

ĐÁNH GIÁ CHẤT LƯỢNG ĐỀ THI BẰNG CÁC CHỈ SỐ ĐO LƯỜNG VÀ GIẢI PHÁP NÂNG CAO HIỆU QUẢ CÔNG TÁC KHẢO THÍ TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP QUẢNG NINH

Nguyễn Thu Hiền^{1,*}, Nguyễn Thị Phương¹

¹Trường Đại học Công nghiệp Quảng Ninh

*Email: hientn123@gmail.com

TÓM TẮT

Nghiên cứu nhằm xây dựng quy trình đánh giá chất lượng đề thi dựa trên chỉ số định lượng toán học, góp phần nâng cao tính khách quan và hiệu quả trong công tác phân tích đề thi sau mỗi kỳ thi kết thúc học phần. Các chỉ số được sử dụng gồm độ khó (P), độ phân biệt (D) giúp đánh giá mức độ phù hợp của từng câu hỏi trong việc phân loại năng lực người học. Phương pháp nghiên cứu dựa trên dữ liệu kết quả thi kết thúc học phần, áp dụng thuật toán phân nhóm thí sinh theo bách phân vị (Upper 27%, Lower 27%) để tính toán độ phân biệt, đồng thời ứng dụng hiệu quả công cụ Excel nhằm tự động hoá quy trình xử lý dữ liệu. Kết quả nghiên cứu cho thấy các chỉ số P và D giúp nhận diện chính xác những câu hỏi có độ khó hoặc độ phân biệt không đạt yêu cầu, từ đó đưa ra các khuyến nghị giữ lại, chỉnh sửa hoặc loại bỏ. Nghiên cứu khẳng định việc áp dụng các chỉ số toán học đơn giản nhưng hợp lý giúp tiết kiệm thời gian, giảm thiểu sai sót chủ quan, nâng cao chất lượng ngân hàng câu hỏi và hiệu quả công tác khảo thí tại trường Đại học Công nghiệp Quảng Ninh.

Từ khóa: Đánh giá đề thi, Độ khó, Độ phân biệt, bách phân vị, ngân hàng câu hỏi.

1. ĐẶT VẤN ĐỀ

Trong bối cảnh giáo dục đại học đổi mới theo hướng chuẩn đầu ra và kiểm định chất lượng, công tác kiểm tra, đánh giá kết quả học tập đóng vai trò then chốt trong việc phản ánh năng lực thực tế của người học nhằm đáp ứng theo chuẩn đầu ra được công bố. Với định hướng phát triển chương trình đào tạo theo hướng ứng dụng và hội nhập, trường Đại học Công nghiệp Quảng Ninh (ĐHCNQN) xác định nâng cao chất lượng khảo thí là nhiệm vụ trọng tâm, trong đó công tác xây dựng và đánh giá đề thi học phần được đặc biệt quan tâm. Tuy nhiên, thực tế cho thấy hoạt động đánh giá chất lượng đề thi tại trường hiện chủ yếu dựa vào nhận định chủ quan và các thống kê đơn giản (điểm trung bình, tỷ lệ đạt...) mà chưa đi sâu vào phân tích chất lượng từng câu hỏi. Điều này dẫn đến những hạn chế khi một số câu hỏi quá dễ, quá khó, hoặc thiếu khả năng phân loại năng lực người học.

Để khắc phục, việc ứng dụng các chỉ số thống kê như độ khó (P), độ phân biệt (D) là cần

thiết nhằm đánh giá mức độ phù hợp của từng câu hỏi trong đề thi [1], [2], [3]. Bên cạnh đó, bách phân vị được sử dụng để phân tích sâu hơn kết quả thi, so sánh nhóm sinh viên có kết quả cao và nhóm sinh viên có kết quả thấp nhằm đánh giá mức độ phân hoá của từng câu hỏi. Đồng thời, việc tự động hoá quy trình đánh giá thông qua các thuật toán toán học không những nâng cao tính khách quan, tiết kiệm thời gian và hỗ trợ hiệu quả trong công tác khảo thí. Trên cơ sở đó, nghiên cứu này hướng đến xây dựng quy trình đánh giá chất lượng đề thi kết thúc học phần dựa trên các chỉ số định lượng toán học, kết hợp với các giải pháp công nghệ và tổ chức phù hợp với điều kiện thực tiễn của Nhà trường. Qua đó, từng bước chuẩn hoá công tác kiểm tra, đánh giá, nâng cao chất lượng ngân hàng đề thi và đáp ứng phù hợp với yêu cầu của chuẩn đầu ra chương trình đào tạo, chương trình dạy học và kiểm định chất lượng giáo dục.

2. CƠ SỞ LÝ LUẬN

2.1. Chất lượng đề thi và yêu cầu đánh giá

Chất lượng đề thi được hiểu là mức độ đáp ứng các yêu cầu về nội dung kiến thức, hình thức trình bày và đặc biệt là khả năng phân hóa năng lực người học một cách khách quan, chính xác. Một đề thi đạt chất lượng cần đáp ứng yêu cầu về nội dung, hình thức và khả năng phân hóa, phù hợp với chuẩn đầu ra và bao phủ các mức độ tư duy theo ma trận, có độ khó hợp lý và phản ánh đúng năng lực người học. Ngoài ra, kết quả phân tích chất lượng đề thi còn là căn cứ quan trọng để cải tiến nội dung và phương pháp giảng dạy [2], [4].

2.2. Lý thuyết trắc nghiệm cổ điển (Classical Test Theory - CTT) [3], [4].

CTT là lý thuyết nền tảng được ứng dụng rộng rãi trong đo lường giáo dục, cung cấp các chỉ số cơ bản nhưng có tính ứng dụng cao để đánh giá chất lượng từng câu hỏi trong đề thi, trong đó quan trọng nhất là độ khó (P) và độ phân biệt (D).

2.2.1. Độ khó (P - Difficulty Index) [2], [3].

+ **Đối với hình thức thi trắc nghiệm:** Độ khó của câu hỏi/ đề thi được đo bằng tỉ số giữa số người học trả lời đúng và tổng số người học tham gia trả lời.

$$P = \frac{n}{N} \times 100, \% \quad (1)$$

Trong đó: P: độ khó của câu hỏi/ đề thi, %

n: số người học trả lời đúng;

N: Tổng số người học tham gia trả lời.

+ **Đối với các hình thức thi khác:** Độ khó của đề thi là tỉ số giữa điểm trung bình của tất cả sinh viên tham dự so với điểm tối đa mà sinh viên đó có thể đạt được.

$$P = \frac{\bar{X}}{M} \times 100, \% \quad (2)$$

Trong đó: P: độ khó của câu hỏi/ đề thi, %

$\bar{X}$: điểm trung bình;

M: điểm tối đa.

Giá trị P càng cao, câu hỏi càng dễ; P càng thấp, câu hỏi càng khó. Mức P lý tưởng thường nằm trong khoảng 20% – 80% để đảm bảo câu hỏi có khả năng phân loại tốt. Cụ thể như sau:

Độ khó P, %	Mức độ
≥ 80	Dễ
$60 \leq P < 80$	Trung bình

$20 \leq P < 60$	Khó
< 20	Rất khó

2.2.2. Độ phân biệt (D - Discrimination Index)

Độ phân biệt (D) là một chỉ số quan trọng trong Lý thuyết trắc nghiệm Cổ điển (CTT), dùng để đánh giá khả năng một câu hỏi có thể phân biệt giữa người học có năng lực cao và người học có năng lực thấp. Một câu hỏi có độ phân biệt cao sẽ giúp nhận diện rõ ràng sự khác biệt về năng lực, qua đó phản ánh giá trị của câu hỏi trong bài thi [2], [3].

+ **Đối với hình thức thi trắc nghiệm:** Độ phân biệt là tỉ số giữa hiệu số người học nhóm có kết quả cao trả lời đúng và người học nhóm có kết quả thấp trả lời đúng với tổng số người học mỗi nhóm.

$$D = P_U - P_L \quad (3)$$

Trong đó: D: độ phân biệt của câu hỏi;

P_U : Tỷ lệ người học ở nhóm điểm cao (Upper Group) trả lời đúng câu hỏi.

P_L : Tỷ lệ người học ở nhóm điểm thấp (Lower Group) trả lời đúng câu hỏi.

+ **Đối với các hình thức thi khác:** Độ phân biệt là tỉ số giữa hiệu số điểm trung bình của người học thuộc nhóm có kết quả cao và người học thuộc nhóm có kết quả thấp với điểm số tối đa của câu hỏi/ đề thi đó.

$$D = \frac{\bar{X}_U - \bar{X}_L}{M} \quad (4)$$

Trong đó: D: độ phân biệt của câu hỏi;

$\bar{X}_U$: điểm trung bình của người học thuộc nhóm có kết quả cao;

$\bar{X}_L$: điểm trung bình của người học thuộc nhóm có kết quả thấp;

M: điểm tối đa.

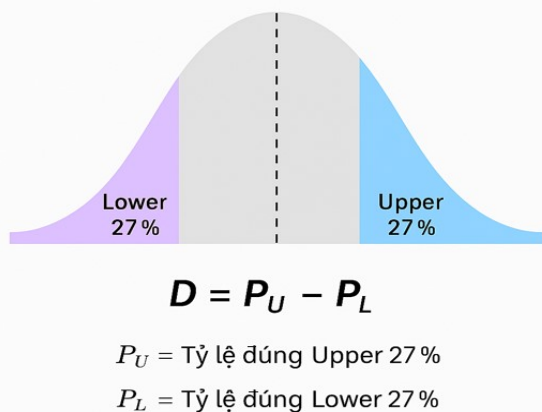
Độ phân biệt là chỉ số cốt lõi phản ánh khả năng phân hóa của đề thi. Chỉ số độ phân biệt càng cao chứng tỏ đề thi có khả năng phân loại rõ ràng giữa người học có năng lực cao và thấp, từ đó đảm bảo tính phân hóa và giá trị đánh giá thực chất của bài kiểm tra.

Độ phân biệt D	Mức độ
$\geq 0,4$	Rất tốt
$0,3 \leq P < 0,4$	Tốt

$0,2 \leq P < 0,3$	Trung bình
$< 0,2$	Kém

2.2.3. Phương pháp phân nhóm người học tính độ phân biệt D

Cách lựa chọn nhóm người học điểm cao và điểm thấp ảnh hưởng trực tiếp đến giá trị phân biệt D. Trong thực tiễn phân tích, phương pháp phân nhóm theo bách phân vị 27%-27% được sử dụng phổ biến vì tính đơn giản và độ tin cậy.



Hình 1: Độ phân biệt theo bách phân vị

Bách phân vị là giá trị chia tập dữ liệu thành 100 phần bằng nhau. Khi phân tích chất lượng đề thi, bách phân vị thường được dùng để chia người học thành ba nhóm:

- + Nhóm cao (Upper) 27%: Thí sinh có tổng điểm cao nhất.
- + Nhóm thấp (Lower) 27%: Thí sinh có tổng điểm thấp nhất.
- + Nhóm giữa (Middle) 46%: Các thí sinh còn lại.

Sự khác biệt về tỷ lệ trả lời đúng giữa nhóm cao và nhóm thấp chính là cơ sở để tính toán độ phân biệt của từng câu hỏi. Chính vì ưu điểm vừa đơn giản, vừa đáng tin cậy, phân nhóm theo bách phân vị 27%-27% đã trở thành chuẩn mực trong nhiều nghiên cứu và hệ thống khảo thí.

3. Công tác đánh giá chất lượng đề thi của Trường Đại học Công nghiệp Quảng Ninh

3.1. Thực trạng công tác đánh giá chất lượng đề thi của Nhà trường

Các nghiên cứu quốc tế và trong nước như của Kelley (1939), Haladyna (1999), Nitko (2001), Trần Bá Hoàng (2005), Lê Văn Hào

(2015) đã khẳng định vai trò quan trọng của độ khó (P), độ phân biệt (D) trong việc đánh giá chất lượng đề thi. Trong đó, phương pháp phân nhóm người học theo bách phân vị 27%-27% được xem là cách tiếp cận đơn giản nhưng hiệu quả, giúp đo lường chính xác giá trị phân biệt của từng câu hỏi [2], [3], [4], [5], [6].

Tại trường ĐHCNQN, công tác đánh giá đề thi hiện nay chủ yếu dừng ở mức độ thống kê mô tả như điểm trung bình, tỷ lệ sinh viên đạt điểm trên trung bình hoặc phổ điểm chung của cả lớp, chưa áp dụng các chỉ số phân tích chuyên sâu về độ khó (P), độ phân biệt (D). Hệ quả nhiều đề thi tuy đảm bảo nội dung kiến thức, nhưng vẫn tồn tại câu hỏi quá dễ, quá khó hoặc kém khả năng phân loại [1]. Bên cạnh đó, việc đánh giá còn phụ thuộc vào kinh nghiệm cá nhân của giảng viên, thiếu tiêu chí định lượng và quy trình chuẩn hoá, làm giảm tính khách quan và minh bạch.

Để khắc phục hạn chế nêu trên, cần thiết xây dựng một giải pháp tổng thể, kết hợp giữa các chỉ số toán học đo lường định lượng (CTT) và các công cụ công nghệ thông tin nhằm chuẩn hóa quy trình đánh giá.

3.2. Xây dựng bộ công cụ hỗ trợ đánh giá tự động

Để nâng cao hiệu quả công tác đánh giá chất lượng đề thi, dựa trên nền tảng Excel, việc phát triển một bộ công cụ hỗ trợ đánh giá tự động là rất cần thiết, giúp tiết kiệm thời gian, đảm bảo tính chính xác và phù hợp với điều kiện thực tiễn tại Trường Đại học Công nghiệp Quảng Ninh. Bộ công cụ đánh giá chất lượng đề thi được thiết kế gồm 2 bước cơ bản:

Bước 1: Thu thập dữ liệu sau thi (kết quả chi tiết từng câu hỏi của từng sinh viên).

Bước 2: Hệ thống tự động phân nhóm thí sinh theo bách phân vị 27% - 27%, tính toán các chỉ số P, D cho từng câu hỏi, và xuất báo cáo kết quả phân loại [6].

3.3. Kết quả thực nghiệm tại Trường Đại học Công nghiệp Quảng Ninh

Dựa trên số liệu thu thập từ 34 sinh viên tham gia thi học phần "Toán cao cấp 2 đề 07" học kỳ II năm học 2024-2025, việc phân tích

chất lượng từng câu hỏi trong đề thi được thực hiện theo phương pháp phân nhóm bách phân vị 27% - 27%, các chỉ số về độ khó (P) và độ phân biệt (D) của từng câu hỏi được tính toán nhằm phản ánh mức độ phù hợp cũng như khả năng phân loại của đề thi. Kết quả phân tích được thể hiện trong bảng 1 như sau:

Bảng 1. Kết quả phân tích D, P của đề thi

Mã câu hỏi/ đề thi	Độ khó (P)	Mức độ khó	Độ phân biệt (D)	Mức độ phân biệt	Nhận xét
Câu 1	0,7	Trung bình	0,5	Rất Tốt	Đạt
Câu 2	0,5	Khó	0,9	Rất Tốt	Đạt
Câu 3	0,9	Dễ	0,5	Rất Tốt	Đạt
Câu 4	0,2	Khó	0,5	Rất Tốt	Đạt
Câu 5	0,08	Rất Khó	0,3	Tốt	Cần xem xét

Kết quả cho thấy cấu trúc đề thi có sự phân bố đa dạng về độ khó: từ dễ ($P=0,9$ ở Câu 3) đến câu rất khó ($P=0,08$ ở Câu 5). Phân tích chi tiết về tỷ lệ câu hỏi theo mức độ cho thấy: 20% câu hỏi ở mức "Dễ", 20% câu hỏi ở mức "Trung bình", 40% ở mức "Khó" và 20% ở mức "Rất khó". Cấu trúc này tương đối phù hợp với nguyên tắc xây dựng đề thi, vừa đảm bảo kiểm tra kiến thức cơ bản, vừa có khả năng phân loại sinh viên theo năng lực. Tuy nhiên theo khuyến nghị chuẩn, đề thi nên tăng tỷ lệ câu dễ lên khoảng 50% để nâng cao tính bao quát kiến thức cơ bản, giảm tỷ lệ câu rất khó, đồng thời giữ cân bằng giữa các mức độ còn lại.

Về chỉ số phân biệt, hầu hết các câu hỏi đều đạt mức "Rất tốt" ($D \geq 0,4$), chứng tỏ đề thi có khả năng tách biệt rõ ràng giữa nhóm thí sinh giỏi và yếu. Câu 2 nổi bật với chỉ số phân biệt cao ($D=0,9$), khẳng định đây là câu hỏi chất lượng, có tính phân loại mạnh. Câu 3, mặc dù ở mức độ "Dễ", vẫn đạt độ phân biệt $D=0,5$. Điều này cho thấy ngay cả những câu hỏi kiểm tra kiến thức cơ bản cũng góp phần phân loại người học. Tuy nhiên, Câu 5 có độ khó quá thấp ($P=0,08$), chỉ số phân biệt ở mức "Tốt" ($D=0,3$) cho thấy câu hỏi này có nguy cơ gây nhiễu và

làm giảm tính ổn định chung của đề thi. Điều này cần xem xét cách ra đề hoặc mức độ yêu cầu của câu hỏi để đảm bảo tính hợp lý.

Tổng thể, kết quả thực nghiệm khẳng định đề thi có chất lượng tương đối tốt, với đa số các câu hỏi đạt chuẩn cả về độ khó và độ phân biệt. Tuy nhiên để nâng cao hơn nữa tính ổn định và tính công bằng, cần điều chỉnh tỷ lệ câu hỏi quá dễ và quá khó, đồng thời giữ lại những câu hỏi có độ phân biệt cao.

4. MỘT SỐ GIẢI PHÁP ĐỊNH HƯỚNG ÁP DỤNG TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP QUẢNG NINH

4.1. Chuẩn hoá quy trình đánh giá chất lượng đề thi

Hiện nay, trường Đại học Công nghiệp Quảng Ninh hiện đã ban hành quy định xây dựng quản lý và sử dụng ngân hàng đề thi. Tuy nhiên, kết quả thực nghiệm cho thấy để đảm bảo tính khách quan, minh bạch và khoa học, quy trình đánh giá chất lượng đề thi cần được chuẩn hoá hơn nữa. Việc chuẩn hoá này bao gồm: thiết lập các chỉ số đo lường định lượng cơ bản (độ khó P và độ phân biệt D), thiết kế quy trình phân tích dữ liệu và lập báo cáo kiểm định, đồng thời tích hợp linh hoạt với điều kiện thực tế của từng khoa, bộ môn [2], [4].

4.2. Phát triển bộ công cụ hỗ trợ đánh giá tự động thành phần mềm nội bộ

Hiện nay, bộ công cụ đánh giá được vận hành trên nền tảng Microsoft Excel. Tuy nhiên, việc vận hành công cụ hiện nay vẫn còn phụ thuộc nhiều vào thao tác thủ công, chưa đáp ứng được nhu cầu tự động hoá cao trong bối cảnh khối lượng dữ liệu công tác khảo thí ngày càng lớn, yêu cầu thời gian xử lý và lập báo cáo ngày càng nhanh chóng, linh hoạt. Điều này đặt ra yêu cầu cấp thiết phải nâng cấp bộ công cụ hiện có thành một phần mềm nội bộ chuyên biệt, được thiết kế tối ưu quy trình đánh giá chất lượng đề thi một cách khép kín và tự động hoá hoàn toàn [1].

Phần mềm này cần được xây dựng theo hướng thân thiện với người sử dụng, giao diện trực quan, đơn giản, phù hợp với thực tiễn sử dụng tại các khoa, bộ môn trong Nhà trường.

Các chức năng chính của phần mềm bao gồm:

- Tự động phân nhóm người học theo bách phân vị 27% - 27%, làm cơ sở tính toán độ khó (P) và độ phân biệt (D) của từng câu hỏi.

- Tính toán và phân tích chỉ số P, D chính xác, tự động xuất kết quả phân tích chi tiết từng câu hỏi, hỗ trợ giảng viên nhận diện các câu hỏi chưa đạt chuẩn.

- Xuất báo cáo đánh giá đề thi: Hệ thống cho phép trích xuất báo cáo tổng hợp, bảng phân tích chi tiết, biểu đồ minh hoạ (Histogram, biểu đồ phân bố P, D), giúp bộ môn dễ dàng đánh giá, lập kế hoạch cải tiến đề thi.

- Lưu trữ, quản lý dữ liệu lịch sử để phục vụ kiểm tra, đánh giá lâu dài.

Về mặt kỹ thuật, phần mềm có thể được phát triển dựa trên các ngôn ngữ lập trình đơn giản, chi phí thấp như hiệu quả như Python kết hợp với giao diện ứng dụng thân thiện. Đây là những giải pháp công nghệ phù hợp với điều kiện nguồn lực hiện có của Nhà trường, đồng thời có khả năng mở rộng, nâng cấp khi cần thiết. Việc phát triển phần mềm nội bộ không chỉ hỗ trợ đánh giá chất lượng đề thi mà còn tạo nền tảng công nghệ cho Nhà trường trong xây dựng hệ thống quản lý đào tạo thông minh, đáp ứng yêu cầu chuyển đổi số.

4.3. Đào tạo, tập huấn giảng viên và cán bộ khảo thí

Trong tiến trình chuẩn hoá và hiện đại hoá công tác kiểm tra, đánh giá kết quả học tập, năng lực chuyên môn của đội ngũ giảng viên và cán bộ khảo thí đóng vai trò quyết định tới hiệu quả triển khai thực tiễn. Việc ứng dụng các chỉ số đo lường định lượng như độ khó (P), độ phân biệt (D) cùng các công cụ hỗ trợ phân tích dữ liệu đòi hỏi đội ngũ thực hiện phải có trình độ kỹ năng phù hợp, nắm vững các nguyên lý đo lường giáo dục và thao tác thành thạo với các phần mềm hỗ trợ.

Tại Trường Đại học Công nghiệp Quảng Ninh, mặc dù đội ngũ giảng viên có nhiều kinh nghiệm trong xây dựng đề thi, song kỹ năng áp dụng các công cụ phân tích định lượng, đặc biệt là các chỉ số đánh giá chất lượng câu hỏi còn hạn chế. Để đảm bảo phương pháp đánh giá

này được áp dụng một cách đồng bộ, hiệu quả và bền vững, Nhà trường cần tổ chức các chương trình tập huấn chuyên đề nhằm nâng cao năng lực chuyên môn cho đội ngũ nhân sự liên quan. Nội dung chương trình tập huấn phải được thiết kế theo hướng vừa đảm bảo tính khoa học, vừa đáp ứng được nhu cầu ứng dụng thực tiễn, tập trung vào ba nhóm kỹ năng cốt lõi: kỹ năng xử lý và phân tích dữ liệu sau thi, kỹ năng tính toán và phân tích các chỉ số độ khó (P), độ phân biệt (D) và kỹ năng sử dụng thành thạo các công cụ hỗ trợ đánh giá chất lượng đề thi như file Excel tự động hoặc phần mềm nội bộ.

Bên cạnh các chương trình tập huấn ngắn hạn, Nhà trường cần xây dựng kế hoạch bồi dưỡng dài hạn nhằm cập nhật các xu hướng công nghệ mới trong lĩnh vực khảo thí, đo lường giáo dục. Hoạt động tập huấn không chỉ dừng lại ở việc trang bị kiến thức, kỹ năng mà còn tạo được môi trường trao đổi, chia sẻ kinh nghiệm giữa các giảng viên, chuyên gia khảo thí, giúp hình thành đội ngũ chuyên trách có năng lực chuyên môn cao, sẵn sàng đáp ứng yêu cầu đổi mới trong kiểm tra, đánh giá kết quả học tập.

Việc đẩy mạnh công tác đào tạo, tập huấn sẽ tạo tiền đề vững chắc cho quá trình chuẩn hoá quy trình đánh giá chất lượng đề thi tại trường ĐHCNQN, đồng thời góp phần xây dựng môi trường khảo thí chuyên nghiệp, hiện đại và minh bạch phù hợp với các tiêu chuẩn kiểm định giáo dục hiện hành.

4.4. Xây dựng ngân hàng câu hỏi chuẩn hoá

Ngân hàng câu hỏi là nền tảng của hệ thống kiểm tra, đánh giá, đảm bảo đề thi khách quan, công bằng và phù hợp với chuẩn đầu ra. Để phát huy hiệu quả, mỗi câu hỏi cần được kiểm chứng chất lượng bằng các chỉ số định lượng, đồng thời được phân loại theo chuẩn đầu ra và mức độ vận dụng kiến thức.

Xuất phát từ thực tiễn, Nhà trường cần xây dựng một ngân hàng câu hỏi chuẩn hoá, được quản lý khoa học, có quy trình rà soát, cập nhật định kỳ và ứng dụng công nghệ thông tin trong vận hành. Việc xây dựng một ngân hàng câu hỏi đạt chuẩn không chỉ nâng cao chất lượng công tác kiểm tra đánh giá mà còn cung cấp dữ liệu

phục vụ cho công tác tự đánh giá, kiểm định và cải tiến chương trình đào tạo. Đây cũng là công cụ cốt lõi giúp Nhà trường hiện đại hoá công tác khảo thí, phù hợp với xu hướng chuyển đổi số trong giáo dục.

4.5. Tích hợp với công tác đảm bảo chất lượng giáo dục toàn trường

Việc chuẩn hóa quy trình đánh giá chất lượng đề thi và xây dựng ngân hàng câu hỏi đạt chuẩn cần được xem như là một cấu phần chiến lược trong hệ thống đảm bảo chất lượng giáo dục của Nhà trường. Các dữ liệu trong việc đánh giá chất lượng đề thi, các báo cáo phân tích về độ khó, độ phân biệt sẽ trở thành những minh chứng cụ thể, có giá trị, phục vụ cho hoạt động tự đánh giá và kiểm định chất lượng chương trình đào tạo. Sự tích hợp này giúp trường Đại học Công nghiệp Quảng Ninh từng bước hình thành hệ thống đảm bảo chất lượng giáo dục bền vững, nâng cao uy tín, thương hiệu Nhà trường, đồng thời đáp ứng yêu cầu ngày càng cao của xã hội về chất lượng nguồn nhân lực.

5. KẾT LUẬN

Nghiên cứu đã chứng minh được rằng việc ứng dụng các chỉ số toán học như độ khó (P), độ phân biệt (D) kết hợp với phương pháp phân tích bách phân vị là một hướng tiếp cận hiệu

quả trong đánh giá chất lượng đề thi [1], [3], [6]. Kết quả thực nghiệm tại trường Đại học Công nghiệp Quảng Ninh cho thấy cách tiếp cận này giúp nhận diện chính xác các câu hỏi chưa đạt chuẩn, từ đó hỗ trợ điều chỉnh, hoàn thiện cấu trúc đề thi theo hướng khách quan, khoa học và phù hợp với chuẩn đầu ra đào tạo.

Trên cơ sở đó, nghiên cứu đề xuất một số giải pháp trọng tâm: (i) chuẩn hoá quy trình đánh giá đề thi dựa trên các chỉ số đo lường định lượng; (ii) phát triển bộ công cụ hỗ trợ đánh giá tự động tiến tới phần mềm nội bộ; (iii) tăng cường đào tạo, bồi dưỡng đội ngũ giảng viên, cán bộ khảo thí và (iv) xây dựng ngân hàng câu hỏi chuẩn hoá, tích hợp với hệ thống đảm bảo chất lượng giáo dục toàn trường.

Kết quả nghiên cứu khẳng định rằng việc áp dụng phương pháp đo lường định lượng kết hợp giải pháp công nghệ không chỉ góp phần nâng cao chất lượng khảo thí tại Trường đại học Công nghiệp Quảng Ninh mà còn là bước đi chiến lược trong xây dựng hệ thống đảm bảo chất lượng bền vững, đáp ứng yêu cầu kiểm định và xu thế chuyển đổi số trong giáo dục đại học hiện nay.

TÀI LIỆU THAM KHẢO

1. Khánh, N. C. (2019). Phân tích chỉ số độ khó, độ phân biệt và hệ số tương quan điểm - nhị phân trong kiểm định chất lượng câu hỏi trắc nghiệm khách quan. *Tạp chí khoa học giáo dục*, số 164, tr. 12-17.
2. Hoành, T. B. (2005). Đo lường và đánh giá trong giáo dục. *Nhà xuất bản Đại học Sư phạm, Hà Nội*.
3. Hào, L. V. (2015). Ứng dụng Lý thuyết trắc nghiệm cổ điển (CTT) và Lý thuyết đáp ứng câu hỏi (IRT) trong đánh giá chất lượng câu hỏi trắc nghiệm. *Tạp chí khoa học - Trường Đại học Sư phạm TP. Hồ Chí Minh*, số. 65, pp. 45-55.
4. Nitko, A. J. & Brookhart, S. M. (2014). *Educational Assessment of Students (7th Edition)*. Pearson Education, Inc.
5. Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items (3rd Edition)*. Lawrence Erlbaum Associates.
6. Kelley, T. L. (1939). *The Selection of Upper and Lower Groups for the Validation of Test Items*. *Journal of Educational Psychology*, 30(1), 17-24.

Thông tin của tác giả:

ThS. Nguyễn Thu Hiền

Phòng Thanh tra và KĐCL, Trường Đại học Công nghiệp Quảng Ninh

Điện thoại: +(84).862078861 - Email: hientn123@gmail.com

ThS. Nguyễn Thị Phương

Phòng Thanh tra &ĐBCL, Trường Đại học Công nghiệp Quảng Ninh

Điện thoại: +(84).904544205 - Email: phuongnt@qui.edu.vn

ASSESSING EXAM QUALITY THROUGH MEASUREMENT INDICES AND SOLUTIONS FOR IMPROVING TESTING EFFECTIVENESS AT QUANG NINH UNIVERSITY OF INDUSTRY

Information about authors:Nguyen Thu Hien, M.S., Quang Ninh University of Industry. Email: hientn123@gmail.comNguyen Thi Phuong, M.S., Quang Ninh University of Industry. Email: phuongnt@qui.edu.vn**ABSTRACT:**

This study aims to develop a procedure for evaluating exam quality based on mathematical indices, thereby enhancing objectivity and efficiency in the post-exam analysis of end-of-course assessments. The indices employed include Difficulty Index (P), Discrimination Index (D), and Percentile Rank, which provide a basis for assessing the appropriateness of each test item in differentiating student performance levels. The research method relies on data from end-of-course examinations, applying a percentile-based grouping algorithm (Upper 27%, Lower 27%) to calculate discrimination, while also developing an Excel-based tool to automate the data processing workflow. The findings indicate that indices P and D effectively identify items with unsatisfactory difficulty or discrimination values, thus informing recommendations for retention, revision, or elimination. The study affirms that applying simple yet rational mathematical indices not only saves time and reduces subjective bias but also improves the quality of item banks and the overall effectiveness of assessment practices at Quang Ninh University of Industry.

Keywords: Exam evaluation, Difficulty Index, Discrimination Index, Percentile rank, Item bank.

REFERENCES

1. Khanh, N. C.(2019). Analysis of Difficulty Index, Discrimination Index, and Point-Biserial Correlation Coefficient in the Validation of Multiple-Choice Test Items. *Journal of Educational Science*, No. 164, pp. 12-17.
2. Hoanh, T. B. (2005). *Measurement and Evaluation in Education*. Hanoi National University of Education Publishing House.
3. Hao, L. V. (2015). Application of Classical Test Theory (CTT) and Item Response Theory (IRT) in Evaluating the Quality of Multiple-Choice Test Items. *Journal of Science – Ho Chi Minh City University of Education*, No. 65, pp. 45-55.
4. Nitko, A. J. & Brookhart, S. M. (2014). *Educational Assessment of Students (7th Edition)*. Pearson Education, Inc.
5. Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items (3rd Edition)*. Lawrence Erlbaum Associates.
6. Kelley, T. L. (1939). *The Selection of Upper and Lower Groups for the Validation of Test Items*. *Journal of Educational Psychology*, 30(1), 17-24.

Ngày nhận bài: 27/08/2025;

Ngày nhận bài sửa: 16/09/2025;

Ngày chấp nhận đăng: 26/09/2025.