

## BỘ TIÊU CHÍ/BẢNG HỎI ĐÁNH GIÁ LỚP HỌC PHẦN TẠI TRƯỜNG ĐẠI HỌC NGOẠI THƯƠNG: NGHIÊN CỨU SO SÁNH VỀ ĐỘ TIN CẬY VÀ TÍNH GIÁ TRỊ

Dương Thị Thanh Thuỷ

Trường Đại học Ngoại thương

Email: [thuydt@ftu.edu.vn](mailto:thuydt@ftu.edu.vn)

### Article history

Received: 27/8/2025

Accepted: 29/9/2025

Published: 20/12/2025

### Keywords

Course evaluation,  
quantitative, reliability,  
criteria set, validity,  
questionnaire

### ABSTRACT

To collect learners' evaluation of various aspects of a course, higher education institutions often build a set of criteria/checklist using a multi-level Likert scale. This study was conducted to compare the reliability and validity of two sets of criteria/checklist for course evaluation in a public university in Vietnam. One set is currently used by its Centre of Testing and Quality Assurance for all courses of various subjects and the other is exclusively designed for a foreign language course for specific purposes (ESP2) at Foreign Trade University. The primary data was collected via a paper-based questionnaire using a 5-level Likert scale. The study showed that many low-quality variables had been dismissed from both sets of criteria/checklist and the set exclusively designed for ESP2 prevailed both in terms of reliability and validity. The study concludes with a set of criteria/questionnaire officially recommended for the evaluation of ESP2 courses meeting two basic standards of reliability and validity. It also makes some methodological implications for scientifically and objectively evaluating a set of criteria/checklist with quantitative tools. The paper not only provides an empirical basis for selecting a suitable set of assessment criteria for courses but also contributes methodologically to research and practice in ensuring the quality of higher education.

### 1. Mở đầu

Đánh giá lớp học phần (HP) là một công tác được chú trọng đặc biệt tại các cơ sở giáo dục đại học trên thế giới. Công cụ thường dùng để đánh giá lớp HP là bộ tiêu chí/bảng hỏi sử dụng thang đo Likert nhiều bậc được xây dựng nhằm đạt được những mục đích đánh giá nhất định. Ở Việt Nam, các cơ sở giáo dục đại học tuân theo các quy định của cơ quan quản lý nhà nước về giáo dục và các bộ phận trực thuộc xây dựng bộ tiêu chí/bảng hỏi đánh giá lớp HP chủ yếu nhằm mục tiêu đảm bảo chất lượng và nâng cao hiệu quả dạy và học HP thông qua việc lấy ý kiến đánh giá của người học (Cục Nhà giáo và Cán bộ quản lý giáo dục, 2013; Bộ GD-ĐT, 2021, 2024). Các trung tâm đảm bảo chất lượng của các đại học thường là đơn vị chuyên trách thực hiện việc đánh giá lớp HP này. Các nghiên cứu trên thế giới chỉ ra rằng bộ tiêu chí/bảng hỏi đánh giá lớp HP sử dụng thang đo Likert cần đạt được hai tiêu chuẩn cần bản gồm: (1) độ tin cậy của thang đo và (2) tính giá trị của bảng hỏi để các kết quả đánh giá thực sự khoa học và chính xác (Kember và Leung, 2008; Kogan và Shea, 2007; Papadogiannis và cộng sự, 2024; Schiekirka và Raupach, 2015). Để đánh giá tính giá trị của bộ tiêu chí/bảng hỏi, phân tích nhân tố (factor analysis) là phương pháp phổ biến nhất giúp xác định chiều hướng thang đo (scale dimensionality). Để đánh giá độ tin cậy của thang đo cần ước lượng sự nhất quán nội bộ (internal consistency) của các tiêu chí/câu hỏi trong bộ tiêu chí/bảng hỏi bằng hệ số Cronbach's Alpha của các nhân tố/biến tổng với nhiều biến quan sát thành phần. Đây là yếu tố mà các nhà đánh giá giáo dục đặc biệt chú ý khi xây dựng công cụ mới nhằm đảm bảo tính nhất quán các điểm đánh giá của các tiêu chí trong những thang đo phức hợp (Copeland và Hewson, 2000; Litzelman và cộng sự, 2005; Steiner và cộng sự, 2003).

Bối cảnh của nghiên cứu này là việc đánh giá hai bộ tiêu chí/bảng hỏi đánh giá lớp HP tại một đại học công lập ở Việt Nam. Một bộ đang được Trung tâm khảo thí và đảm bảo chất lượng (TTKTĐBCL) của trường sử dụng chung cho tất cả các lớp HP, không phân biệt thuộc HP nào và một bộ được thiết kế riêng cho một HP ngoại ngữ chuyên ngành (kí hiệu: ESP2) ở một nghiên cứu trước đó. Bài báo đánh giá và so sánh: (1) độ tin cậy của thang đo; (2) tính giá trị của bảng hỏi của hai bộ tiêu chí trên khi áp dụng chúng lấy ý kiến đánh giá của sinh viên (SV) các lớp HP có

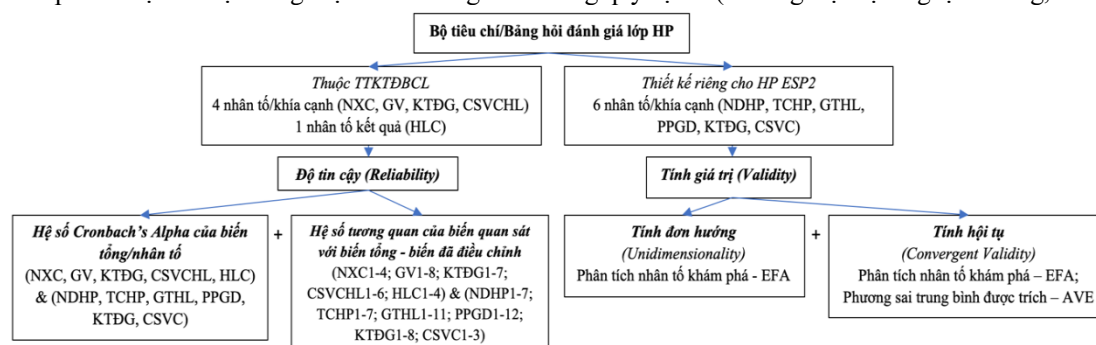
mã số ESP121 thuộc ESP2 trong thực tiễn; từ đó rút ra một bộ tiêu chí đáp ứng cả hai tiêu chuẩn nói trên đề xuất chính thức áp dụng cho các lớp HP ESP121 thuộc ESP2 của trường trong thời gian tới.

## 2. Kết quả nghiên cứu

### 2.1. Khung đánh giá Bộ tiêu chí/bảng hỏi đánh giá lớp học HP

Đo lường định lượng độ tin cậy (reliability) và tính giá trị (validity) của bộ tiêu chí/bảng hỏi đánh giá lớp HP sử dụng thang đo Likert nhiều bậc là một khía cạnh quan trọng trong cách thức mà chủ thể đánh giá triển khai việc đánh giá với một/một số lớp HP nhất định. Crossley và cộng sự (2002) lập luận rằng, cần có một nghiên cứu thử nghiệm để đánh giá tính khả thi, sự chấp nhận được, tính giá trị (có liên quan tới mục đích dự kiến) và độ tin cậy (có liên quan tới phạm vi khái quát hoá dự kiến) của một bộ tiêu chí/bảng hỏi được thiết kế để lấy ý kiến đánh giá của người học với những mục đích nhất định. Greenwald (1997) đặt vấn đề về tính giá trị của các đánh giá của người học. Bốn khía cạnh cần được xem xét gồm: cấu trúc khái niệm (đánh giá đơn hướng hay đa hướng), tính hội tụ (các đánh giá tương quan chặt chẽ mức nào với những chỉ số khác nhau về hiệu quả giảng dạy), tính phân biệt (các đánh giá có bị ảnh hưởng bởi các nhân tố ngoài hiệu quả giảng dạy hay không), và tính hệ quả (các đánh giá có được sử dụng hiệu quả trong đánh giá và phát triển nhân sự hay không). Kogan và Shea (2007) cho rằng, tính giá trị của bộ tiêu chí/bảng hỏi liên quan tới sự diễn giải và dùng các điểm số gán cho công cụ đánh giá. Tính giá trị của bộ tiêu chí/bảng hỏi thể hiện ở chỗ đo được cái cần đo, xác định được điểm đánh giá theo chiều hướng nào. Về độ tin cậy của thang đo, các tác giả cho rằng tính nhất quán nội bộ sẽ phản ánh các tiêu chí/câu hỏi trong một thang đo/nhân tố đo lường cho một ý tưởng/khái niệm duy nhất. Beckman và cộng sự (2003) tổng quan hệ thống trong số 21 công cụ đánh giá giảng dạy lâm sàng trong giáo dục y khoa thì phân tích nhân tố là phương pháp phổ biến nhất để xác định chiều hướng thang đo và ước lượng tính nhất quán nội bộ của các tiêu chí/câu hỏi bằng hệ số Cronbach's Alpha.

Trên cơ sở các nghiên cứu trước về đánh giá lớp HP sử dụng bộ tiêu chí/bảng hỏi với thang đo Likert nhiều bậc, tác giả xây dựng khung đánh giá độ tin cậy và tính giá trị cho hai bộ tiêu chí/bảng hỏi thuộc nghiên cứu trong hình 1. Lớp HP thuộc nghiên cứu này được định nghĩa là: “Lớp được tổ chức cho những SV cùng học một HP theo cùng thời khoá biểu, cùng giảng đường và được tổ chức dựa trên kết quả đăng kí khối lượng học tập của SV ở từng học kì. Mỗi lớp HP được kí hiệu bằng một mã số riêng do Trường quy định” (Trường Đại học Ngoại thương, 2021).



Hình 1. Khung đánh giá Bộ tiêu chí/bảng hỏi đánh giá lớp HP (Nguồn: Đề xuất của tác giả)

### 2.2. Cấu trúc Bộ tiêu chí và phương pháp nghiên cứu

#### 2.2.1. Cấu trúc Bộ tiêu chí/bảng hỏi

Bộ tiêu chí của TTKTĐBCL gồm: 4 nhân tố/ khía cạnh chính của lớp HP - Nhận xét chung (NXC), Giảng viên (GV), Kiểm tra, đánh giá (KTĐG), Cơ sở vật chất học liệu (CSVCHL). Mỗi nhân tố gồm những biến quan sát thành phần như sau: NXC1-4, GV1-8, KTĐG1-7, CSVCHL1-6. Tổng cộng 25 biến quan sát. Ngoài ra, nhân tố Hải lòng chung - HLC - thuộc loại nhân tố kết quả, gồm 4 biến quan sát HLC1-4 ứng với sự hài lòng với từng nhân tố nêu trên. Bộ tiêu chí sử dụng thang đo Likert 5 bậc gồm: Hoàn toàn không đồng ý (1); Không đồng ý (2); Trung lập (3); Đồng ý (4); Hoàn toàn đồng ý (5). Thiết kế của Bộ tiêu chí cho thấy mục đích chính để đánh giá mức độ hài lòng của người học về các khía cạnh của lớp HP.

Bộ tiêu chí thiết kế riêng cho ESP2 gồm: 6 nhân tố - Nội dung học phần (NDHP), Tổ chức học phần (TCHP), Giáo trình học liệu (GTHL), Phương pháp giảng dạy của GV (PPGD), KTĐG, và Cơ sở vật chất (CVC). Các biến quan sát thành phần thuộc mỗi nhân tố như sau: NDHP1-7, TCHP1-7, GTHL1-11, PPGD1-12, KTĐG1-8, CVC1-3. Tổng cộng 48 biến quan sát. Bộ tiêu chí sử dụng thang đo Likert 5 bậc gồm: Rất không đồng ý (1); Không đồng ý (2); Trung lập (3); Đồng ý (4); Rất đồng ý (5). Thiết kế của Bộ tiêu chí cho thấy mục đích chính lấy ý kiến đánh

giá của người học về các khía cạnh của lớp HP nhằm cải tiến nâng cao chất lượng dạy và học ESP2. Mục tiêu này tương đồng với nghiên cứu của Alturkistani và cộng sự (2020).

### 2.2.2. Dữ liệu định lượng

Dữ liệu sơ cấp thu thập bằng bảng hỏi giấy phát trực tiếp cho SV thuộc 06 lớp HP ESP121 của ESP2 tại Trường Đại học Ngoại thương (giai đoạn 1/Kì 2/năm học 2024-2025) sau khi học xong và kiểm tra giữa kì (Bộ của TTKTĐBCL) và sau khi thi kết thúc HP (Bộ thiết kế cho ESP2). Tỷ lệ SV trả lời bảng hỏi trung bình của 06 lớp là 76,4% (TTKTĐBCL) và 89,58% (thiết kế cho ESP2) ứng với tổng số phiếu thu về 418/547 và 490/547 phiếu phát ra. Tỷ lệ phiếu loại tương ứng là 1,2% và 0,2% sau nhập liệu, rà soát và làm sạch. Tổng số SV hoàn thành ESP2 giai đoạn này là 1.952 SV. Quy mô mẫu ở cả hai bộ tiêu chí cao hơn nhiều mẫu tối ưu (Slovin, 1960) với độ tin cậy 95% (chỉ cần 332 phiếu).

### 2.2.3. Các phân tích định lượng

Trên cơ sở khung đánh giá bộ tiêu chí/bảng hỏi đánh giá lớp HP (hình 1), các phân tích định lượng gồm: (1) Kiểm tra độ tin cậy của thang đo: hệ số Cronbach's Alpha và hệ số tương quan với biến tổng; (2) Kiểm tra tính giá trị của bảng hỏi: thể hiện ở tính đơn hướng và tính hội tụ: Tính đơn hướng cho biết các biến quan sát có xu hướng chỉ phản ánh tính chất của 1 biến tổng - phân tích nhân tố khám phá EFA (Exploratory Factor Analysis); Tính hội tụ cho biết các biến quan sát của 1 biến tổng có tương quan thuận chiều với nhau không và sự tương quan thuận này mạnh tới mức độ nào - sử dụng EFA và hệ số phương sai trung bình được trích - AVE (Average Variance Extracted). Giá trị tham chiếu AVE > 0,5; biến có AVE < 0,5 sẽ bị loại khỏi mô hình; (3) Kiểm định lại độ tin cậy của thang đo sau EFA lần cuối (sau điều chỉnh mô hình và loại biến xấu).

## 2.3. Kết quả phân tích định lượng

### 2.3.1. Kiểm tra độ tin cậy của thang đo

- *Bộ tiêu chí TTKTĐBCL*: Hệ số Cronbach's Alpha của 5 biến tổng/nhân tố đều > 0,7 (giá trị tham chiếu) (NXC 0,841; GV 0,926; KTĐG 0,907; CSVCHL 0,858; HLC 0,886) và lớn hơn hệ số này ở các biến quan sát thành phần (NXC1-4; GV1-8; KTĐG1-7; CSVCHL1-6; HLC1-4). Hệ số tương quan với biến tổng của các biến quan sát thành phần đều thỏa mãn > 0,3 (mức tham chiếu). Hệ số này càng cao thể hiện tính giải thích càng lớn. Như vậy, các thang đo là tin cậy và không cần loại biến quan sát thành phần nào.

- *Bộ tiêu chí Thiết kế cho ESP2*: Hệ số Cronbach's Alpha của 6 biến tổng/nhân tố đều > 0,7 (giá trị tham chiếu) (NDHP 0,945; TCHP 0,940; GTHL 0,945; PPGD 0,961; KTĐG 0,955; CSVC 0,869) và lớn hơn hệ số này ở các biến quan sát thành phần (NDHP1-7; TCHP1-7; GTHL1-11; PPGD1-12; KTĐG1-8; CSVC1-3) ngoại trừ PPGD12 (chênh lệch 0,001 - 0,961 so với 0,962) và CSVC2 (chênh lệch 0,002 - 0,869 so với 0,871). Hệ số tương quan với biến tổng của các biến quan sát thành phần đều thỏa mãn > 0,3 (mức tham chiếu). Như vậy, các thang đo về cơ bản là tin cậy và tạm thời không loại biến quan sát nào tại thời điểm này.

### 2.3.2. Kiểm tra tính giá trị của bảng hỏi

- *Kiểm tra tính đơn hướng*: 02 giả thuyết nghiên cứu được đặt ra trong phân tích EFA: (1) Giả thuyết  $H_0$ : Các biến trong tổng thể không có tương quan với nhau; (2) Giả thuyết  $H_1$ : Các biến trong tổng thể có tương quan với nhau.

Kiểm định Bartlett cho thấy giữa các biến trong tổng thể có mối tương quan với nhau ( $\text{sig.} = 0,000 < 0,05$ ) dẫn tới bác bỏ  $H_0$ , chấp nhận  $H_1$ .

Tác giả sử dụng phương pháp trích PCA (*Principal Component Analysis*) với phép quay varimax tại điểm dừng khi trích các nhân tố có Eigenvalue > 1. Thang đo nào có tổng phương sai trích  $\geq 50\%$  được chấp nhận (Gerbing và Anderson, 1988). Các biến có hệ số tải (*factor loading*) < 0,5 sẽ bị loại. Tại mỗi biến quan sát có chênh lệch hệ số tải lớn nhất và bất kì phải đạt  $\geq 0,3$  (Jabnoun và Al-Tamimi, 2003). Với EFA, điều kiện cần là hệ số KMO có giá trị lớn ( $0,5 \leq \text{KMO} \leq 1$ ) thể hiện phân tích EFA là phù hợp. Theo Hoàng Trọng và Chu Nguyễn Mộng Ngọc (2008), KMO  $\geq 0,9$  rất tốt;  $0,9 \geq \text{KMO} \geq 0,8$  tốt; KMO trong biên độ  $[0,7 - 0,8]$  được; KMO  $\geq 0,6$  tạm được; KMO trong khoảng  $[0,5, 0,6]$  thấp và < 0,5 - không được chấp nhận. Mức tham chiếu cơ bản KMO > 0,5 (Kaiser, 1974).

Dữ liệu từ Bộ tiêu chí TTKTĐBCL có KMO = 0,897 (tốt) và Bộ tiêu chí Thiết kế cho ESP2 có KMO = 0,957 (rất tốt), chứng tỏ phân tích nhân tố để nhóm biến lại với nhau là thích hợp và dữ liệu phù hợp với phân tích EFA.

Phương sai trích (%) sau EFA lần đầu cho thấy các biến tổng/nhân tố trong mô hình giải thích được 67,424% (TTKTĐBCL) và 76,8% (Thiết kế cho ESP2) thay đổi trong đánh giá của người học. Tỷ lệ này sau EFA lần cuối đạt 72,42% (TTKTĐBCL) và 81,05% (Thiết kế cho ESP2).

Hệ số tải nhân tố của các biến quan sát trong mô hình thể hiện trong ma trận xoay nhân tố. Thực hiện loại biến xấu, điều chỉnh mô hình theo 03 nguyên tắc cơ bản sau: (1) loại biến rỗng, (2) loại các biến trong cùng nhân tố nếu hệ số tải  $< 0,5$ ; nếu cùng  $< 0,5$  thì loại biến lạc nhóm trước; loại biến có hệ số tải nhân tố nhỏ hơn trước; (3) loại biến nếu chênh lệch hệ số tải trong cùng một hàng  $< 0,3$ . Với bộ tiêu chí TTKTĐBCL, sau EFA lần 2 đã loại được toàn bộ biến xấu (12/25 biến) và sau 22 lần cho Bộ tiêu chí Thiết kế cho ESP2 (25 biến xấu/48 biến bị loại). Kết quả, Bộ tiêu chí TTKTĐBCL còn lại 13 biến quan sát thuộc 3 nhân tố tác động gồm: GV(2,3,4,5); KTĐG(2,3,4,5); CSVCHL(1,2,4,5,6) và 3 biến thuộc nhân tố kết quả HLC (HLC2,3,4); Nhân tố NXC cùng các biến quan sát thành phần bị loại toàn bộ. Bộ tiêu chí Thiết kế cho ESP2 còn lại 23 biến thuộc 5 nhân tố gồm: NDHP(1,2,3,4,5); GTHL(7,8,9,10,11); PPGD(1,2,3,4,5); KTĐG (1,2,3,7,8); CSVC(1,2,3); Nhân tố TCHP cùng các biến quan sát thành phần bị loại toàn bộ.

Bảng 1 và bảng 2 trình bày hệ số tải nhân tố của hai Bộ tiêu chí sau lần chạy EFA cuối cùng.

*Bảng 1. Hệ số tải nhân tố (sau EFA lần cuối) (TTKTĐBCL)*

	Nhân tố		
	1	2	3
GV4	0,846		
GV5	0,834		
GV2	0,833		
GV3	0,826		
CSVCHL4		0,790	
CSVCHL5		0,760	
CSVCHL2		0,740	
CSVCHL6		0,713	
CSVCHL1		0,706	
KTĐG4			0,848
KTĐG5			0,818
KTĐG2			0,757
KTĐG3			0,754

*Bảng 2. Hệ số tải nhân tố (sau EFA lần cuối) (Thiết kế cho ESP2)*

	Nhân tố				
	1	2	3	4	5
GTHL10	0,837				
GTHL7	0,817				
GTHL8	0,805				
GTHL9	0,792				
GTHL11	0,671				
PPGD2		0,797			
PPGD3		0,782			
PPGD5		0,774			
PPGD4		0,756			
PPGD1		0,736			
NDHP1			0,806		
NDHP5			0,800		
NDHP2			0,769		
NDHP3			0,716		
NDHP4			0,669		
KTĐG2				0,768	
KTĐG1				0,735	
KTĐG3				0,728	
KTĐG8				0,714	
KTĐG7				0,681	
CSV2					0,846
CSV3					0,766
CSV1					0,729



- *Kiểm tra tính hội tụ*: Để kiểm tra tính hội tụ trong các biến tổng/nhân tố thì sau khi chạy EFA loại biến xấu lần cuối cần chạy EFA cho từng nhân tố và tính toán hệ số phương sai trung bình được trích - AVE. Bảng 3 và bảng 4 lần lượt trình bày ma trận AVE của Bộ tiêu chí\_TTKTĐBCL và của Bộ tiêu chí\_Thiết kế cho ESP2.

*Bảng 3. Ma trận AVE (TTKTĐBCL)*

	AVE	GV	KTĐG	CSVCHL
GV	0,81046	0,9003		
KTĐG	0,75558	0,572***	0,8692	
CSVCHL	0,62263	0,488***	0,523***	0,7891

*Chú thích: \*\*\* Mức ý nghĩa 1% (hai phía)*

Bảng 3 cho thấy, cả 3 giá trị trên đường chéo của ma trận AVE đều lớn hơn 0,5 và lớn hơn các hệ số tương quan giữa các biến thành phần ở mức ý nghĩa 1% cho thấy các biến quan sát trong từng nhân tố GV, KTĐG, CSVCHL (Bộ tiêu chí\_TTKTĐBCL) đều có tính hội tụ.

*Bảng 4. Ma trận AVE (Thiết kế cho ESP2)*

	AVE	NDHP	GTHL	PPGD	KTĐG	CSVC
NDHP	0,81279	0,9015				
GTHL	0,76892	0,606***	0,8769			
PPGD	0,81994	0,691***	0,637***	0,9055		
KTĐG	0,78327	0,728***	0,627***	0,722***	0,8850	
CSVC	0,80441	0,623***	0,534***	0,547***	0,610***	0,8969

*Chú thích: \*\*\* Mức ý nghĩa 1% (hai phía)*

Bảng 4 cho thấy, cả 5 giá trị trên đường chéo của ma trận AVE đều lớn hơn 0,5 và lớn hơn các hệ số tương quan giữa các biến thành phần ở mức ý nghĩa 1% cho thấy các biến quan sát trong từng nhân tố NDHP, GTHL, PPGD, KTĐG, và CSVC (Bộ tiêu chí\_Thiết kế cho ESP2) đều có tính hội tụ.

### 2.3.3. Kiểm định lại độ tin cậy của thang đo (sau EFA lần cuối)

- *Bộ tiêu chí\_TTKTĐBCL*: Hệ số Cronbach's Alpha của 4 biến tổng/nhân tố đều > 0,7 (giá trị tham chiếu) (GV 0,922; KTĐG 0,891; CSVCHL 0,833; HLC 0,834) và lớn hơn hệ số này ở các biến quan sát thành phần (GV<sub>2,3,4,5</sub>; KTĐG<sub>2,3,4,5</sub>; CSVCHL<sub>1,2,4,5,6</sub>; HLC<sub>2,3,4</sub>) ngoại trừ chênh lệch rất nhỏ 0,005 ở biến HLC4 (0,834 so với 0,839). Tuy nhiên, biến HLC4 thuộc nhân tố HLC - nhân tố kết quả - vẫn được giữ lại vì có hệ số tải nhân tố đạt yêu cầu (> 0,5). Như vậy, độ tin cậy của thang đo ở Bộ tiêu chí\_TTKTĐBCL sau điều chỉnh mô hình và loại biến xấu vẫn được đảm bảo.

- *Bộ tiêu chí\_Thiết kế cho ESP2*: Hệ số Cronbach's Alpha của 5 biến tổng/nhân tố đều > 0,7 (giá trị tham chiếu) (NDHP 0,940; GTHL 0,924; PPGD 0,945; KTĐG 0,931; CSVC 0,869) và lớn hơn hệ số này ở các biến quan sát thành phần (NDHP<sub>1,2,3,4,5</sub>; GTHL<sub>7,8,9,10,11</sub>; PPGD<sub>1,2,3,4,5</sub>; KTĐG<sub>1,2,3,7,8</sub>; CSVC<sub>1,2,3</sub>) ngoại trừ chênh lệch rất nhỏ (0,002) ở biến CSVC2 (0,869 so với 0,871). Tuy nhiên, biến CSVC2 thuộc nhân tố CSVC vẫn được giữ lại vì có hệ số tải nhân tố đạt yêu cầu (> 0,5). Như vậy, độ tin cậy của thang đo ở Bộ tiêu chí\_Thiết kế cho ESP2 sau điều chỉnh mô hình và loại biến xấu vẫn được đảm bảo.

## 3. Kết luận

Bài báo thực hiện so sánh độ tin cậy của thang đo và tính giá trị của bảng hỏi của hai Bộ tiêu chí/bảng hỏi đánh giá lớp HP - một Bộ của TTKTĐBCL đang sử dụng đại trà cho tất cả các lớp HP - không phân biệt thuộc HP nào và một Bộ được thiết kế riêng cho HP ESP2 ở một nghiên cứu trước đó. Kết quả phân tích định lượng cho thấy sự vượt trội của Bộ tiêu chí\_Thiết kế cho ESP2 (sau điều chỉnh mô hình, loại biến xấu - gồm 23 biến quan sát) ở cả hai tiêu chuẩn nêu trên. Cụ thể, Bộ tiêu chí\_Thiết kế cho ESP2 có: (1) Hệ số Cronbach's Alpha cao hơn, phản ánh độ tin cậy của thang đo cao hơn; (2) Hệ số KMO cao hơn thể hiện bộ tiêu chí phù hợp với phân tích EFA ở mức độ cao hơn; (3) Phương sai trích (%) cao hơn thể hiện tính giải thích lớn hơn của các nhân tố/biến tổng trong thay đổi đánh giá của người học; (4) Sau nhiều lần phân tích EFA điều chỉnh mô hình và loại biến xấu, tỉ lệ % phương sai trích của Bộ tiêu chí\_Thiết kế cho ESP2 tăng lên tới 81,05% (cao hơn so với mức 72,42% của Bộ tiêu chí\_TTKTĐBCL) - thể hiện tính giải thích lớn hơn; (5) Tính giá trị của bảng hỏi phản ánh ở tính đơn hướng và tính hội tụ cũng cao hơn. Bộ tiêu chí\_Thiết kế cho ESP2 (gồm 23 biến quan sát) đảm bảo về độ tin cậy và tính giá trị vừa bám sát được bối cảnh cụ thể của HP đặc biệt về GTHL, vì vậy, được đề xuất tới Nhà trường chính thức áp dụng cho các lớp HP ESP121 thuộc ESP2 trong thời gian tới - thay cho Bộ tiêu chí dùng chung hiện nay.

Nghiên cứu cũng đưa ra một số hàm ý về cách thức đánh giá một bộ tiêu chí/bảng hỏi đánh giá lớp HP sử dụng thang đo Likert nhiều bậc bằng các công cụ định lượng khoa học và khách quan. Tác giả nhấn mạnh sự kết hợp phân tích hệ số Cronbach's Alpha của các nhân tố/biến tổng và của từng biến quan sát thành phần với phân tích EFA, đối

chiếu hệ số AVE để đánh giá tính giá trị của bảng hỏi một cách toàn diện thông qua tính đơn hướng và tính hội tụ. Việc chưa loại ngay biến quan sát thành phần sau bước đánh giá độ tin cậy của thang đo mà chờ đến khi có kết quả phân tích EFA và AVE và kiểm định lại độ tin cậy của thang đo sau các điều chỉnh thực sự có ý nghĩa, không làm mất đi biến quan sát thành phần vừa đảm bảo độ vững của mô hình và kết quả nghiên cứu. Hơn nữa, một bộ tiêu chí nếu được xây dựng để áp dụng chung cho tất cả các lớp HP sẽ khó cá biệt hoá tới từng HP, đặc biệt về GTHL. Vì vậy, các khoa chuyên môn được đề xuất xây dựng các tiêu chí thành phần thuộc nhân tố GTHL của từng HP mà khoa phụ trách và đưa vào đánh giá thực nghiệm về độ tin cậy và tính giá trị để chốt lại các tiêu chí phù hợp cho nhân tố GTHL của từng HP được xem xét. Với cách thức đó, TTKTĐBCL sẽ có được những bộ tiêu chí đánh giá lớp HP được xây dựng riêng cho từng HP khi cần lấy ý kiến đánh giá của người học.

**Lời cảm ơn:** Tác giả cảm ơn sự tài trợ của Trường Đại học Ngoại thương với Đề tài Khoa học và Công nghệ cấp Cơ sở “Đánh giá lớp học phần Tiếng Anh chuyên ngành 2 - Giao tiếp kinh doanh - trong khuôn khổ Chương trình Tiếng Anh áp dụng cho sinh viên từ khoá 60 của Trường Đại học Ngoại thương”, mã số: NTCS 2023-13.

### Tài liệu tham khảo

- Alturkistani, A., Lam, C., Foley, K., Stenfors, T., Blum, R. E., Velthoven, H. V. M., & Meinert, E. (2020). Massive Open Online Course. Evaluation Methods: Systematic Review. *Journal of Medical Internet Research*, 22(4), 1-14. <https://doi.org/10.2196/13851>
- Beckman, T. J., Lee, M. C., Rohren, C. H., & Pankratz, V. S. (2003). Evaluating an instrument for the peer review of impatient teaching. *Medical Teacher*, 25(2), 131-135. <https://doi.org/10.1080/0142159031000092508>
- Bộ GD-ĐT (2021). *Thông tư số 08/2021/TT-BGDĐT ngày 18/3/2021 ban hành Quy chế đào tạo trình độ đại học*.
- Bộ GD-ĐT (2024). *Thông tư số 01/2024/TT-BGDĐT ngày 05/02/2024 ban hành Chuẩn cơ sở giáo dục đại học*.
- Copeland, H. L., & Hewson, M. G. (2000). Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Academic Medicine*, 75(2), 161-166.
- Crossley, J., Humphris, G., & Jolly, B. (2002). Assessing health professionals. *Medical Education*, 36, 800-804.
- Cục Nhà giáo và Cán bộ quản lý giáo dục (2013). *Hướng dẫn số 7324/BGDĐT-NGCBQLGD ngày 08/10/2013 về tổ chức lấy ý kiến phản hồi từ người học về hoạt động giảng dạy của giảng viên*.
- 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.
- Greenwald, A. G. (1997). Validity Concerns and Usefulness of Student Ratings of Instruction. *American Psychologist*, 52(11), 1182-1186.
- Hoàng Trọng, Chu Nguyễn Mộng Ngọc (2008). *Phân tích dữ liệu nghiên cứu với SPSS*. NXB Hồng Đức.
- Jabnoun, N., & Hassan Al-Tamimi, H. A. (2003). Measuring Perceived Service Quality at UAE Commercial Banks. *International Journal of Quality and Reliability Management*, 47-55. <https://doi.org/10.1108/02656710310468614>
- Kember, D., & Leung, D. Y. P. (2008). Establishing the validity and reliability of course evaluation questionnaires. *Assessment & Evaluation in Higher Education*, 33(4), 341-353. <https://doi.org/10.1080/02602930701563070>
- Kogan, J. R., & Shea, J. A. (2007). Course evaluation in medical education. *Teaching and Teacher Education*, 23(3), 251-264. <https://doi.org/10.1016/j.tate.2006.12.020>
- Litzelman, D. K., Shea, J. A., Kogan, J. R., & Wales, P. S. (2005). Evaluation of the Clerkship: Clinical Teachers and Programs. In *Alliance for Clinical Education (Chapter 7)* (3rd ed.). <https://familymed.uthscsa.edu/ACE/chapter7.htm>
- Papadogiannis, I., Vassilakis, C., Wallace, M., & Katsis, A. (2024). On the Quality and Validity of Course Evaluation Questionnaires Used in Tertiary Education in Greece. *Trends in Higher Education*, 3(2), 221-234. <https://doi.org/10.3390/higheredu3020013>
- Schiekirk, S., & Raupach, T. (2015). A systematic review of factors influencing student ratings in undergraduate medical education course evaluations. *BMC Medical Education*, 15(30), 1-9. <https://doi.org/10.1186/s12909-015-0311-8>
- Slovin, E. (1960). *Slovin's formula for sampling technique*. <https://prudencexd.weebly.com/>
- Steiner, I. P., Yoon, P. W., Kelly, K. D., Diner, B. M., Donoff, M. G., & Mackey, D. S. (2003). Resident evaluation of clinical teachers based on teachers' certification. *Academic Emergency Medicine*, 10(7), 731-737.
- Trường Đại học Ngoại thương (2021). *Quyết định số 3188/QĐ-ĐHNT ngày 21/12/2021 ban hành Quy chế đào tạo trình độ Đại học của Trường Đại học Ngoại thương*.