

Thiết kế bài kiểm tra trắc nghiệm khách quan đánh giá sự tiến bộ của sinh viên Ngôn ngữ Anh năm thứ nhất tại Học viện Tài chính

Trần Thu Hoài*

*Khoa Ngoại ngữ, Học viện Tài chính

Received: 14/9/2023; Accepted: 20/9/2023; Published: 25/9/2023

Abstract: It has been easy to assume that testing is an important tool in educational research and for program evaluation. That is why the study is aimed to design a progress achievement English for Finance and Accounting objective test for 1st year students of Faculty of Foreign Languages at Academy of Finance. The test is considered a periodic test of students' progress before they finish the first semester of their first year. Then the results of the test will be analyzed, evaluated, and interpreted to see whether or not the test satisfies the qualities of a good test.

Keywords: Discrimination, matching, multiple-choice, objective test, practicality, progress achievement test, reliability, true/false, validity.

1. Giới thiệu

Gần đây, kiểm tra đánh giá thường được coi là một công cụ quan trọng trong nghiên cứu giáo dục và đánh giá chương trình. Là một giảng viên dạy tiếng Anh chuyên ngành Tài chính Kế toán, sự quan tâm của tôi đối với chủ đề này xuất phát từ nền tảng kiến thức cũng như quan điểm của tôi về kiểm tra khách quan. Vì vậy, mục tiêu của nghiên cứu là thiết kế một bài kiểm tra khách quan đo lường sự tiến bộ của sinh viên (SV) trong việc học tiếng Anh chuyên ngành Tài chính Kế toán trong năm đầu tiên tại Khoa Ngoại ngữ, Học viện Tài chính. Trước khi học kỳ đầu tiên của năm thứ nhất kết thúc, bài kiểm tra được coi là đánh giá định kỳ về sự tiến bộ của SV. Sau đó, bài kiểm tra sẽ được nghiên cứu, đánh giá và diễn giải để xác định xem nó có đáp ứng các tiêu chí là một bài kiểm tra tốt hay không. Việc đánh giá về độ tin cậy (reliability), tính giá trị (validity), tính thực tiễn (practicality) và khả năng phân loại (discrimination) của bài kiểm tra khách quan này sẽ hỗ trợ giảng viên trong việc tìm ra những cách hiệu quả hơn để đánh giá trình độ của SV tiếng Anh chuyên ngành Kế toán và Tài chính ở cả hiện tại và tương lai.

2. Nội dung nghiên cứu

2.1. Thực trạng

Sinh viên Khoa Ngoại ngữ, Học viện Tài chính dù học chuyên ngành ngôn ngữ nhưng nền tảng tiếng Anh của các em vẫn có sự khác biệt đáng kể do đến từ nhiều địa phương trên cả nước, trong khi trước khi tốt nghiệp phải đạt chuẩn đầu ra trình độ C1 hoặc tương đương. Trên thực tế, sinh viên chỉ có cơ hội

học 6 học phần tiếng Anh theo giáo trình Intelligent Business (3 cấp độ: Pre-Intermediate, Intermediate và Upper-Intermediate), mỗi học phần có 6 tín chỉ. Dù học tập chăm chỉ nhưng nhiều em vẫn thi trượt hoặc đạt kết quả kém sau mỗi kỳ thi. Một trong những nguyên nhân quan trọng dẫn đến tình trạng trên là vấn đề bài kiểm tra. Đôi khi, các bài kiểm tra ngày nay cho thấy sự khác biệt giữa những gì sinh viên học trên lớp và các dạng câu hỏi trong bài kiểm tra, tức là các bài kiểm tra có những câu hỏi không quen thuộc hoặc quá khó đối với các em.

2.2. Phương pháp nghiên cứu

Để thu thập dữ liệu cho nghiên cứu, một bài kiểm tra khách quan gồm 50 mục về tiếng Anh chuyên ngành Tài chính Kế toán đã được giao cho 60 SV năm thứ nhất của khoa Ngoại ngữ, Học viện Tài chính được chọn ngẫu nhiên. Các SV chuyên ngữ này làm bài thi tại 2 phòng học dưới sự giám sát của 2 giáo viên trong khung thời gian nhất định (60 phút). Khi làm bài kiểm tra này, SV đã hoàn thành 7 bài của giáo trình "Intelligent Business – Pre-Intermediate" (Johnson C., 2007). Nội dung của bài kiểm tra liên quan trực tiếp đến những gì được dạy trong giáo trình. Sau đó, các bài kiểm tra được thu lại và được chấm theo thang điểm từ 0-10. Cuối cùng, kết quả của các bài kiểm tra được phân tích và diễn giải để tìm ra số SV làm bài tốt và số SV làm bài kém. Kết quả cũng cho biết điểm số thường xuyên nhất mà người làm bài kiểm tra đạt được, cách tính điểm của những người này và sự chênh lệch so với điểm trung bình, v.v.

2.3. Kết quả và thảo luận

2.3.1. Tỉ trọng tương đối của các phần khác nhau trong bài kiểm tra

Bài kiểm tra gồm 50 câu hỏi trắc nghiệm được chia thành 5 phần, mỗi phần chiếm tỉ trọng 20%.

- *Phần 1:* gồm 10 câu, mỗi câu khuyết 1 chỗ trống với 4 đáp án gợi ý để kiểm tra về ngữ pháp và từ vựng.

- *Phần 2:* gồm 1 đoạn văn ngắn khoảng 150 từ, đọc để lấy ý chính/ thông tin cụ thể và từ vựng, với 10 câu hỏi đi kèm 4 đáp án gợi ý cho mỗi câu.

- *Phần 3:* kiểm tra về từ vựng, ngữ pháp và chức năng với 10 câu lựa chọn Đúng/Sai.

- *Phần 4:* kiểm tra về thuật ngữ chuyên ngành, ngữ pháp và chức năng với 10 cặp câu để ghép đôi (10 matching)

- *Phần 5:* kiểm tra về diễn đạt, cấu trúc ngữ pháp và từ vựng với 10 câu có các từ bị xáo trộn kèm 4 phương án lựa chọn cho mỗi câu.

2.3.2. Phân tích kết quả bài kiểm tra

Điểm của 60 bài thi được tổng hợp và phân tích theo các hạng mục mà Heaton đã từng đề cập (2000) như sau:

a. Phổ điểm (frequency distribution):

SV đạt các mức điểm khác nhau từ 3 đến 10 (điểm thấp nhất là 3 và cao nhất là 10). Sự phân bố điểm khá không đồng đều, chẳng có học sinh nào đạt điểm 4,5, 7,5 và 9,5; số điểm mà nhiều SV đạt được nhất là 6.

b. Xu hướng tập trung (central tendency):

Để đo lường xu hướng tập trung của điểm kiểm tra, chúng tôi dùng 3 tham số:

1. The mode (số yếu vị): 6,0 - điểm mà hầu hết người tham gia kiểm tra đạt được;

2. The median (số trung vị): 6,5 - điểm đạt được ở giữa theo thứ tự thành tích;

3. The mean (M): điểm trung bình, được tính theo công thức: $M = \sum fx/N \approx 6,3$ (trong đó \sum là tổng, N là số SV tham gia làm bài, x là điểm số và f là tần suất các mốc điểm xuất hiện)

Có thể thấy rằng trong trường hợp này có sự tương ứng khá chặt chẽ giữa giá trị trung bình (6,3) và trung vị (6,5). Sự tương ứng chặt chẽ như vậy không phải lúc nào cũng phổ biến và đã xảy ra trong trường hợp này vì điểm số có xu hướng tập trung đối xứng xung quanh một điểm trung tâm.

c. Sự phân tán (dispersion):

Có ba cách chính để kiểm tra độ phân tán: mức thấp-cao (low-high), phạm vi (range) và độ lệch chuẩn (standard deviation).

d. Mức thấp cao:

Thấp-cao liên quan đến việc tìm giá trị thấp nhất và giá trị cao nhất trong một tập hợp số. Khi sắp xếp điểm mà SV đạt được theo thứ tự từ cao xuống thấp, ta thấy ngay giá trị thấp nhất là 3 và giá trị cao nhất là 10. Như vậy, thấp-cao là 3 -10.

e. *Phạm vi:* Phạm vi là một cách đơn giản để đo lường sự phân bố của điểm, là sự khác biệt giữa điểm cao nhất và điểm thấp nhất, tức là giá trị cao nhất trừ đi điểm thấp nhất. Công thức tính khoảng điểm như sau:

Range = H - L = 10 - 3 = 7 (trong đó H - điểm cao nhất, L - điểm thấp nhất)

Bài kiểm tra có phạm vi rộng chứng tỏ rằng những người được thử nghiệm có nhiều khả năng khác nhau.

d. Độ lệch chuẩn (SD):

Độ lệch chuẩn là một cách khác để diễn tả về phổ điểm, với công thức tính như sau:

$$SD = \sqrt{\frac{\sum D^2}{N}} = \sqrt{\frac{\sum (X-M)^2}{N}} = 1.08$$

Trong đó: SD là độ lệch chuẩn, X là điểm số, M là điểm trung bình và N là số SV tham gia.

Như đã thấy từ kết quả trên, chúng ta có độ lệch chuẩn lớn (1,08), cho thấy: thứ nhất, sự phân bố điểm của bài kiểm tra rất rộng; thứ hai, bài kiểm tra đã phân loại được học sinh và thứ ba, có khoảng cách khá lớn về trình độ của SV.

2.3.3. Phân tích về độ khó và khả năng phân loại của bài kiểm tra

Tất cả 50 câu hỏi của bài kiểm tra khách quan đều được phân tích về độ khó và khả năng phân loại trình độ SV dựa theo các chỉ số được liệt kê bởi Heaton (2000) như sau:

a. Chỉ số độ khó (item difficulty)

- Công thức tính chỉ số độ khó là:

- FV = R/N (trong đó FV là chỉ số độ khó, R là số câu trả lời đúng và N là số SV tham gia làm bài)

- Thang đo độ khó của các câu hỏi là:

+ FV = 0,81 ÷ 1 → ve (rất dễ) (tức là 81 đến 100% học sinh làm đúng)

+ FV = 0,61 ÷ 0,8 → e (dễ) (tức là 61 đến 80% SV làm đúng)

+ FV = 0,41 ÷ 0,6 → ok (tức là 41 đến 60% SV làm đúng)

+ FV = 0,21 ÷ 0,4 → d (khó) (tức là 21 đến 40% SV làm đúng)

+ FV = 0 ÷ 0,2 → vd (rất khó) (tức là 0 đến 20% SV làm đúng)

- Kết quả thu thập được cho thấy các câu 8, 16, 34, 40, 44 đều khó với chỉ số độ khó dao động từ 0,25 đến 0,40; và các câu 28, 32, 36, 37, 47 đều rất khó với chỉ số độ khó dao động từ 0,08 đến 0,20. Trong khi đó, các câu 1, 5, 6, 7, 10, 11, 14, 20, 21, 27, 35 lại rất dễ vì có chỉ số độ khó trên 0,80 và các câu 3, 13, 18, 19, 23, 25, 30, 33, 45, 46, 50 có thể được coi là dễ dàng với chỉ số độ khó nằm trong khoảng từ 0,61 đến 0,80. Với chỉ số độ khó dao động từ 0,41 đến 0,60, các câu 2, 4, 9, 12, 15, 17, 22, 24, 26, 29, 31, 38, 39, 41, 42, 43, 48, 49 đều phù hợp với trình độ SV.

b. Khả năng phân loại của từng câu hỏi

- Ta có công thức tính như sau:

$D = (CU - CL) / N$ (trong đó CU là số người trả lời đúng của nửa trên và CL là số người trả lời đúng của nửa dưới).

- Thang đo khả năng phân loại trình độ là:

$D = 0,6 \div 1 \rightarrow$ gd (phân biệt tốt)

$D = 0,3 \div 0,59 \rightarrow$ md (độ phân biệt trung bình)

$D = 0 \div 0,29 \rightarrow$ ld (độ phân biệt thấp)

$D < 0 \rightarrow$ bi (câu hỏi chưa tốt)

Kết quả cho thấy hầu hết các câu hỏi đều có mức độ phân loại trình độ thấp, ngoại trừ câu 9, 15, 18, 21, 29, 49 có mức độ phân loại ở mức trung bình và các câu 2, 3, 17, 40, 50 được coi là chưa tốt vì có chỉ số phân loại quá thấp, có nghĩa là chúng không thể phân loại được trình độ của SV.

2.3.4. Đánh giá độ tin cậy

- Có nhiều cách để tính độ tin cậy của bài kiểm tra, truy nhiên công thức 21 của Kuder-Richardson là phổ biến nhất:

$r = 1 - M(n-M)/ns^2$ (trong đó có $M = 6,3$ là điểm trung bình, $n = 50$ là tổng số câu hỏi và s là độ lệch chuẩn $SD = 1,08$)

- Thang đo độ tin cậy là:

0,8 \rightarrow 1,0 độ tin cậy cao

0,6 \rightarrow 0,8 độ tin cậy trung bình

0,4 \rightarrow 0,6 độ tin cậy thấp

0,2 \rightarrow 0,4 độ tin cậy rất thấp

Sử dụng công thức trên, chúng ta thu được kết quả $r = 0,91$, cho thấy bài kiểm tra có độ tin cậy cao.

2.4. Đánh giá chung

Tóm lại, để đánh giá bài kiểm tra khách quan trên có tốt hay không, chúng ta nên quay lại với 4 tiêu chí: độ tin cậy (reliability), tính giá trị (validity), khả năng phân loại (discrimination) và tính thực tiễn (practicality).

- Rõ ràng, độ tin cậy và tính giá trị là rất quan trọng đối với bất kỳ bài kiểm tra nào và được coi là

chất lượng đo lường thiết yếu. Có một mối quan hệ giữa họ. Một mặt, một bài kiểm tra có thể đáng tin cậy nhưng không có giá trị. Mặt khác, nếu bài kiểm tra không đáng tin cậy thì nó không thể có giá trị gì cả. Việc tính toán độ tin cậy ở trên cho thấy bài kiểm tra là đáng tin cậy, do đó bây giờ có thể nói rằng bài kiểm tra cũng có giá trị.

- Ngoài ra, xét về khả năng phân loại trình độ các SV nhằm phản ánh sự chênh lệch về trình độ của từng cá nhân trong nhóm, các câu hỏi trong bài kiểm tra đều có thang độ khó rộng, từ câu rất dễ đến câu rất khó, tức là bài kiểm tra không quá dễ cũng không quá khó. Vì vậy, bài kiểm tra khách quan này là bài kiểm tra có thể hiện thực hóa mục đích phân loại trình độ giữa các SV.

- Hơn nữa, để chuẩn bị cho việc tiến hành cho SV làm bài kiểm tra, tính thực tế luôn được xem xét kỹ trong quá trình tiến hành kiểm tra. Tất cả những thứ cần chuẩn bị đều đã có sẵn ví dụ: phòng thi (2 phòng học), người chấm bài (các thành viên trong nhóm nghiên cứu) và in ấn bài kiểm tra không tốn nhiều chi phí cũng như không mất nhiều thời gian.

- Từ tất cả các phân tích ở trên, có thể nói rằng đây là một bài kiểm tra khá tốt mặc dù để trở thành một bài kiểm tra hoàn hảo thì còn khoảng cách khá xa.

3. Kết luận

Cuối cùng, bài nghiên cứu này được thực hiện với mong muốn đóng góp một số kỹ thuật giúp cho không chỉ giảng viên khoa Ngoại ngữ tại Học viện Tài chính mà tất cả các giảng viên dạy tiếng Anh tại các trường cao đẳng, đại học khác có thể nâng cao khả năng thiết kế bài thi tiếng Anh nói chung và bài kiểm tra tiếng Anh trắc nghiệm khách quan nói riêng. Ngoài ra, chúng tôi hy vọng rằng các giáo viên dạy tiếng Anh có thể sử dụng những gì chúng tôi đã giới thiệu làm tài liệu tham khảo trong quá trình giảng dạy của mình.

Tài liệu tham khảo

1. Alderson, J.C., et al., (2001), *Language Test Construction and Evaluation*, Cambridge: Cambridge University Press

2. Bachman, L. F., & Palmer, A. S. (2010), *Language assessment in practice*, Oxford: Oxford University Press

3. Heaton, J.B., (2000), *Writing English Language Tests (new edition)*, Beijing: Foreign Language Teaching and Research Press.

4. Johnson C., 2007, *Intelligent Business Course book – Pre-Intermediate*, England: Pearson Education Limited