫu tối thiểu theo công thức Slovin với  $n = N/(1+N*\sigma^2)$ ; trong đó  $n$  là kích thước mẫu,  $N$  là kích thước tổng thể đã biết,  $\sigma$  là sai số trung bình của dữ liệu và thường được chọn bằng 5% (tương ứng với độ tin cậy 95%). Ngoài ra, mẫu được sử dụng trong bước nghiên cứu này cũng cần đảm bảo phân phối chuẩn với giá trị trung bình và trung vị gần bằng nhau, giá trị tuyệt đối của độ nghiêng skewness nhỏ hơn 2 và giá trị tuyệt đối của độ nhọn kurtosis nhỏ hơn 2 (Dash & Paul, 2021). Khi đã đảm bảo các yêu cầu về đầu vào của mẫu nghiên cứu, mô hình đo lường sẽ được kiểm định bằng phương pháp phân tích nhân tố khẳng định – CFA (Confirmatory Factor Analysis). Trong CFA, nhà nghiên cứu chỉ định một mô hình lý thuyết đưa ra giả thuyết về cách các biến được đo có liên quan đến cấu trúc cơ bản của năng lực cần đánh giá của học sinh. Ở đây, mô hình năng lực của học sinh có thể biểu diễn dưới hai dạng CFA bậc 1 và bậc 2 (như đã đề cập ở Hình 1). Sau đó, mô hình được kiểm tra bằng cách so sánh dữ liệu quan sát được (dữ liệu về các mức độ đạt được của các chỉ số hành vi của học sinh) với ma trận hiệp phương sai của mô hình. Mức độ phù hợp của mô hình được đánh giá bằng các chỉ số phù hợp khác nhau. Nếu thỏa mãn tất cả các tiêu chí đánh giá thì có thể kết luận rằng mô hình cấu trúc năng lực được nghiên cứu là phù hợp, qua đó sẽ phản ánh mức độ hiểu biết của nhà nghiên cứu về bản chất của cấu trúc năng lực đó.

- **Đánh giá độ tin cậy của thang đo**

Để đánh giá độ tin cậy của thang đo các nhà khoa học căn cứ vào hai chỉ số là hệ số  $\alpha$  và độ tin cậy tổng hợp – CR (composite reliability). Độ tin cậy tổng hợp tốt nhất khi có giá trị lớn hơn 0,7 (Hair, Black, Babin, & Anderson, 2019). Khi đánh giá tính ổn định nội tại của thang đo, nên xem xét cả hai tiêu chí của độ tin cậy nhất quán nội tại với hệ số  $\alpha$  là giới hạn dưới và CR là giới hạn trên (Hair, Hult, Ringle, & Sarstedt, 2021). Hệ số độ tin cậy tổng hợp được tính theo công thức của Jöreskog (1971) dựa trên hệ số hồi quy chuẩn hóa của các biến quan sát.

- **Đánh giá giá trị hội tụ**

Giá trị hội tụ cho biết mức độ hội tụ hay chia sẻ một tỉ lệ phương sai chung của các chỉ báo cho một khái niệm, nghĩa là để đạt được giá trị hội tụ thì tập hợp các biến đo lường một khái niệm nghiên cứu phải có mối tương liên hệ tương quan cao (Kline, 2016). Điều này được thể hiện qua các hệ số hồi quy chuẩn hóa đối với từng biến của khái niệm tiềm ẩn nếu nó là thang đo đơn hướng. Để một thang đo đạt giá trị hội tụ thì hệ số hồi quy chuẩn hóa của các biến thành phần của thang đo phải lớn hơn 0,5 và có ý nghĩa thống kê (Gerbing & Anderson, 1988). Ngoài ra, giá trị hội tụ cũng được đánh giá qua phương sai trích trung bình AVE (average variance extracted). Phương sai trích trung bình phản ánh lượng biến thiên

chung của các biến quan sát được tính toán bởi biến tiềm ẩn và phải đạt giá trị lớn hơn 0,5 tương ứng với mỗi khái niệm (Hair, Black, Babin, & Anderson, 2019). Phương sai trích trung bình được tính theo công thức của Fornell-Larcke dựa vào hệ số hồi quy chuẩn hóa của các biến quan sát.

**- Đánh giá giá trị phân biệt**

Giá trị phân biệt cho biết mức độ mà một khái niệm này thật sự phân biệt về mặt thống kê với một khái niệm khác cả về sự tương quan lẫn các biến đo lường trong mỗi khái niệm (Hair, Black, Babin, & Anderson, 2019). Cách tiếp cận truyền thống đánh giá tính phân biệt là tiêu chí Fornell-Larcker, được xác định bằng điều kiện hệ số căn bậc hai của AVE -  $\sqrt{AVE}$  (square root of AVE) phải lớn hơn hệ số tương quan giữa các khái niệm (inter-construct correlations) (Fornell & Larcker, 1981). Ngoài tiêu chí Fornell-Larcker, còn có tiêu chí đánh giá khác là tiêu chí tương quan giữa hai khái niệm phải nhỏ hơn 0,85 và phương sai riêng lớn nhất MSV (maximum shared variance) nhỏ hơn chỉ số AVE (Kline, 2016).

**- Kiểm định sự phù hợp của mô hình đo lường**

Trước đây,  $\chi^2$  là tiêu chuẩn đánh giá mức độ phù hợp của mô hình cổ điển thông qua phản ánh sự khác biệt giữa ma trận phương sai-hiệp phương sai của mô hình ước lượng và ma trận ban đầu của các biến quan sát. Khi  $\chi^2$  không có ý nghĩa thống kê thì có nghĩa ma trận phương sai-hiệp phương sai của mô hình ước lượng và ma trận ban đầu của các biến quan sát tương đương nhau. Điều này phản ánh rằng mô hình lí thuyết là sự mô phỏng mối quan hệ của phương sai-hiệp phương sai mẫu. Tuy nhiên, tiêu chí này rất nhạy với cỡ mẫu, và với cỡ mẫu lớn (trên 200)  $\chi^2$  thường có ý nghĩa về mặt thống kê. Do đó, với phương pháp phân tích thống kê với cỡ mẫu lớn như SEM thì tiêu chí này thường không thỏa mãn và được đánh giá là quá khắt khe khi yêu cầu đánh giá mức độ phù hợp hoàn hảo của mô hình. Thay vào đó, để đo lường mức độ phù hợp của mô hình, các nhà nghiên cứu đề xuất các tiêu chuẩn như: độ phù hợp tuyệt đối (absolute fit), độ phù hợp tương đối (relative fit), sự phù hợp được hiệu chỉnh cẩn thận (parsimonious fit) (Nguyen, Nguyen, Dang, & Nguyen, 2020; Hair, Black, Babin, & Anderson, 2019). Độ phù hợp tuyệt đối là các chỉ số đánh giá mức độ khớp giữa mô hình năng lực được đề xuất và dữ liệu thực tế bằng cách so sánh các giá trị dự đoán của mô hình với các giá trị quan sát thực tế. Độ phù hợp tương đối là các chỉ số đánh giá mức độ khớp giữa mô hình và dữ liệu thực tế bằng cách so sánh mô hình hiện tại với một mô hình tham chiếu (null model). Độ phù hợp được hiệu chỉnh cẩn thận là các chỉ số đánh giá mức độ khớp giữa mô hình và dữ liệu thực tế bằng cách so sánh mô hình hiện tại với mô hình tham chiếu đã được hiệu chỉnh cho số lượng tham số trong mô hình được ước lượng. Bảng 1 cung cấp các tiêu chuẩn đánh giá độ phù hợp của mô hình với các chỉ số cơ bản.

**Bảng 1.** Các tiêu chuẩn đánh giá độ phù hợp của mô hình năng lực của học sinh

Tiêu chí đánh giá	Chỉ số	Tiêu chuẩn	Đánh giá
	p-value	$\geq 0,05$	Hoàn hảo
	Bậc tự do của mô hình df	Càng nhỏ càng tốt	
Đánh giá sự phù hợp tuyệt đối	$\chi^2/df$	$\leq 5,00$ $\leq 2,00$	Chấp nhận được Tốt
	GFI	$\geq 0,80$ $\geq 0,90$	Chấp nhận được Tốt
		$\geq 0,95$	Rất tốt
	RMSEA	$\leq 0,08$ $\leq 0,05$	Tốt Rất tốt
		SRMR	$\leq 0,10$
	Đánh giá sự phù hợp tương đối	IFI	$\geq 0,90$
TLI		$\geq 0,90$	Phù hợp
CFI		$\geq 0,95$	Phù hợp
RFI		$\geq 0,05$	Phù hợp
Đánh giá sự hiệu chỉnh cẩn thận	PGFI	$> 0,50$	Phù hợp
	PNFI	$> 0,50$	Phù hợp
	PCFI	$> 0,50$	Phù hợp

**- Kiểm định sự phù hợp của mô hình đo lường**

Thủ tục Bootstrap dùng để đánh giá tính ổn định của độ tin cậy của các hệ số ước lượng cũng như có thể đánh giá được tính bền vững của mô hình cấu trúc năng lực của học sinh. Đây là thủ tục lấy mẫu lặp lại có thay thế từ mẫu ban đầu, trong đó mẫu ban đầu đóng vai trò là tổng thể (Efron, 2000). Số lần lấy mẫu lặp lại trong nghiên cứu này được chọn là N lần, kết quả ước lượng với N lần từ n mẫu được tính trung bình và giá trị này có xu hướng gần với ước lượng của tổng thể. Nếu kết quả phân tích cho thấy độ chệch của ước lượng (Bias) và sai lệch chuẩn (Se) có giá trị rất nhỏ, không có ý nghĩa thống kê ( $p > 0,05$ ) thì có thể kết luận các ước lượng của các nhân tố trong mô hình SEM là tin cậy.

**2.2.3. Sự đáp ứng các tiêu chuẩn về độ tin cậy và độ giá trị của quy trình**

Đối chiếu với các tiêu chuẩn nghiêm ngặt về độ giá trị và độ tin cậy, thiết kế nghiên cứu kì vọng sẽ giúp các nhà nghiên cứu phát triển và chuẩn hóa được các thang đo năng lực của học sinh chất lượng trên cơ bản đã đảm bảo toàn diện về các hiệu lực như giá trị của nội dung, giá trị về đo lường, giá trị về cấu trúc, tính ổn định, tính tương đương và tính nhất quán (Bảng 2).

**Bảng 2.** Sự đáp ứng tiêu chuẩn về độ tin cậy và độ giá trị của quy trình xây dựng và chuẩn hóa thang đo năng lực của học sinh

Quy trình mô hình hóa khung đánh giá năng lực của học sinh	Độ giá trị			Độ tin cậy		
	Giá trị nội dung	Giá trị đo lường	Giá trị cấu trúc	Tính ổn định	Tính tương đương	Tính nhất quán
Bước 1	Tổng quan lí thuyết	<input checked="" type="checkbox"/>				
Bước 2	Delphi	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	
Bước 3	Cronbach's Alpha					<input checked="" type="checkbox"/>
	Phân tích nhân tố khám phá		<input checked="" type="checkbox"/>			
Bước 4	Độ tin cậy tổng hợp CR					<input checked="" type="checkbox"/>
	Hệ số hồi quy chuẩn hóa		<input checked="" type="checkbox"/>			
	Phương sai trích trung bình AVE		<input checked="" type="checkbox"/>			
	Tiêu chí Fornell-Larcker		<input checked="" type="checkbox"/>			
	Kiểm định sự phù hợp của mô hình phân tích nhân tố khẳng định		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
	Phân tích Bootstrapping				<input checked="" type="checkbox"/>	

### 3. Kết luận

Tóm lại, thông qua tổng quan tài liệu về đánh giá năng lực và mô hình cấu trúc tuyến tính, bài báo này đã cung cấp sơ bộ về các bước chính trong quá trình hình thành và phát triển thang đo năng lực cho người học, cũng như diễn giải các cơ sở khoa học cho phương pháp luận này. Đối chiếu với các tiêu chuẩn nghiêm ngặt về độ giá trị và độ tin cậy, thì quy trình được đề xuất cho thấy sự đáp ứng toàn diện về các hiệu lực như giá trị của nội dung, giá trị về đo lường, giá trị về cấu trúc, tính ổn định, tính tương đương và tính nhất quán. Qua đó, bài viết này cho thấy tiềm năng mở rộng ứng dụng của mô hình cấu trúc tuyến tính trong các nghiên cứu về giáo dục. Đối với hoạt động kiểm tra đánh giá, đề xuất này được kì vọng sẽ hỗ trợ các nhà nghiên cứu trong quá trình phát triển thêm các thang đo năng lực chuẩn hóa cho đối tượng học sinh trong tương lai.

- ❖ **Tuyên bố về quyền lợi:** Tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.
- ❖ **Lời cảm ơn:** Nghiên cứu này được tài trợ Chương trình học bổng đào tạo thạc sĩ, tiến sĩ trong nước của Quỹ Đổi mới sáng tạo Vingroup (VINIF), mã số VINIF.2022.ThS.097.

### TÀI LIỆU THAM KHẢO

- Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, 84, 191-215.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Front Public Health*, 6, Article 149. doi:10.3389/fpubh.2018.00149
- Bradley, J. M., & Keating, C. B. (2014). Systems Theory for a Competency Model Framework. *9th International Conference on System of Systems Engineering (SOSE)* (pp. 61-66). Adelaide, Australia: IEEE. doi:10.1109/SYSOSE.2014.6892464
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education (6th ed.)*. New York, NY, USA: Routledge. Retrieved from <https://gtu.ge/Agro-Lib/RESEARCH%20METHOD%20COHEN%20ok.pdf>
- Colby, R. L. (2019). *Competency-Based Education: A New Architecture for K-12 Schooling*. Cambridge, Cambridgeshire, England: Harvard Education Press.
- Comrey, A. L., & Lee, H. B. (2013). *A First Course in Factor Analysis (2nd ed.)*. New York, NY, US: Psychology Press.
- Dang, T. H. (2017). Modeling in Educational Research [Mo hình hoa trong nghiên cứu giáo dục]. *Vietnam Science and Education Journal*, 137, 12-16. Retrieved from <http://vjes.vnies.edu.vn/vi/mo-hinh-hoa-trong-nghien-cuu-giao-duc>
- Dash, G., & Paul, J. (2021). CB-SEM and PLS-SEM Methods for Research in Social Sciences and Technology Forecasting. *Technological Forecasting & Social Change*, 173(3), Article 121092. doi:10.1016/j.techfore.2021.121092
- Devellis, R., & Thorpe, C. T. (2021). *Scale Development Theory and Applications (5th ed.)*. Thousand Oaks, CA, US: SAGE Publications.
- Efron, B. (2000). The Bootstrap and Modern Statistics. *Journal of the American Statistical Association*, 95(452), 1293-1296. doi:10.1080/01621459.2000.10474333
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39-50. doi:10.1177/002224378101800104
- Gehring, E. F. (2017). Self-Assessment to Improve Learning and Evaluation. *2017 ASEE Annual Conference and Exposition*, (p. 19411). Columbus, Ohio, US. doi:10.18260/1-2--28816
- Gerbing, D. W., & Anderson, J. C. (1988). An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment. *Journal of Marketing Research*, 25(2), 186-192. doi:10.2307/3172650
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of Sample Size to The Stability of Component Patterns. *Psychological Bulletin*, 103(2), 265-275. doi:10.1037/0033-2909.103.2.265
- Guion, R. M. (1977). Content Validity: The Source of My Discontent. *Applied Psychological Measurement*, 1(1), 1-10. doi:10.1177/014662167700100103
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis (8th ed.)*. Cengage Learning, EMEA.

- Hair, J. F., Hult, G. T., Ringle, C. M., & Sarstedt, M. (2021). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) (3rd ed.)*. Thousand Oaks, CA, US: SAGE Publications.
- Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. *Psychological Assessment*, 7(3), 238-247. doi:10.1037/1040-3590.7.3.238
- Hoang, H. B. (2015). Competence and Assessment by Competence [Nang luc va danh gia theo nang luc]. *Ho Chi Minh City University of Education Journal of Science*, 6(71), 21-32. doi:10.54607/hcmue.js.0.6(71).667.658(2015)
- In'am, A., & Sutrisno, E. S. (2021). Strengthening Students' Self-efficacy and Motivation in Learning Mathematics through the Cooperative Learning Model. *International Journal of Instruction*, 14(1), 395-410. doi:10.29333/iji.2021.14123
- Joppe, M. (2000). *The Research Process*. Retrieved from <https://www.uoguelph.ca/hftm/research-process>
- Jöreskog, K. G. (1971). Simultaneous Factor Analysis in Several Populations. *Psychometrika*, 36, 409-426. doi:10.1007/BF02291366
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling (4th ed.)*. The Guilford Press.
- Kruger, J., & Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134. doi:10.1037/0022-3514.77.6.1121
- Landis, J. R., & Koch, G. G. (1977). An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics*, 33(2), 363-374. doi:10.2307/2529786
- McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument Development in the Affective Domain (3rd ed.)*. New York, NY, US: Springer. doi:10.1007/978-1-4614-7135-6
- MOET. (2018). *General Education Curriculum [Chuong trinh giao duc pho thong tong the]*. Hanoi.
- Nunnally, J. C., & Bernstein, I. H. (1994). The Assessment of Reliability. In *Psychometric Theory* (Vol. 3, pp. 248-292). New York, NY, US: McGraw-Hill.
- Nguyen, V. B., Nguyen, A. T., Dang, V. S., & Nguyen, T. T. (2020). Reliability and Validity an Instrument to Assess Creative Competency in Engineering Design on STEM Education [Xay dung cong cu danh gia nang luc sang tao thiet ke ki thuat trong giao duc STEM]. *HNUE Journal of Science*, 65(1), 151-162. doi:10.18173/2354-1075.2020-0015
- Pham, D. V., & Le, T. M. (2022). Designing a Rubric to Assess Scientific Study Competence in Teaching Natural Sciences in Junior High School [Thiet ke khungkiem tra, danh gia nang luc nghien cuu khoa hoc trong day hoc mon Khoa hoc tu nhien trung hoc co so]. *Ho Chi Minh City University of Education Journal of Science*, 19(2), 240-250. doi:10.54607/hcmue.js.19.2.3009(2022)
- Rychen, D. S., & Salganik, L. H. (2002). *Definition and Selection of Competencies (DESECO): theoretical and conceptual foundations. Strategy paper*. Neuchatel, Switzerland: Swiss Federal Statistical Office. Retrieved from <http://hdl.voced.edu.au/10707/156754>
- Ta, T. T., & Nguyen, T. N. (2022). A Comparison of Using CB-SEM and PLS-SEM to Assess Training Effectiveness Evaluation Model for Teacher's Online Continuing Professional Development [So sanh phuong phapphan tich CB-SEM va PLS-SEM trongkiem dinh mo hinh danh gia hieu qua cua hinh thuc boi duong giao vien truc tuyen]. *Ho Chi Minh City University of Education Journal of Science*, 19(2), 213-228. doi:10.54607/hcmue.js.19.2.3306(2022)
- Ta, T. T., Tran, T. X., Nguyen, P. U., & Nguyen, T. N. (2022). Construct and Standardize the STEM Competency Assessment Tool for High School Students in Ho Chi Minh City [Xay dung va chuan hoa cong cu danh gia nang luc STEM cua hoc sinh trung hoc pho thong tai Thanh pho

- Ho Chi Minh]. *Ho Chi Minh City University of Education Journal of Science*, 19(8), 1255-1270. doi:10.54607/hcmue.js.19.8.3408(2022)
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48, 1273-1296. doi:10.1007/s11165-016-9602-2
- Tavakol, M., & Dennick, R. (2011). Making Sense of Cronbach's Alpha. *International Journal of Medical Education*, 2, 53-55. doi:10.5116/ijme.4dfb.8dfd
- Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C. H. (2011). 5 Development of Competencies Across the Life Span. *Zeitschrift für Erziehungswissenschaft*, 14, 67-86. doi:10.1007/s11618-012-0265-0
- Winter, G. (2000). A Comparative Discussion of the Notion of 'Validity' in Qualitative and Quantitative Research. *The Qualitative Report*, 4(3), 1-14. doi:10.46743/2160-3715/2000.2078
- 

**PROPOSING THE PROCESS OF DEVELOPING  
A SCALE FOR STUDENTS' COMPETENCY:  
A STRUCTURAL EQUATION MODELING APPROACH**

*Tạ Thanh Trung*

*Saigon Education and Science joint stock Company*

*Corresponding author: Tạ Thanh Trung – Email: trungtphysics@gmail.com*

*Received: April 26, 2023; Revised: August 10, 2023; Accepted: August 24, 2023*

**ABSTRACT**

*With the trend of shifting from content-based to student-centered teaching, developing and standardizing a student competency assessment framework is an important task to ensure the quality of education. Although many recent studies have focused on developing frameworks to assess students' competencies using various methods, only some studies have demonstrated the reliability and validity of such assessment frameworks/tools. As a result, this article presents a research design to standardize a student competency assessment framework, which includes the following stages: (1) developing a theoretical scale; (2) preliminary qualitative research using the Delphi method; (3) preliminary quantitative research using the EFA method; and (4) formal quantitative research using the CFA method. Furthermore, the study demonstrates the process can meet the rigorous standards of testing the reliability and validity of a student assessment framework. This research design can assist educators in standardizing complex competency frameworks and assist teachers in selecting behavioral indicators to assess student learning outcomes.*

**Keywords:** CFA; competency assessment; EFA; SEM; systems theory