THIẾT KẾ VÀ XÂY DỰNG CÔNG CỤ ĐÁNH GIÁ KỸ NĂNG NGHE HIỂU

Nguyễn Quang Thuấn

Kiểm tra-đánh giá ngày càng có vai trò đặc biệt quan trọng trong dạy và học tiếng nước ngoài. Song không phải tất cả những người dạy ngoại ngữ đều nắm chắc được lý thuyết cũng như kỹ thuật kiểm tra-đánh giá, nhất là thiết kế và xây dựng một trắc nghiệm. Thực tế cho thấy rằng, phần lớn giáo viên ngoại ngữ kể cả giảng viên đại học thiết kế và xây dựng một trắc nghiệm chủ yếu là dựa vào kinh nghiệm, thậm chí còn "lúng túng" khi được giao nhiệm vụ ra một đề thi hay đề kiểm tra. Hậu quả là đánh giá việc dạy và học thiếu chính xác và ảnh hưởng ít nhiều không tốt đến kết quả đào tạo. Trong khuôn khổ bài viết này, chúng tôi muốn đề cập sâu hơn, cụ thể hơn những nguyên tắc, những kỹ thuật liên quan đến đánh giá một kỹ năng cụ thể. Chúng tôi sẽ trình bày một cách cụ thể nguyên tắc và các bước tiến hành xây dựng một công cụ kiểm tra-đánh giá một kỹ năng nghe hiểu, chính xác hơn là một trắc nghiệm dùng để đo năng lực hiểu một văn bản nghe bằng tiếng nước ngoài của người học.

Trước hết chúng ta cần phân biệt sự khác nhau giữa ***đo*** "mesure" ***trắc nghiệm "test"*** và ***đánh giá*** (evaluation).

Những thuật ngữ trên thường được sử dụng một cách đồng nghĩa. Thực vậy trong thực tế nó đều ám chỉ đến một hoạt động giống nhau. Ví dụ khi người ta muốn đánh giá kỹ năng ngôn ngữ hay kỹ năng giao tiếp của một người học, người ta thường căn điểm số mà người học nhận được từ một ***trắc nghiệm***. Việc chỉ chú ý vào sự giống nhau bề ngoài của các thuật ngữ này có thể làm mất đi những đặc tính riêng biệt của chúng, những đặc tính cho phép phân biệt giữa chúng với nhau. Tuy nhiên, việc phân biệt chúng có ý nghĩa lớn trong việc phát triển và sử dụng các ***trắc nghiệm*** ngôn ngữ.

Đo "mesure" trong lĩnh vực khoa học xã hội và nhân văn là quá trình định lượng con người hay cá nhân theo các biện pháp và các qui tắc tường minh rõ ràng. Định nghĩa này bao hàm ba nét khác nhau: định lượng, các đặc tính và các thủ pháp tường minh.

Trắc nghiệm theo Carrolle (1968: 46) được định nghĩa như sau:

"*Một **trắc nghiệm** trong lĩnh vực tâm lý hay trong lĩnh vực giáo dục là một biện pháp được điều chỉnh sao cho tạo nên một phản ứng nào đó và từ các ứng xử này người ta có thể suy diễn về một số đặc tính của một người hay một cá nhân*"(1)

Từ định nghĩa này, người ta có thể hiểu rằng ***trắc nghiệm*** là một công cụ đo được thiết kế và xây dựng nhằm tạo ra tập tính của một cá nhân. Với tư cách là một loại hình ***đo, trắc nghiệm*** định lượng cần thiết các đặc tính của con người theo các biện pháp tường minh. Điều này phân biệt một ***trắc nghiệm*** với các loại hình *đo* khác ở chỗ nó được thiết kế để thu nhận một tập hợp đặc biệt các phản ứng. Các ***trắc nghiệm*** ngôn ngữ cung cấp cho ta các phương tiện để tập trung vào các kỹ năng ngôn ngữ hay các kỹ năng giao tiếp mà chúng ta quan tâm. Các ***trắc nghiệm*** ngôn ngữ có thể được đánh giá là các công cụ tốt nhất bảo đảm rằng mẫu ngôn ngữ (Échantillon) đủ cho phép ***đo*** dự kiến ngay

cả người ta muốn quan tâm đến các kỹ năng chung hay tổng quát. Thực tế, không ít người quan niệm rằng ***trắc nghiệm*** đồng nghĩa với *trắc nghiệm khách quan*. Thực chất, trắc nghiệm được phân làm hai loại: *trắc nghiệm khách quan* và *trắc nghiệm tự luận*, người ta cũng gọi là: *trắc nghiệm chấm khách quan* và *trắc nghiệm chấm chủ quan*, vì cũng không có trắc nghiệm nào là khách quan tuyệt đối cả. Mỗi một loại hình có chức năng và nhiệm vụ riêng, ví dụ trắc nghiệm khách quan thường hay dùng để kiểm tra-đánh giá kỹ năng hiểu hơn là dùng để đánh giá kỹ năng sản sinh.

Trong khi mà *đo* thường dựa trên quan sát phản ứng trong môi trường tự nhiên về một thời kỳ nào đó như là cho điểm của người thầy, những quan sát như vậy có thể không chứa các mẫu biểu hiện các kỹ năng hay phẩm chất đặc thù.

Đánh giá có thể được định nghĩa như là tìm kiếm, thu thập một cách hệ thống các thông tin để ra quyết định (Weiss 1972) xác suất ra một quyết định đúng trong các tình huống không chỉ tuỳ thuộc vào người ra quyết định mà còn tuỳ thuộc vào chất lượng thông tin làm cơ sở cho việc ra quyết định đó. Mọi con cò đều trắng, song thông tin càng tin cậy, càng chính xác và càng xác đáng, xác suất ra các quyết định đúng càng cao. Chính vì vậy, đánh giá không phải nhất thiết lúc nào cũng đòi hỏi phải làm ***trắc nghiệm***. Tương tự, bản thân ***trắc nghiệm*** không mang tính đánh giá. Các loại ***trắc nghiệm*** thường được sử dụng với mục đích sư phạm hay như một phương tiện để động viên khích lệ học sinh, sinh viên học, hay cũng như phương tiện để xem lại nội dung, phương pháp dạy, trong trường hợp này ra quyết định không dựa trên kết quả của ***trắc nghiệm***. Người ta cũng dùng ***trắc nghiệm*** với mục đích miêu tả. Chỉ khi ***trắc nghiệm*** được sử dụng làm cơ sở để ra các quyết định mà người ta nói ***đánh giá***. Người ta thường phân biệt ba loại *đánh giá* chủ yếu: *đánh giá tổng kết*, (evaluation sommative)- thường được thực hiện sau một quá trình đào tạo, như kết thúc một học phần hay kết thúc một môn học và có chức năng phân loại, xác nhận trình độ của người học hay thí sinh, *đánh giá điều chỉnh quá trình đào tạo* (evaluation formative) thường được thực hiện trong quá trình đào tạo và có chức năng điều chỉnh việc dạy và học, *đánh giá chẩn đoán* (evaluation pronostique) thường được thực hiện trước quá trình đào tạo và có chức năng phát hiện và nhận biết các mặt mạnh và yếu của người học từ đó cho phép tổ chức và sắp xếp việc dạy và học hiệu quả hơn.

Nói về các đặc tính chủ yếu của đánh giá, người ta không thể không nói đến hai đặc tính quan trọng của ***trắc nghiệm*** là độ tin cậy và tính giá trị. Đây là hai đặc tính quan trọng nhất của khi thiết kế và xây dựng một ***trắc nghiệm***. Độ tin cậy cho phép đo chính xác các kết quả của ***trắc nghiệm*** và kết quả của ***trắc nghiệm*** được gọi là có độ tin cậy là kết quả không có sai về ***đo***. Độ tin cậy có quan hệ mật thiết với ***đo***. Trong bất cứ hoàn cảnh ***trắc nghiệm*** nào, cũng có sai về *đo*. Do vậy, quan tâm đầu tiên đến độ tin cậy của ***trắc nghiệm*** là phải nhận ra trước hết các nguồn sai và sau đó sử dụng các biện pháp để đánh giá ảnh hưởng của các nguồn sai này.

Tính giá trị là chất lượng quan trọng nhất của ***trắc nghiệm*** và sử dụng ***trắc nghiệm***. Nó cho phép chúng ta xác định sử dụng ***trắc nghiệm*** khi nào và như thế nào

và kết quả của có thích hợp và có ích lợi không. Trong mọi trường hợp, hai đặc tính đặc biệt quan trọng trên đây cho phép chúng ta quyết định có sử dụng ***trắc nghiệm*** hay không. Tự nó, ***trắc nghiệm*** sẽ không có giá trị nếu thiếu một trong hai đặc tính này. Độ tin cậy bảo đảm chất lượng của một ***trắc nghiệm*** còn tính giá trị bảo đảm một ***trắc nghiệm*** có sử dụng hay không.

Nhìn chung, quá trình làm ***trắc nghiệm*** trong dạy và học ngoại ngữ gồm 7 bước:

- Bước 1: Xác định mục đích **của *trắc nghiệm*,**
- Bước 2: Kế hoạch hoá ***trắc nghiệm***,
- Bước 3: Lựa chọn các *tiểu mục* và nhiệm vụ,
- Bước 4: Tiến hành ***trắc nghiệm***.
- Bước 5: Chấm và cho điểm.
- Bước 6: Phân tích *tiểu mục* của ***trắc nghiệm*** và nhiệm vụ của từng tiểu mục của trắc nghiệm.
- Bước 7: Tổng kết kết quả của ***trắc nghiệm.***

Về đánh giá nghe hiểu cũng tuân theo bảy bước này, cụ thể như sau:

Bước 1: Xác định mục đích của ***trắc nghiệm***

Nhìn chung, trong đánh giá nghe hiểu, mục đích của ***trắc nghiệm*** là kiểm tra năng lực, trình độ hiểu văn bản nghe thực. Tuy nhiên ***trắc nghiệm*** cũng có thể nhằm vào đánh giá quá trình có nghĩa là các kỹ năng, các chiến lược mà người nghe sử dụng để hiểu văn bản nghe hoặc hiểu văn bản như một sản phẩm có nghĩa là kết quả hiểu văn bản mà không cần biết tiến trình, phương pháp, kỹ năng, chiến lược sử dụng để hiểu.

Bước 2: Kế hoạch hoá ***trắc nghiệm***,

Kế hoạch hoá một ***trắc nghiệm*** nói chung, kỹ năng nghe hiểu nói riêng là khâu cực kỳ quan trọng cho phép xây dựng một trắc nghiệm chuẩn, chất lượng. Sau khi xác định mục tiêu, mục đích của kiểm tra-đánh giá, phải xác định nội dung cần đánh giá, thông qua việc xây dựng ma trận đề. Nhìn chung, chính ở giai đoạn này mà người ta quyết định xem ***trắc nghiệm*** sẽ nhằm đánh giá cấp độ hiểu nào: hiểu khái quát, hiểu sâu hay chỉ là nhận biết các loại hình văn bản, vv Người ta thường dựa vào phân loại kiến thức, năng lực của Bloom để thiết kế, xây dựng trắc nghiệm, để đánh giá. Đối với đánh giá kỹ năng nghe hiểu, điều quan trọng là phải nhớ lại người ta nghe như thế nào trong các hoàn cảnh giao tiếp hàng ngày. Nhìn chung, người ta biết rằng cái mà người ta sẽ nghe và nghe vì mục đích gì. Vậy thích hợp nhất là khi kế hoạch hoá ***trắc nghiệm*** phải nghĩ đến các phương tiện gì cần sử dụng để chuẩn bị cho người học nghe văn bản.

Bước 3: Lựa chọn các *tiểu mục* (Item) và nhiệm vụ

Pha 1: Lựa chọn "ngữ liệu"

Trong giai đoạn lựa chọn ngữ liệu sử dụng để đánh giá, người ta cố gắng lựa chọn ngữ liệu thực nhất có thể sao cho nó phản ánh được các đặc tính của ngôn ngữ nói. Có nghĩa là ngôn ngữ được nói với một nhịp độ bình thường, tự nhiên với âm điệu, ngữ điệu bình thường. Ngữ liệu được lựa chọn thường lấy từ những buổi thời sự, ca nhạc, trò chơi ở đài phát thanh hay đài truyền hình của các nước bản ngữ, các buổi hội thảo, nói chuyện do người bản ngữ trình bày. Tuy nhiên người ta cũng có thể sử dụng các tài liệu ghi chất lượng để không gây nhiều khó khăn cho người nghe và cũng không giảm tính hiệu quả của ***trắc nghiệm***. Người ta cũng có thể ghi đi ghi lại nhiều lần sao cho đạt được chất

lượng như mong muốn. Nếu như các tài liệu ghi với mục đích duy nhất là cho các ***trắc nghiệm***, thì cẩn thận làm cho nó có thể tự nhiên nhất có thể. Thí dụ, ngôn ngữ nói thường được đặc trưng là hay có nói đi nói lại thì phải giữ nguyên mà không nên bỏ đi.

Độ dài của tài liệu nghe cũng cần được tính đến và tuỳ theo mục đích của trắc nghiệm.

Pha 2: Chuẩn bị nhiệm vụ kiểm tra - đánh giá

Nhiệm vụ đánh giá tương ứng với cái mà người học hay thí sinh phải thực hiện. Nhiệm vụ cũng phải xác định cái mà người học hay học sinh phải thực hiện để đạt được mục tiêu của mình.

Soạn thảo các *tiểu mục*

Điều đầu tiên phải làm là tiến hành nghe ngữ liệu đã được chọn bằng cách ghi lại. Ví dụ, người ta sẽ nghe một trích đoạn của một hội thảo hay trình bày chuyên đề bằng cách ghi chép để sau đó viết cái mà người học hoặc thí sinh sẽ phải và có khả năng trả lời các câu hỏi, tức là hiểu được văn bản. Tiếp theo người ta tiến hành các item nhằm kiểm tra xem người học hoặc thí sinh có thể hiểu được cái mà họ phải có khả năng hiểu từ văn bản nghe. Thủ pháp này cũng sẽ có thể sử dụng đối với các văn bản nghe ngắn hơn. Điều quan trọng là phải giữ lại các *tiểu mục* đủ cách biệt riêng rẽ giữa *tiểu mục* này với *tiểu mục* khác để sao cho thí sinh có may mắn được đánh giá đối với từng *tiểu mục*. Các *tiểu mục* phải sử dụng tất cả các *từ-chìa khoá* xuất hiện đồng thời trong các *tiểu mục* và trong văn bản nghe. Ví dụ, một *tiểu mục* có thể đòi hỏi thí sinh điểm quan trọng thứ hai mà một giáo sư nhấn mạnh trong báo cáo của mình tại một hội thảo và thí sinh sẽ nghe thấy "*Điểm quan trọng thứ hai mà tôi muốn nhấn mạnh…*". Câu hỏi không nhất thiết và không nên giống hoàn toàn như thông tin chứa trong văn bản nghe, nhưng các thông tin chứa trong văn bản nghe phải đủ cho phép thí sinh trả lời được câu hỏi đó.

Ý đồ nghe và các câu hỏi có thể được biên soạn bằng tiếng mẹ đẻ của thí sinh. Thậm chí thí sinh cũng có thể sử dụng tiếng mẹ đẻ của mình để trả lời các câu hỏi (Bolton, 1987). Các nhà thiết kế và xây dựng đề thi và bài kiểm tra phải dự đoán được thời gian mà thí sinh cần để trả lời các câu hỏi.

Việc soạn các tiểu mục có ý nghĩa rất quan trọng vì nó góp phần quyết định chất lượng của một trắc nghiệm. Chính các tiểu mục cho phép đánh giá tính phân loại (niveau de discrimination), độ khó của trắc nghiệm.

Dưới đây là một số dạng câu hỏi hay kỹ thuật đặt câu hỏi:

- Câu hỏi đúng-sai-không biết hoặc không thể trả lời được (Vrai-Faux-Impossible)

Trước đây người ta chỉ dùng kỹ thuật loại "Đúng/Sai" song do nhược điểm của loại kỹ thuật này là thí sinh có thể trả lời đúng với xác suất 50% mà không cần nghe văn bản. Ngày nay người ta phải thêm vào "*?*" hoặc "*Không thể trả lời được*" để giảm sắc xuất trả lời đúng một cách ngẫu nhiên. Loại kỹ thuật này thường đòi hỏi thí sinh khẳng định hoặc bác bỏ thông tin chứa trong câu hỏi và thường kèm theo "?" hoặc "không biết", có nghĩa là thông tin chứa trong đoạn văn nghe không cho phép thí

sinh thể khẳng định hoặc bác bỏ câu hỏi. Ngoài ra, ưu diểm của loại kỹ thuật này có thể giúp cho người dạy bổ sung điều chỉnh nội dung giảng dạy.

- Câu hỏi nhiều sự lựa chọn (Questions à choix multiples)

Đây là loại câu hỏi có nhiều lựa chọn, có nghĩa là thí sinh phải chọn một câu trả lời trong một số câu trả lời gợi ý mà trong các câu trả lời gợi ý nay chỉ có một câu đúng hoặc đúng hơn cả. Khó khăn ở đây là các câu hỏi và các câu trả lời gợi ý cũng sẽ đòi hỏi thí sinh phải hiểu ngoài việc phải hiểu văn bản đọc. Các câu hỏi có thể không phản ánh hiểu văn bản. Vả lại, khi mà thí sinh trả lời sai sẽ khó có thể biết được lý do thí sinh không hiểu văn bản hay thí sinh không hiểu câu hỏi. Loại kỹ thuật này cho phép đo kiểm tra sản phẩm hơn là quá trình. Phương pháp cho điểm khách quan, bởi vì người chấm không phải đánh giá gì thêm.

- Câu hỏi với câu trả lời ngắn (Questions à réponse brève)

Trong đánh giá nghe hiểu, sẽ là lý tưởng nếu các tiểu mục ngắn tức là câu trả lời ngắn. Điều này sẽ giúp thí sinh chỉ phải viết ít nhất có thể.

- Chuyển giao thông tin (Transfert d'information)

Kỹ thuật này rất có ích trong đánh giá nghe hiểu như trong đánh giá đọc hiểu bởi vì để trả lời các câu hỏi thí sinh phải viết rất ít. Nó rất thích hợp với các hoạt động sau đây: nhận biết một sơ đồ hay một hình ảnh, điền một tờ khai hải quan chẳng hạn, hoặc chỉ đường theo một bản đồ có sẵn.

- Ghi lại thông tin (Prise de notes)

Kỹ thuật này đòi hỏi thí sinh sản sinh lại các thông tin mà họ đã nghe được ở một buổi phát thanh truyền hình, một buổi hội thảo hay một tài liệu nghe của một trắc nghiệm.

Bước 4: Thực hiện trắc nghiệm (administrer le test)

Giống như các trắc nghiệm hiểu, điều quan trọng là phải cung cấp cho người học hoặc thí sinh thông tin ngữ cảnh mà họ cần để tiếp cận văn bản nghe. Cũng sẽ là quan trọng là phải bảo đảm cho họ hiểu rõ nhiệm vụ mà họ phải làm, tức là hiểu rõ câu hỏi và bảo đảm cho họ có đủ thời gian để trả lời các câu hỏi đó. Tuỳ theo trình độ mà người ta sẽ quyết định các câu hỏi hay yêu cầu thi bằng tiếng nước ngoài hay tiếng mẹ đẻ của người học hay thí sinh.

Bước 5: Chấm và cho điểm

Đánh giá năng lực nghe hiểu cũng như năng lực đọc hiểu, các lỗi ngữ pháp, chính tả không tính, có nghĩa là người ta không cho trừ điểm các lỗi này. Người ta chỉ quan tâm đến câu trả lời đúng về phương diện ngữ nghĩa, tức là quan tâm đến thông tin, nghĩa của thông tin. Đối với câu hỏi dạng "Đúng/Sai, nếu đòi hỏi phải chứng minh, thì cho thêm điểm.

Bước 6: Phân tích các tiểu mục của trắc nghiệm và các nhiệm vụ

Nhất thiết phải phân tích độ khó của trắc nghiệm. Độ khó của trắc nghiệm tương ứng với độ dễ hay độ khó của các tiểu mục. Điều này sẽ cho phép sử dụng loại tiểu mục nào thì thích hợp. Ví dụ, nếu quá nhiều thí sinh gặp khó khăn với trắc nghiệm điền chỗ trống (Test de closure) có thể ta sẽ dùng loại "bảy từ gạch một từ" thay vì "5 từ gạch một từ", vv Độ khó của trắc nghiệm được tính theo công thức: độ khó = số học sinh làm đúng các tiểu mục chia cho tổng số học sinh trả lời các tiểu mục. Độ khó càng cao thì tiểu mục càng dễ.

Ví dụ trong một trắc nghiệm đã cho, 10 trên tổng số 20 học sinh trả lời đúng câu hỏi hay xử lý đúng một tiểu mục nào đó, độ khó sẽ là 10/20 = 0,50 - độ khó trung bình, tiểu mục khó vừa phải;

Trong một trắc nghiệm khác, 10 trên tổng số 100 học sinh trả lời đúng một câu hỏi hay xử lý đúng một tiểu mục nào đó, độ khó sẽ là 10/100 = 0,1 - độ khó nhỏ, tiểu mục rất khó;

Hoặc trong một trắc nghiệm khác nữa, 50 trên tổng số 50 học sinh trả lời đúng một câu hỏi hay xử lý đúng một tiểu mục nào đó, độ khó sẽ là 50/50 = 1 - độ khó quá lớn, tiểu mục quá khó.

Người ta vẫn có thể giữ lại một trắc nghiệm mà độ khó rất nhỏ, tức là độ khó của trắc nghiệm rất lớn tuỳ theo mục đích của đánh giá. Về mặt lý thuyết, độ khó của tiểu mục được phép dao động từ 0,4 đến 0,7.

Mặt khác cũng tuỳ thuộc vào mục tiêu của đánh giá: đánh giá tiêu chí hay đánh giá chuẩn. Chính vì vậy ta phải loại bỏ và thay thế các tiểu mục "kém chất lượng", phải xem lại nhiệm vụ của thí sinh căn cứ vào mục tiêu dạy hay mục tiêu đánh giá. Thậm chí có thể thay thế cả văn bản nghe nếu thấy cần thiết.

Về độ phân loại của trắc nghiệm, đây là một đặc tính rất quan trọng của một trắc nghiệm. Độ phân loại lý tưởng của một trắc nghiệm là 1. Tuy nhiên trong thực tế ít khi đạt được độ phân loại này (về công thức tính độ phân loại tương đối phức tạp, chúng tôi sẽ trình bày trong một bài viết khác). Về lý thuyết, độ khó nên dao động từ 0,3 đến 0,8.

Bước 7: Tổng kết, đánh giá kết quả trắc nghiệm

Sẽ là rất quan trọng và cần thiết xem lại các trắc nghiệm đã sử dụng, tìm ra được mặt mạnh, mặt yếu của nó giúp cho việc thiết kế và xây dựng các trắc nghiệm sau tốt hơn. Và sẽ là ích lợi khi sử dụng lại các trắc nghiệm đã sử dụng bằng cách thay đổi ít nhiều cho phù hợp, hơn thế nữa, sau một, hai năm trắc nghiệm đã được sử dụng đưa ra sử dụng lại vẫn bảo đảm độ tin cậy và giá trị của nó, vì người học, thí sinh không phải là các học sinh, thí sinh trước nữa.

Trên đây, chúng tôi vừa trình bày một số vấn đề mang tính lý luận và các bước tiến hành cụ thể khi xây dựng các công cụ kiểm tra-đánh giá, cụ thể là các nguyên tắc, các kỹ thuật và các bước tiến hành thiết kế một trắc nghiệm đánh giá kỹ năng nghe hiểu. Để xây dựng các trắc nghiệm có chất lượng, đòi hỏi các tác giả soạn thảo đề thi hay kiểm tra, giáo viên phải nắm chắc lý luận kiểm tra-đánh giá đặc biệt là nắm vững các nguyên tắc và kỹ thuật thiết kế, xây dựng các loại hình trắc nghiệm khác nhau, soạn thảo các loại tiểu mục khác nhau. Trong các bài viết tiếp, chúng tôi sẽ cố gắng trình bày các nguyên tắc, các kỹ thuật và các bước tiến hành thiết kế và xây dựng các trắc nghiệm đánh giá các kỹ năng đọc hiểu, diễn đạt nói, diễn đạt viết, vv

TÀI LIỆU THAM KHẢO

1. Bachman, L.F., *Fundamental considerations in language testing*, Oxford: Oxford University Press, 1990.
2. Bolton, S., *Evaluation de la compétence communicative en langue étrangère, Traduit de l'allemand par Bertrand*, Paris: Hatier, 1987.
3. Carroll, J.B., The Psychology of language testing, In A. Davis (Ed*.), Language Testing Syposium, A psycholinguistic perspective,* London: Oxford University Press, 1968, pp.46-69.
4. Cornaire, C., and Bayliss, D., *Les stimuli des choix multiples aux niveaux intermédiaire fort et avancé: doit-on privilégier la question ou l'énoncé*? *AQEFLS, vol.17, N$_0$ **3-4**,* 1996, p.132-139.
5. Dénommé, J. and Roy, M., *Pour une pédagogie interactive*, Montréal: Gaëtan Morin, 1998.
6. Gonlund, N.E., *Mesurement and Evaluation in teaching*, New York: Macmillan, 1976.
7. Holmes, B., and Roser, N., *Five ways to assess Readers' Prior Knowledge*, The Teaching Teacher, 40**(7),** 1987, p.646-650.
8. Lussier, D., *Evaluer les apprentissages dans une approche communicative*, Paris: Nathan, 1992.
9. Scallon, G., *L'évaluation des apprentissages dans une approche par compétences*, Editions du Renouveau pédagogique Inc, 2004.
10. Ministère de l'Education, Direction Générale du développement pédagogique, *Guide docimologique: Conseils pratiques pour la construction d'un instrument de mesure*, Québec: Gouvernement du Québec, No**5**, 1981.