VIETNAM NATIONAL UNIVERSITY, HANOI

**UNIVERSITY OF LANGUAGES AND INTERNATIONAL STUDIES**

**NGUYỄN THỊ TỐ LOAN**

**THE DEVELOPMENT AND ARGUMENT-BASED VALIDATION OF**

**A SCALE FOR MEASURING VIETNAMESE EFL LECTURERS’ LANGUAGE ASSESSMENT LITERACY**

(Thiết kế và xác trị thang đo năng lực đánh giá ngôn ngữ của giảng viên

dạy tiếng Anh ở Việt Nam theo hướng dựa trên các lập luận)

Chuyên ngành: Lý luận và phương pháp dạy học Bộ môn tiếng Anh

Mã: 9140231.01

**TÓM TẮT LUẬN ÁN TIẾN SĨ**

**HÀ NỘI - 2024**

Công trình được hoàn thành tại: Trường Đại học Ngoại ngữ,

Đại học Quốc gia Hà Nội

Người hướng dẫn khoa học: **1. PGS. TS. Nguyễn Văn Trào**

**2. TS. Dương Thu Mai**

Phản biện 1: …………………………………………

Phản biện 2: …………………………………………

Phản biện 3: …………………………………………

Luận án sẽ được bảo vệ trước Hội đồng chấm luận án tiến sĩ họp tại:

Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội

vào hồi giờ ngày tháng năm 2024

**Có thể tìm hiểu luận án tại:**

- Thư viện Quốc gia Việt Nam.

- Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội.

# CHƯƠNG 1. GIỚI THIỆU

Chương giới thiệu nêu lý do thực hiện nghiên cứu, ý nghĩa của nghiên cứu và hoàn cảnh nghiên cứu ở Việt Nam có liên quan tới luận án. Chương này cũng đặt ra câu hỏi nghiên cứu, giới hạn phạm vị nghiên cứu và giới thiệu bố cục của luận án.

## 1.1. Lý do thực hiện nghiên cứu

## Trong những thập kỷ gần đây, các cải cách giáo dục toàn cầu đã nhấn mạnh vào các đánh giá thường xuyên để nâng cao việc học của học sinh (Inbar-Lourie, 2008; Davison & Leung, 2009). Sự thay đổi này đã mở rộng vai trò của giáo viên, đòi hỏi họ phải tích hợp các phương pháp đánh giá khác nhau vào việc giảng dạy, thu hút học sinh tích cực, sử dụng dữ liệu đánh giá để cung cấp thông tin cho việc giảng dạy và liên kết các báo cáo với các tiêu chuẩn bên ngoài (Inbar-Lourie, 2008; Rea-Dickins, 2008). Do đó, các đánh giá hiện chiếm một phần đáng kể thời gian của giáo viên, khoảng một phần tư đến một phần ba (Cheng và cộng sự, 2004).

## Kiến thức đánh giá (AL), được định nghĩa là khả năng tiến hành các đánh giá hợp lệ và đáng tin cậy, ngày càng được công nhận là cần thiết cho việc giảng dạy hiệu quả. Năng lực đánh giá mạnh cho phép giáo viên liên kết các đánh giá với các mục tiêu hướng dẫn, trong khi năng lực đánh giá không đầy đủ có nguy cơ đánh giá không đáng tin cậy, có khả năng gây hiểu lầm cho các bên liên quan như học sinh, phụ huynh và quản trị viên (DeLuca & Klinger, 2010; Xu, 2019b). Năng lực đánh giá cũng tăng cường hiệu quả giảng dạy bằng cách cho phép giáo viên diễn giải và sử dụng dữ liệu đánh giá để đáp ứng các nhu cầu đa dạng của học sinh (Pastore & Andrade, 2019).

## Trong giáo dục ngôn ngữ, năng lực đánh giá ngôn ngữ (LAL) đặc biệt quan trọng đối với giáo viên giảng dạy tiếng Anh, những người vừa là người hướng dẫn vừa là người đánh giá. năng lực kiểm tra đánh giá ngôn ngữ bao gồm kiến ​​thức và kỹ năng cụ thể cần thiết để đánh giá người học ngôn ngữ một cách hiệu quả (Inbar-Lourie, 2013; Vogt và cộng sự, 2020). Năng lực đánh giá ngôn ngữ cao tương quan với việc cải thiện hướng dẫn ngôn ngữ, vì giáo viên có thể đưa ra quyết định sáng suốt để hỗ trợ sự tiến bộ của học sinh (Harding & Kremmel, 2016; Hakim, 2015).

## Mặc dù có tầm quan trọng, các công cụ đáng tin cậy để đo lường năng lực đánh giá ngôn ngữ, đặc biệt là đối với bối cảnh giảng dạy tiếng Anh, vẫn còn khan hiếm. Nhiều giáo viên ngôn ngữ có năng lực đánh giá hạn chế, dẫn đến sự không phù hợp giữa đánh giá và mục tiêu học tập (Inbar-Lourie, 2008; Xu & Brown, 2016). Việc không tập trung đầy đủ vào đào tạo đánh giá trong đào tạo giáo viên càng làm trầm trọng thêm vấn đề này (Stabler-Havener, 2018).

## Trong giảng dạy tiếng Anh ở bậc đại học của Việt Nam, khoảng cách này rất rõ rệt, với ít nghiên cứu về năng lực đánh giá ngôn ngữ trong số các giảng viên giảng dạy tiếng Anh của Việt Nam. Sự mở rộng nhanh chóng của giáo dục tiếng Anh, được thúc đẩy bởi các cải cách kinh tế xã hội (Ngo, 2018, 2021), làm tăng nhu cầu về các giám định viên ngôn ngữ có năng lực. Mặc dù có nhiều dự án nâng cao năng lực chuyên môn ​​như Đề án Ngoại ngữ Quốc gia 2008 - 2020, các khuôn khổ hoặc công cụ đánh giá ngôn ngữ toàn diện dành cho giáo viên bậc đại học vẫn còn thiếu.

## Nghiên cứu này giải quyết những khoảng cách này bằng cách phát triển và xác nhận thang đo năng lực đánh giá ngôn ngữ phù hợp cho giáo dục đại học Việt Nam, nhằm mục đích nâng cao các chương trình đào tạo giáo viên và cải thiện chất lượng giảng dạy tiếng Anh tại Việt Nam.

## 1.2. Vấn đề nghiên cứu

## Lĩnh vực đánh giá ngôn ngữ đã thu hút sự chú ý đáng kể kể từ khi Brindley (2001) đề xuất tích hợp năng lực đánh giá vào các hoạt động đánh giá ngôn ngữ. Mặc dù có nhiều nghiên cứu đã được thực hiện, nhiều vấn đề về năng lực đánh giá ngôn ngữ vẫn chưa được giải quyết triệt để (Giraldo, 2018). Các nghiên cứu lý thuyết và thực nghiệm trong lĩnh vực này còn hạn chế (Fulcher, 2012; Giraldo, 2018; Hakim, 2015), cho thấy năng lực đánh giá ngôn ngữ vẫn đang trong quá trình phát triển và cần được khám phá thêm (Yan & Fan, 2020). Dù Harding và Brunfaut (2020) đã ghi nhận sự tiến bộ trong nghiên cứu về năng lực này, song vẫn chưa có sự đồng thuận về định nghĩa của nó, dẫn đến các nghiên cứu và thảo luận đa dạng về bản chất và phạm vi.

## Việc đánh giá kỹ năng ngôn ngữ đòi hỏi năng lực đánh giá chuyên biệt, phù hợp với các bối cảnh giáo dục và xã hội cụ thể, nhằm đảm bảo độ tin cậy và tính hợp lệ trong quy trình đánh giá. Một khoảng trống đáng kể trong tài liệu là thiếu các công cụ đo lường được xác thực cho năng lực đánh giá ngôn ngữ, nhất là khi so sánh với lĩnh vực giáo dục phổ thông vốn đã có nhiều công cụ sẵn có. Các nghiên cứu cho thấy năng lực đánh giá ngôn ngữ là một khái niệm đa diện, chịu ảnh hưởng mạnh mẽ bởi bối cảnh, với các khía cạnh chưa được xác định rõ ràng (Coombe và cộng sự, 2020; Giraldo, 2021). Điều này nhấn mạnh nhu cầu cấp thiết trong việc xác định rõ và phát triển các công cụ đo lường năng lực đánh giá ngôn ngữ cho giáo viên tiếng Anh.

## Trong bối cảnh giáo dục đại học tại Việt Nam, nhu cầu nâng cao trình độ tiếng Anh của sinh viên sau khi tốt nghiệp đã làm gia tăng yêu cầu đánh giá ngôn ngữ chính xác và hiệu quả (Bodewig & Badiani-Magnusson, 2014; Tran, 2019). Tuy nhiên, vẫn còn thiếu các khuôn khổ toàn diện xác định các năng lực cần thiết cho giáo viên tiếng Anh trong lĩnh vực này (Nguyen, 2017; Pham, 2017). Dù nhiều cải cách, như Đề án ngoại ngữ quốc gia 2020, đã được triển khai nhằm cải thiện giảng dạy và đánh giá tiếng Anh, hoạt động đánh giá ở bậc đại học vẫn chủ yếu mang tính tổng kết và hướng đến kiểm tra (Nguyen, 2013; Ngo, 2018). Sự hạn chế về năng lực đánh giá của giáo viên, cùng các rào cản vật chất và thiếu sự chủ động, đã cản trở việc áp dụng các phương pháp đánh giá sáng tạo (Phung & Tran, 2017; Vu, 2017). Đồng thời, hiện nay chưa có các công cụ hợp lệ, nhạy cảm với ngữ cảnh để đo lường năng lực đánh giá ngôn ngữ của giáo viên tiếng Anh tại Việt Nam, đặc biệt trong giáo dục đại học (Ngo, 2021; Giraldo, 2018).

## Mặc dù đã có nỗ lực khái niệm hóa năng lực đánh giá ngôn ngữ từ góc nhìn của cả nhà nghiên cứu và giáo viên (Butler và cộng sự, 2021), trọng tâm hiện vẫn chủ yếu hướng đến các giáo viên đang công tác với kinh nghiệm và chuyên môn đa dạng. Nghiên cứu cách giáo viên tiếng Anh trong môi trường đại học định nghĩa năng lực đánh giá ngôn ngữ là vấn đề đặc biệt quan trọng. Hơn nữa, năng lực này thường được thể hiện qua các hành vi có thể quan sát và đo lường, phản ánh khả năng kiểm tra và đánh giá ngôn ngữ của giáo viên (Griffin, 2014, 2018; Masters, 2003).

## Với những thiếu hụt về công cụ đo lường hợp lệ và lập luận về tính hợp lệ, đặc biệt ở Việt Nam, nghiên cứu này đặt mục tiêu phát triển và xác thực một thang đo đánh giá năng lực kiểm tra, đánh giá ngôn ngữ (LALS) dành riêng cho giảng viên tiếng Anh tại Việt Nam. Nghiên cứu này kỳ vọng đáp ứng nhu cầu về công cụ đo lường năng lực đánh giá ngôn ngữ đáng tin cậy, đồng thời cung cấp nền tảng cho các nghiên cứu tương lai về sự phát triển năng lực này trong bối cảnh giáo dục Việt Nam và các quốc gia có điều kiện tương tự.

**1.3. Mục tiêu của nghiên cứu**

Nghiên cứu này nhằm đáp ứng những khoảng trống trong tài liệu hiện có về Năng lực đánh giá ngôn ngữ bằng cách phát triển và xác thực một mô hình toàn diện, nắm bắt các năng lực cốt lõi cần thiết cho giảng viên tiếng Anh như ngôn ngữ thứ hai tại Việt Nam trong lĩnh vực đánh giá ngôn ngữ. Kết quả nghiên cứu sẽ góp phần cải thiện chương trình đào tạo giáo viên và phát triển chuyên môn tại Việt Nam thông qua việc cung cấp những thông tin chi tiết về thực trạng kiểm tra, đánh giá của giảng viên.

Nghiên cứu có hai mục tiêu chính:

+ Phát triển thang đo để đánh giá toàn diện năng lực kiểm tra đánh giá ngôn ngữ của giảng viên tiếng Anh.

Mục tiêu đầu tiên là xây dựng và hiệu chỉnh một khung lý thuyết đại diện cho các năng lực kiểm tra đánh giá ngôn ngữ thiết yếu, dựa trên các nghiên cứu trước đây và dữ liệu thực tế từ giảng viên tiếng Anh tại Việt Nam trong bối cảnh giáo dục đại học. Công cụ này sẽ tập trung vào các kiến thức, kỹ năng và năng lực cần thiết để thực hiện đánh giá ngôn ngữ một cách hiệu quả.

+ Xác thực thang đo theo phương pháp luận dựa trên lập luận.

Mục tiêu thứ hai là xác thực các kết quả của thang đo bằng cách áp dụng khung xác thực dựa trên lập luận của Kane (2006), được Chapelle và cộng sự (2008) mở rộng. Quá trình xác thực sẽ thu thập bằng chứng định tính và định lượng để đánh giá bốn suy luận về tính hợp lệ, bao gồm suy luận miền, suy luận đánh giá, suy luận tổng quát và suy luận giải thích. Những suy luận này tạo nền tảng toàn diện để đánh giá độ tin cậy và tính chính xác của LALS trong đo lường năng lực đánh giá ngôn ngữ của giảng viên tiếng Anh tại Việt Nam. Những phát hiện từ nghiên cứu sẽ góp phần cung cấp cơ sở cho việc thiết kế chương trình giảng dạy và các hoạt động phát triển chuyên môn, từ đó nâng cao năng lực đánh giá ngôn ngữ trong giáo dục đại học tại Việt Nam.

**1.4. Câu hỏi nghiên cứu**

Phù hợp với mục tiêu chính của nghiên cứu - phát triển và xác thực thang đo năng lực đánh giá ngôn ngữ dành cho giảng viên tiếng Anh theo phương pháp xác trị dựa trên lập luận - các câu hỏi nghiên cứu sau sẽ định hướng cho nghiên cứu:

Câu hỏi nghiên cứu 1: Tiêu chí nào trong thang đo năng lực đánh giá ngôn ngữ (LAL) của giảng viên tiếng Anh ở Việt Nam?

Câu hỏi nghiên cứu 2: Thang đo năng lực kiểm tra đánh giá ngôn ngữ là công cụ có giá trị và đáng tin cậy để đánh giá năng lực kiểm tra đánh giá ngôn ngữ của giảng viên tiếng Anh Việt Nam ở mức độ nào?

Câu hỏi này chia ra thành 4 câu hỏi nhỏ:

Câu hỏi nghiên cứu 1: Các mục trong thang đo năng lực kiểm tra đánh giá ngôn ngữ phản ánh chính xác đến mức nào về kiến thức, kỹ năng và năng lực cần thiết của giảng viên tiếng Anh, dựa trên phản hồi từ chuyên gia và mức độ phù hợp của nội dung? (Suy luận mô tả lĩnh vực).

Câu hỏi nghiên cứu 2: Các thành phần của thang đo có đáp ứng các tiêu chí trắc nghiệm tâm lý như độ khó, khả năng phân biệt và tính nhất quán nội bộ để đảm bảo độ tin cậy và tính hợp lệ trong đánh giá không? (Suy luận đánh giá).

Câu hỏi nghiên cứu 3: Điểm số của thang đo có thể được khái quát hóa đến mức nào cho các nhóm giảng viên tiếng Anh tại Việt Nam với các đặc điểm nhân khẩu học và trình độ kinh nghiệm khác nhau? (Suy luận khái quát).

Câu hỏi nghiên cứu 4: Mức độ mà thang đo phản ánh cấu trúc cơ bản của năng lực kiểm tra đánh giá ngôn ngữ cũng như khả năng của các mục trong thang đo này trong việc phân biệt năng lực của giảng viên ở các cấp độ khác nhau được đánh giá ra sao? (Suy luận giải thích).

**1.5. Các cân nhắc về phương pháp luận**

Nghiên cứu này áp dụng thiết kế phương pháp hỗn hợp tuần tự thăm dò (Creswell & Plano Clark, 2018) để phát triển và xác thực thang đo năng lực kiểm tra đánh giá dành cho giảng viên tiếng Anh tại Việt Nam. Quá trình nghiên cứu bắt đầu với việc thu thập dữ liệu định tính qua các cuộc phỏng vấn bán cấu trúc và phản hồi từ chuyên gia, tạo nền tảng cho giai đoạn nghiên cứu định lượng tiếp theo. Thiết kế này cho phép đi sâu khám phá các thành phần của năng lực kiểm tra đánh giá ngôn ngữ trước khi chuyển sang giai đoạn đo lường, với việc tích hợp dữ liệu theo khung xác thực dựa trên lập luận.

Giai đoạn định tính tập trung vào việc thu thập ý kiến từ giảng viên tiếng Anh về tính rõ ràng, mức độ liên quan và tính khả thi của các mục trong thang đo năng lực kiểm tra đánh giá ngôn ngữ. Đồng thời, phản hồi từ các chuyên gia cũng được thu thập để đánh giá mức độ phù hợp của các mục trong việc phản ánh năng lực của giảng viên. Sau đó, giai đoạn định lượng được triển khai, bao gồm việc đánh giá phản hồi của giảng viên đối với thang đo đã thử nghiệm, nhằm cung cấp bằng chứng về tính hợp lệ trắc nghiệm tâm lý của công cụ.

Quá trình phát triển thang đo năng lực kiểm tra đánh giá ngôn ngữ gồm nhiều bước, bắt đầu từ việc xem xét tài liệu để tạo bản thảo thang đo. Bản thảo này được xem xét qua nhóm tập trung để thu thập phản hồi định tính về tính phù hợp và độ chính xác, từ đó điều chỉnh thang đo theo ý kiến của người tham gia. Tiếp theo, thang đo trải qua hai vòng thảo luận với chuyên gia: vòng đầu tiên tập trung vào tính rõ ràng, mức độ liên quan và tính đại diện; vòng thứ hai nhấn mạnh vào sự sắp xếp phát triển và tính logic của các mục.

Trong giai đoạn định lượng cuối cùng, thang đo đã được điều chỉnh và thử nghiệm với các giảng viên tiếng Anh, qua đó phản hồi của họ được phân tích để xác định tính hợp lệ trắc nghiệm tâm lý. Một số phát hiện từ giai đoạn định tính đã được chuyển thành các biến định lượng, giúp thực hiện phân tích toàn diện kết hợp dữ liệu từ cả hai giai đoạn. Phương pháp này đảm bảo các phát hiện định tính ban đầu góp phần quan trọng cho đánh giá định lượng, phù hợp với các nguyên tắc xác thực cấu trúc (ví dụ: Kane, 2006; Messick, 1995).

Cách tiếp cận hệ thống này kết hợp các nguyên tắc lý thuyết và trắc nghiệm tâm lý, tạo lập khung tham chiếu tiêu chí phân chia năng lực kiểm tra đánh giá ngôn ngữ thành các năng lực cốt lõi, hành vi quan sát được và tiêu chí đánh giá. Thông qua quy trình phương pháp luận chặt chẽ, đa giai đoạn này, thang đo năng lực kiểm tra đánh giá ngôn ngữ trở thành một công cụ đáng tin cậy, phù hợp với bối cảnh để đánh giá năng lực năng lực kiểm tra đánh giá ngôn ngữ của giảng viên tiếng Anh tại Việt Nam, hỗ trợ việc nâng cao năng lực đánh giá trong hệ thống giáo dục đại học của Việt Nam.

**1.6. Ý nghĩa của nghiên cứu**

Nghiên cứu này có những đóng góp quan trọng về mặt lý thuyết, phương pháp luận và thực tiễn cho lĩnh vực nghiên cứu năng lực kiểm tra đánh giá ngôn ngữ, đặc biệt là trong bối cảnh giảng viên giảng dạy tiếng Anh Việt Nam.

Về mặt lý thuyết, nghiên cứu này giải quyết một khoảng trống nghiên cứu quan trọng bằng cách làm rõ khái niệm hóa năng lực kiểm tra đánh giá ngôn ngữ theo quan điểm của giáo viên giảng dạy tiếng Anh bậc đại học đang công tác, một nhóm thường bị bỏ qua trong các tài liệu hiện có. Các nghiên cứu trước đây chủ yếu tập trung vào giáo viên sắp ra trường (Mertler & Campbell, 2005) và chủ yếu được tiến hành trong bối cảnh phương Tây. Bằng cách tập trung vào các giảng viên giảng dạy tiếng Anh đại học đang công tác tại Việt Nam, nghiên cứu này mở rộng phạm vi nghiên cứu năng lực kiểm tra đánh giá ngôn ngữ, cung cấp sự hiểu biết toàn diện hơn về các thành phần của nó. Điều này góp phần vào khối lượng kiến ​​thức năng lực kiểm tra đánh giá ngôn ngữ rộng hơn bằng cách làm sáng tỏ cách năng lực kiểm tra đánh giá ngôn ngữ được khái niệm hóa và vận hành trong bối cảnh giáo dục đại học không phải phương Tây. Những hiểu biết sâu sắc từ nghiên cứu này sẽ cung cấp một khuôn khổ có giá trị cho các nghiên cứu trong tương lai và giúp thúc đẩy các nỗ lực cải thiện giáo dục ngôn ngữ và các hoạt động đánh giá.

Về mặt phương pháp luận, nghiên cứu này áp dụng cả phương pháp tiếp cận phát triển để phát triển thang đo và khuôn khổ xác thực dựa trên lập luận của Kane (Kane, 2004, 2013). Mặc dù các khuôn khổ này đã được sử dụng rộng rãi trong kiểm tra ngôn ngữ để đảm bảo thiết kế và chất lượng bài kiểm tra mạnh mẽ, nhưng ứng dụng của chúng trong nghiên cứu năng lực kiểm tra đánh giá ngôn ngữ vẫn còn hạn chế. Việc sử dụng các khuôn khổ này trong nghiên cứu này góp phần vào tính nghiêm ngặt về phương pháp luận của nghiên cứu năng lực kiểm tra đánh giá ngôn ngữ, đưa ra phương pháp tiếp cận có cấu trúc để phát triển và xác thực thang đo theo ngữ cảnh cụ thể để đánh giá năng lực kiểm tra đánh giá ngôn ngữ.

Về mặt thực tế, nghiên cứu này cung cấp một thang đo năng lực kiểm tra đánh giá ngôn ngữ đã được xác thực, sẵn sàng sử dụng có thể được sử dụng cho mục đích sư phạm và nghiên cứu. Giảng viên đại học giảng dạy tiếng Anh Việt Nam, cũng như giảng viên trong các bối cảnh tương tự, có thể sử dụng thang đo để tự đánh giá hoặc phát triển chuyên môn, xác định những lỗ hổng trong kiến ​​thức đánh giá ngôn ngữ của họ và đặt ra các mục tiêu phù hợp. Những phát hiện từ nghiên cứu này có thể cung cấp thông tin cho việc thiết kế cả chương trình đào tạo giáo viên trước khi vào nghề và các sáng kiến ​​phát triển chuyên môn trong khi đang làm việc, giải quyết các nhu cầu năng lực kiểm tra đánh giá ngôn ngữ cụ thể của giảng viên. Hơn nữa, nghiên cứu này cung cấp những hiểu biết có giá trị về trình độ đánh giá năng lực kiểm tra đánh giá ngôn ngữ hiện tại trong số các giảng viên giảng dạy tiếng Anh của Việt Nam, có thể cung cấp thông tin cho các quyết định chính sách và đóng góp vào sự phát triển rộng rãi hơn của năng lực kiểm tra đánh giá ngôn ngữ trong khu vực.

**1.7. Cấu trúc của nghiên cứu**

Luận án có 09 chương, bắt đầu với **Chương 1** giới thiệu về chủ đề nghiên cứu. **Chương 2** mô tả cấu trúc của năng lực đánh giá ngôn ngữ, **Chương 3** tập trung vào việc phát triển thang đo đánh giá. **Chương 4** trình bày các phương pháp xác trị thang đo, tiếp theo là **Chương 5** mô trả các phương pháp nghiên cứu. Từ **Chương 6** đến **Chương 8** lần lượt trình bày các kết quả xác thực thang đo kiến thức đánh giá ngôn ngữ của giảng viên tiếng Anh tại Việt Nam qua các suy luận mô tả miền, đánh giá và tổng quát hóa, cũng như suy luận giải thích. Cuối cùng, **Chương 9** đưa ra kết luận và kiến nghị cho nghiên cứu.

# CHƯƠNG 2. CẤU TRÚC CỦA NĂNG LỰC ĐÁNH GIÁ NGÔN NGỮ

# Chương này giới thiệu về cấu trúc năng lực đánh giá ngôn ngữ trong giáo dục, tập trung vào cách hiểu và phát triển năng lực này theo các nền tảng lý thuyết và thực tiễn. Qua việc phân tích các quan điểm và cách tiếp cận khác nhau, chương làm rõ những thành phần chính của năng lực đánh giá ngôn ngữ, từ đó đề xuất các mô hình phát triển và các công cụ đánh giá phù hợp, đặc biệt trong bối cảnh giáo dục đa ngôn ngữ như Việt Nam.

# 2.1. Hiểu về đánh giá năng lực trong giáo dục

# Theo truyền thống, năng lực đọc viết đề cập đến các kỹ năng cơ bản về đọc, viết và tính toán. Theo thời gian, năng lực đọc viết đã mở rộng để bao gồm các năng lực như tư duy phản biện, giải quyết vấn đề và thành thạo kỹ thuật số (Taylor, 2013). Trong giáo dục, năng lực đọc viết (AL), theo định nghĩa của Stiggins (1991), trang bị cho các nhà giáo dục khả năng hiểu, thiết kế và triển khai các đánh giá nhằm nâng cao kết quả học tập. Có hai quan điểm thống trị về năng lực đọc viết: quan điểm phân đôi, dán nhãn cá nhân là biết chữ hoặc mù chữ, và quan điểm tiếp cận liên tục, coi năng lực đọc viết là một loạt các năng lực có liên quan đến bối cảnh văn hóa hoặc giáo dục cụ thể.

# Năng lực đánh giá bao gồm đánh giá, đo lường, kiểm tra và đánh giá—mỗi mục đích có mục đích riêng biệt trong bối cảnh giáo dục (Brown & Abeywickrama, 2010). Đánh giá là một khái niệm rộng, bao gồm các quy trình để theo dõi và cải thiện việc học, trong khi đo lường gán các giá trị định lượng cho kết quả học tập. Kiểm tra đề cập đến các công cụ cụ thể, như kỳ thi, đánh giá năng lực của học sinh và đánh giá diễn giải những kết quả này để tinh chỉnh các hoạt động hướng dẫn. Cùng nhau, các thành phần này cho phép các nhà giáo dục áp dụng hiệu quả các kỹ năng của họ trong nhiều bối cảnh.

# 2.2. Các giả thuyết về bản thể luận và nhận thức luận trong việc định nghĩa năng lực kiểm tra đánh giá ngôn ngữ

# Năng lực kiểm tra đánh giá ngôn ngữ, một tập hợp con chuyên biệt của năng lực kiểm tra đánh giá, tích hợp các kỹ năng, kỹ thuật, ngữ cảnh và đạo đức trong đánh giá ngôn ngữ. Ba quan điểm lý thuyết chính góp phần vào sự hiểu biết của chúng ta về năng lực kiểm tra đánh giá ngôn ngữ:

# *2.2.1. Lý thuyết xã hội văn hóa và năng lực ngữ cảnh*

# Dựa trên lý thuyết xã hội văn hóa của Vygotsky, năng lực được đồng xây dựng thông qua các tương tác xã hội trong các bối cảnh cụ thể. Đối với đánh giá ngôn ngữ, lý thuyết xã hội văn hóa cho rằng các nhà giáo dục phát triển năng lực kiểm tra đánh giá ngôn ngữ bằng cách tham gia vào các cộng đồng chuyên nghiệp và học hỏi từ các đồng nghiệp. Vùng phát triển gần nhất (ZPD) của Vygotsky nhấn mạnh rằng năng lực kiểm tra đánh giá ngôn ngữ phát triển thông qua các tương tác có hướng dẫn, trong đó các nhà giáo dục mở rộng năng lực của mình thông qua hỗ trợ và cộng tác. Khung này nhấn mạnh bản chất nhạy cảm với ngữ cảnh của năng lực kiểm tra đánh giá ngôn ngữ, đặc biệt là trong các bối cảnh đa dạng, nơi các đánh giá cần được điều chỉnh theo nền tảng ngôn ngữ và văn hóa của học sinh.

# *2.2.2. Lý thuyết kiến ​​tạo và học tập theo kinh nghiệm*

# Lý thuyết kiến ​​tạo, được Piaget (1977) và Dewey (1938) ủng hộ, coi việc xây dựng kiến ​​thức là một quá trình chủ động, theo kinh nghiệm. Các nguyên tắc kiến ​​tạo cho rằng các nhà giáo dục xây dựng năng lực kiểm tra đánh giá ngôn ngữ thông qua thực hành phản ánh, liên tục tinh chỉnh các phương pháp đánh giá của họ dựa trên kinh nghiệm thực tế trong lớp học. Phương pháp phản ánh này nhấn mạnh rằng năng lực kiểm tra đánh giá ngôn ngữ linh hoạt và dễ thích ứng, cho phép các nhà giáo dục điều chỉnh các hoạt động của mình khi họ hiểu rõ hơn về nhu cầu và kết quả học tập của học sinh.

# *2.2.3. Lý thuyết trắc nghiệm tâm lý và khả năng đo lường năng lực*

# Lý thuyết trắc nghiệm tâm lý cung cấp một khuôn khổ có cấu trúc, có thể đo lường cho năng lực kiểm tra đánh giá ngôn ngữ bằng cách áp dụng các mô hình như Rasch và Lý thuyết phản hồi mục (IRT). Các mô hình này cho phép các nhà giáo dục đánh giá năng lực của họ theo từng bước, định lượng năng lực kiểm tra đánh giá ngôn ngữ khi nó phát triển. Ví dụ, mô hình Rasch liên kết các cấp độ kỹ năng của cá nhân với mức độ phức tạp của nhiệm vụ, cho phép các tổ chức theo dõi sự phát triển của các nhà giáo dục một cách có hệ thống. Phương pháp tiếp cận này hỗ trợ năng lực kiểm tra đánh giá ngôn ngữ như một năng lực có thể đo lường được dựa trên bằng chứng, đảm bảo độ tin cậy và tính nhất quán trong các đánh giá.

# Các lý thuyết này cùng nhau định hình năng lực kiểm tra đánh giá ngôn ngữ như một năng lực phát triển, nhạy cảm với bối cảnh, kết hợp các kỹ năng kỹ thuật với nhận thức xã hội - văn hóa, có thể thích ứng với nhiều môi trường giáo dục khác nhau.

# 2.3. Các cách tiếp cận để xác định cấu trúc của năng lực kiểm tra đánh giá ngôn ngữ

# Một số khuôn khổ thông báo cấu trúc của năng lực kiểm tra đánh giá ngôn ngữ, nêu bật nó vừa là một bộ kỹ năng kỹ thuật vừa là một năng lực đang phát triển:

# *2.3.1. Cách tiếp cận theo danh sách năng lực*

# Các mô hình theo danh sách năng lực, như mô hình của AFT, NCME & NEA (1990), định nghĩa năng lực kiểm tra đánh giá ngôn ngữ thông qua các năng lực riêng biệt cần thiết cho đánh giá hiệu quả. Chúng bao gồm các kỹ năng như lựa chọn phương pháp đánh giá và nhận ra các cân nhắc về mặt đạo đức, nhưng chúng đã bị chỉ trích vì tập trung quá nhiều vào đánh giá tổng kết. Brookhart (2011) sau đó đã điều chỉnh mô hình để bao gồm đánh giá hình thành, nhận ra nhu cầu về một bộ kỹ năng có thể thích ứng, tập trung vào giảng dạy trong năng lực kiểm tra đánh giá ngôn ngữ.

# *2.3.2. Cách tiếp cận ba thành phần*

# Các mô hình ba thành phần, chẳng hạn như khuôn khổ Kiến thức, Kỹ năng và Nguyên tắc của Davies (2008), mở rộng năng lực kiểm tra đánh giá ngôn ngữ vượt ra ngoài các năng lực kỹ thuật để bao gồm các chiều hướng đạo đức và bối cảnh. Fulcher (2012) cũng đề xuất một cấu trúc phân cấp, cho rằng năng lực kiểm tra đánh giá ngôn ngữ tích hợp các kỹ năng cơ bản, các nguyên tắc chỉ đạo và nhận thức về tác động giáo dục rộng hơn. Mô hình này nhấn mạnh rằng đánh giá ngôn ngữ hiệu quả phải xem xét không chỉ các kỹ năng kỹ thuật mà còn cả nhận thức về đạo đức và ngữ cảnh.

# *2.3.3. Phương pháp tiếp cận theo thang bậc*

# Các mô hình theo thang bậc khái niệm hóa năng lực kiểm tra đánh giá ngôn ngữ như một chuỗi liên tục, với các nhà giáo dục tiến triển qua các giai đoạn phát triển từ trình độ cơ bản đến trình độ nâng cao. "Thang bậc thang trình độ biết chữ" của Pill và Harding (2013) cung cấp một sự tiến triển có cấu trúc, phân loại các kỹ năng theo các cấp độ biết chữ. Khung của Taylor (2013) bổ sung các chiều cho các bên liên quan có yêu cầu về trình độ biết chữ khác nhau, nhấn mạnh rằng các vai trò khác nhau trong giáo dục đòi hỏi các năng lực năng lực kiểm tra đánh giá ngôn ngữ khác nhau.

# *2.3.4. Phương pháp tiếp cận “đàm phán”*

# Phương pháp tiếp cận đàm phán, được minh họa bằng mô hình của Xu và Brown (2016), định vị năng lực kiểm tra đánh giá ngôn ngữ như một kỹ năng xã hội-văn hóa năng động được phát triển thông qua tương tác trong môi trường giảng dạy. Mô hình này định vị năng lực kiểm tra đánh giá ngôn ngữ như một năng lực được hình thành bởi bản sắc nghề nghiệp của các nhà giáo dục và phản ánh một chuỗi liên tục cân bằng các kỹ năng kỹ thuật với khả năng thích ứng theo ngữ cảnh cụ thể. Khung của Xu và Brown nhấn mạnh vai trò của bản sắc nghề nghiệp và sự nhạy cảm theo ngữ cảnh trong việc định hình các hoạt động đánh giá ngôn ngữ hiệu quả.

# 2.4. Các nghiên cứu trước đây về phát triển năng lực đánh giá ngôn ngữ

# Nghiên cứu về năng lực kiểm tra đánh giá ngôn ngữ cho thấy nhiều phương pháp khác nhau, mỗi phương pháp phản ánh nhu cầu giáo dục riêng. Trên thế giới, các nghiên cứu thường sử dụng các công cụ dựa trên tình huống để mô phỏng đánh giá lớp học, mặc dù các công cụ này có thể thiếu trọng tâm cụ thể về ngôn ngữ cần thiết cho bối cảnh giảng dạy tiếng Anh. Tại Việt Nam, Tu (2023) đã phát triển Đánh giá năng lực kiểm tra, đánh giá ngôn ngữ - Điều chỉnh dành cho Việt Nam (LAL-RV), được điều chỉnh để đáp ứng nhu cầu văn hóa và giáo dục của giáo viên giảng dạy tiếng Anh người Việt Nam. Sự điều chỉnh này làm nổi bật tầm quan trọng của các công cụ năng lực kiểm tra đánh giá ngôn ngữ nhạy cảm với ngữ cảnh, đặc biệt là trong môi trường đa ngôn ngữ, đa văn hóa. Mặc dù có những tiến bộ trong phép đo năng lực kiểm tra đánh giá ngôn ngữ, việc xác thực hạn chế các công cụ trong nhiều bối cảnh khác nhau nhấn mạnh nhu cầu về các công cụ đáng tin cậy, có thể điều chỉnh được, phù hợp với các bối cảnh giáo dục cụ thể.

# Chương này xem xét năng lực kiểm tra đánh giá ngôn ngữ như một năng lực đa diện, tích hợp các kỹ năng kỹ thuật, cân nhắc về đạo đức và hiểu biết về văn hóa xã hội. Nó khám phá các mô hình cốt lõi và quan điểm lý thuyết, nhấn mạnh năng lực kiểm tra đánh giá ngôn ngữ như một bộ kỹ năng năng động và phụ thuộc vào bối cảnh. Cuộc thảo luận đã đặt nền tảng cho việc tạo ra một khuôn khổ năng lực kiểm tra đánh giá ngôn ngữ đáp ứng phù hợp với các nhà giáo dục EFL Việt Nam, đáp ứng cả nhu cầu địa phương và các tiêu chuẩn chuyên môn. Chương tiếp theo sẽ tập trung phát triển thang đo năng lực kiểm tra đánh giá ngôn ngữ tùy chỉnh cho các nhà giáo dục ngôn ngữ (tiếng Anh) ở Việt Nam, kết hợp các phương pháp xác nhận nghiêm ngặt để đảm bảo tính phù hợp, độ tin cậy và phù hợp với nhu cầu lý thuyết và thực tiễn của năng lực kiểm tra đánh giá ngôn ngữ.

**CHƯƠNG 3: PHÁT TRIỂN THANG ĐO ĐÁNH GIÁ**

# Chương này phác thảo quá trình phát triển Thang đánh giá năng lực đọc viết ngôn ngữ (LALS) được thiết kế để đo lường năng lực kiểm tra đánh giá ngôn ngữ (LAL) của giảng viên tiếng Anh như ngôn ngữ thứ hai của Việt Nam. Chương này trình bày chi tiết về nền tảng lý thuyết, các bước thực hành và các quy trình trắc nghiệm tâm lý được sử dụng để đảm bảo tính hợp lệ, độ tin cậy và tính phù hợp theo ngữ cảnh của thang đánh giá.

# 3.1. Các phương pháp tiếp cận lý thuyết để phát triển thang đo năng lực kiểm tra đánh giá ngôn ngữ

# Việc phát triển thang đo năng lực kiểm tra đánh giá ngôn ngữ dựa trên một số phương pháp tiếp cận, mỗi phương pháp đều đóng góp những hiểu biết độc đáo:

# Phương pháp tiếp cận năng lực nhấn mạnh các kỹ năng có thể chuyển giao, chẳng hạn như khả năng thích ứng và tư duy phản biện, rất cần thiết để quản lý các tình huống đánh giá đa dạng. Tuy nhiên, do thiếu các chuẩn mực có cấu trúc để tiếp thu kỹ năng nên phương pháp này cần được tích hợp với các phương pháp tiếp cận khác để áp dụng cụ thể.

# Phân tích chức năng tập trung vào các năng lực cụ thể của nhiệm vụ, chia nhỏ các vai trò chuyên môn thành các kỹ năng riêng biệt cần thiết cho việc đánh giá hiệu quả, như thiết kế các mục kiểm tra và diễn giải kết quả. Mặc dù cung cấp các chuẩn mực rõ ràng, nhưng nó có thể thiếu khả năng thích ứng của các phương pháp tiếp cận khác.

# Phương pháp tiếp cận đồng thuận thu thập phản hồi của chuyên gia để xác định các năng lực cốt lõi. Mặc dù tốn nhiều nguồn lực, nhưng nó thúc đẩy sự phù hợp trong các bối cảnh giáo dục, kết hợp các quan điểm liên ngành.

# Phương pháp tiếp cận phát triển nhiều phương pháp tích hợp các yếu tố của từng phương pháp tiếp cận, cân bằng các năng lực cụ thể với tính linh hoạt, hỗ trợ cả nhu cầu giảng dạy tức thời và sự phát triển chuyên môn lâu dài.

# Phương pháp tiếp cận toàn diện này lý tưởng để phát triển một khuôn khổ năng lực kiểm tra đánh giá ngôn ngữ năng động phù hợp với các nhu cầu giáo dục phức tạp.

# 3.2. Phát triển thang đánh giá năng lực kiểm tra đánh giá ngôn ngữ

# Quá trình phát triển thang đánh giá năng lực kiểm tra đánh giá ngôn ngữ tuân theo một phương pháp tiếp cận có hệ thống, bắt đầu bằng định nghĩa cấu trúc và sơ đồ cấu trúc. Dựa trên Wilson (2011) và Wolfe & Smith (2007a), sơ đồ cấu trúc này phác thảo năng lực kiểm tra đánh giá ngôn ngữ như một quá trình tiến triển của các kỹ năng, phản ánh các mức độ năng lực đánh giá khác nhau. Tính hợp lệ được kiểm tra thông qua phân tích Rasch, so sánh quá trình tiến triển năng lực kiểm tra đánh giá ngôn ngữ theo giả thuyết với hiệu suất thực tế, theo gợi ý của Woods & Griffin (2013). Ngoài ra, ý kiến ​​đóng góp từ các chuyên gia đã được tích hợp để đảm bảo thể hiện toàn diện cấu trúc.

# Một mô hình phát triển đã được lựa chọn để minh họa quá trình phát triển của năng lực năng lực kiểm tra đánh giá ngôn ngữ theo thời gian, với các mục được sắp xếp từ kỹ năng cơ bản đến kỹ năng nâng cao. Mô hình tín chỉ một phần (Masters, 1982) được chọn để chấm điểm nhằm nắm bắt phạm vi thành thạo rộng với ít mục hơn, cho phép đo lường sắc thái các cấp độ kỹ năng trong khi vẫn duy trì khả năng diễn giải. Mô hình này giúp thiết lập tính hợp lệ thực chất bằng cách liên kết các dự đoán lý thuyết về năng lực với các kết quả quan sát được.

# Để tạo ra các chỉ số có thể đo lường được, các hướng dẫn về hành vi có thể quan sát được của Griffin, Gillis và Calvitto (2004) đã được sử dụng. Khung này phân định phạm vi năng lực kiểm tra đánh giá ngôn ngữ thành các năng lực, hành vi và tiêu chí thành thạo, tạo ra một đánh giá tham chiếu theo tiêu chí bắt nguồn từ công trình của Glaser (1981). Quá trình phát triển thang đi năng lực kiểm tra đánh giá ngôn ngữ diễn ra qua bốn giai đoạn: soạn thảo, nhóm tập trung và hai vòng thảo luận chuyên gia, trong đó tinh chỉnh khung thông qua phản hồi lặp lại.

# *3.2.1. Soạn thảo nhóm mục*

# Giai đoạn soạn thảo bắt đầu bằng việc xem xét tài liệu để xác định cấu trúc và xác định các thành phần năng lực kiểm tra đánh giá ngôn ngữ cụ thể. Các nguồn bao gồm các khung năng lực kiểm tra đánh giá ngôn ngữ nổi bật của các học giả như Brindley (2001), Davies (2008) và Fulcher (2012). Đánh giá này đã thông báo về một khuôn khổ dự thảo cho giáo viên Việt Nam, với các năng lực được sắp xếp theo thứ bậc như Griffin và cộng sự (2007) đề xuất.

# Ba chiều của năng lực kiểm tra đánh giá ngôn ngữ xuất hiện từ khuôn khổ này: chuẩn bị đánh giá, tiến hành đánh giá và sửa đổi đánh giá. Mỗi chiều bao gồm 13 năng lực, tổng cộng 65 chỉ số hiệu suất có nguồn gốc từ nhiều mô hình năng lực kiểm tra đánh giá ngôn ngữ khác nhau. Tiêu chí chất lượng cho từng chỉ số được xây dựng dựa trên các nguyên tắc thiết kế tiêu chí của Griffin, Nguyen và Gillis (2004) và mô hình tiếp thu kỹ năng Dreyfus và Dreyfus (1986), đảm bảo rằng mỗi tiêu chí đại diện cho các mức độ thành thạo riêng biệt mà không chồng chéo. Phương pháp tiếp cận theo thang phát triển này cho phép giáo viên tự đánh giá năng lực của mình theo định dạng có cấu trúc và tiến bộ.

# *3.2.2. Bảng chuyên gia*

# Giai đoạn bảng chuyên gia bao gồm hai vòng đánh giá của chuyên gia, rất quan trọng để tinh chỉnh và xác thực nội dung của thang đánh giá. Các chuyên gia về đánh giá ngôn ngữ, đo lường giáo dục và giáo dục đại học Việt Nam đã xem xét các mục về tính phù hợp, rõ ràng và đầy đủ.

# Trong vòng đầu tiên, các chuyên gia đã đánh giá tính phù hợp của các mục và nhấn mạnh việc đưa các năng lực đánh giá kỹ thuật số vào, do sự gia tăng của các hoạt động đánh giá trực tuyến và đánh giá quá trình. Các mục đã được sửa đổi để kết hợp kiến ​​thức số và đảm bảo tính rõ ràng, đặc biệt là đối với giảng viên ở những khu vực có nguồn lực hạn chế. Vòng thứ hai tập trung vào việc tinh chỉnh tiến trình phát triển trong các tiêu chí chất lượng, điều chỉnh ngôn ngữ và đảm bảo các mục thể hiện chính xác một chuỗi liên tục từ năng lực cơ bản đến nâng cao. Quá trình lặp đi lặp lại này đã tạo ra một tập hợp các mục được sắp xếp hợp lý, hợp lý, phù hợp để chấm điểm phát triển.

# *3.2.3. Thử nghiệm thang đo*

# Giai đoạn thử nghiệm đã kiểm tra thực nghiệm các đặc tính trắc nghiệm tâm lý của thang đo bằng cách sử dụng mẫu gồm 177 giảng viên tiếng Anh như một ngoại ngữ của Việt Nam từ nhiều khu vực khác nhau. Các chiến lược lấy mẫu phân tầng và cụm đảm bảo tính đại diện trên khắp các tổ chức, nắm bắt nhiều bối cảnh giáo dục khác nhau.

# Những người tham gia đã hoàn thành thang đo như một bài tự đánh giá, với các mục phản ánh chuỗi liên tục phát triển của trình độ thành thạo năng lực kiểm tra đánh giá ngôn ngữ. Lý thuyết kiểm tra cổ điển (CTT) và mô hình Rasch đã phân tích độ tin cậy và tính hợp lệ của cấu trúc. CTT cho thấy tính nhất quán nội bộ cao, trong khi Mô hình tín chỉ một phần của Rasch xác thực thang đo phát triển bằng cách căn chỉnh độ khó của mục với tiến trình dự kiến ​​về năng lực của giảng viên. Thống kê phù hợp cho thấy tính đơn chiều, khẳng định rằng các mục đại diện chung cho năng lực năng lực kiểm tra đánh giá ngôn ngữ và sơ đồ mục-người xác nhận phạm vi độ khó phù hợp để phân biệt năng lực của giảng viên.

# 3.3. Lập bản đồ cấu trúc của năng lực kiểm tra đánh giá ngôn ngữ

# Dựa trên nghiên cứu trước đây, bản đồ cấu trúc định hình năng lực kiểm tra đánh giá ngôn ngữ như một năng lực năng động, nhạy cảm với ngữ cảnh, được chia thành ba giai đoạn: Chuẩn bị đánh giá (WHAT), Tiến hành đánh giá (HOW) và Xem lại đánh giá (WHY). Mỗi giai đoạn đều xây dựng trên các năng lực được xác định bởi các mô hình năng lực kiểm tra đánh giá ngôn ngữ cơ bản, cung cấp cho giảng viên tiếng Anh như ngôn ngữ thứ hai của Việt Nam một phương pháp tiếp cận có hệ thống để phát triển các kỹ năng đánh giá.

# Chuẩn bị đánh giá (WHAT): Bao gồm việc hiểu các nguyên tắc đánh giá chính và đặt ra các mục tiêu rõ ràng, có thể đo lường được phù hợp với tiêu chuẩn giáo dục.

# Tiến hành đánh giá (HOW): Bao gồm các bước thực tế để quản lý, chấm điểm và diễn giải các đánh giá, đảm bảo phù hợp với các mục tiêu học tập và tính công bằng.

# Xem lại đánh giá (WHY): Tập trung vào việc đánh giá liên tục và tinh chỉnh các hoạt động đánh giá để đảm bảo tính phù hợp và hiệu quả.

# Khung dựa trên quy trình này hỗ trợ sự phát triển chuyên môn liên tục, cho phép các nhà giáo dục ngữ cảnh hóa năng lực kiểm tra đánh giá ngôn ngữ trong môi trường giảng dạy độc đáo của họ và phản ánh về vai trò đang phát triển của họ với tư cách là người đánh giá.

# 3.4. Đánh giá chuyên gia

# Giai đoạn đánh giá bao gồm hai vòng đánh giá của chuyên gia, tinh chỉnh khung năng lực kiểm tra, đánh giá để đảm bảo tính phù hợp, rõ ràng và phù hợp với nhu cầu của giảng viên tiếng Anh Việt Nam. Các chuyên gia đánh giá tiếng Anh, đo lường giáo dục và giáo dục đại học Việt Nam đánh giá thang đo, tập trung vào độ chính xác về mặt kỹ thuật và sự phù hợp với bối cảnh. Ở vòng đầu tiên, các chuyên gia đã đánh giá dự thảo ban đầu gồm 65 hạng mục trên ba khía cạnh: chuẩn bị, tiến hành và sửa đổi đánh giá. Phản hồi nhấn mạnh sự cần thiết phải chú trọng nhiều hơn đến đánh giá kỹ thuật số và thực hành hình thành, dẫn đến việc đưa vào các mục mới và sửa đổi để cải thiện tính rõ ràng, đặc biệt đối với các nhà giáo dục ở nông thôn. Vòng thứ hai tập trung vào quá trình phát triển, đơn giản hóa ngôn ngữ để phân biệt các cấp độ năng lực và sắp xếp lại các mục để tăng cường tính logic. Cuối cùng, thang đo này phản ánh một công cụ có cấu trúc tốt và phù hợp với ngữ cảnh dành cho giảng viên tiếng Anh Việt Nam.

# 3.5. Thử nghiệm

# Giai đoạn thử nghiệm khung năng lực kiểm tra đánh giá được thử nghiệm với 177 giảng viên tiếng Anh ở Việt Nam từ các bối cảnh giảng dạy khác nhau trên cả nước. Sử dụng phương pháp lấy mẫu phân tầng, cụm và lấy mẫu thuận tiện, nghiên cứu đã thu thập dữ liệu từ các tổ chức có nguồn lực công nghệ khác nhau. Những người tham gia đã hoàn thành bài tự đánh giá mức độ thành thạo năng lực kiểm tra đánh giá và phân tích dữ liệu kết hợp Lý thuyết kiểm tra cổ điển và mô hình Rasch. Tính nhất quán nội tại cao (Cronbach’s alpha > 0,80) và tính đơn hướng đã khẳng định độ tin cậy và sự phù hợp của thang đo với năng lực kiểm tra đánh giá. Phân tích của Rasch đã chứng minh nhiều độ khó của từng mục, giúp phân biệt hiệu quả khả năng của giảng viên, trong khi các câu trả lời theo nhóm nhân khẩu học cho thấy tính nhất quán và khả năng thích ứng. Phản hồi đã thúc đẩy những điều chỉnh nhỏ đối với năng lực kỹ thuật số, đảm bảo tính toàn diện. Các phát hiện đã xác nhận thang đo này là một thước đo đáng tin cậy, phù hợp với bối cảnh, có khả năng hỗ trợ việc tự đánh giá, phát triển chuyên môn và đánh giá thể chế cho giảng viên tiếng Anh ở Việt Nam.

# Chương 3 trình bày quá trình phát triển thang đo như một phương pháp tiếp cận có cấu trúc, được xác thực theo kinh nghiệm để đo lường năng lực kiểm tra đánh giá ngôn ngữ cho giảng viên tiếng Anh như ngôn ngữ thứ hai của Việt Nam. Thang đo, được hỗ trợ bởi phản hồi của chuyên gia và phân tích trắc nghiệm tâm lý nghiêm ngặt, cung cấp một công cụ đáng tin cậy, phù hợp với ngữ cảnh để đánh giá và bồi dưỡng năng lực tiếng Anh như ngôn ngữ thứ hai. Thang đo này hỗ trợ tự đánh giá, phát triển chuyên môn và đánh giá của tổ chức, phù hợp với nhu cầu cụ thể của các nhà giáo dục tiếng Anh như ngôn ngữ thứ hai trong các bối cảnh giáo dục đa dạng của Việt Nam. Các bước tiếp theo bao gồm tinh chỉnh ứng dụng của thang đo này vào các bối cảnh giảng dạy khác nhau và tận dụng nó cho các chương trình cải thiện tiếng Anh như ngôn ngữ thứ hai có mục tiêu trên khắp Việt Nam.

# CHƯƠNG 4: CÁC PHƯƠNG PHÁP XÁC TRỊ THANG ĐO

# Chương này trình bày chi tiết về phương pháp xác nhận cho Thang đánh giá năng lực kiểm tra đánh giá ngôn ngữ (LALS), tập trung vào khuôn khổ dựa trên lập luận (Kane, 1992, 2013; Chapelle và cộng sự, 2008) được hỗ trợ bởi mô hình Rasch. Quá trình xác nhận nhấn mạnh vào cách diễn giải điểm số dựa trên bằng chứng hơn là bản thân bài kiểm tra, diễn ra theo hai giai đoạn. Đầu tiên, Lập luận diễn giải và sử dụng (IUA) được phát triển để phác thảo cách diễn giải và cách sử dụng điểm số dự định. Thứ hai, một lập luận về tính hợp lệ được xây dựng để đánh giá sự hỗ trợ thực nghiệm cho các cách diễn giải này. Các suy luận cốt lõi - mô tả miền, đánh giá, khái quát hóa và giải thích - được sử dụng để xây dựng bằng chứng hỗ trợ thang đo. Cấu trúc này cung cấp một phương pháp tiếp cận chặt chẽ, dựa trên bằng chứng để xác nhận LALS, đặt nền tảng để giải quyết các suy luận ngoại suy và sử dụng trong các nghiên cứu trong tương lai.

# 4.1. Các phương pháp tiếp cận để xác nhận đánh giá

# *4.1.1. Quan điểm về tính hợp lệ đương đại*

# Khái niệm về tính hợp lệ đã phát triển từ các định nghĩa đơn giản thành các phương pháp tiếp cận tích hợp (AERA, APA và NCME, 1985, 2014; Kane, 2013; Sireci, 2013). Quan điểm truyền thống phân loại tính hợp lệ thành ba loại chính: tính hợp lệ nội dung, liên quan đến tiêu chí và tính hợp lệ cấu trúc. Tính hợp lệ nội dung đánh giá xem các mục kiểm tra có đại diện chính xác cho phạm vi hay không; tính hợp lệ liên quan đến tiêu chí tập trung vào mối tương quan giữa điểm kiểm tra với các chuẩn mực bên ngoài, trong khi tính hợp lệ cấu trúc, được Cronbach và Meehl (1955) giới thiệu, liên quan đến việc đảm bảo bài kiểm tra đo lường một đặc điểm lý thuyết theo ý định. Sau đó, lý thuyết thống nhất của Messick (1989) đã hợp nhất những điều này thành một khuôn khổ duy nhất, coi tính hợp lệ là một phán đoán đánh giá được hỗ trợ bởi nhiều loại bằng chứng khác nhau. Mô hình thống nhất này, được phản ánh trong Tiêu chuẩn kiểm tra giáo dục và tâm lý, nhấn mạnh quan điểm toàn diện về tính hợp lệ, xem xét nội dung, cấu trúc và hậu quả xã hội của bài kiểm tra.

# *4.1.2. Phát triển các khuôn khổ xác thực hiện đại*

# Các khuôn khổ xác thực hiện đại, bao gồm lý thuyết hợp lệ thống nhất của Messick, lập luận sử dụng đánh giá (AUA) của Bachman và Palmer (2010), phương pháp tiếp cận dựa trên lập luận của Kane (1992, 2006, 2013), thiết kế lấy bằng chứng làm trung tâm (ECD) của Mislevy và Yin (2012) và khuôn khổ nhận thức xã hội của Weir (2005), cung cấp các phương pháp tiếp cận toàn diện tích hợp bằng chứng lý thuyết và thực nghiệm để xác thực các cách diễn giải và sử dụng điểm kiểm tra.

# Lý thuyết hợp lệ thống nhất của Messick tích hợp tất cả các hình thức hợp lệ, mặc dù bản chất trừu tượng của nó hạn chế hướng dẫn thực tế.

# AUA của Bachman và Palmer cấu trúc xác thực thông qua mô hình dựa trên lập luận tập trung vào việc sử dụng đánh giá, khiến nó trở nên có giá trị đối với các ứng dụng thực tế.

# Phương pháp tiếp cận dựa trên lập luận của Kane tổ chức xác thực thông qua suy luận và bằng chứng hỗ trợ, nhấn mạnh các khía cạnh thực tế và có hậu quả. Mô hình này, có thể thích ứng với nhiều bối cảnh kiểm tra khác nhau, đặc biệt phù hợp với các đánh giá có rủi ro cao và đánh giá trong lớp học.

# Khung ECD của Mislevy và Yin nhúng các nhu cầu về bằng chứng vào thiết kế bài kiểm tra, sắp xếp các nhiệm vụ với các cấu trúc ngay từ đầu, mặc dù nó phù hợp hơn với các bối cảnh chuẩn hóa.

# Khung xã hội-nhận thức của Weir tích hợp các cân nhắc về nhận thức và xã hội, có lợi cho các đánh giá ngôn ngữ có rủi ro cao, trong đó tính xác thực của nhiệm vụ được ưu tiên.

# Phương pháp của Kane, với xác thực tập trung vào suy luận, đặc biệt hữu ích cho các đánh giá giáo dục và kiểm tra ngôn ngữ. Tính linh hoạt của khung của ông hỗ trợ các lập luận xác thực có cấu trúc, khiến nó phù hợp với xác thực LALS trong các bối cảnh khác nhau.

# 4.2. Xây dựng khung xác trị của nghiên cứu

# *4.2.1. Cơ sở lý luận cho việc lựa chọn khung*

# Cách tiếp cận dựa trên lập luận của Kane (1992, 2013) được sử dụng làm nền tảng cho nghiên cứu này do quy trình có cấu trúc, dựa trên suy luận, liên kết một cách có hệ thống điểm kiểm tra với cách diễn giải và cách sử dụng chúng. Nó cung cấp một khuôn khổ rõ ràng để xây dựng Đối số sử dụng diễn giải (IUA) và Đối số hợp lệ, đảm bảo sự mạch lạc về mặt logic và hỗ trợ thực nghiệm cho từng suy luận, chẳng hạn như mô tả, đánh giá và khái quát hóa miền.

# Khả năng thích ứng của Kane với các bối cảnh đánh giá khác nhau khiến nó phù hợp để xác nhận thang đo năng lực kiểm tra đánh giá, nhằm mục đích đánh giá cấu trúc phức tạp, đa chiều của năng lực kiểm tra đánh giá trên các môi trường giáo dục đa dạng của Việt Nam. Không giống như các mô hình giá trị truyền thống, cách tiếp cận của Kane nhấn mạnh ý nghĩa thực tế, phù hợp với bản chất văn hóa xã hội của năng lực kiểm tra đánh giá và đảm bảo kết quả kiểm tra có ý nghĩa và có thể áp dụng được cho các bên liên quan. Sự nhấn mạnh này hỗ trợ sự phù hợp của thang đo với thực tiễn giảng dạy và đánh giá trong thế giới thực. Bằng cách kết hợp khuôn khổ của Kane với những hiểu biết sâu sắc của Bachman, Palmer và Chapelle, nghiên cứu này đảm bảo quy trình xác nhận nghiêm ngặt, phù hợp với ngữ cảnh, thiết lập thang đo như một công cụ đáng tin cậy để đánh giá năng lực kiểm tra đánh giá của các giảng viên tiếng Anh tại Việt Nam.

# *4.2.2. Cấu trúc dựa trên đối số để xác thực thang đo*

# *4.2.2.1. Mô hình lập luận của Toulmin*

# Mô hình lập luận của Toulmin (1958) tạo thành nền tảng của khung xác nhận dựa trên lập luận, cung cấp cách tiếp cận có cấu trúc để liên kết các tuyên bố với bằng chứng. Nó bao gồm sáu thành phần: dữ liệu, yêu cầu bồi thường, bảo đảm, ủng hộ, vòng loại và bác bỏ. Mô hình này nhấn mạnh mối liên hệ hợp lý giữa dữ liệu và các tuyên bố, được hỗ trợ bởi các bảo đảm và sự ủng hộ của chúng, đồng thời giải quyết các lập luận phản biện thông qua các phản bác. Khuôn khổ của Toulmin đã có ảnh hưởng rộng rãi trong nhiều lĩnh vực khác nhau, bao gồm đo lường giáo dục, trong đó Kane (2013) đã điều chỉnh nó để xác nhận giá trị của bài kiểm tra, nhấn mạnh sự gắn kết giữa thực tiễn và lý thuyết trong việc diễn giải và sử dụng điểm kiểm tra.

# *4.2.2.2. Khung dựa trên lập luận trong nghiên cứu này*

# Cách tiếp cận dựa trên lập luận của Kane khái niệm hóa tính hợp lệ như một thuộc tính của việc giải thích và sử dụng điểm kiểm tra, chứ không phải bản thân bài kiểm tra. Việc xác nhận bao gồm việc xây dựng hai đối số: Đối số sử dụng diễn giải (IUA), trong đó chỉ định cách giải thích và sử dụng điểm số theo dự kiến, và Đối số hiệu lực, đánh giá tính mạch lạc và hợp lý của IUA với bằng chứng thực nghiệm. Khung này đã được Bachman, Palmer và Chapelle mở rộng để bao gồm các khía cạnh thực dụng, nhận thức và hệ quả, đảm bảo sự phù hợp với các nhiệm vụ và bối cảnh ngôn ngữ trong thế giới thực. Cách tiếp cận dựa trên lập luận tích hợp cả cấu trúc lý thuyết và cân nhắc thực tế, sử dụng bằng chứng định lượng và định tính để tạo ra một quy trình xác nhận toàn diện.

# *4.2.3. Các bước trong ứng dụng khung*

# *4.2.3.1. Giải thích và sử dụng lập luận*

# IUA cấu trúc trình tự logic từ kết quả kiểm tra được quan sát đến các tuyên bố về năng lực LAL của giáo viên tiếng Anh. Điều này liên quan đến việc chỉ định các giả định về căn chỉnh miền của thang đo, độ chính xác của điểm số, tính nhất quán của điểm số và phản ánh cấu trúc. Mỗi suy luận đều được hỗ trợ bởi các bảo đảm - các tuyên bố biện minh cho các suy luận dựa trên các giả định về quá trình đánh giá. Ví dụ: suy luận mô tả miền được hỗ trợ bởi bằng chứng cho thấy các mục của thang đo phù hợp với các thành phần năng lực kiểm tra đánh giá thiết yếu dành cho giáo viên tiếng Anh.

# *4.2.3.2. Đối số hợp lệ*

# Lập luận về tính hợp lệ, giai đoạn thứ hai, đánh giá tính mạch lạc của IUA, thu thập bằng chứng để hỗ trợ cho từng suy luận. Quá trình này đánh giá xem điểm thang đo năng lực kiểm tra đánh giá ngôn ngữ có đại diện chính xác cho cấu trúc năng lực kiểm tra đánh giá ngôn ngữ hay không, có được chấm điểm đáng tin cậy và có thể khái quát hóa trong nhiều bối cảnh giáo dục khác nhau hay không. Lập luận về tính hợp lệ của thang đo năng lực kiểm tra đánh giá ngôn ngữ bao gồm hỗ trợ thực nghiệm thông qua mô hình Rasch, đánh giá của chuyên gia và phân tích trắc nghiệm tâm lý, đảm bảo một khuôn khổ xác thực toàn diện.

# 4.3. Yêu cầu về bằng chứng cho các suy luận

# Phương pháp tiếp cận dựa trên lập luận chỉ định các loại bằng chứng để chứng minh cho từng suy luận:

# - Bằng chứng mô tả miền: Đánh giá của chuyên gia và phân tích liên kết thiết lập sự liên quan của nội dung với năng lực kiểm tra đánh giá ngôn ngữ.

# - Bằng chứng đánh giá: Tính nhất quán của điểm được chứng minh thông qua độ tin cậy giữa những người đánh giá và mô hình Rasch, xác nhận độ chính xác của điểm thang đo năng lực kiểm tra đánh giá ngôn ngữ.

# - Bằng chứng khái quát hóa: Các nghiên cứu về độ tin cậy, bao gồm alpha của Cronbach, hỗ trợ tính nhất quán của điểm trong nhiều bối cảnh khác nhau.

# - Bằng chứng giải thích: Tính hợp lệ của cấu trúc được xác nhận thông qua phân tích nhân tố và bản đồ Wright, xác minh sự liên kết giữa điểm thang đo năng lực kiểm tra đánh giá ngôn ngữ và các năng lực năng lực kiểm tra đánh giá ngôn ngữ cơ bản.

# 4.4. Trình bày các lập luận diễn giải và tính hợp lệ cho thang đo năng lực kiểm tra đánh giá ngôn ngữ

# Xác thực thang đo năng lực kiểm tra đánh giá ngôn ngữ ưu tiên mô tả miền, đánh giá, khái quát hóa và suy luận giải thích, căn chỉnh bằng chứng thực nghiệm với từng miền để chứng minh cho các diễn giải về điểm số dự định của thang đo năng lực kiểm tra đánh giá ngôn ngữ. IUA chỉ định các giả định, suy luận và bằng chứng hỗ trợ, trong khi lập luận về tính hợp lệ thu thập dữ liệu thực nghiệm để xác nhận logic của IUA. Ví dụ, suy luận giải thích liên quan đến bằng chứng từ phân tích nhân tố và thống kê phù hợp Rasch để xác minh rằng điểm số thang đo năng lực kiểm tra đánh giá ngôn ngữ phản ánh cấu trúc năng lực kiểm tra đánh giá ngôn ngữ dự định.

# Chương 4 mô tả quá trình xác thực thang đo năng lực kiểm tra đánh giá ngôn ngữ bằng cách sử dụng khuôn khổ dựa trên lập luận của Kane, nhấn mạnh vào cấu trúc Lập luận diễn giải và sử dụng và Lập luận về tính hợp lệ. Bằng cách tập trung vào các suy luận cốt lõi và thu thập bằng chứng có liên quan, phương pháp này thiết lập thang đo năng lực kiểm tra đánh giá ngôn ngữ là một công cụ đáng tin cậy và hợp lệ để đánh giá năng lực kiểm tra đánh giá ngôn ngữ trong số các giáo viên tiếng Anh như ngôn ngữ thứ hai của Việt Nam. Khuôn khổ xác thực dựa trên bằng chứng này, với các suy luận bổ sung dành riêng cho nghiên cứu trong tương lai, xác nhận rằng thang đo năng lực kiểm tra đánh giá ngôn ngữ phù hợp với các ứng dụng dự định của nó trong các bối cảnh giáo dục đa dạng.

# CHƯƠNG 5: PHƯƠNG PHÁP NGHIÊN CỨU

# Chương này trình bày chi tiết phương pháp để xây dựng và xác thực Thang đo Kiến thức Đánh giá Ngôn ngữ (LALS) dành cho giảng viên giảng dạy tiếng Anh tại Việt Nam trong hệ thống giáo dục đại học. Phương pháp nghiên cứu bao gồm bối cảnh nghiên cứu, câu hỏi nghiên cứu, thiết kế, lấy mẫu, công cụ thu thập dữ liệu, phương pháp phân tích và quy trình xác thực. Cách tiếp cận có cấu trúc này tạo nền tảng cho tính thực tiễn và tính chất tâm lý đáng tin cậy của thang đo năng lực kiểm tra đánh giá ngôn ngữ.

# 5.1. Bối cảnh nghiên cứu

# Nghiên cứu được thực hiện trong hệ thống giáo dục đại học tại Việt Nam, nơi các cải cách giáo dục gần đây và sự nhấn mạnh vào tiếng Anh cho hội nhập toàn cầu đã làm gia tăng nhu cầu về năng lực kiểm tra đánh giá ngôn ngữ đối với giảng viên giảng dạy tiếng Anh. Mặc dù Bộ Giáo dục và Đào tạo (MOET) đã đưa ra các chính sách và tiêu chuẩn, nhưng thực tiễn đánh giá thường dựa trên các kỳ thi có tính chất quyết định cao, chú trọng đến việc ghi nhớ hơn là kỹ năng giao tiếp, tạo ra một hệ thống thi cử trung tâm. Bên cạnh đó, các hạn chế từ phía cơ sở giáo dục – như quy mô lớp học lớn, chênh lệch tài nguyên và khối lượng giảng dạy nặng nề – đã hạn chế cơ hội đánh giá thường xuyên, đặc biệt là ở khu vực nông thôn. Về văn hóa, di sản Khổng giáo của Việt Nam nhấn mạnh mối quan hệ thứ bậc giữa giáo viên và sinh viên, gây khó khăn cho các phương pháp tiếp cận lấy học sinh làm trung tâm và mang tính xây dựng. Các giáo viên giảng dạy tiếng Anh thường có ít đào tạo về các kỹ thuật đánh giá đa dạng, với sự phụ thuộc lớn vào các phương pháp truyền thống, đánh giá tổng kết. Những yếu tố này nhấn mạnh sự cần thiết của một khung năng lực kiểm tra đánh giá ngôn ngữ địa phương hóa để hướng dẫn các giáo viên giảng dạy tiếng Anh Việt Nam áp dụng các thực tiễn đánh giá cân bằng và hiệu quả hơn.

# 5.2. Xem lại câu hỏi nghiên cứu

# Nghiên cứu đề cập đến hai mục tiêu chính:

# Phát triển thang đo năng lực kiểm tra đánh giá ngôn ngữ: Đánh giá các năng lực quan trọng đối với năng lực kiểm tra đánh giá ngôn ngữ của giảng viên giảng dạy tiếng Anh Việt Nam, tập trung vào kiến thức, kỹ năng và khả năng trong đánh giá.

# Xác thực thang đo năng lực kiểm tra đánh giá ngôn ngữ: Sử dụng khung xác thực dựa trên lập luận của Kane (2006), nghiên cứu xác thực bốn suy luận – mô tả miền, đánh giá, tổng quát hóa và giải thích:

Câu hỏi nghiên cứu 1: Các mục trong thang đo năng lực kiểm tra đánh giá ngôn ngữ phản ánh chính xác đến mức nào về kiến thức, kỹ năng và năng lực cần thiết của giảng viên tiếng Anh, dựa trên phản hồi từ chuyên gia và mức độ phù hợp của nội dung? (Suy luận mô tả lĩnh vực).

Câu hỏi nghiên cứu 2: Các thành phần của thang đo có đáp ứng các tiêu chí trắc nghiệm tâm lý như độ khó, khả năng phân biệt và tính nhất quán nội bộ để đảm bảo độ tin cậy và tính hợp lệ trong đánh giá không? (Suy luận đánh giá)

Câu hỏi nghiên cứu 3: Điểm số của thang đo có thể được khái quát hóa đến mức nào cho các nhóm giảng viên tiếng Anh tại Việt Nam với các đặc điểm nhân khẩu học và trình độ kinh nghiệm khác nhau? (Suy luận khái quát).

Câu hỏi nghiên cứu 4: Mức độ mà thang đo phản ánh cấu trúc cơ bản của năng lực kiểm tra đánh giá ngôn ngữ cũng như khả năng của các mục trong thang đo này trong việc phân biệt năng lực của giảng viên ở các cấp độ khác nhau được đánh giá ra sao? (Suy luận giải thích).

# 5.3. Thiết kế nghiên cứu

# Thiết kế phương pháp hỗn hợp tuần tự thăm dò đã được áp dụng, tích hợp dữ liệu định tính và định lượng thông qua khung xác thực dựa trên lập luận. Giai đoạn định tính, bao gồm các nhóm tập trung và phản hồi từ chuyên gia, đã cung cấp thông tin cho giai đoạn định lượng tiếp theo. Thiết kế này cho phép khám phá chi tiết về năng lực kiểm tra đánh giá ngôn ngữ trước khi chuyển sang đo lường, và tạo điều kiện liên kết các hiểu biết định tính với đánh giá định lượng.

# 5.4. Đối tượng nghiên cứu và lấy mẫu

# *5.4.1. Giai đoạn định tính:* Lấy mẫu có chủ đích được sử dụng cho các nhóm tập trung và đánh giá từ chuyên gia. Trong nhóm tập trung, 12 người tham gia, bao gồm giảng viên giảng dạy tiếng Anh và các chuyên gia đánh giá ngôn ngữ, đã cung cấp những hiểu biết về sự rõ ràng và tính ứng dụng của khung. Sáu chuyên gia đánh giá, với nền tảng rộng lớn trong lĩnh vực đánh giá ngôn ngữ, đã tinh chỉnh khung nhằm đảm bảo phù hợp với bối cảnh giáo dục đại học Việt Nam.

# *5.4.2. Giai đoạn định lượng:* Chiến lược lấy mẫu phân tầng theo các khu vực Bắc, Trung và Nam của Việt Nam đảm bảo một mẫu đại diện cho việc xác thực định lượng. Mẫu cuối cùng gồm 177 người tham gia, đáp ứng các yêu cầu cho phân tích Rasch và cho phép ước lượng tham số mục đáng tin cậy. Cách tiếp cận lấy mẫu này cân bằng tính khả thi về mặt logistic với sự bao phủ toàn diện trên bối cảnh giáo dục đại học Việt Nam.

# 5.5. Công cụ thu thập dữ liệu

# Nghiên cứu sử dụng ba công cụ chính:

# Hướng dẫn nhóm tập trung: Một hướng dẫn với 12 câu hỏi mở đã tạo điều kiện cho các cuộc thảo luận sâu về sự liên quan của nội dung khung. Hướng dẫn này đã được kiểm tra với các giáo viên trung học để cung cấp định dạng có cấu trúc nhưng linh hoạt nhằm thu thập phản hồi từ người tham gia.

# Phỏng vấn bán cấu trúc cho hội đồng chuyên gia: Trong hai vòng, sáu chuyên gia đã cung cấp phản hồi chi tiết về sự rõ ràng, tính liên quan và cấu trúc của khung. Những ý kiến này giúp tinh chỉnh các chỉ số và căn chỉnh khung với các nguyên tắc phát triển trong năng lực kiểm tra đánh giá ngôn ngữ.

# Bảng câu hỏi trực tuyến để thử nghiệm: Bảng câu hỏi bao gồm các câu hỏi về nhân khẩu học và các mục tự đánh giá dựa trên khung thang đo năng lực kiểm tra đánh giá ngôn ngữ, đã được kiểm tra để đảm bảo sự rõ ràng. Phản hồi từ người tham gia giúp hoàn thiện các mục, đảm bảo tự đánh giá chính xác trong một mẫu lớn hơn.

# 5.6. Phân tích dữ liệu

# Nghiên cứu sử dụng các kỹ thuật phân tích định tính và định lượng để trả lời các câu hỏi nghiên cứu và hỗ trợ suy luận xác thực.

# *5.6.1. Phân tích định tính*

# Phân tích chủ đề, theo phương pháp sáu giai đoạn của Braun và Clarke (2012), xác định các mẫu và chủ đề từ các nhóm tập trung và đánh giá chuyên gia. Mã hóa qua MAXQDA cho phép tinh chỉnh các chủ đề, đặc biệt liên quan đến suy luận Mô tả Miền, đảm bảo rằng khung thang đo năng lực kiểm tra đánh giá ngôn ngữ phản ánh chính xác các năng lực quan trọng trong hệ thống giáo dục đại học Việt Nam.

# *5.6.2. Phân tích định lượng*

# Phân tích Rasch: Mô hình Tín chỉ Một phần (PCM) được sử dụng để hiệu chỉnh các mục và xác định các tương tác giữa mục và người, cung cấp hiểu biết tinh tế về độ khó và khả năng phân biệt của mục.

# Phân tích độ tin cậy: Cả độ nhất quán nội bộ (hệ số Cronbach’s alpha) và các chỉ số dựa trên mô hình Rasch đều xác nhận độ tin cậy. Các chỉ số phân tách mục và người hỗ trợ khả năng phân biệt các cấp độ năng lực khác nhau của thang đo.

# Độ phù hợp mô hình và phân biệt mục: Thống kê độ phù hợp (INFIT và OUTFIT) đo lường sự căn chỉnh của các mục với cấu trúc năng lực kiểm tra đánh giá ngôn ngữ, với các giá trị mean-square và thống kê t định hướng tinh chỉnh mục. Các chỉ số phân tách mục và tương quan điểm-biserial xác nhận khả năng phân biệt hiệu quả của mục.

# Đường cong đặc trưng mục (ICCs) và bản đồ Wright: ICCs thể hiện mức độ phân biệt năng lực của các mục, trong khi bản đồ Wright hiển thị phân phối các mục, xác nhận liên tục kỹ năng năng lực kiểm tra đánh giá ngôn ngữ từ cơ bản đến nâng cao.

# Chương này đã trình bày cách tiếp cận toàn diện trong việc phát triển và xác thực thang đo năng lực kiểm tra đánh giá ngôn ngữ thông qua thiết kế phương pháp hỗn hợp dựa trên lập luận xác thực. Quy trình nghiên cứu giải quyết có hệ thống từng suy luận xác thực, phù hợp với các phương pháp tốt nhất cho đánh giá giáo dục trong giáo dục đại học Việt Nam. Phương pháp này, kết hợp những hiểu biết định tính và độ chính xác định lượng, hỗ trợ khả năng áp dụng và độ tin cậy của thang đo năng lực kiểm tra đánh giá ngôn ngữ như một công cụ để đánh giá năng lực kiểm tra đánh giá ngôn ngữ cho các giảng viên giảng dạy tiếng Anh tại Việt Nam. Chương tiếp theo sẽ phân tích thêm các đặc điểm tâm lý của thang đo năng lực kiểm tra đánh giá ngôn ngữ, trình bày kết quả xác thực để củng cố độ tin cậy và tính thực tiễn của nó.

**CHƯƠNG 6: KẾT QUẢ XÁC THỰC THANG ĐO KIẾN THỨC ĐÁNH GIÁ NGÔN NGỮ CỦA GIẢNG VIÊN GIẢNG DẠY TIẾNG ANH TẠI VIỆT NAM - SUY LUẬN MÔ TẢ MIỀN**

Chương này trình bày quy trình xác thực cho thang đo năng lực kiểm tra đánh giá ngôn ngữ thông qua suy luận mô tả miền, tập trung vào việc xác định tính liên quan và khả năng áp dụng của thang đo trong hệ thống giáo dục đại học Việt Nam.

**6.1. Bằng chứng về tính hợp lệ ban đầu từ quá trình xây dựng thang đo**

Việc phát triển thang đo năng lực kiểm tra đánh giá ngôn ngữ dành cho giảng viên giảng dạy tiếng Anh Việt Nam tuân theo bốn bước để đảm bảo tính liên quan với bối cảnh giảng dạy. Các năng lực bao gồm ba miền chính, mỗi miền gắn với các giai đoạn đánh giá quan trọng: chuẩn bị, thực hiện và điều chỉnh đánh giá.

- Soạn thảo: Sử dụng tài liệu và phân tích bối cảnh (từ Chương 3 và 4), giai đoạn này đã phác thảo ba giai đoạn đánh giá:

* + Chuẩn bị đánh giá: Kiến thức về các khung năng lực, logistics đánh giá và xây dựng bài kiểm tra.
  + Thực hiện đánh giá: Quản lý bài kiểm tra, chấm điểm, cung cấp phản hồi và báo cáo.
  + Điều chỉnh đánh giá: Đánh giá liên tục để cải thiện chất lượng dựa trên phản hồi và dữ liệu kết quả.

Những giai đoạn này tương ứng với mô hình tiếp thu kỹ năng của Dreyfus (Dreyfus & Dreyfus, 1986), phân loại sự tiến triển năng lực từ người mới đến chuyên gia. Cách tiếp cận có cấu trúc này đảm bảo giảng viên có kiến thức từ cơ bản đến nâng cao về đánh giá, phù hợp với các năng lực năng lực kiểm tra đánh giá ngôn ngữ trong lớp học giảng dạy tiếng Anh Việt Nam.

- Phát triển năng lực và chỉ số: Sáu mươi lăm chỉ số hiệu suất đã được xác định dựa trên tài liệu và nhu cầu bối cảnh. Các chỉ số trong giai đoạn Chuẩn bị đánh giá bao gồm hiểu các khái niệm đánh giá, đặt mục tiêu phù hợp với chương trình giảng dạy đại học và chọn các phương pháp đánh giá phù hợp. Các chỉ số trong Thực hiện đánh giá bao gồm kỹ năng thực hiện như nhất quán trong chấm điểm, cung cấp phản hồi hiệu quả và truyền đạt kết quả đánh giá. Trong Điều chỉnh đánh giá, giảng viên được mong đợi đánh giá chất lượng bài kiểm tra dựa trên thực tế lớp học, như được các học giả như Davies (2008) và Fulcher (2012) đề cập.

- Phát triển tiêu chí chất lượng: Mô hình Dreyfus cũng là cơ sở cho tiêu chí chất lượng, được tổ chức theo mức độ thành thạo tăng dần từ ghi nhớ cơ bản đến áp dụng trong bối cảnh giảng dạy thực tế. Ví dụ, giảng viên tiến bộ từ định nghĩa các thuật ngữ cơ bản đến đánh giá và áp dụng các khung đánh giá ngôn ngữ trong bối cảnh đại học. Cách tiếp cận này hỗ trợ đánh giá đa chiều và có cấu trúc cho năng lực kiểm tra đánh giá ngôn ngữ của giảng viên.

- Kết quả nhóm tập trung: Phân tích dữ liệu từ nhóm tập trung đã cho thấy một số chủ đề:

* + Điều chỉnh khung: Các chuyên gia khuyến nghị đơn giản hóa các chỉ số để dễ hiểu và sắp xếp lại chúng để phù hợp hơn với các giai đoạn đánh giá thực tiễn.
  + Tính phù hợp văn hóa: Khung tích hợp các tiêu chuẩn giáo dục Việt Nam, nâng cao khả năng áp dụng trong bối cảnh địa phương.
  + Cấu trúc năng lực: Cấu trúc ba giai đoạn đảm bảo một đại diện toàn diện cho các quy trình năng lực kiểm tra đánh giá ngôn ngữ cốt lõi của giảng viên giảng dạy tiếng Anh.
  + Tính thực tiễn của chỉ số: Các chỉ số hiệu suất được tinh chỉnh để phản ánh thực tế lớp học, bao gồm các bổ sung cho các loại đánh giá chẩn đoán, hình thành và tổng kết.

Những hiểu biết từ nhóm tập trung đã xác nhận tính phù hợp của thang đo năng lực kiểm tra đánh giá ngôn ngữ và giúp cải thiện để sử dụng hiệu quả hơn cho giảng viên giảng dạy tiếng Anh Việt Nam.

**6.2. Hội đồng chuyên gia để xác thực**

Các đánh giá của chuyên gia được thực hiện trong hai vòng, tập trung vào tính rõ ràng, tính liên quan và tính đại diện của các mục.

***6.2.1. Vòng đầu tiên của đánh giá chuyên gia***

Vòng này cung cấp thông tin chi tiết về một số khía cạnh xác thực, chẳng hạn như tính dễ hiểu, tính đại diện, cấu trúc, căn chỉnh bên ngoài, tính ứng dụng và tính tổng quát.

* Tính dễ hiểu: Các chuyên gia thấy hầu hết các mô tả rõ ràng, mặc dù họ khuyến nghị bổ sung thêm các ví dụ thực tiễn để hỗ trợ hiểu các mục.
* Tính liên quan và tính đại diện: Các chuyên gia xác thực rằng các nhóm năng lực bao phủ đầy đủ các thành phần thiết yếu của năng lực kiểm tra đánh giá ngôn ngữ, nhưng nhấn mạnh sự cần thiết của kỹ năng số và thực hành đánh giá hình thành.
* Khía cạnh cấu trúc: Các tinh chỉnh được đề xuất bao gồm làm rõ mối quan hệ giữa các chỉ số và tinh gọn ngôn ngữ để tránh sự dư thừa.
* Cấu trúc bên ngoài: Các chuyên gia đề xuất căn chỉnh thuật ngữ với các tiêu chuẩn quốc tế và điều chỉnh nội dung để phản ánh các yêu cầu đặc thù của Việt Nam.
* Tính ứng dụng: Các đề xuất làm rõ các thuật ngữ mơ hồ và đảm bảo chỉ số hiệu suất quan sát được đã tăng cường tính thực tiễn cho giảng viên.
* Tính tổng quát: Các chuyên gia nhấn mạnh nhu cầu phân biệt năng lực để phù hợp với giảng viên có trình độ và bối cảnh tổ chức khác nhau, nhấn mạnh khả năng thích ứng cho cả kiến thức lý thuyết và thực tiễn.

***6.2.2. Vòng thứ hai của đánh giá chuyên gia***

Các cải tiến tiếp theo tập trung vào tính rõ ràng, tính khả dụng và sự sẵn sàng cho thử nghiệm thực địa.

* Tính rõ ràng và cụ thể: Các chỉnh sửa làm rõ các chỉ số, đặc biệt là các thuật ngữ liên quan đến các thành phần năng lực ngôn ngữ, tăng cường tính thực tiễn.
* Luồng logic và tính khả dụng: Các năng lực được sắp xếp lại theo trình tự logic, bắt đầu từ các kỹ năng nền tảng, giúp dễ dàng điều hướng.
* Độ sâu và tính đại diện: Các chỉ số phụ bổ sung làm tăng độ sâu, trong khi các chuyên gia cảnh báo chống lại các cấu trúc quá phức tạp có thể làm giảm tính thực tiễn.
* Sự nhất quán về thuật ngữ: Thuật ngữ được chuẩn hóa, căn chỉnh với các khung quốc tế, cải thiện tính rõ ràng và áp dụng giữa các bối cảnh.
* Sự sẵn sàng cho thử nghiệm thực địa: Công cụ được đánh giá là sẵn sàng cho triển khai rộng rãi hơn, với các chỉnh sửa nhỏ được đề xuất để tăng cường tính áp dụng trong thực tế.

**6.3. Đánh giá tổng thể về suy luận mô tả miền**

Việc xác thực suy luận mô tả miền dựa trên hai giả định chính, mỗi giả định đều được hỗ trợ bởi bằng chứng từ phản hồi của chuyên gia và dữ liệu thực nghiệm.

* Đại diện cho kỹ năng và kiến thức: Bằng chứng từ các tài liệu tham khảo và ý kiến chuyên gia xác nhận rằng thang đo năng lực kiểm tra đánh giá ngôn ngữ đại diện hiệu quả cho các năng lực năng lực kiểm tra đánh giá ngôn ngữ thiết yếu cần thiết cho giảng viên giảng dạy tiếng Anh Việt Nam. Các giả định về kiến thức và kỹ năng năng lực kiểm tra đánh giá ngôn ngữ cần thiết đã được chứng minh, khẳng định nền tảng của thang đo dựa trên nghiên cứu thực nghiệm.
* Các mục là chỉ báo của các chiều năng lực kiểm tra đánh giá ngôn ngữ: Các đánh giá lặp đi lặp lại từ chuyên gia và các cuộc thảo luận nhóm tập trung đã tinh chỉnh các mục để đảm bảo chúng phản ánh chính xác từng miền năng lực kiểm tra đánh giá ngôn ngữ. Quá trình này chứng minh rằng các mục trong thang đo nắm bắt hiệu quả các khía cạnh đa chiều của năng lực kiểm tra đánh giá ngôn ngữ một cách phù hợp với bối cảnh.

Những phát hiện này xác thực suy luận mô tả miền, khẳng định rằng thang đo năng lực kiểm tra đánh giá ngôn ngữ là một công cụ mạnh mẽ, thích nghi với văn hóa để đo lường kiến thức đánh giá ngôn ngữ của giảng viên giảng dạy tiếng Anh tại Việt Nam.

Chương này đã trình bày quy trình xác thực cho suy luận mô tả miền trong thang đo năng lực kiểm tra đánh giá ngôn ngữ, thiết lập một nền tảng vững chắc cho việc đánh giá năng lực kiểm tra đánh giá ngôn ngữ của giảng viên giảng dạy tiếng Anh Việt Nam. Phản hồi từ chuyên gia và các cuộc thảo luận nhóm tập trung đã cho phép cải tiến lặp đi lặp lại, nâng cao tính dễ hiểu, tính đại diện và tính phù hợp với văn hóa của thang đo. Suy luận mô tả miền được hỗ trợ đầy đủ, khẳng định rằng thang đo năng lực kiểm tra đánh giá ngôn ngữ phù hợp với nhu cầu giáo dục cụ thể của giảng viên giảng dạy tiếng Anh Việt Nam. Việc xác thực thêm thông qua thử nghiệm thực địa sẽ cung cấp thêm thông tin về độ tin cậy và khả năng áp dụng của thang đo, với các chương tiếp theo sẽ khám phá các suy luận đánh giá và tổng quát hóa.

**CHƯƠNG 7: KẾT QUẢ XÁC THỰC THANG ĐO KIẾN THỨC ĐÁNH GIÁ NGÔN NGỮ CỦA GIẢNG VIÊN GIẢNG DẠY TIẾNG ANH TẠI VIỆT NAM - SUY LUẬN ĐÁNH GIÁ VÀ TỔNG QUÁT HÓA**

Chương này trình bày các kết quả hỗ trợ cho suy luận đánh giá và tổng quát hóa trong khung xác thực của Thang đo Kiến thức Đánh giá Ngôn ngữ (năng lực kiểm tra đánh giá ngôn ngữS), nhằm đánh giá các năng lực đánh giá ngôn ngữ của giảng viên giảng dạy tiếng Anh Việt Nam.

**7.1. Suy luận đánh giá**

***7.1.1. Bằng chứng về tính hợp lệ 1: Quá trình xây dựng thang đo và đồng thuận của chuyên gia***

Thang đo năng lực kiểm tra đánh giá ngôn ngữ được phát triển thông qua một quy trình lặp đi lặp lại, có sự tham gia của các chuyên gia để đảm bảo phù hợp với các năng lực mục tiêu. Các bước chính trong quá trình phát triển thang đo bao gồm:

* Nền tảng lý thuyết: thang đo năng lực kiểm tra đánh giá ngôn ngữ được xây dựng dựa trên tài liệu phong phú, tổng hợp các năng lực cần thiết cho đánh giá ngôn ngữ từ các mô hình lý thuyết đã được công nhận.
* Đánh giá và tinh chỉnh bởi chuyên gia: Các nhóm tập trung đã cung cấp phản hồi về tính phù hợp với nhu cầu của giảng viên giảng dạy tiếng Anh Việt Nam, đảm bảo tính liên quan, rõ ràng và tính đại diện của các mục.
* Tài liệu đồng thuận: Các thay đổi mục, lý do điều chỉnh và thang đo cuối cùng gồm 42 mục được xây dựng dựa trên khuyến nghị của chuyên gia, được trình bày trong Bảng 6.5.

***7.1.2. Bằng chứng về tính hợp lệ 2: Các đặc điểm tâm lý***

Đánh giá tâm lý học bao gồm một loạt các phân tích, đánh giá 177 câu trả lời hợp lệ thông qua Lý thuyết Kiểm tra Cổ điển (CTT) và các mô hình Rasch. Ba mươi sáu câu trả lời đã bị loại do không đủ điều kiện.

* Ngưỡng Thurstonian: Các mục thể hiện các ước lượng ngưỡng phù hợp (dao động từ -1,71 đến +3,15 logits), xác nhận tính phù hợp cho việc đo lường các mức độ năng lực khác nhau.
* Tương quan mục-tổng: Hầu hết các mục cho thấy khả năng phân biệt tốt, với các chỉ số trên 0,4, cho thấy khả năng phân biệt hiệu quả giữa các năng lực của người trả lời.
* Độ phù hợp mục và độ tin cậy: Các giá trị INFIT mean square dao động từ 0,77 đến 1,25, cho thấy sự nhất quán của các mục với kỳ vọng của mô hình Rasch.
* Tương quan mục-tổng: Tương quan mục-tổng dao động từ 0,29 đến 0,62, cho thấy khả năng phân biệt phù hợp. Các mục có tương quan thấp hơn (ví dụ: Q214 ở mức 0,29) được lưu ý để xem xét chỉnh sửa, mặc dù hầu hết các mục đóng góp có ý nghĩa cho cấu trúc của thang đo.
* Phân tích Rasch: Phân tích Rasch cho các giá trị MNSQ trọng số nằm trong khoảng 0,6 đến 1,4, với INFIT trung bình ở mức 1,02, xác nhận sự phù hợp của hầu hết các mục với mô hình Rasch. Một vài mục, như Q213 và Q215, cho thấy sự không phù hợp nhẹ nhưng vẫn nằm trong giới hạn chấp nhận được.

**7.2. Kết quả tính hợp lệ của suy luận tổng quát hóa**

Suy luận này xem xét liệu điểm số thang đo năng lực kiểm tra đánh giá ngôn ngữ có nhất quán qua các khu vực, giới tính và cấp độ kinh nghiệm hay không. Các chỉ số độ tin cậy cao và phân tách mục hỗ trợ tính mạnh mẽ của thang đo năng lực kiểm tra đánh giá ngôn ngữ qua các mẫu.

***7.2.1. Bằng chứng về tính hợp lệ 1: Thống kê độ phù hợp MNSQ trọng số qua các mẫu***

* Khu vực: Phần lớn các mục cho thấy các giá trị MNSQ nằm trong phạm vi chấp nhận được. Sự không phù hợp nhẹ ở các khu vực Trung và Nam cho thấy một số biến thể khu vực nhỏ.
* Giới tính: Các mục hoạt động tốt cho cả hai giới, tuy nhiên Mục 27 cho thấy mức độ không phù hợp cao hơn trong mẫu nam.
* Cấp độ kinh nghiệm: Các mục nhất quán trên các nhóm kinh nghiệm, mặc dù các mục như Q215 được lưu ý để xem xét trong các nhóm kinh nghiệm cụ thể (6-10 năm).

Độ phù hợp trọng số của các mục theo Khu vực, Giới tính và Nhóm Kinh nghiệm cho thang đo năng lực kiểm tra đánh giá ngôn ngữ. Các mục 4, 5 và 27 cho thấy sự không phù hợp ở một số nhóm do sự không liên quan hoặc các nhiệm vụ mang tính logistic hơn là dựa trên năng lực.

***7.2.2. Bằng chứng về tính hợp lệ 2: Phân tích độ tin cậy và phân tách mục (Sau khi điều chỉnh thang đo)***

* Độ tin cậy phân tách mục: Hệ số cao (r = 0,892) xác nhận sự phân bố mục qua các cấp độ khó.
* Độ tin cậy EAP/PV: Một chỉ số ở mức r = 0,914 cho thấy khả năng đo lường năng lực ổn định qua các mẫu.
* Hệ số Cronbach’s Alpha: Hệ số alpha cao là 0,907 càng củng cố tính mạch lạc của các mục.

***7.2.3. Đánh giá tổng thể về suy luận tổng quát hóa***

Tất cả các mục nhìn chung hỗ trợ cho suy luận tổng quát hóa trên các nhóm, chỉ cần một số điều chỉnh nhỏ.

Chương này đã xác nhận tính hợp lệ của thang đo năng lực kiểm tra đánh giá ngôn ngữ thông qua việc phát triển có sự tham gia của chuyên gia, các đặc điểm tâm lý và tính nhất quán trên các nhóm đa dạng. Chương tiếp theo sẽ giải quyết suy luận giải thích, tích hợp các kết quả từ các suy luận trước đó.

**CHƯƠNG 8: KẾT QUẢ XÁC THỰC THANG ĐO KIẾN THỨC ĐÁNH GIÁ NGÔN NGỮ CỦA GIẢNG VIÊN GIẢNG DẠY TIẾNG ANH TẠI VIỆT NAM - SUY LUẬN GIẢI THÍCH**

Chương này giải quyết suy luận giải thích, xem xét liệu điểm số của giảng viên giảng dạy tiếng Anh Việt Nam có phù hợp với các cấu trúc lý thuyết của thang đo năng lực kiểm tra đánh giá ngôn ngữ hay không. Phân tích suy luận này dựa trên hai giả định: thứ nhất, rằng các mục của thang đo phản ánh chính xác các thành phần dự định, được xác minh qua phân tích tâm lý học; và thứ hai, rằng thang đo phân biệt được các mức độ năng lực, được đánh giá thông qua phân tích Rasch và bản đồ Wright.

**8.1. Bằng chứng về tính hợp lệ 1: Diễn giải bản đồ biến số**

Bản đồ biến số minh họa sự liên kết giữa khả năng của giảng viên với các thông số độ khó của mục, đánh giá liệu độ khó của các mục có phù hợp với năng lực của giảng viên hay không. Bản đồ hiển thị:

* Biểu diễn trên thang logit: Cả khả năng của giảng viên và độ khó của các mục được lập bản đồ trên thang đo logit, với năng lực trải dài từ -2 đến +3 logits cho giảng viên và -1,74 đến +3,20 logits cho các mục. Sự khớp gần giữa trung bình của mục và trung bình của người (trung bình mục = 0; trung bình người = 0,21) cho thấy các mục phù hợp với năng lực của giảng viên.
* Thứ bậc hành vi: Các mục nằm ở phần thấp của bản đồ đại diện cho các năng lực mới bắt đầu, trong khi các mục cao hơn phản ánh năng lực thành thạo, như được thấy trong bản đồ Wright của thang đo năng lực kiểm tra đánh giá ngôn ngữ. Bản đồ hỗ trợ tính chất xác suất của các năng lực, nơi những người thực hiện tốt hơn có khả năng thể hiện các năng lực nền tảng, với các mục được xác nhận bởi mô hình tín chỉ một phần.

Bản chất phân cấp, phát triển của bản đồ phù hợp với nền tảng lý thuyết của thang đo năng lực kiểm tra đánh giá ngôn ngữ, mô tả tiến trình của giảng viên từ các năng lực mới bắt đầu đến thành thạo, với các mô tả chi tiết về trình độ thành thạo cho từng cấp độ được cung cấp trong Bảng 8.2 đến 8.5.

**8.2. Bằng chứng về tính hợp lệ 2: So sánh giữa tiếp tục thực nghiệm và dự kiến**

Việc so sánh giữa tiếp tục thực nghiệm và dự kiến hỗ trợ khả năng diễn giải bằng cách xác nhận sự phù hợp với các cấp độ lý thuyết (từ mới bắt đầu đến thành thạo). Các điều chỉnh bao gồm:

* Định vị lại năng lực đã chỉnh sửa: Các mục như 33.1 (sự tham gia của các bên liên quan) được nâng lên để nhấn mạnh các năng lực tiên tiến, trong khi các mục nền tảng như 3.2 và 25.2 được điều chỉnh lại về cấp độ thấp hơn, phản ánh sự phù hợp nền tảng.
* Nhấn mạnh thành thạo cao cấp: Các mục 23.2 (tính nhất quán trong mô tả trình độ thực hiện), 11.2 (xác định các khu vực cần cải thiện), và 35.2 (cải thiện độ tin cậy thông qua đào tạo người chấm điểm) được nâng lên cấp độ 4, nhấn mạnh vai trò quan trọng của chúng trong thực tiễn đánh giá cao cấp.
* Thách thức tích hợp công nghệ: Mục 19.2 (phương pháp số trong đánh giá) được định vị là thách thức cao nhất, thừa nhận những khó khăn thích ứng với công nghệ đang diễn ra trong các thực tiễn đánh giá.

**8.3. Đánh giá tổng thể về suy luận giải thích**

Như đã chỉ ra trong Bảng 8.8, cả hai giả định của suy luận giải thích rằng các mục đại diện cho các thành phần mục tiêu và rằng điểm thang đo thay đổi theo mức độ năng lực đều được hỗ trợ. Phân tích thống kê xác nhận sự phù hợp của các mục thang đo với cấu trúc lý thuyết, và sự biến đổi điểm số qua các khu vực và cấp độ kinh nghiệm phản ánh tính hợp lệ của cấu trúc thang đo năng lực kiểm tra đánh giá ngôn ngữ, ghi nhận các mức độ thành thạo khác nhau.

Chương này đã xác thực suy luận giải thích, khẳng định tính hợp lệ cấu trúc của thang đo năng lực kiểm tra đánh giá ngôn ngữ và khả năng phân biệt giữa các mức độ thành thạo. Các phân tích tâm lý học và nội dung, cùng với mô hình Rasch, cho thấy rằng thang đo năng lực kiểm tra đánh giá ngôn ngữ đại diện chính xác cho các năng lực đánh giá ngôn ngữ và phân biệt năng lực của giảng viên. Tiến trình của thang đo năng lực kiểm tra đánh giá ngôn ngữ phù hợp với mô hình kỹ năng của Dreyfus và Dreyfus, xác nhận việc đo lường đáng tin cậy kiến thức đánh giá ngôn ngữ của giảng viên giảng dạy tiếng Anh Việt Nam từ cấp độ mới bắt đầu đến thành thạo. Chương tiếp theo sẽ tổng hợp các kết quả xác thực này trong đánh giá khung tổng thể.

**CHƯƠNG 9: KẾT LUẬN VÀ KIẾN NGHỊ**

Chương cuối tóm tắt quá trình phát triển và xác thực thang đo năng lực kiểm tra, đánh giá ngôn ngữ cho giảng viên tiếng Anh tại Việt Nam, tổng hợp các phát hiện chính từ các chương trước. Nó đánh giá các điểm mạnh và hạn chế của nghiên cứu, đồng thời đề xuất những khuyến nghị thực tiễn để cải thiện công tác đào tạo và hỗ trợ giáo viên. Cuối cùng, chương nhấn mạnh tác động của nghiên cứu đối với việc nâng cao năng lực đánh giá ngôn ngữ và đưa ra hướng nghiên cứu trong tương lai.

**9.1. Tóm tắt kết quả nghiên cứu và lập luận xác trị**

***9.1.1. Thang đo đã phát triển***

Thang đo Kiến thức Đánh giá Ngôn ngữ (LALS) cho giảng viên giảng dạy tiếng Anh Việt Nam đã được phát triển và xác thực để đáp ứng các nhu cầu giáo dục đặc thù của Việt Nam. Thang đo đánh giá các năng lực của giáo viên qua ba giai đoạn: chuẩn bị, thực hiện và điều chỉnh đánh giá, bao gồm các kỹ năng thiết yếu từ việc hiểu các khái niệm cơ bản đến việc cải thiện các nhiệm vụ đánh giá. Dựa trên mô hình tiếp thu kỹ năng của Dreyfus & Dreyfus (1986), cách tiếp cận dựa trên năng lực của thang đo này theo dõi sự phát triển của giáo viên từ mức độ mới bắt đầu đến chuyên gia, cung cấp các hành vi có thể quan sát ở mỗi giai đoạn. Thang đo gồm 39 mục, mỗi mục có mô tả qua bốn cấp độ thành thạo, mang đến một khung toàn diện để giáo viên tự đánh giá năng lực kiểm tra đánh ngôn ngữ của mình.

***9.1.2. Điểm khác biệt của thang đo năng lực kiểm tra đánh giá của Việt Nam***

Thang đo của Việt Nam được thiết kế độc đáo cho bối cảnh giáo dục của Việt Nam, tích hợp tính phù hợp về văn hóa, sự điều chỉnh chính sách và tính nhạy cảm về nguồn lực. Nó phù hợp với các hướng dẫn của Bộ GDĐT, KNLNNVN và NFLP, đảm bảo sự gắn kết với các mục tiêu quốc gia.

Thang đo này giải quyết các hạn chế của địa phương như quy mô lớp học lớn và công nghệ hạn chế, thúc đẩy các đánh giá thực tế và thực hành đạo đức phù hợp với Việt Nam. Mô hình phát triển của nó hỗ trợ các giảng viên tiếng Anh trong việc hoàn thiện năng lực và bồi dưỡng năng lực liên văn hóa.

Bằng cách cân bằng giữa hiện đại hóa với tính khả thi và đảm bảo chất lượng thông qua xác nhận tâm lý, thang đo này cung cấp một khuôn khổ thực tế, bản địa hóa cho các nhà giáo dục Việt Nam trong bối cảnh nguồn lực hạn chế.

**9.2. Bằng chứng về tính hợp lệ của thang đo: Tóm tắt**

***9.2.1. Suy luận mô tả miền***

Các mục trong thang đo năng lực kiểm tra đánh giá ngôn ngữ đại diện chính xác cho các năng lực LAL thiết yếu cần thiết cho giảng viên giảng dạy tiếng Anh Việt Nam, được xác nhận qua phản hồi của chuyên gia, tính liên quan của nội dung và sự phù hợp với các lý thuyết năng lực kiểm tra đánh giá ngôn ngữ đã được thiết lập. Các lần tham vấn chuyên gia lặp đi lặp lại đã tinh chỉnh các mục để đảm bảo tính phù hợp thực tiễn trong bối cảnh Việt Nam, đáp ứng các nhu cầu cụ thể về kiến thức đánh giá ngôn ngữ.

***9.2.2. Suy luận đánh giá***

Thang đo năng lực kiểm tra đánh giá ngôn ngữ thể hiện các đặc điểm tâm lý mạnh mẽ, xác nhận khả năng đo lường đáng tin cậy. Phân tích Rasch xác thực các mức độ khó của mục, các chỉ số phân biệt và độ nhất quán nội bộ. Thống kê độ phù hợp và ước lượng phân biệt mục-tổng khẳng định khả năng của thang đo trong việc đo lường năng lực kiểm tra đánh giá ngôn ngữ một cách nhất quán, với độ tin cậy phân tách mục cao (r=0,92), độ tin cậy phân tách người ở mức trung bình (r=0,68), và chỉ số độ tin cậy EAP/PV là 0,70.

***9.2.3. Suy luận tổng quát hóa***

Điểm số thang đo năng lực kiểm tra đánh giá ngôn ngữ có thể tổng quát hóa cho giảng viên giảng dạy tiếng Anh Việt Nam với các nền tảng nhân khẩu học và cấp độ kinh nghiệm khác nhau. Sự nhất quán trong hiệu suất mục qua các nhóm này hỗ trợ tính thích ứng của thang đo. Mặc dù sự khác biệt khu vực chưa được đại diện đầy đủ do khó khăn trong lấy mẫu, các thống kê độ phù hợp và độ tin cậy phân tách mục cho thấy hiệu quả của thang đo trên các nhóm phụ.

***9.2.4. Suy luận giải thích***

Thang đo năng lực kiểm tra đánh giá ngôn ngữ phản ánh hiệu quả cấu trúc lý thuyết của năng lực kiểm tra đánh giá ngôn ngữ, phân biệt độ thành thạo một cách đáng tin cậy. Mô hình Rasch và bản đồ Wright xác nhận sự phù hợp mạnh mẽ giữa khả năng của giáo viên và độ khó của các mục. Các đặc điểm tâm lý học xác thực khả năng của thang đo trong việc nắm bắt tiến trình năng lực của giáo viên, hỗ trợ tính hợp lệ cấu trúc trong việc phân biệt các mức độ thành thạo của giáo viên.

**9.3. Đóng góp của đề tài**

Những phát hiện của nghiên cứu này có ý nghĩa quan trọng đối với lý thuyết, phương pháp luận và sư phạm trong đánh giá ngôn ngữ trong bối cảnh giảng dạy tiếng Anh Việt Nam.

***9.3.1. Đóng góp về lý thuyết***

Cách tiếp cận dựa trên năng lực của nghiên cứu này đối với năng lực kiểm tra đánh giá ngôn ngữ xác định năng lực kiểm tra đánh giá ngôn ngữ là một tập hợp kỹ năng động và phát triển thay vì một cấu trúc tĩnh. Việc bản địa hóa năng lực kiểm tra đánh giá ngôn ngữ trong bối cảnh Việt Nam đã tiết lộ các năng lực độc đáo, chẳng hạn như việc điều chỉnh thực tiễn đánh giá để phù hợp với chính sách địa phương. Các kết quả nhấn mạnh tầm quan trọng của phát triển chuyên môn liên tục trong năng lực kiểm tra đánh giá ngôn ngữ, kết hợp cả các nguyên tắc đánh giá truyền thống và các sáng kiến kỹ thuật số.

***9.3.2. Đóng góp về phương pháp luận***

Nghiên cứu này đóng góp vào các phương pháp đánh giá phát triển và phát triển chuyên môn. Áp dụng mô hình tiếp thu kỹ năng của Dreyfus, nghiên cứu này phác thảo các năng lực năng lực kiểm tra đánh giá ngôn ngữ tiến bộ, từ người mới bắt đầu đến chuyên gia, phù hợp với các ứng dụng thực tế trong giáo dục giảng dạy tiếng Anh tại Việt Nam. Việc tích hợp mô hình Rasch cung cấp những hiểu biết thực nghiệm về tính hợp lệ của thang đo, củng cố tiềm năng ứng dụng của nó trong các lĩnh vực khác trong giáo dục.

***9.3.3. Đóng góp về sư phạm***

*9.3.3.1. Hỗ trợ tự đánh giá và thực hành phản ánh*

Thang đo năng lực kiểm tra đánh giá ngôn ngữ cho phép giảng viên giảng dạy tiếng Anh Việt Nam tự đánh giá năng lực của mình, thúc đẩy thực hành phản ánh nhằm nâng cao tính minh bạch trong đánh giá và các đánh giá tập trung vào sinh viên.

*9.3.3.2. Định hướng phát triển chuyên môn*

Khung thang đo năng lực kiểm tra đánh giá ngôn ngữ có thể được tích hợp vào các chương trình phát triển giáo viên, cung cấp một lộ trình cho tiến trình kỹ năng. Các tiêu chí rõ ràng của thang đo cho phép các chiến lược cải thiện mục tiêu phù hợp với nhu cầu đánh giá ngôn ngữ đang phát triển.

*9.3.3.3. Nâng cao chương trình đào tạo giáo viên*

Thang đo năng lực kiểm tra đánh giá ngôn ngữ có tiềm năng cho các chương trình đào tạo giáo viên trước và trong khi giảng dạy, hỗ trợ tư vấn cá nhân hóa, phân bổ tài nguyên và phát triển liên tục qua các cấp độ năng lực khác nhau.

**9.4. Hạn chế của nghiên cứu**

Việc lấy mẫu theo địa lý của nghiên cứu giới hạn tính tổng quát của nó trên toàn Việt Nam. Phần lớn người tham gia là từ miền Bắc Việt Nam do rào cản thu thập dữ liệu, điều này có thể giới hạn khả năng áp dụng các phát hiện cho các khu vực khác. Việc sử dụng các biện pháp tự báo cáo, mặc dù có giá trị và độ tin cậy, có thể không phản ánh đầy đủ năng lực kiểm tra đánh giá ngôn ngữ của giáo viên trong thực tế, kêu gọi nghiên cứu tương lai kết hợp quan sát và đánh giá bên ngoài để có cái nhìn toàn diện hơn.

**9.5. Các lĩnh vực cần nghiên cứu tiếp tục**

Các nghiên cứu tiếp theo nên khám phá ứng dụng của năng lực kiểm tra đánh giá ngôn ngữ trong các bối cảnh giáo dục đa dạng và xem xét ảnh hưởng của các yếu tố ngữ cảnh, chẳng hạn như nền tảng giáo viên và chính sách khu vực. Việc tiếp tục xác thực thông qua mô hình Rasch và việc bổ sung các nguồn dữ liệu khác sẽ tăng cường độ tin cậy của thang đo, với cơ hội mở rộng khung này sang các lĩnh vực khác.

**9.6. Hướng dẫn sử dụng thang đo năng lực kiểm tra đánh giá ngôn ngữ cho giảng viên giảng dạy tiếng Anh Việt Nam**

Thang đo năng lực kiểm tra đánh giá ngôn ngữ cung cấp cho giảng viên giảng dạy tiếng Anh Việt Nam một công cụ phát triển để đánh giá và nâng cao kỹ năng đánh giá ngôn ngữ của họ, với Mức 2 (Sơ khởi) là mục tiêu thành thạo tối thiểu ở các giai đoạn đánh giá. Các cơ sở giáo dục được khuyến khích tích hợp thang đo này vào các chương trình trước và trong khi giảng dạy, khuyến khích tự đánh giá thường xuyên và phát triển chuyên môn phù hợp với các tiêu chuẩn giáo dục đang thay đổi.

|  |  |  |
| --- | --- | --- |
|  | **DANH MỤC CÁC BÀI BÁO, CÔNG TRÌNH KHOA HỌC**  **ĐÃ CÔNG BỐ CÓ LIÊN QUAN ĐẾN LUẬN ÁN** |  |
| 1. | **Nguyen Thi To Loan** (2021). *Conceptualizing language teachers’ language assessment literacy: A review of literature*. Proceedings of 2021 International Graduate Research Symposium (IGRS), pp. 898-906. Vietnam National University Press, Hanoi. |  |
| 2. | **Nguyễn Thị Tố Loan** (2022). *Năng lực khảo thí ngôn ngữ của giáo viên dạy ngoại ngữ: Những khoảng trống trong nghiên cứu*. Kỷ yếu Hội thảo Ngữ học toàn quốc 2022 - Ngôn ngữ học Ứng dụng trong Xu hướng Hội nhập Quốc tế, tr. 952-959. Nhà xuất bản Đại học Quốc gia Hà Nội, Hà Nội. |  |
| 3. | **Nguyen Thi To Loan**, Duong Thu Mai (2023). Assessment literacy in language education: the story so far. *VNU Journal of Foreign Studies*, 39(4), pp. 59-72. |  |
| 4. | **Nguyen Thi To Loan**, Vu Thi Quynh Dung (2023). *Defining the language assessment literacy framework required for EFL language teachers*. Proceedings of the 11th OpenTESOL International Conference 2023, pp. 229-255. Ho Chi Minh City Open University. |  |
| 5. | Cutri, J., Yip, S.Y., **Nguyen, T. T. L.** (2024). Global Teachers on the Move: A Discussion of Teacher Identity and Intercultural Pedagogical Practices in Cross-cultural Contexts. In: Chowdhury, R., Tuan, H. A. (Eds), *Engaging with Australasia: Comparative Research on ELT and English Teacher Education* (pp. 379-403). Palgrave Macmillan, Singapore. |  |
| 6. | **Nguyen Thi To Loan**, Ta Thi Mai Huong, Ngo Thi Thanh and Ngoc-Tung Vu (2024). Vietnamese EFL Teachers’ Conceptions of Assessment: A Model to Navigate Accountability in the Digital Transformation of Education. In: Yen Hoang Phuong, Thanh-Thao Le (Eds.), *Exploring Teacher Beliefs: Insights and Perspectives* (pp. 235-258). Nova Science Publishers, New York, USA. |  |
| 7. | **Nguyen Thi To Loan**, Duong Thu Mai (2024). *Exploring EFL teachers’ assessment literacy in digital educational contexts: Opportunities and challenges*. Proceedings of the 12th OpenTESOL International Conference 2024, pp. 464-476. Ho Chi Minh City Open University. |  |