

VIỆN NGHIÊN CỨU CON NGƯỜI
TS. Nguyễn Công Khanh

ĐÁNH GIÁ VÀ ĐO LƯỜNG

TRONG
KHOA HỌC XÃ HỘI

QUY TRÌNH, KỸ THUẬT THIẾT KẾ, THÍCH NGHI,
CHUẨN HÓA CÔNG CỤ ĐO



NHÀ XUẤT BẢN CHÍNH TRỊ QUỐC GIA

**ĐÁNH GIÁ VÀ
ĐO LƯỜNG**
TRONG
KHOA HỌC XÃ HỘI

VIỆN NGHIÊN CỨU CON NGƯỜI
TS. Nguyễn Công Khanh

ĐÁNH GIÁ VÀ ĐO LƯỜNG TRONG KHOA HỌC XÃ HỘI

**QUY TRÌNH, KỸ THUẬT THIẾT KẾ, THÍCH NGHI,
CHUẨN HÓA CÔNG CỤ ĐO**

NHÀ XUẤT BẢN CHÍNH TRỊ QUỐC GIA
Hà Nội - 2004

LỜI NHÀ XUẤT BẢN

Bước vào thế kỷ XXI, Đảng và Nhà nước ta khẳng định phát triển khoa học công nghệ cùng với phát triển giáo dục và đào tạo là quốc sách hàng đầu, là nền tảng và động lực đẩy mạnh công nghiệp hoá, hiện đại hoá đất nước. Nhiệm vụ chủ yếu của khoa học xã hội và nhân văn là hướng vào việc giải quyết các vấn đề lý luận và thực tiễn, dự báo các xu thế phát triển, cung cấp luận cứ khoa học cho việc hoạch định đường lối, chủ trương, chính sách phát triển kinh tế - xã hội, xây dựng con người.

Trong nghiên cứu khoa học xã hội, việc đánh giá và định lượng các kết quả nghiên cứu và thực nghiệm có vai trò quan trọng đối với việc đưa ra những kết luận và những giải pháp. Nhằm cung cấp cho bạn đọc, đặc biệt là các cán bộ nghiên cứu, nghiên cứu sinh và sinh viên chuyên ngành khoa học xã hội có được một số khái niệm cơ bản và các kỹ năng thực hành nghiên cứu, Nhà xuất bản Chính trị quốc gia xuất bản cuốn sách *"Đánh giá và đo lường trong khoa học xã hội (Quy trình, kỹ thuật thiết kế, thích nghi, chuẩn hoá công cụ đo)"* của TS. Nguyễn Công Khanh, Viện Nghiên cứu con người.

Nội dung cuốn sách trình bày phương pháp luận, quy trình và các nguyên tắc, kỹ thuật thiết kế công cụ đo lường; các phương pháp phân tích item, chọn mẫu, đánh giá độ tin

cây, độ hiệu lực, thiết kế công cụ đo cũng như các bước cơ bản thực hành các kỹ năng thu thập, xử lý, thích nghi hoá dữ liệu đó. Ngoài ra, trong phần phụ lục còn đưa ra các mô hình xử lý số liệu và bảng phiếu hỏi mẫu để bạn đọc tham khảo.

Xin giới thiệu cuốn sách với bạn đọc.

Tháng 4 năm 2004

NHÀ XUẤT BẢN CHÍNH TRỊ QUỐC GIA

THAY LỜI TỰA

Thực tiễn xã hội luôn có những vấn đề cần phải nghiên cứu hay những câu hỏi cần phải giải đáp. Thực hành các nghiên cứu đánh giá và đo lường trong khoa học xã hội thực chất là xem xét giải quyết những vấn đề (hay những câu hỏi) của thực tiễn trên các bình diện: lý thuyết, phương pháp nghiên cứu và xử lý phân tích kết quả nghiên cứu.

Bất kỳ thủ pháp nào được sử dụng để tập hợp, thu thập thông tin (quan sát, phỏng vấn, nghiên cứu hồ sơ, tư liệu, sản phẩm,... bảng hỏi, trắc nghiệm) về đối tượng cần đánh giá, nhằm mục đích nào đó (làm quyết định, so sánh với chuẩn...) đều được gọi là đánh giá. Đánh giá trong khoa học xã hội thực chất là mô tả, lý giải sự có mặt hay vắng mặt (tần suất, mức độ quan trọng... của những biểu hiện cụ thể) của đặc tính cần đánh giá (kiến thức, thái độ, kỹ năng, nhu cầu hứng thú... hay đặc điểm nhân cách).

Đo lường trong khoa học xã hội thường liên quan đến việc phạm trù hoá hay lượng hoá các hiện tượng hay thuộc tính theo tiêu chí/ bản chất (định tính/đo lường về chất lượng) hoặc liên quan đến việc sử dụng các con số vào quá trình lượng hoá các hiện tượng hay thuộc tính

đó (định lượng/ đo lường về số lượng) để mô tả hoặc đánh giá. Để đo lường người ta sử dụng các thang đo như: phiếu trưng cầu, bảng hỏi, phiếu điều tra, bảng nghiệm kê, bảng liệt kê... trắc nghiệm nhằm phục vụ cho mục tiêu đánh giá (như đo lường kiến thức, thái độ, kỹ năng, thuộc tính hay phẩm chất).

Do đối tượng đo lường trong khoa học xã hội thường không giống như trong khoa học tự nhiên, thường phức tạp, khó chính xác, khó đo lường trực tiếp; nên công việc thiết kế công cụ đo lường không thể tiến hành tùy tiện, đặc biệt là thiết kế trắc nghiệm. Quá trình thiết kế công cụ đo lường đòi hỏi phải tuân thủ các nguyên tắc, quy trình, kỹ thuật thiết kế. Người thiết kế phải biết cách phân tích, đánh giá kiểm tra các đặc tính thiết kế, các đặc tính đo lường của công cụ trước khi dùng những công cụ này đánh giá, thu thập số liệu.

Để giúp cho sinh viên, nghiên cứu sinh, cán bộ nghiên cứu có được những kỹ năng cơ bản thực hành nghiên cứu, chúng tôi biên soạn cuốn sách *"Đánh giá và đo lường trong khoa học xã hội (Quy trình kỹ thuật, thiết kế, thích nghi, chuẩn hoá công cụ đo)"*. Cuốn sách này đề cập: cơ sở phương pháp luận, quy trình nghiên cứu; các nguyên tắc, quy trình kỹ thuật thiết kế công cụ đo; các phương pháp phân tích item; phương pháp chọn mẫu; phương pháp đánh giá độ tin cậy; phương pháp đánh giá độ hiệu lực. Đặc biệt, cuốn sách còn cung cấp các kỹ năng thực hành thiết kế một phép đo dùng cho việc đánh giá thực trạng; kỹ năng

thích nghi và chuẩn hoá một trắc nghiệm.

Mặc dù đã rất cố gắng, song cuốn sách khó tránh khỏi những hạn chế, do vậy tác giả mong nhận được những ý kiến góp ý để tiếp tục sửa chữa hoàn chỉnh cho lần xuất bản sau.

Hà Nội, tháng 4-2004

TÁC GIẢ

MỤC LỤC

	<i>Trang</i>
<i>Lời Nhà xuất bản</i>	5
<i>Thay lời tựa</i>	7
Chương I: Phương pháp luận và những nguyên tắc kỹ thuật khi thực hành nghiên cứu, trong khoa học xã hội	15
1. Những cơ sở lý luận khi thực hành các nghiên cứu trong khoa học xã hội	15
2. Những nhân tố chi phối, hạn chế, hoặc làm sai lạc kết quả nghiên cứu	19
3. Quy trình thực hành nghiên cứu	21
4. Những câu hỏi hướng dẫn, kiểm tra những thiếu sót, sai lầm khi triển khai các bước của quy trình nghiên cứu	23
Chương II: Đánh giá, đo lường và các kiểu thang đo đặc trưng	29
1. Khái niệm đánh giá và đo lường	29
2. Các kiểu đo lường đặc trưng	30
3. Trắc nghiệm và các phép đo không phải là trắc nghiệm	35
Chương III: Quy trình và các nguyên tắc, kỹ thuật thiết kế công cụ đo lường	40
1. Các bước cơ bản của quy trình thiết kế công cụ đo lường	40

2. Những vấn đề và các nguyên tắc thiết kế công cụ đo	47
3. Những kỹ thuật thiết kế item	58
Chương IV: Các kỹ thuật phân tích item	70
1. Phân tích độ khó của item	70
2. Phân tích độ phân biệt của item	75
3. Phân tích năng lực gây nhiễu của những lựa chọn thay thế	84
Chương V: Các phương pháp chọn mẫu	87
1. Lý thuyết chọn mẫu	87
2. Các phương pháp chọn mẫu cơ bản	89
Chương VI: Một số phương pháp cơ bản đánh giá độ tin cậy của công cụ đo lường	103
1. Mô hình lý thuyết về độ tin cậy	103
2. Các phương pháp cơ bản đánh giá độ tin cậy của phép đo	107
3. Những yếu tố ảnh hưởng đến độ tin cậy của trắc nghiệm	112
Chương VII: Độ hiệu lực và các phương pháp đánh giá độ hiệu lực	115
1. Các lý thuyết làm cơ sở cho sự phân tích độ hiệu lực	115
2. Các kiểu phân tích độ hiệu lực	117
3. Các phương pháp cơ bản đánh giá độ hiệu lực	120
4. Những yếu tố ảnh hưởng đến độ hiệu lực	121
Chương VIII: Thích nghi và chuẩn hoá công cụ	123
1. Chuẩn hoá công cụ	123
2. Thích nghi hoá công cụ	133
Chương IX: Thực hành các kỹ năng thiết kế công cụ đo	135
1. Mục tiêu, nội dung, đối tượng điều tra và	

phương pháp chọn mẫu	135
2. Các khái niệm công cụ, các câu trắc và các chỉ số	137
3. Các nguyên tắc thiết kế công cụ	141
4. Quy trình thiết kế công cụ	142
5. Mô tả bộ công cụ điều tra	146
6. Nguyên tắc và các phương pháp xử lý số liệu	148
7. Điều tra thử, kiểm tra các đặc tính thiết kế và các đặc tính đo lường của Bộ công cụ	149
8. Đánh giá độ tin cậy và độ hiệu lực của công cụ	159
Chương X: Thực hành các kỹ năng thích nghi hoá công cụ đo	176
1. Các khái niệm công cụ	176
2. Xuất xứ trắc nghiệm	178
3. Các thông số chuẩn hoá của trắc nghiệm	181
4. Quá trình thiết kế và xây dựng trắc nghiệm	184
5. Quy trình thích nghi hoá trắc nghiệm	192
6. Đánh giá lại các tham số định tâm của trắc nghiệm	194
7. Đánh giá lại độ tin cậy của trắc nghiệm	196
8. Đánh giá lại độ hiệu lực của trắc nghiệm	201
Chương XI: Thực hành các kỹ năng xây dựng trắc nghiệm đánh giá kiến thức kỹ năng	211
1. Mục tiêu của trắc nghiệm	211
2. Nội dung đánh giá	211
3. Các bước của quy trình thiết kế trắc nghiệm	213
Phụ lục 1. Mô hình xử lý số liệu	224
Phụ lục 2. Các công đoạn xử lý số liệu	226
Phụ lục 3. Phiếu hỏi	232
Tài liệu tham khảo	253

Chương I

PHƯƠNG PHÁP LUẬN VÀ NHỮNG NGUYÊN TẮC KỸ THUẬT KHI THỰC HÀNH NGHIÊN CỨU, ĐÁNH GIÁ TRONG KHOA HỌC XÃ HỘI

1. Những cơ sở lý luận khi thực hành các nghiên cứu trong khoa học xã hội

1.1. Lý thuyết

Xác định vấn đề cần được nghiên cứu, câu hỏi cần được giải đáp là cái gì, nhằm đạt tới đâu, những nội dung nào cần tập trung, những phạm vi nào cần giới hạn... Điều này có nghĩa là phải xác định rõ mục đích nghiên cứu, các mục tiêu nghiên cứu cụ thể, giới hạn các nội dung nghiên cứu, làm rõ các khái niệm... các cấu trúc cụ thể cần đo lường...

1.2. Phương pháp nghiên cứu

Chọn lựa các phương pháp nghiên cứu phù hợp với mục tiêu, nội dung, nhiệm vụ nghiên cứu. Trong đó quan trọng nhất là phải thiết kế hay chọn lựa được các công cụ đo lường có độ tin cậy, có độ hiệu lực, phải phù hợp với khách thể nghiên cứu... để có thể thu thập được những số liệu đáng tin cậy nhằm đáp ứng các mục tiêu và nhiệm vụ

ngiên cứu.

Thiết kế công cụ nghiên cứu là công việc khó khăn, mất thời gian, tốn kém nhiều công sức, do vậy thông thường người ta thích sử dụng các bộ công cụ của nước ngoài. Việc du nhập những bộ công cụ của nước ngoài là cách làm ít tốn kém, nhưng việc sử dụng những bộ trắc nghiệm dịch từ nước ngoài không qua quá trình thích nghi hoá - chuẩn hoá lại (trên mẫu đại diện của Việt Nam) nhằm xác định lại độ tin cậy, độ hiệu lực, những sai số của sự đo lường... là điều rất nguy hiểm (bởi vì tất cả những trắc nghiệm bằng lời đều có bản chất văn hoá-xã hội, sẽ không thích hợp nếu dùng cho đối tượng ở các nền văn hoá-xã hội khác mà không qua thích nghi và chuẩn hoá lại).

Công việc thích nghi hoá trắc nghiệm không phải chỉ là dịch nghĩa, dịch bản hướng dẫn sử dụng (manual) mà cái chính là phải biết được xuất xứ, mục tiêu đo lường, các tiêu chuẩn thiết kế, các thông số kỹ thuật (độ tin cậy, độ hiệu lực...), các mẫu dùng để chuẩn hoá... Tức là phải hiểu "nguồn gốc": trắc nghiệm này xuất phát từ đâu, được thiết kế để đo cái gì, dựa trên mô hình lý thuyết nào, gồm những miền đo/tiêu chí nào, dựa trên những nguyên tắc hay phương pháp nào để thiết kế và hiệu lực hoá các item, từng item có nội dung đo lường cụ thể nào... đã áp dụng kỹ thuật chọn mẫu nào, sử dụng những nhóm mẫu nào để chuẩn hoá, sử dụng những kỹ thuật nào để phân tích... từ đó tìm cách *chuyển ngữ tương đương, xác định lại các thông số kỹ thuật...* đưa ra những cảnh báo thận trọng khi sử dụng kết quả (điểm số của trắc nghiệm) để đánh giá,

chẩn đoán.

Nhiều nghiên cứu không tìm được những bộ công cụ đo lường thích hợp sẵn có, do vậy người nghiên cứu phải tự thiết kế. Nếu tự thiết kế, tùy mục đích nghiên cứu, mục tiêu đo lường, công việc thiết kế có mức độ phức tạp khác nhau, nhưng để bảo đảm tính khoa học, việc thiết kế công cụ nhất thiết phải tuân thủ các nguyên tắc, quy trình thiết kế và phải đáp ứng các tiêu chuẩn thiết kế, tiêu chuẩn đo lường.

1.3. Chọn mẫu

Mẫu có vai trò quan trọng đặc biệt, quyết định giá trị khoa học của những kết luận rút ra từ số liệu thu thập trên mẫu đó. Kết quả nghiên cứu trên một mẫu nào đó chỉ có thể khái quát hoá khi mẫu đó có tính đại diện cho một tổng thể (population) nào đó.

1.4. Thống kê - xử lý phân tích dữ liệu

Dựa theo mô hình lý thuyết (tùy mục đích nghiên cứu) chọn lựa mô hình xử lý thích hợp. Việc chọn lựa được các chiến lược xử lý - phân tích phù hợp nhất làm tăng giá trị khoa học của công trình nghiên cứu. Bất kỳ mô hình, chiến lược hay một kiểu xử lý cụ thể nào được chọn lựa cũng phải cân nhắc, đánh giá các điều kiện cơ bản, thoả mãn các giả thuyết điều kiện. Chẳng hạn kích cỡ đủ lớn của mẫu, tính đại diện của mẫu, tính chuẩn của phân bố điểm trên mẫu, tính đồng nhất của phương sai... khi dùng phép thử t-Student để đánh giá sự khác biệt của trị số trung bình của hai mẫu độc lập. Cũng vậy khi muốn đánh

giá hệ số tin cậy alpha của phép đo bằng phép thử độ phù hợp của item, phải thoả mãn các điều kiện: thang đo khoảng hay thang tỷ lệ, các item phải có nhiều mức độ, được điểm hoá (cách cho điểm thống nhất theo cùng một thứ bậc), các item phải có độ khó tương đương với nhau... trước khi sử dụng mô hình tương quan alpha của Cronbach.

Số liệu sau khi vào máy tính, phải được kiểm tra lỗi nhập dữ liệu, phải được làm sạch (loại bỏ những phần tử có điểm bất thường) trước khi xử lý - phân tích. Chỉ khi nào nhà nghiên cứu thoả mãn hoặc “không vi phạm” các điều kiện giả thiết của một phép kiểm định thống kê nào đó thì những số liệu từ kết quả phân tích xử lý mới có giá trị để rút ra những kết luận hoặc suy đoán.

Để số liệu có giá trị khoa học khi công bố, công trình nghiên cứu sử dụng công cụ đo đạc nào (dù đã được chuẩn hoá) nếu khách thể nghiên cứu có những đặc tính khác với khách thể dùng để chuẩn hoá phép đo, thì khi phân tích số liệu phải khẳng định lại các thông số về độ tin cậy và độ hiệu lực của phép đo...

Nhà nghiên cứu phải có kỹ năng đọc hiểu các số liệu làm cho các dữ liệu “biết nói”. Khi phân tích phải dựa trên số liệu thực tế thu thập hay đo lường được đối chiếu với mô hình lý thuyết để xem số liệu thoả mãn đến mức nào mục tiêu, nhiệm vụ nghiên cứu, điều này phụ thuộc đáng kể vào trình độ kinh nghiệm và tay nghề nghiên cứu.

Thực tế tham gia xử lý phân tích số liệu cho rất nhiều đề tài, dự án, chúng tôi nhận thấy khâu này còn nhiều bất

cập, chẳng hạn như:

- Không có mô hình xử lý và phân tích số liệu dựa trên những giả thuyết hợp lý đáp ứng mục tiêu đo lường ngay từ lúc thiết kế trắc nghiệm (ví dụ xem Phụ lục mô hình xử lý số liệu Đề tài KX-05-06).

- Bỏ qua hoặc ẩn bớt các công đoạn xử lý. Chẳng hạn không xem công đoạn làm sạch số liệu có vai trò quan trọng đặc biệt (xem Phụ lục các công đoạn xử lý và phân tích số liệu Đề tài KX-05-06).

- Không đánh giá lại các đặc tính đo lường của công cụ trước khi sử dụng những số liệu được thu thập bằng công cụ này.

- Sử dụng các phép toán thống kê không thích hợp, bỏ qua những điều kiện cần và đủ để những phép toán thống kê này được phép áp dụng.

- Thiếu kỹ năng đọc hiểu số liệu, giải thích số liệu một cách khoa học.

Những bất cập này chỉ có thể khắc phục triệt để bằng cách gửi họ đi đào tạo. Các chương trình đào tạo đại học, sau đại học cần dạy môn học này.

2. Những nhân tố chi phối, hạn chế, hoặc làm sai lạc kết quả nghiên cứu

Tại sao cùng một vấn đề của thực tiễn cần phải giải đáp, các nhà nghiên cứu lại đưa ra những số liệu khác nhau hoặc cũng một số liệu thu được họ lại đưa ra những nhận định, kết luận rất khác nhau (thậm chí mâu thuẫn cao độ). Câu trả lời có thể là: do các nhà

ngiên cứu đã sử dụng những cách tiếp cận khác nhau, dùng các phương pháp nghiên cứu khác nhau hoặc họ đã bỏ qua hay đơn giản hoá một bước nào đó trong quy trình nghiên cứu hoặc phạm phải một số lỗi trong khi tiến hành nghiên cứu.

Thông thường khi nghiên cứu chúng ta hay phạm phải những sai sót sau :

- Vấn đề nghiên cứu quá rộng hay quá hẹp, không xác định rõ khái niệm, hoặc khái niệm được xác định bao hàm những thành phần không thể đo đạc được.

- Không thao tác hoá khái niệm thành những miền đo cụ thể, thành những chỉ số có thể đo lường được.

- Dùng những phương pháp nghiên cứu đánh giá không phù hợp

- Sử dụng những công cụ đo không đáng tin cậy, không có độ hiệu lực hoặc đo đạc nhầm (đo không trúng cái cần phải đo).

- Không thống nhất trong cách hướng dẫn khi tiến hành đo, hoặc xác định thời điểm đo không hợp lý, số lần đo lặp lại quá nhiều, khoảng cách giữa các lần đo quá gần hoặc quá xa...

- Chọn mẫu không phù hợp, hoặc mẫu không đại diện, mẫu không đủ lớn.

- Áp dụng những mô hình và kỹ thuật xử lý - phân tích số liệu không phù hợp.

- Dựa trên số liệu không đầy đủ, không đại diện, không đủ tin cậy để khái quát hóa rút ra những kết luận, suy đoán.

- Dựa trên những thông tin cảm tính (*made-up information*), thiếu tính lô gích hoặc tính hợp lý để đánh giá, suy đoán.

- Những thành kiến của nhà nghiên cứu ảnh hưởng đến việc thu thập số liệu.

3. Quy trình thực hành nghiên cứu

Quá trình nghiên cứu phải được xây dựng như một quy trình “công nghệ” gồm nhiều bước nhằm trả lời các câu hỏi do mục tiêu nghiên cứu, nhiệm vụ và nội dung nghiên cứu đặt ra. Theo các chuyên gia, quy trình nghiên cứu thường gồm những bước cơ bản sau (xem sơ đồ quy trình nghiên cứu):

1. Tất cả các nghiên cứu đều bắt đầu với vấn đề cần nghiên cứu (ý tưởng) hay câu hỏi cần trả lời. Câu hỏi rất đa dạng, nó có thể có hình thức rất cụ thể, dân dã kiểu “tôi muốn biết cái gì có thể xảy ra nếu...?”. Câu hỏi có thể có hình thức khá trừu tượng, mơ hồ, kiểu khám phá những cấu trúc tâm lý: “Tôi muốn biết thực trạng chất lượng đạo đức và lối sống của sinh viên hiện nay như thế nào?”. Câu hỏi có thể có hình thức “chính thống” như một giả thuyết về kết quả nghiên cứu được mong đợi (còn gọi là giả thuyết khoa học) như: “Liệu rối nhiễu tâm lý tuổi vị thành niên có quan hệ mật thiết với xung đột trong quan hệ mẹ con...?”.

2. Sau khi câu hỏi nghiên cứu đã được xác định thì nhà nghiên cứu phải xem xét các lý thuyết, các kết quả nghiên cứu hiện có đã tiến triển tới đâu, liên quan như thế nào đến câu hỏi nghiên cứu? Đây còn gọi là nghiên cứu tổng quan về lý luận hay nghiên cứu lịch sử vấn đề để làm rõ cơ

sở lý luận của vấn đề cần nghiên cứu (khái niệm, cấu trúc...), làm rõ mục đích, mục tiêu nghiên cứu, phạm vi nghiên cứu... Nội dung nghiên cứu cần được chi tiết hoá, tốt nhất là lập thành một sơ đồ cụ thể hoá nội dung nghiên cứu thành những cấu trúc cần phải và có thể đo lường được.

3. Bước tiếp theo nhà nghiên cứu phải quyết định bằng cách nào để thu thập được số liệu, tức là những số liệu nào sẽ được thu thập và dùng những phương pháp nào để thu thập số liệu (những phép đo nào sẽ được sử dụng).

4. Thiết kế công cụ đo (nếu không có sẵn những công cụ đo thích hợp) theo một quy trình chặt chẽ, bám sát nội dung nghiên cứu và nội dung đo lường. Kiểm tra các đặc tính thiết kế của công cụ, đánh giá các đặc tính đo lường trên mẫu thử. Hiệu chỉnh công cụ đo.

5. Chọn mẫu và tiến hành đo đạc, thu thập số liệu như đã lập kế hoạch. Người nghiên cứu phải xem xét chọn kiểu mẫu nào là phù hợp với mục đích nghiên cứu (xem bài phương pháp chọn mẫu). Nếu là những nghiên cứu điều tra khảo sát thực trạng thì mẫu phải đủ lớn, phải có tính đại diện. Việc tiến hành đo đạc phải có huấn luyện trước.

6. Đưa số liệu vào xử lý phân tích. Người nghiên cứu phải định hướng mô hình, cách thức khai thác số liệu (xem Phụ lục mô hình xử lý số liệu Đề tài KX-05-06, KX-05-07). Việc xử lý và phân tích số liệu thường phải dựa theo mô hình. Lúc này nhà nghiên cứu cần sử dụng hợp lý các phép toán thống kê mô tả và thống kê suy luận để mô tả, kiểm định, phân tích, so sánh... theo

những mô hình lý thuyết khi thiết kế nhằm tìm ra câu trả lời có căn cứ khoa học cho các câu hỏi nghiên cứu.

7. Lý giải kết quả nghiên cứu. Nhà nghiên cứu không những chỉ tìm cách trả lời câu hỏi nghiên cứu mà còn đưa ra những định hướng ứng dụng kết quả nghiên cứu (thường gọi là ý nghĩa thực tiễn và lý luận của kết quả nghiên cứu).

4. Những câu hỏi hướng dẫn, kiểm tra những thiếu sót, sai lầm khi triển khai các bước của quy trình nghiên cứu

Để phát hiện những sai sót, nhà nghiên cứu cần đặt ra những câu hỏi khác nhau gắn liền với nội dung cần phải nghiên cứu ở từng giai đoạn. Sau đây là một số câu hỏi cơ bản cho một vài giai đoạn :

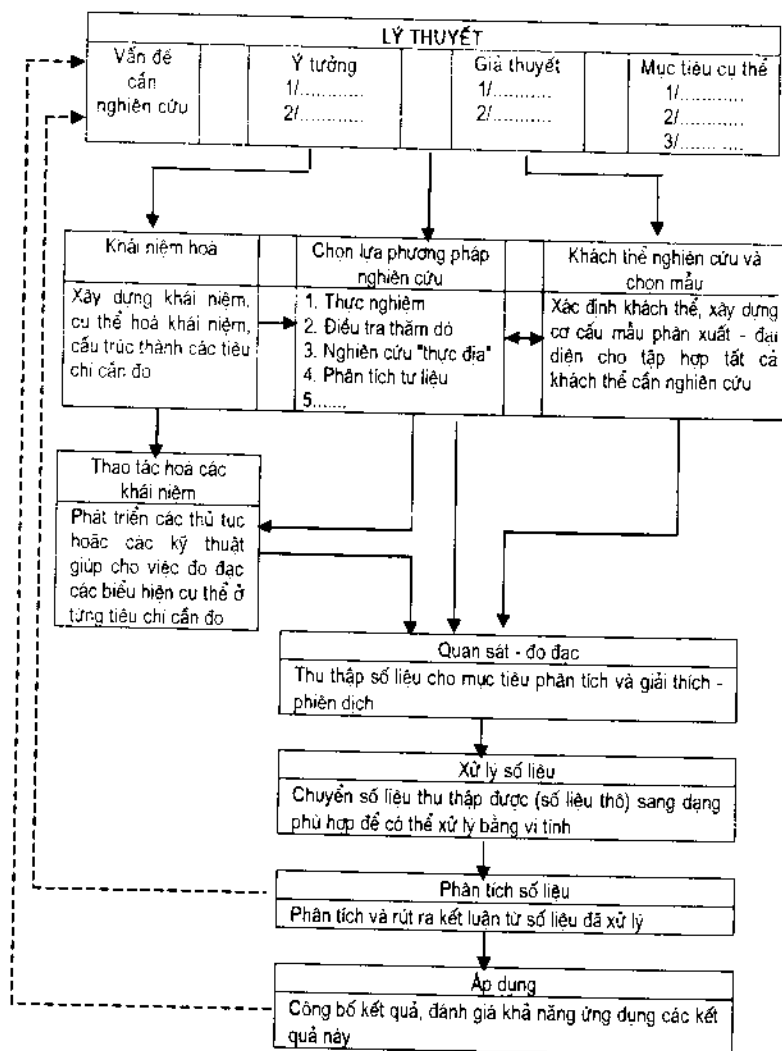
4.1. Tìm cơ sở lý luận phục vụ việc soạn thảo đề cương nghiên cứu

- Lịch sử vấn đề?
- Những khái niệm lý thuyết cơ bản?
- Mục tiêu của nghiên cứu là gì? (ví dụ : khám phá, mô tả, giải thích hay phối hợp). Các mục tiêu cụ thể tập trung nghiên cứu là gì?

Với từng mục tiêu cụ thể có những nội dung nghiên cứu nào có thể triển khai đo đạc được?

- Khách thể nghiên cứu là gì? Đơn vị phân tích là gì (cá nhân hay nhóm)? Liệu nó có phù hợp với mục tiêu cần đạt được không?...

Sơ đồ Quy trình nghiên cứu



4.2. Chọn lựa các phương pháp nghiên cứu

- Liệu các phương pháp nghiên cứu đã chọn có phù hợp với mục tiêu và nội dung nghiên cứu? Nên sử dụng những phương pháp cụ thể nào? Liệu có phương pháp nào khả dĩ tốt hơn? Chẳng hạn dùng phương pháp thực nghiệm có đối chứng: tại sao lại chọn phương pháp này? Những ảnh hưởng nào mà nhà nghiên cứu mong tìm? Liệu có những kiểu sai sót nào có thể ảnh hưởng đến kết quả thực nghiệm, ảnh hưởng đến độ tin cậy, độ hiệu lực của công cụ đo vì vậy làm sai lệch số liệu? Nhóm đối chứng và nhóm thực nghiệm được lựa chọn có phù hợp (các nghiệm thể có được xếp một cách ngẫu nhiên vào hai nhóm, có được đánh giá trước khi thực nghiệm...)?...

- Các phương pháp đo lường: những loại công cụ cụ thể nào được dùng để đo đạc? Nếu sử dụng các công cụ của người khác, liệu độ tin cậy, độ hiệu lực của các công cụ này có bảo đảm? Nếu là các bộ công cụ dịch của nước ngoài thì liệu các công cụ này đã được thích nghi? Tính khái quát của số liệu nghiên cứu nên được giới hạn đến đâu?

4.3. Thiết kế công cụ đo đạc

Nếu phải tự thiết kế công cụ đo thì từ mô hình lý thuyết định hướng, liệu nhà nghiên cứu có triển khai thành các bước cơ bản của quy trình thiết kế công cụ? Chẳng hạn như:

- Có xác định rõ khái niệm, nội hàm khái niệm, thao

tác hóa khái niệm thành các cấu trúc bộ phận với các miền đo cụ thể, gồm các tiêu chí cụ thể để có thể đo đạc được (xác định cái cần đo, các biểu hiện cụ thể có thể đo được ở từng tiêu chí hay chỉ số cụ thể)?

- Có xây dựng “bể item”, “ngân hàng item” (tức là có sử dụng các chuyên gia để viết item, có nảy sinh nhiều nhất các item cho từng chỉ số đo, phạm vi đo lường của các item có đủ bao trùm hết miền đo)?

- Có chọn lựa được các kiểu item phù hợp? Liệu hình thức biểu đạt của từng item có chỗ nào chưa hợp lý?

- Có thực hành việc hiệu lực hoá các item theo một quy trình với những tiêu chuẩn khách quan (chất lọc để từng item có tính đại diện, đơn nghĩa, dễ hiểu, thích hợp với trình độ đọc hiểu của nghiệm thể)? Liệu mỗi item có đảm bảo chỉ đo một biểu hiện cụ thể, đo đúng cái cần đo ở từng chỉ số?... Liệu số lượng các item có đủ nhiều để bao quát đầy đủ các khía cạnh của tiêu chí hay chỉ số muốn đo?

- Liệu lời hướng dẫn có đủ rõ ràng?

- Liệu cách tính điểm cho các item đã hợp lý? Điểm số có ý nghĩa gì, nó được sử dụng để làm gì?

- Khi đo trên mẫu thử cho kết quả đánh giá sơ bộ thế nào? Độ phù hợp, độ khó, độ phân biệt của từng item?...

- Mức độ đọc hiểu của công cụ liệu có phù hợp với đối tượng (khách thể)?

- Thời gian trung bình cho nghiệm thể hoàn thành trắc nghiệm là bao nhiêu phút?

- Trắc nghiệm được thử nghiệm cho kết quả thế nào

về các thông số đo lường (độ tin cậy, độ hiệu lực...).

4.4. Chọn mẫu

- Mẫu được chọn theo phương pháp nào?
- Tính đại diện của mẫu có bảo đảm?
- Mẫu có đủ lớn thích hợp cho các phép toán thống kê?

- Tính đồng nhất, tính chuẩn của mẫu có bảo đảm?...

4.5. Xử lý - phân tích số liệu

- Liệu các kỹ thuật mã hóa các thông tin về đối tượng trên từng tiêu chí đã hợp lý, có khả năng tách ra thành các nhóm để so sánh?

- Đã kiểm tra loại bỏ các lỗi khi nhập số liệu?

- Nếu có câu hỏi mở, thì chúng được mã hóa theo phương pháp nào ?...

- Liệu số liệu đã được làm sạch (đây là công đoạn rất quan trọng nhưng nhiều người lại xem nhẹ) để giảm những sai số trong thống kê phân tích?

- Liệu tính chuẩn của mẫu có được thoả mãn (phân bố điểm có gần với đường cong chuẩn...), sai số của mẫu có nằm trong phạm vi cho phép?

- Mô hình xử lý và phân tích số liệu dựa trên những giả thuyết nào, có hợp lý không? Những phép toán thống kê nào được dùng để kiểm tra tương quan, so sánh sự khác biệt điểm số giữa các nhóm đối tượng, mức độ có ý nghĩa?... Những giả thiết điều kiện của các phép toán thống kê này có bị vi phạm không?

- Liệu nhà nghiên cứu rút ra kết luận có căn cứ trên

các số liệu thu được? Phạm vi ứng dụng số liệu nghiên cứu?

- Những cảnh báo nào cần đưa ra? Những điểm cần lưu ý, cần thận trọng khi sử dụng số liệu là gì ?...

Tóm lại, mỗi bước trong quy trình nghiên cứu đều ẩn chứa những sai sót tiềm năng có thể làm hỏng hoặc làm giảm giá trị khoa học của công trình nghiên cứu. Trên thực tế không thể, hoặc khó có thể thực hiện hoàn hảo, chuẩn xác tất cả các bước của quy trình nghiên cứu. Vì vậy nhà nghiên cứu, người đánh giá, phản biện công trình nghiên cứu cần phân biệt những loại sai sót nào là nghiêm trọng không thể chấp nhận được (ví dụ: chọn sai phương pháp nghiên cứu, sử dụng những công cụ không đáng tin cậy, không có hiệu lực để đo hoặc vi phạm các nguyên tắc chọn mẫu, vi phạm những điều kiện của các phép kiểm định thống kê trong xử lý phân tích, đưa ra những kết luận khái quát không dựa trên số liệu thu được...) và những sai sót nào là không nghiêm trọng, có thể bỏ qua.

Chương II

ĐÁNH GIÁ, ĐO LƯỜNG VÀ CÁC KIỂU THANG ĐO ĐẶC TRƯNG

1. Khái niệm đánh giá và đo lường

Đánh giá (assessment) và đo lường (measurement) là hai khái niệm có nội hàm gần giống nhau, vì vậy thường dùng thay thế cho nhau, tuy nhiên vẫn có vài điểm khác. Bất kỳ thủ pháp nào được sử dụng để tập hợp, thu thập thông tin (quan sát, phỏng vấn, nghiên cứu hồ sơ, nghiên cứu văn bản, nghiên cứu sản phẩm,... bảng hỏi, trắc nghiệm) về đối tượng cần đánh giá, nhằm mục đích nào đó (làm quyết định, xác định hiệu quả tốt xấu, so sánh với chuẩn...) đều được gọi là đánh giá. Đánh giá trong khoa học xã hội thực chất là mô tả, lý giải sự có mặt hay vắng mặt (tần suất, mức độ quan trọng... của những biểu hiện cụ thể) của đặc tính cần đánh giá (kiến thức, thái độ, kỹ năng, nhu cầu, hứng thú... hay đặc điểm nhân cách).

Đánh giá hành vi của con người và các quá trình tâm thần thường là một quá trình thu thập thông tin nhờ sử dụng các phương pháp, thủ tục hay kỹ thuật như: quan sát, phỏng vấn, nghiên cứu hồ sơ, nghiên cứu sản

phẩm... hoặc qua các thang đo như: phiếu trưng cầu, bảng hỏi, phiếu điều tra, bản nghiệm kê, bảng liệt kê... và trắc nghiệm. Như vậy mọi sự đánh giá thường bắt đầu từ sự đo lường.

Đo lường trong các khoa học xã hội là sử dụng những thủ pháp hay kỹ thuật nhằm lượng hoá sự vật, hiện tượng phục vụ cho mục tiêu đánh giá (chẳng hạn, đo lường sự hiểu biết, kiến thức, kỹ năng..., cấu trúc, thuộc tính hay phẩm chất). Đo lường liên quan đến việc phạm trù hoá các sự kiện, hiện tượng hay thuộc tính theo bản chất/tiêu chí (định tính/đo lường về chất lượng) hoặc liên quan đến việc sử dụng các con số vào quá trình lượng hoá các sự kiện, hiện tượng hay thuộc tính (định lượng/đo lường về số lượng). Đối tượng đo lường trong các khoa học xã hội thường không giống như trong các khoa học tự nhiên, thường phức tạp, khó chính xác, khó đo lường trực tiếp. Chẳng hạn không có “thuớc đo” nào để đo lòng hiếu thảo của con cái đối với cha mẹ. Vì vậy các nhà nghiên cứu phải đặt ra các cách thức, phương pháp đặc trưng để biểu thị kết quả đo đạc.

2. Các kiểu đo lường đặc trưng

Những công cụ đo lường của các nghiên cứu hành vi và các quá trình tâm thần nói chung gồm bốn kiểu thang đo đặc trưng sau đây:

2.1. Thang định danh

Thang định danh (nominal scale) là một kiểu đánh giá sự vật, hiện tượng hay đặc tính theo tên gọi, theo

danh hiệu/nhãn hiệu, hoặc theo phạm trù. Chẳng hạn khi người ta muốn phân loại mọi người theo giới tính (nam/nữ), theo học lực (kém, yếu, trung bình, khá, giỏi), theo các nhóm tuổi (mẫu giáo, học sinh tiểu học...), hoặc khi người ta muốn phân loại đồ vật theo màu sắc (xanh, đỏ, vàng...) người ta sử dụng thang định danh.

Thang định danh là phép đo có tính khái quát, không nhằm cung cấp những thông tin chính xác về sự khác biệt cá nhân, mà chỉ nhằm đánh giá sự có mặt hay vắng mặt của sự vật, hiện tượng hoặc bản chất của thuộc tính.

Các con số trong phép đo định danh chỉ đơn thuần là con số mã hoá, biểu thị phạm trù (chẳng hạn: nam =1, nữ = 2 hoặc nông thôn =1, thành phố =2), các con số này không nói lên thứ bậc, mức độ, khối lượng của đặc tính cần đánh giá. Vì vậy thang định danh không phải là phép đo số lượng.

Các biến giới tính, khu vực, thành phần, chủng tộc... là các phép đo định danh. Tuy nhiên thang định danh cũng hay được chọn để đo các biến khác, chẳng hạn có thể phân loại một nhóm học sinh được đo trí tuệ (IQ) vào các nhóm điểm dưới trung bình, trung bình, trên trung bình.

Chú ý, khi ta chuyển điểm từ các thang đo khác vào thang định danh thì các thông tin trở nên khái quát hơn, nhưng cũng ít chính xác hơn.

2.2. Thang định hạng

Thang định hạng (ordinal scale) là một kiểu đánh giá phân loại sự vật, hiện tượng hay đặc tính theo thứ

bậc hay trật tự. Chẳng hạn khi người ta muốn phân loại năng lực học tập của một nhóm học sinh theo thứ hạng (xếp điểm tổng kết hay điểm thi môn toán theo thứ hạng từ học sinh có số điểm cao nhất-hạng nhất, đến học sinh có số điểm thấp nhất - hạng bét), hoặc khi muốn phân loại đồ vật theo kích cỡ, trọng lượng (từ to nhất/nặng nhất đến bé nhất/nhẹ nhất...) người ta sử dụng thang định hạng. Nếu 3 học sinh có số điểm bằng nhau, chỉ đứng sau 2 học sinh khác (tức là có thứ hạng 3, 4, 5) thì chúng được xếp cùng vào hạng 4 (trung bình cộng của 3, 4, 5).

Thang định hạng cũng là phép đo có tính khái quát, không nhằm cung cấp những thông tin chính xác về mức độ khác biệt giữa các cá nhân, mà chỉ nhằm chỉ ra vị trí của mỗi thành viên trong mối tương quan với các thành viên khác, hoặc vị trí của một sự vật, hiện tượng trong mối tương quan thứ bậc với các sự vật, hiện tượng khác.

Các con số trong phép đo định hạng đơn thuần chỉ biểu thị thứ hạng theo một kiểu xếp nào đó (chẳng hạn từ học sinh có điểm tổng kết môn toán cao nhất đến thấp nhất), các con số này không nói lên mức độ khác biệt, chất lượng cụ thể của đặc tính cần đánh giá. Chẳng hạn khi cần tuyển 120 thí sinh có điểm thi cao nhất trong kỳ thi tuyển vào đại học trong số hàng nghìn thí sinh thi, người ta chỉ cần biết điểm của 120 người có điểm cao nhất mà không quan tâm đến điểm cụ thể của từng người là bao nhiêu hoặc sự khác biệt giữa họ thế nào.

2.3. Thang định khoảng

Thang định khoảng (interval scale) là một kiểu đánh

giá phân loại sự vật, hiện tượng hay đặc tính theo những đơn vị phân loại bằng nhau ở bất kỳ khoảng nào trên thang đo. Ví dụ, phép đo chiều cao là một kiểu của thang định khoảng: sự khác biệt giữa người cao 160 cm - 165 cm và người cao 145 cm - 150 cm đều ở một khoảng như nhau là 5 cm.

Sự khác nhau theo những khoảng bằng nhau (biểu thị dưới dạng điểm số) trên thang đo ngầm chỉ sự khác nhau (theo những khoảng bằng nhau) ở những đặc điểm hay thuộc tính được đo. Ví dụ trên trắc nghiệm tiếng Anh, học sinh A đạt 9 điểm, học sinh B được 7 điểm, sự khác nhau này ngầm chỉ “năng lực tiếng Anh” của A hơn B một khoảng tương đương 2 điểm.

Hệ thống điểm của hầu hết các thang đo hành vi, các thang đo tâm thần được thiết kế theo thang định khoảng. Ví dụ, bốn học sinh A, B, C và D lần lượt nhận được các điểm trắc nghiệm là 65, 55, 45, 45 trên một trắc nghiệm nào đó (lo âu hay trầm cảm). Nếu đây là thang đo khoảng thì ta có thể rút ra những kết luận sau đây về điểm số (điểm số của từng học sinh cho phép xác định mức độ tương đối về mức độ lo lắng hay trầm cảm của họ) của 4 học sinh này: (1) *A có mức lo lắng cao hơn B, B có mức lo lắng cao hơn C, C có mức lo lắng ngang bằng với D*; (2) *Có thể nói sự khác nhau giữa A và B (10 điểm) cũng tương đương sự khác nhau giữa B và C (10 điểm)*; (3) *Có thể nói sự khác nhau giữa A và C là (20 điểm) gấp 2 lần giữa A và B.*

Mặc dù điểm của hầu hết các thang đo tâm lý, giáo

dục được coi là thang khoảng thì vẫn có rất nhiều ý kiến tranh luận về những giả thiết của kiểu thang đo này. Chẳng hạn, liệu tất cả các câu hỏi (item) của trắc nghiệm có độ khó ngang nhau? Liệu từ 60-70 điểm có thể đòi hỏi mức độ tương đối dễ hơn là từ 80-90 điểm, mặc dù cả 2 cùng khác là 10 điểm. Một trắc nghiệm trí tuệ (IQ test) mặc dù có hệ thống điểm theo thang khoảng, nhưng không có nghĩa rằng người có IQ là 110 và 120 cùng cách nhau một khoảng 10 điểm như người có IQ là 80 và 90 (do vậy nên xem đây là một phép đo định hạng).

Để tiện thực hành rất nhiều thang đo lượng hoá hành vi cần được giả thiết là thang định khoảng nhưng thực tế không nên tuyệt đối hoá khoảng cách như là thước đo cơ học (thước mét).

2.4. Thang định tỷ lệ

Khi cần phân loại những sự vật, hiện tượng với thang đo khoảng mà thang đo có điểm không thực sự người ta thường dùng thang định tỷ lệ (ratio scale). Như vậy thang định tỷ lệ cũng là một thang đo khoảng nhưng có điểm zero thực sự. Một điểm zero thực sự chỉ ra rằng tại đó vắng thiếu đặc tính được đo. Chẳng hạn, thang đo vận tốc (km/h) là một ví dụ về thang đo khoảng. Số km mỗi giờ ô tô đi (40km/h) nói lên tốc độ chuyển động, mỗi km trên giờ chỉ sự gia tăng tốc độ theo thang khoảng, còn ở 0 km/h nói rằng ô tô đang đứng yên.

Sự có mặt của điểm 0 giúp ta thiết lập được tỷ lệ

giữa các điểm số thu được. Chẳng hạn ta có thể xác định chính xác vận tốc 10km/h lớn gấp 2 lần 5km/h. Thang định tỷ lệ hiếm khi được dùng trong các phép đo hành vi, hay các phép đo cả quá trình tâm thần, vì người ta không thể xác định được điểm 0 thực sự cho hầu hết các đặc tính tâm lý. Liệu có ai đó đạt điểm 0 trên một trắc nghiệm đo IQ (IQ test) có nên được phân loại như là hoàn toàn không có trí tuệ. Khi một cá nhân nhận được điểm 0 trên một trắc nghiệm ta nên hiểu người này đạt điểm 0 trên trắc nghiệm đó, chứ không nên nói đặc tính được đo bằng 0. Điểm 0 trên các trắc nghiệm đánh giá nhận thức, thái độ, năng lực, hành vi của một cá nhân thực sự là “điểm 0 tự đặt” (arbitrary zero point).

Tóm lại nhà nghiên cứu phải hiểu biết số liệu được thu thập theo cách nào, theo kiểu thang đo nào để chọn lựa các kiểm định thống kê thích hợp.

3. Trắc nghiệm và các phép đo không phải là trắc nghiệm

3.1. Trắc nghiệm (test) là gì?

Trắc nghiệm là một kiểu đánh giá hay đo lường có sử dụng những thủ pháp/những kỹ thuật cụ thể, có tính hệ thống nhằm thu thập thông tin và chuyển những thông tin này thành các con số hoặc điểm để lượng hoá cái cần đo. Trắc nghiệm có sự khác biệt với các kỹ thuật đánh giá khác như quan sát, phỏng vấn chủ yếu là *mức độ kiểm soát được dùng trong suốt quá trình thu thập thông tin*. Trong trắc nghiệm có những câu hỏi, đề mục

(item) hay những tình huống phải giải quyết, những phương án phải lựa chọn hay những nhiệm vụ phải hoàn thành... và một thủ tục hướng dẫn, cho điểm thống nhất cho từng câu trả lời. Trắc nghiệm có thể được thiết kế dưới dạng ngôn ngữ hay phi ngôn ngữ, dưới dạng khách quan hay chủ quan (còn gọi là trắc nghiệm phóng chiếu).

3.2. Những đặc tính của trắc nghiệm

Một trắc nghiệm tốt phải có những *đặc tính thiết kế* tốt (tức là được thiết kế theo đúng quy trình và tuân thủ các nguyên tắc thiết kế), đồng thời phải có các *đặc tính đo lường* tốt (tức là được đánh giá về độ tin cậy, độ hiệu lực xem thực tế nó có đưa ra được những thông tin chính xác, phù hợp với mục tiêu đo lường và những thông tin đó có nhiều lợi ích hay không?).

Những *đặc tính thiết kế* thường liên quan đến giai đoạn phác thảo thiết kế trắc nghiệm, do vậy ngay khi bắt đầu xây dựng, người thiết kế phải tính đến có những khả năng nào để tạo ra các *đặc tính* này. Theo các chuyên gia thiết kế trắc nghiệm, có bốn *đặc tính cơ bản* của một trắc nghiệm được thiết kế tốt:

- Có mục đích được xác định rõ ràng: để xác định rõ mục đích đo lường, từng mục tiêu cụ thể, người thiết kế trắc nghiệm phải trả lời các câu hỏi: *test này được thiết kế để đo cái gì? Ai sẽ làm test này? Điểm của test được sử dụng như thế nào?* Câu hỏi đầu tiên liên quan đến mục tiêu đo lường không chỉ của cả phép đo mà còn của

từng item, liên quan đến cấu trúc của phép đo, miền đo... Câu hỏi thứ ba liên quan đến các kiểu item, đến các hình thức cho điểm...

- Có nội dung đo lường cụ thể và được chuẩn hóa: đây chính là sự cụ thể hoá các miền đo, cụ thể hoá các chỉ số đo, cụ thể hoá nội dung đo lường của từng item. Từng item chỉ đo một nội dung rất cụ thể, ở một phạm vi rất hẹp, nhưng tất cả các item được chọn lựa phải đủ rộng bao quát khắp miền đo, phải được người làm test biết hoặc hiểu ở mức có thể, phải tính đến những hạn chế của nghiệm thể... Nội dung phải được chuẩn hoá có nghĩa là tất cả những người làm test đều được đo lường trên cùng một vấn đề, cùng một nội dung, trên từng nội dung cụ thể cùng hiểu theo một cách thống nhất với người thiết kế.

- Có thủ tục hướng dẫn làm test được chuẩn hoá: tức là có sự thống nhất cách hướng dẫn, cách giải đáp thắc mắc như nhau cho mọi nghiệm thể. Trong mọi trường hợp, tất cả các nghiệm thể phải nhận được sự hướng dẫn cách làm như nhau.

- Cách cho điểm được chuẩn hoá: thủ tục cho điểm phải giống nhau với tất cả các nghiệm thể. Những trắc nghiệm phóng chiếu phải thống nhất cách cho điểm theo các phạm trù, mức độ... có sự giải thích chi tiết, thống nhất để sao cho điểm số trắc nghiệm có “nhiều nhất” tính khách quan.

Trái với những đặc tính thiết kế được xác định ngay từ giai đoạn đầu của quá trình thiết kế phép đo, đặc

tính *đo lường* của công cụ thường liên quan đến các giai đoạn sau: đo trên mẫu thử và thử trên các nhóm mẫu khác nhau và đo trên mẫu chính thức để đánh giá độ tin cậy, độ hiệu lực và đánh giá các đặc tính của từng item (độ khó, độ phân biệt, tính đại diện, tính hiệu lực...). Các đặc tính đo lường không chỉ phụ thuộc vào các đặc tính thiết kế, mà còn phụ thuộc đáng kể vào các đặc tính của mẫu (như độ lớn của mẫu, tính đại diện của mẫu...). Mặt khác các đặc tính đo lường còn phụ thuộc vào các phép toán thống kê được dùng để phân tích kiểm định độ tin cậy, độ hiệu lực (phải thoả mãn các điều kiện khi sử dụng một phép toán thống kê nào đó...).

3.3. Các kiểu trắc nghiệm

Có nhiều kiểu phân loại trắc nghiệm:

- Phân loại theo tiêu chuẩn cách thức giải quyết: chẳng hạn, các nhà nghiên cứu thường chia trắc nghiệm thành hai loại chủ quan và khách quan.

- Phân loại theo dấu hiệu hay nội dung: “trắc nghiệm nghiên cứu cái gì”. Chẳng hạn phân trắc nghiệm thành hai loại: các trắc nghiệm bộc lộ tối đa (maximal performance tests) thường gọi tắt là các trắc nghiệm năng lực (ability tests) và các trắc nghiệm bộc lộ đặc trưng (typical performance tests) thường gọi là các trắc nghiệm về nhân cách (personality tests).

- Phân loại theo tiêu chuẩn kiến tạo, chuẩn hoá, phạm vi sử dụng: chẳng hạn phân trắc nghiệm thành hai loại: các trắc nghiệm đã chuẩn hoá dùng cho các

nhóm mẫu (standardised tests) và các trắc nghiệm chưa được chuẩn hoá dùng cho lớp học (non standardised classroom tests).

3.4. Các phép đo khác

Các phiếu hỏi hay bảng hỏi (questionnaires), các phiếu trưng cầu ý kiến (opinionaires): thường gồm một loạt các câu hỏi (item) khách quan (câu hỏi đóng) hoặc chủ quan (câu hỏi mở) như câu hỏi đúng/sai, câu hỏi nhiều lựa chọn (chọn câu đúng nhất hoặc chọn câu phù hợp), câu hỏi điển thế, câu hỏi trả lời ngắn, câu hỏi đánh giá thứ bậc, câu hỏi đánh giá mức độ, câu hỏi ma trận, câu hỏi tự luận... Các thang đo kiểu này thường kết hợp các kiểu item khác nhau thường trong cùng một thang đo, thậm chí các kiểu item khác nhau cũng được kết hợp trong cùng một miền đo.

Các bảng nghiệm kê hay kiểm kê nhân cách (personality inventories), các bảng liệt kê hành vi (behavior checklists) hay các thang đo tỷ lệ (rating scales) thường được thiết kế đặc trưng cho mục đích đo lường thái độ, hứng thú, tình cảm, thói quen, phong cách, thuộc tính tâm lý... Các thang đo kiểu này thường chỉ gồm các item khách quan, các item có cùng hình thức (đúng, sai hoặc nhiều lựa chọn...), cùng tiêu chuẩn cho điểm, không có sự kết hợp các kiểu item khác nhau trong cùng một thang đo.

Chương III

QUY TRÌNH VÀ CÁC NGUYÊN TẮC, KỸ THUẬT THIẾT KẾ CÔNG CỤ ĐO LƯỜNG

Trong khoa học xã hội, các nhà nghiên cứu thường xuất phát từ một mệnh đề có tính giả thiết: "Bất cứ cái gì thực sự đang tồn tại đều có thể đo được", hoặc chúng ta hoàn toàn có thể đo được những cái gì nếu ta chỉ ra được nó đang thực sự tồn tại.

Tuy nhiên nhà nghiên cứu muốn lượng hoá một cái gì đó, họ phải đương đầu với không ít những khó khăn để đi từ những ý tưởng mơ hồ về những gì họ muốn nghiên cứu, đến nhận biết được nó là cái gì và làm thế nào để đo được nó ở thế giới thực.

1. Các bước cơ bản của quy trình thiết kế công cụ đo lường

Khó có thể thiết kế được những bộ công cụ đo lường tốt, nếu nhà nghiên cứu vẫn chưa rõ vấn đề họ muốn nghiên cứu thực chất nó là gì, những câu hỏi nghiên cứu nào phải trả lời, những mục tiêu cụ thể nào phải đạt. Do vậy để thiết kế được một bộ công cụ đo lường thật khoa học, nhà nghiên cứu cần phải tuân thủ quy trình thiết kế gồm những giai đoạn cơ bản như sơ đồ dưới đây.

1.1. Khái niệm hoá

Bước đầu tiên của quy trình thiết kế công cụ đo (phiếu trưng cầu/phiếu hỏi/bảng nghiệm kê... trắc nghiệm) là khái niệm hoá. Khái niệm hoá là quá trình cụ thể hoá những ý tưởng mơ hồ, trừu tượng thành các khái niệm nghiên cứu, các khái niệm sẽ đo lường, xác định các kiểu quan sát, các kiểu đo lường phù hợp nhất cho mục tiêu nghiên cứu.

Khái niệm hoá cũng liên quan đến quá trình chính xác hoá/cụ thể hoá những thuật ngữ trừu tượng (liên quan đến câu hỏi nghiên cứu hay mục tiêu nghiên cứu) thành những khái niệm định danh, có nội dung, phạm vi nghiên cứu cụ thể. Ví dụ, câu hỏi nghiên cứu phải trả lời là: *liệu phụ nữ có giàu tình thương hơn đàn ông? Sẽ không thể trả lời được câu hỏi nghiên cứu này nếu ta không xác định "chính xác", thuật ngữ có tính khái niệm tình thương là cái gì, trong phạm vi nghiên cứu nào.*

Chúng ta thường dựa vào những quan sát và những kinh nghiệm để gọi tên sự vật, hiện tượng. Việc dùng thuật ngữ để gọi tên sự vật, hiện tượng là quá trình khái quát hoá để tạo ra một khái niệm. Các khái niệm thực chất chỉ là những ngôn từ, không tồn tại trong thế giới thực, do vậy chúng không thể đo đạc trực tiếp được. Nhưng chúng ta có thể đo đạc được những thứ, nhờ đó khái niệm được tạo ra hay được khái quát lên. Ví dụ trầm cảm hay thành kiến là những thuật ngữ chỉ khái niệm. Chúng thực chất chỉ là những từ, gồm những âm tiết tạo thành, nhưng chúng lại mang nghĩa "thực", khái

quát từ những hành vi được quan sát, hoặc từ những trải nghiệm của các cá nhân tập hợp thành. Như vậy các nhà khoa học hoàn toàn có cơ sở để tin rằng: nếu biết được một khái niệm chỉ sự vật, hiện tượng tồn tại như thế nào (có tồn tại hay không tồn tại, có mặt hay không có mặt... qua những biểu hiện cụ thể nào) thì hoàn toàn có thể đo đạc được khái niệm đó.

Thực tế, việc thiết lập các khái niệm trong nghiên cứu khoa học phụ thuộc vào các kiểu định nghĩa. Thường có ba kiểu định nghĩa sự vật, hiện tượng như sau:

(1) Định nghĩa "thực". Là kiểu định nghĩa đưa ra một tuyên ngôn khái quát hay một mệnh đề bao quát "bản chất thực" hoặc "những đặc tính cơ bản" của sự vật, hiện tượng. Tuy nhiên khái niệm bản chất thực lại thường là một khái niệm mơ hồ, trừu tượng không tiện lợi cho mục đích nghiên cứu vì nó khó đo lường.

(2) Định nghĩa định danh. Là một kiểu định nghĩa nhằm phân định phạm vi ranh giới cho một thuật ngữ chỉ khái niệm (thuật ngữ này thực sự biểu thị nó là cái gì). Đây chính là một kiểu định nghĩa làm việc do nhà nghiên cứu tạo ra (nó có thể khác với thực tế), nhằm đạt mục tiêu nghiên cứu vì thường không có sự nhất trí khi định nghĩa khái niệm hoặc có sự nhầm lẫn về bản chất khái niệm (giữa cái thực chất và cái bề mặt). Do vậy nhà nghiên cứu phải giới hạn sự vật, hiện tượng muốn nghiên cứu trong những định nghĩa làm việc.

(3) Định nghĩa thao tác. Là sự cụ thể hoá hơn nữa

định nghĩa làm việc. Đây là kiểu định nghĩa mô tả khái niệm thành những hoạt động, những cấu trúc có thể đo lường được.

Tuy nhiên tầm quan trọng của việc định nghĩa khái niệm, làm rõ khái niệm phụ thuộc vào mục đích nghiên cứu. Nếu mục đích chính của nghiên cứu là nhằm mô tả thì việc định rõ khái niệm là nhiệm vụ đặc biệt quan trọng. Ví dụ: Câu hỏi nghiên cứu phải trả lời là có bao nhiêu % người thất nghiệp trong thành phố. Khi đó việc định nghĩa thế nào là những người thất nghiệp đóng vai trò quyết định cho việc thiết kế các phép đo.

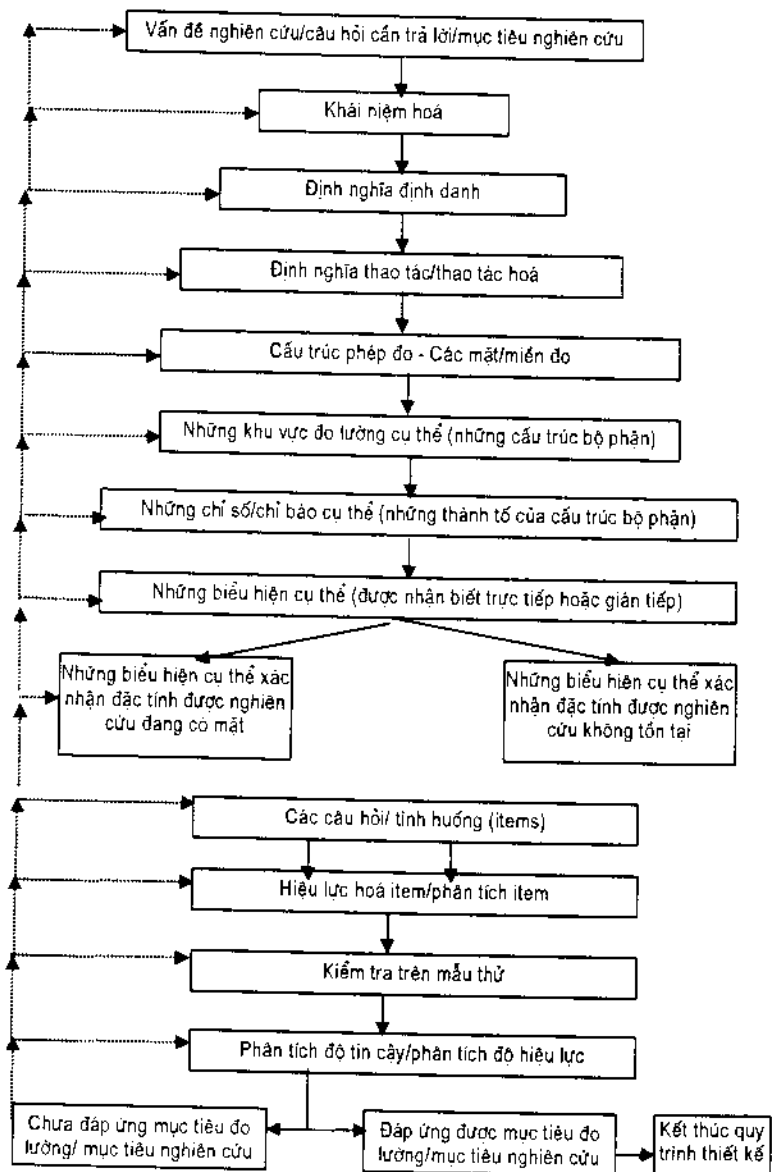
1.2. Thao tác hoá khái niệm

Thao tác hoá khái niệm là sự mở rộng, cụ thể hoá hơn nữa quá trình khái niệm hoá. Tuy nhiên có sự khác biệt giữa hai quá trình này. Quá trình khái niệm hoá là quá trình chính xác hoá và cụ thể hoá các khái niệm trừu tượng. Còn thao tác hoá khái niệm chính là quá trình phát triển những thủ tục nghiên cứu cụ thể, chuyển những khái niệm đã được định nghĩa thành những cấu trúc với những tiêu chí, những hoạt động cụ thể để có thể đo lường được. Tuy nhiên đây là công việc không dễ dàng, dù ta đã xác định rõ mục tiêu, khái niệm.

Để chuyển những khái niệm đã được định nghĩa thành những cấu trúc có thể đo lường, người thiết kế thường phải thực hiện một loạt các bước có tính kỹ thuật sau đây:

- Trước hết xác định những biến cần tìm hiểu, đưa ra những định nghĩa thao tác cho các biến cần đo.

Quy trình thiết kế



- Xác định rõ các miền đo (những nội dung cần đo), các cấu trúc đo lường cụ thể.
- Xác định những cấu trúc trọng tâm (đặc biệt quan trọng) nói lên sự có mặt hay vắng mặt của khái niệm.
- Xác định các tiêu chí, chỉ số, chỉ báo.
- Xác định những biểu hiện cụ thể nói lên sự có mặt hay vắng mặt của các chỉ số.
- Từ mỗi biểu hiện cụ thể nói lên sự có mặt hay vắng mặt của đặc tính được nghiên cứu, phác thảo các item (câu hỏi, nhận định, tình huống...), chọn lựa hình thái thích hợp cho các item. Để có được nhiều item tốt, có tính đại diện, đủ bao quát khắp miền đo, nhà nghiên cứu cần nảy sinh một số lượng lớn các item (càng nhiều càng tốt) để tạo ra một "bể item". Bể item gồm rất nhiều item phục vụ cho bước tiếp theo là hiệu lực hoá item, chọn lọc ra những item có đủ các đặc tính đo lường.

1.3. Hiệu lực hoá item

Từ một "bể" gồm vô số item được nảy sinh từ những cấu trúc, những chỉ số, chỉ báo, nhà nghiên cứu bắt đầu chọn lọc các item. Hiệu lực hoá item chính là quá trình chọn lọc tìm ra những item phù hợp nhất đo đúng từng nội dung cụ thể và đáp ứng được mục tiêu của sự đo lường. Bể item phải là một tập hợp "nhiều nhất" những item làm thành một mẫu đủ lớn bao quát hết nội dung cần đo. Quá trình hiệu lực hoá item phải tuân thủ những thủ tục có tính nguyên tắc và thường gồm những bước cơ bản sau:

- Đánh giá bản chất, nội dung của từng item: mỗi item chỉ nên tập trung vào một nội dung thật cụ thể (biểu hiện cụ thể) và chỉ bao quát một phạm vi rất hẹp nằm trong một chỉ số cụ thể, thuộc nội dung của một miền đo nào đó. Căn cứ theo nội dung, mục tiêu của phép đo xác định các item có nội dung không cụ thể, không thích hợp (không thuộc nội dung cần đo) để loại bỏ, những item có nội dung gần trùng nhau được kết hợp lại thành một item có tính đại diện hơn... Người thiết kế có thể đặt ra những chuẩn “khách quan” để đánh giá các item. Kết thúc bước này, chỉ những item đáp ứng được các tiêu chuẩn nào đó mới được giữ lại để chọn lọc tiếp.

- Đánh giá kiểu, hình thức item: những kiểu item nào thích hợp nhất, mức độ đọc hiểu của item có phù hợp, tính đơn nghĩa của item, mức độ mong muốn trả lời thật các item... những item có nội dung tốt nhưng có hình thức hỏi không thích hợp (đa nghĩa, khó hiểu, thân chủ không muốn trả lời...) cần được viết lại.

- Đánh giá độ phù hợp của từng item (trên cơ sở phân tích tính đồng nhất cấu trúc): nếu phép đo chỉ đo một cấu trúc duy nhất thì các item của nó phải có tính đồng nhất biểu thị bằng sự tương quan chặt chẽ với các item cùng đo cấu trúc đó. Nếu phép đo được thiết kế theo mô hình đa diện, đa cấu trúc, thì các item được thiết kế đo cùng một cấu trúc nên được nhóm lại. Các item trong từng tiểu cấu trúc được kiểm tra tính đồng nhất... Sử dụng các phép toán thống kê (như phân tích

tương quan, phân tích yếu tố) để chọn ra những item phản ánh “rõ nhất” cấu trúc của cái định đo.

- Đánh giá số lượng item phù hợp cho từng tiểu cấu trúc và toàn bộ phép đo: những cấu trúc trọng tâm của phép đo phải có số lượng item nhiều hơn, thành phần trọng tâm của từng tiểu cấu trúc cũng gồm nhiều item hơn các thành phần không trọng tâm. Tuy nhiên số lượng item không thể quá nhiều, phải căn cứ vào đặc điểm tâm sinh lý của nghiệm thể để xác định một thời lượng thích hợp.

Các bước trên đây của quá trình hiệu lực hoá item phải do các chuyên gia hay những người am hiểu thực hiện với sự trợ giúp của các phép toán thống kê.

2. Những vấn đề và các nguyên tắc thiết kế công cụ đo

Trong khi xây dựng, thiết kế công cụ đo, dù ở những hình thức đơn giản nhất như phiếu hỏi (questionnaires), bảng nghiệm kê (checklists) đến những hình thức phức tạp hơn như các thang đo chuẩn hay các trắc nghiệm chuẩn (standardized tests), có những vấn đề có tính nguyên tắc, buộc tất cả những người nghiên cứu, thiết kế công cụ phải tuân thủ. Một bộ công cụ đo lường tốt phải được thiết kế khoa học, theo đúng quy trình và các nguyên tắc thiết kế, đồng thời phải được đánh giá về mặt thực tế, kiểm nghiệm bằng thống kê, để khẳng định liệu nó có đưa ra được những thông tin chính xác, có nhiều lợi ích hay không?

Theo các chuyên gia thiết kế công cụ đo lường (Aiken, 1994; Fridenberger, 1995) ở giai đoạn phác thảo, thiết kế công cụ những vấn đề có tính nguyên tắc gồm những phán xét, những quyết định liên quan đến mục đích, nội dung, cách thức cho điểm và các thủ tục hướng dẫn làm trắc nghiệm.

2.1. Xác định rõ mục đích đo lường của công cụ

Vấn đề quan trọng đầu tiên trong thiết kế công cụ là người thiết kế phải nhận biết rõ mục đích, công cụ được thiết kế để đo cái gì? Mục đích khác nhau sẽ dẫn tới lựa chọn kiểu thiết kế item khác nhau. Mục đích của trắc nghiệm thường được xác định bằng những loại câu hỏi sau đây:

- Phép đo được thiết kế để đo cái gì?

Để trả lời câu hỏi này, người thiết kế phải xác định rõ miền đo của trắc nghiệm. Miền đo của trắc nghiệm là sự hiểu biết, kỹ năng hoặc đặc tính... được đánh giá bởi các câu hỏi (item) của trắc nghiệm. Trắc nghiệm có thể được thiết kế để đo các thành tố của sự nhận thức, kết quả học tập, hoặc các kỹ năng, các đặc tính của nhân cách, thái độ, hứng thú của cá nhân... bất cứ thứ gì mà chúng ta có thể gọi tên và chỉ ra nó đang "thực sự tồn tại".

Theo các chuyên gia thiết kế, mục tiêu đo lường ảnh hưởng trực tiếp đến việc lựa chọn các kiểu hình thái thích hợp cho từng item. Chẳng hạn, một trắc nghiệm dùng để đo năng lực (kiến thức, kỹ năng) cần những câu hỏi (item)

trực tiếp có câu trả lời cụ thể, chính xác. Trái lại, một trắc nghiệm đo các thuộc tính của nhân cách hay hứng thú phải sử dụng những câu hỏi gián tiếp, hoặc những tình huống kích thích mơ hồ và người làm trắc nghiệm có thể phải tự lý giải. Mặt khác những trắc nghiệm về năng lực nhấn mạnh đến mục tiêu đo được tối đa khả năng của người làm trắc nghiệm (mức độ tốt nhất có thể của cá nhân) phải gồm những item có độ khó khác nhau (trong đó có những item rất khó) để những người giỏi nhất cũng khó đạt điểm tối đa. Trái lại những trắc nghiệm về nhân cách nhấn mạnh đến mục tiêu đo được những ý nghĩ hành vi, thói quen thường ngày đặc trưng của người làm trắc nghiệm do vậy phải gồm những item có độ phân biệt tốt nhất. Cũng vậy, một trắc nghiệm đánh giá kiến thức chỉ nhằm phát hiện 30% học sinh kém nhất sẽ được thiết kế khác với một trắc nghiệm nhằm phát hiện 30% học sinh giỏi nhất. Các item của trắc nghiệm trước sẽ gồm nhiều item dễ để 70% học sinh có thể vượt qua, còn trắc nghiệm sau sẽ gồm nhiều những item khó, sao cho chỉ 30% học sinh vượt qua.

- Phép đo được thiết kế cho đối tượng nào?

Vấn đề không kém phần quan trọng mà người thiết kế trắc nghiệm phải biết rõ là trắc nghiệm được thiết kế cho đối tượng nào? Đối tượng làm trắc nghiệm (nghiệm thể) là người lớn hay trẻ em? Những đặc điểm của đối tượng, những hạn chế... trình độ đọc hiểu của item của đối tượng? Liệu các câu hỏi của trắc nghiệm nên được trình bày dưới hình thức nào (câu hỏi kín hay câu hỏi mở, câu

hỏi viết hay tranh vẽ...) là thích hợp nhất cho đối tượng. Một trắc nghiệm thiết kế cho người lớn nhất thiết phải khác một trắc nghiệm thiết kế cho trẻ em, dù cùng một mục đích, cùng một miền đo. Chẳng hạn cùng là trắc nghiệm đánh giá trí tuệ (IQ) nhưng trắc nghiệm thiết kế cho trẻ trước tuổi học khác với trắc nghiệm thiết kế cho học sinh tiểu học và cũng khác với học sinh trung học (các item dùng cho trẻ trước tuổi học chủ yếu là những câu hỏi miệng, tranh vẽ, các hình khối... Ngược lại những item dùng cho học sinh lớn có thể chủ yếu là các câu hỏi viết...).

- Phép đo được thiết kế để làm với từng cá nhân hay với nhóm?

Một số trắc nghiệm chỉ thích hợp với đánh giá từng cá nhân riêng lẻ vì nó được thiết kế có phối hợp những đặc điểm của kỹ thuật phỏng vấn, kỹ thuật tưởng tượng, phóng chiếu. Một số trắc nghiệm khác lại thích hợp cho việc đánh giá theo nhóm vì nó được thiết kế để đánh giá cá nhân trong mối tương quan với nhóm. Cũng có những trắc nghiệm được thiết kế thích hợp cho cả hai cách trên. Về nguyên tắc, quy trình thiết kế, cơ bản không có gì khác, nhưng về kỹ thuật thiết kế các item cụ thể có một số điểm khác nhau giữa các loại công cụ này. Chẳng hạn, những trắc nghiệm đánh giá với từng cá nhân thường có thêm các tình huống, nhiệm vụ hay câu hỏi phụ giúp nghiệm viên giải thích bổ sung khi đối tượng tỏ ra lúng túng không hiểu, hoặc khai thác thêm thông tin làm sáng câu trả lời của nghiệm thể, hoặc có "điểm

dừng” - thôi không tiếp tục làm trắc nghiệm nếu phát hiện thấy nghiệm thể không có khả năng trả lời tiếp tục...

2.2. Xác định rõ nội dung và định chuẩn cho từng nội dung cụ thể của công cụ đo

Một bộ công cụ được thiết kế tốt đòi hỏi người thiết kế phải xác định rõ nội dung cần bao quát và từng nội dung cụ thể phải có tính chuẩn. Nói cụ thể hơn, người thiết kế phải nhận biết chính xác những hiểu biết gì, những kỹ năng nào hoặc những đặc tính, những phẩm chất nào các item của trắc nghiệm phải bao quát trong khi đo lường.

Không có sự cụ thể hoá nội dung và định chuẩn ở từng nội dung cụ thể cần đo, sẽ rất khó đánh giá độ hiệu lực của công cụ. Cụ thể hoá nội dung cần phải làm trước khi bắt tay viết các item trắc nghiệm. Người thiết kế nên phác thảo nội dung vừa hữu hạn vừa thật chi tiết, cụ thể những gì trắc nghiệm cần đo. Để làm rõ nội dung và định chuẩn cho từng nội dung cụ thể nhà thiết kế phải làm rõ những câu hỏi cơ bản sau đây:

- *Phép đo nên gồm những miền đo nào? Những miền đo này đã đủ bao quát hết nội dung cần đo?*

- *Mỗi miền đo nên gồm những mặt nào? Những thành tố nào là trọng tâm?*

- *Nên có bao nhiêu item cho mỗi thành phần?*

- *Liệu các item đã đủ bao quát khắp miền đo? Những thành phần quan trọng nhất của miền đo có gồm nhiều*

item hơn không?

- Những loại item nào nên được dùng?

- Độ khó của item có phải là một vấn đề có ý nghĩa cho trắc nghiệm đang thiết kế? (nếu câu trả lời là có thì liệu trắc nghiệm có cần những item có độ khó khác nhau, có cần một số item đặc biệt khó không?)

- Những kiểu item nào được sử dụng khá thành công ở các trắc nghiệm cùng loại trước đó?...

Nội dung cần đánh giá, cần đo lường chính là miền đo của trắc nghiệm. Xác định rõ nội dung đo lường bắt đầu bằng việc xác định, lựa chọn miền đo. Sau đó tiếp tục cụ thể hoá các loại hiểu biết, kỹ năng, hành vi hay thái độ thuộc từng mặt, từng khía cạnh của miền đo thành các item của trắc nghiệm sao cho chúng phải bao quát khắp miền đo đó. Từng nội dung cụ thể của miền đo phải được chuẩn hoá, nghĩa là tất cả những người làm trắc nghiệm đều được kiểm tra, đánh giá trên cùng những thuộc tính hoặc hiểu biết, trả lời trên cùng một bảng câu hỏi với cùng một kiểu cho điểm, cùng một kiểu hướng dẫn cách làm trắc nghiệm.

2.3. Chọn lựa kiểu cho điểm

Kiểu cho điểm thường phụ thuộc vào mục tiêu của công cụ, tức là bằng cách nào trắc nghiệm được dùng để đánh giá người làm trắc nghiệm. Theo các chuyên gia thiết kế trắc nghiệm có hai kiểu cho điểm điển hình: khách quan và chủ quan.

1. Câu hỏi đầu tiên về kiểu cho điểm là: điểm số

trắc nghiệm sẽ được dùng so sánh kết quả của những người làm trắc nghiệm với nhau hoặc chỉ đơn phương đánh giá đặc tính cá nhân của người làm trắc nghiệm?

Phần thứ nhất của câu hỏi này liên quan đến sự phân tích so sánh điểm của từng cá nhân với điểm trung bình của cả nhóm (điểm liên quan đến mẫu từ đó số liệu được thu thập) nhằm chỉ ra vị trí tương đối của từng thành viên trong tương quan giữa họ với nhau. Phần thứ hai của câu hỏi liên quan đến sự so sánh điểm của từng cá nhân đối với điểm chuẩn (điểm liên quan đến nội dung) nhằm chỉ ra vị trí của từng cá nhân trong tương quan đối với nội dung của trắc nghiệm.

a/ Điểm liên quan đến nhóm mẫu chuẩn hoá hay gọi tắt là điểm norm (norm-referenced scores) được tính trên cơ sở điểm của từng cá nhân (điểm tổng) trên các item. Sau đó chuyển điểm tổng này sang hình thức mới gồm những thông tin về kết quả của từng người cùng làm trắc nghiệm. Chẳng hạn điểm phần trăm thứ hạng (percentile rank): A đạt 8 điểm (điểm thô) trong một nhóm những học sinh làm trắc nghiệm đọc hiểu, vậy kết quả điểm trắc nghiệm của A tốt hơn bao nhiêu % so với số người cùng làm trắc nghiệm? Vì điểm của mỗi cá nhân được chuyển sang thang mới dựa trên kết quả điểm của cả nhóm, do vậy thứ hạng của A phụ thuộc không chỉ vào điểm A đạt được mà còn phụ thuộc vào điểm của người khác trong lớp. Nếu điểm của A đúng nhiều và người khác cũng đúng nhiều thì điểm phần trăm thứ hạng của A không thật cao, trái lại A đúng

nhiều hơn và người khác đúng ít thì điểm phần trăm thứ hạng của A sẽ rất cao.

Hầu hết các trắc nghiệm sử dụng điểm norm, điểm này phù hợp cho việc lựa chọn, phân loại người làm trắc nghiệm thành các nhóm khác nhau về hiểu biết và kỹ năng của họ trên đặc tính nghiên cứu (tiện lợi nhất cho các trắc nghiệm có mục đích đánh giá thứ hạng, chọn lựa, đặt chỗ và chẩn đoán).

Với một số trắc nghiệm sử dụng trong phạm vi lớp học, norm đơn giản chỉ là một số người trong lớp. Còn đối với trắc nghiệm đã chuẩn hoá, norm là những nhóm mẫu đại diện (norm groups).

Người thiết kế công cụ cần nhận thấy có sự khác biệt đáng kể giữa các trắc nghiệm đánh giá năng lực, khả năng (các trắc nghiệm đo kết quả bộc lộ tối đa) và các trắc nghiệm đánh giá nhân cách, hứng thú, thái độ (các trắc nghiệm đo kết quả bộc lộ đặc trưng) sử dụng điểm norm. Các trắc nghiệm thuộc nhóm thứ nhất sử dụng điểm norm để phân loại người làm trắc nghiệm thành các nhóm dựa trên trình độ năng lực. Do vậy người thiết kế phải chú trọng đặc biệt đến độ khó của item (nếu trắc nghiệm không gồm một số item khó mà chỉ có những người thực sự có năng lực, có kỹ năng thành thạo mới làm đúng, thì rất khó phân loại những người làm trắc nghiệm vào các trình độ tốt, khá, trung bình, yếu). Ngược lại các trắc nghiệm thuộc nhóm thứ hai sử dụng điểm norm để phân loại người vào các nhóm theo đặc

trưng cá nhân mà không theo trình độ năng lực. Chẳng hạn những người làm trắc nghiệm cần được phân thành các nhóm: trầm cảm, không trầm cảm hay người có khuynh hướng bảo thủ, người có khuynh hướng tự do. Do vậy người thiết kế phải chú trọng đặc biệt đến khả năng phân biệt của item.

b/ Điểm liên quan đến nội dung hay điểm chuẩn (criterion - referenced scores). Mục đích của điểm chuẩn được thiết kế để đánh giá kết quả làm trắc nghiệm của từng cá nhân một cách riêng rẽ trên cơ sở mức độ thông hiểu, làm chủ những kiến thức, kỹ năng thuộc miền đo của trắc nghiệm. Điểm chuẩn chỉ một mức độ cụ thể như là kết quả trả lời đúng các item của trắc nghiệm và ở mức đó sẽ đạt chuẩn. Chẳng hạn trắc nghiệm đánh giá năng lực của học sinh để chọn đầu vào các trường trung học phổ thông sử dụng một tiêu chuẩn 70% đúng, có nghĩa là học sinh A làm đúng trên 70% các item sẽ đạt chuẩn được nhận vào học, học sinh B làm đúng dưới 70% các item sẽ không đạt chuẩn, do đó sẽ không được nhận vào học. Nói chung điểm chuẩn thường chỉ sử dụng với các trắc nghiệm năng lực (trắc nghiệm bộc lộ kết quả tối đa). Những trắc nghiệm dùng đánh giá học sinh trong lớp học đa số dùng điểm chuẩn. Những trắc nghiệm dùng điểm chuẩn thường chỉ có một miền đo.

Theo ý kiến của các chuyên gia, điểm chuẩn được sử dụng thích hợp nhất cho những trắc nghiệm có nội dung được xác định rất cụ thể, rõ ràng. Tuy nhiên rất khó sử dụng điểm chuẩn cho những trắc nghiệm được chuẩn

hoá trên mẫu đại diện. Hiện tại, hầu hết các trắc nghiệm đánh giá năng lực được chuẩn hoá sử dụng điểm norm.

2. Câu hỏi thứ hai liên quan đến kiểu cho điểm: *trắc nghiệm đo một đặc tính duy nhất (trắc nghiệm có một miên đo duy nhất) hay đo một loạt các đặc tính (trắc nghiệm có nhiều miên đo)? Nếu trắc nghiệm đo một loạt các đặc tính (đa tầng đa diện) thì đo riêng rẽ từng đặc tính hay chỉ xếp hạng điểm mạnh tương đối của từng đặc tính?*

Nếu trắc nghiệm đo một đặc tính duy nhất thì mỗi cá nhân chỉ có một điểm trắc nghiệm duy nhất (điểm tổng của các item). Còn trắc nghiệm đo nhiều đặc tính mỗi cá nhân sẽ có những điểm tách biệt đại diện cho từng đặc tính được đo. Tuy nhiên ta cũng có thể chọn lựa cách cho điểm theo thứ bậc nhu cầu nếu mục đích của trắc nghiệm là xác định sức mạnh tương đối của những đặc tính khác nhau. Chẳng hạn một trắc nghiệm đánh giá tính quyết đoán và khả năng lãnh đạo có 25 item. Mỗi item là một cặp gồm các hành vi thể hiện tính quyết đoán và một hành vi thể hiện khả năng lãnh đạo. Người làm trắc nghiệm được hướng dẫn chỉ chọn một hành vi nào thích hợp với họ. Về lý thuyết, từng item chỉ có một lựa chọn, nếu chọn hành vi có tính quyết đoán sẽ ngăn cản chọn lựa hành vi thể hiện khả năng lãnh đạo. Nếu A chọn 14 hành vi liên quan đến tính quyết đoán và chỉ có 11 hành vi liên quan đến khả năng lãnh đạo, thì rõ

ràng A có tính quyết đoán mạnh hơn là khả năng lãnh đạo.

2.4. Chuẩn hoá thủ tục hướng dẫn làm trắc nghiệm

Một trắc nghiệm tốt đòi hỏi phải có thủ tục hướng dẫn chuẩn, tức là mọi người làm trắc nghiệm nhận được cùng một kiểu hướng dẫn như nhau. Có ba hình thức chung nhất là: *hướng dẫn làm trắc nghiệm theo cá nhân, làm theo nhóm, hoặc làm theo những câu hỏi cài trước trong máy tính.*

Có những trắc nghiệm thiết kế chỉ thích hợp cho đánh giá theo từng cá nhân tách biệt. Những trắc nghiệm làm với cá nhân thường dùng item có câu trả lời tự do, câu hỏi bỏ dỏ hoặc các câu hỏi phóng chiếu. Ví dụ trắc nghiệm TAT là trắc nghiệm nhân cách dùng kỹ thuật phóng chiếu, sử dụng những tưởng tượng nhìn mơ hồ như là những item của trắc nghiệm. Theo lý thuyết, khi trả lời những câu hỏi, người làm thực nghiệm phóng chiếu những thành phần vô thức của nhân cách vào trong các mô tả này. Trắc nghiệm này chỉ áp dụng làm với từng cá nhân, khó làm theo nhóm hoặc theo máy tính.

Một số trắc nghiệm cho điểm khách quan thích hợp làm theo nhóm. Đa số các trắc nghiệm được thiết kế để làm theo nhóm, cho dễ thu thập số liệu trên mẫu lớn. Mẫu càng lớn, càng đại diện thì càng có lợi khi tạo ra điểm Norm. Chi phí cho các trắc nghiệm làm theo nhóm cũng đỡ tốn kém. Tuy nhiên trắc nghiệm làm theo nhóm không

thể thiết kế item phù hợp được những đặc điểm của item kiểu phỏng vấn như các trắc nghiệm hướng dẫn làm cá nhân.

3. Những kỹ thuật thiết kế item

Trong chiến lược thiết kế item, người thiết kế có thể áp dụng những cách tiếp cận khác nhau nhằm phạm trù hoá theo miền đo.

3.1. Miền đo là những hiểu biết và kỹ năng

Các trắc nghiệm đo kết quả bộc lộ tối đa (maximal performance tests) gồm các trắc nghiệm đánh giá về năng lực (năng khiếu và thành tích) hay khả năng thường có miền đo là những hiểu biết, kiến thức hay kỹ năng. Các trắc nghiệm này về mặt truyền thống hay sử dụng thang khoảng, cho điểm trên cơ sở câu trả lời đúng. Những hiểu biết, kỹ năng được cụ thể hoá, chuyển thành các item để câu trả lời được nhận dạng như là đúng/sai, tốt/xấu hoặc đúng nhất, thích hợp nhất. Các item phải được thiết kế sao cho những người có hiểu biết, có kỹ năng, trả lời đúng nhiều nhất các câu hỏi và do đó sẽ đạt điểm trắc nghiệm cao hơn. Ngược lại, những người ít hiểu biết, thiếu kỹ năng trả lời sai nhiều hơn, do đó đạt điểm trắc nghiệm thấp hơn.

Các item đánh giá tri thức, kỹ năng có các dạng điển hình là: đúng/sai, nhiều lựa chọn, ma trận, trả lời tự do (viết luận, trả lời ngắn, điền chỗ trống...). Dưới đây là một số dạng tiêu biểu.

Ví dụ: Item có dạng trả lời đúng/sai:

- Hãy xác định những nhận định sau đây là đúng hay sai:

Đ S Văn Miếu được xây dựng vào thời nhà Trần.

Ví dụ: Item có dạng nhiều lựa chọn:

- Hãy chọn lựa câu trả lời đúng cho câu hỏi sau đây:

Văn Miếu được xây dựng vào:

a) Thời Tiền Lê

b) Thời Lý

c) Thời Trần

- Hãy đặt tên phù hợp cho từng cách tiếp cận trị liệu sau đây (vẽ mũi tên chỉ sự thích hợp)

a / Phân tâm..... còn gọi là cách trị liệu tự quan sát nội tâm

b / Nhân văn-hiện sinh..... còn gọi là cách trị liệu bằng hành động

c / Nhận thức-hành vi..... còn gọi là cách trị liệu bằng trò chuyện

- Mục tiêu của trị liệu tâm lý là

a / Làm giảm hoặc mất triệu chứng rối nhiễu ở thân chủ

b / Điều chỉnh lại những mối quan hệ nhân cách bị rối nhiễu

c/ Phát triển các kỹ năng giải quyết vấn đề, tạo khả năng thích nghi tốt hơn ở thân chủ

d/ Ý kiến khác (ghi cụ thể)

- Cách tiếp cận trị liệu tâm lý nào sau đây:

a/ Phân tâm

b/ Nhân văn-hiện sinh

c/ Nhận thức-hành vi

.....cho rằng rối nhiễu tâm lý nào đó tồn tại là do có những điều kiện hiện tại đang duy trì sự rối nhiễu này.

.....cho rằng rối nhiễu tâm lý nào đó xảy ra là do có những xung đột xảy ra trước đó vẫn không được giải quyết.

..... cho rằng rối nhiễu tâm lý nào đó xảy ra là do thiếu sự nhận thức đầy đủ về bản thân, thiếu lòng tự tin và không chấp nhận chính mình.

Ví dụ: Item có dạng trả lời tự do:

-là trường đại học đầu tiên, lâu đời nhất của Việt Nam.

- Mối quan hệ giữa nhà trị liệu và người có rối nhiễu là mối quan hệ lâm sàng trên chủ thể, tức là:

.....
.....

Ví dụ: Item có dạng ma trận, nhiều mức độ:

- Những lý do nào sau đây thúc đẩy bạn vào học trường chuyên:

Đúng Đúng Không
 một phần đúng

a/ Chính sách ưu đãi về học bổng,
 học phí ☐ ☐ ☐

b/ Cơ sở trường lớp, thiết bị dạy và
 học tốt hơn ☐ ☐ ☐

c/ Thầy cô có trình độ, kinh
 nghiệm giảng dạy tốt hơn ☐ ☐ ☐

d/ Có môi trường học tập
 nghiêm túc ☐ ☐ ☐

đ/ Có nhiều bạn giỏi để học hỏi,
 thi đua ☐ ☐ ☐

e/ Đáp ứng nguyện vọng của
 cha mẹ ☐ ☐ ☐

g/ Thoả mãn nguyện vọng của bản
 thân ☐ ☐ ☐

h/ Có danh tiếng "học sinh chuyên" ☐ ☐ ☐

i/ Các lý do khác (ghi cụ thể).....

- Dưới đây là những mệnh đề được dùng để mô tả trạng thái tâm lý của bạn. Xin bạn hãy đọc kỹ từng câu, sau đó khoanh tròn vào một số thích hợp (từ 1 đến 4) biểu thị trạng thái tình cảm CỦA BẠN VÀO LÚC NÀY.

STT	Trạng thái xúc cảm, tình cảm	Mức độ không có	Một chút	Tương đối	Rất nhiều
1	Tôi cảm thấy bình tĩnh	1	2	3	4
2	Tôi cảm thấy an toàn	1	2	3	4
3	Tôi đang căng thẳng	1	2	3	4
4	Tôi cảm thấy mệt mỏi và lo âu	1	2	3	4

Khi thiết kế các item có dạng trả lời đúng/sai nên tránh sử dụng các cụm từ hạn định "luôn luôn", "chưa bao giờ", "đôi khi" vì chúng có thể tạo ra những "gợi ý" cho câu trả lời. Các item nên ngắn gọn, dễ hiểu và đơn nghĩa. Những thành phần quan trọng nhất của miến đo nên có nhiều item hơn.

Khi thiết kế các item có dạng nhiều lựa chọn, mỗi item chỉ nên có *một lựa chọn đúng*, các lựa chọn còn lại là không đúng, nhưng nên có hình thức "ngụy trang" hợp lý để người làm trắc nghiệm ít hiểu biết khó nhận ra. Nên tránh các câu hỏi âm tính, tránh các câu hỏi có cấu trúc ngữ pháp phức tạp, dài, đa nghĩa. Mỗi item chỉ nên tập trung vào một vấn đề thật cụ thể (biểu hiện cụ thể) và chỉ bao quát một phạm vi rất hẹp nằm trong một chỉ số cụ thể, thuộc nội dung của một miến đo nào đó. Khi viết các item có dạng trả lời tự do (như điền vào chỗ trống, hoàn thành câu bỏ dở, câu trả lời ngắn và bài

luận) sử dụng hệ thống cho điểm chủ quan, do vậy thủ tục hướng dẫn cho điểm phải đủ chi tiết, phải thống nhất để hạn chế tối đa sai số. Cần đánh giá độ khó của item để loại bỏ những item quá khó hoặc quá dễ (vì những item này thực sự không đóng góp gì cho độ tin cậy và độ hiệu lực của trắc nghiệm).

3.2. Miến đo là những mẫu ứng xử, hành vi

Một số công cụ đo được thiết kế để đo hoặc dự đoán những mẫu ứng xử, hành vi, kỹ năng sống. Để làm được điều này, người thiết kế phải liệt kê được các nhiệm vụ, các yêu cầu, các công việc, các tình huống đặc trưng, qua đó cá nhân bộc lộ rõ nhất bản chất đặc tính được nghiên cứu. Quá trình cụ thể hoá nội dung đo lường, chính là quá trình phân tích nhiệm vụ, phân tích những yêu cầu công việc. Trong quá trình này công việc hay nhiệm vụ được chia nhỏ thành những thành phần, bộ phận có nội dung cụ thể. Mỗi thành phần này được chuyển thành những hành vi cụ thể dùng cho việc nảy sinh các item. Để làm được điều này, người thiết kế nên sử dụng phương pháp chuyên gia. Chẳng hạn để xác định các mẫu ứng xử đặc trưng ở lứa tuổi vị thành niên, ta có thể nhờ các giáo viên trung học phổ thông, những người có nhiều năm giảng dạy, tiếp xúc hay nghiên cứu về tuổi này phát hiện, rồi trên cơ sở đó nảy sinh các item. Cũng có thể yêu cầu những nhóm học sinh trung học phổ thông khác nhau đưa ra các mẫu ứng xử riêng, trên cơ sở đó phát triển thành các item.

Dưới đây là vài ví dụ về các item được thiết kế để đo những mẫu ứng xử nhận thức, thái độ, hành vi:

- Nếu làm sai pháp luật hoặc một quy định chung nào đó mà không ai biết là bạn yên tâm:

a- Đúng b- Phân vân c- Không đúng

- Nếu như bạn vi phạm một cách không chủ định các quy định ở nơi công cộng thì bạn cho rằng điều đó không sao cả?

a- Đúng b- Phân vân c- Không đúng

- Trẻ em có một số những điều quan tâm lo lắng, chẳng hạn như những khó khăn xảy ra ở lớp, ở trường, nơi ở, nơi làm việc, những xung đột gia đình, những vướng mắc với bạn bè... những nỗi khổ tâm. Để ứng phó giải quyết những khó khăn vướng mắc này, những bạn bè đồng lứa thường sử dụng một loạt những cách thức dưới đây. Xin bạn hãy đọc kỹ từng câu và điền **một** con số chỉ mức độ phù hợp nhất vào đầu câu:

1 = Không dùng cách này, không nghĩ và làm như vậy

2 = Sử dụng rất ít, đôi khi

3 = Sử dụng khá nhiều, khá thường xuyên

4 = Sử dụng rất nhiều, rất thường xuyên

1. Nói chuyện với người khác xem họ xử trí thế nào nếu họ rơi vào hoàn cảnh tương tự.

2. Cố gắng tìm hiểu xem cái gì là nguyên nhân gây ra vấn đề.

3. Chấp nhận, tìm cách thích nghi với hoàn cảnh.

4. Phó mặc cho số phận, cho sự may rủi.

5. Tìm lời khuyên từ một người mình tin cậy.

6. Cảm thấy lo lắng cho tương lai của mình

7. Lúng túng, không biết làm gì cả

- Dưới đây là những hành vi mà một người ở lứa tuổi bạn có thể làm. Xin hãy đọc kỹ từng câu, nghĩ xem mình có làm thường xuyên không và ở mức độ nào?

a) Đánh giá mức độ thực hiện:

- Chưa bao giờ khoan số 0

- Đôi khi khoan số 1

- Thường xuyên khoan số 2

b) Đánh giá mức độ quan trọng:

- Không quan trọng khoan số 0

- Quan trọng khoan số 1

- Rất quan trọng khoan số 2

	Câu hỏi	Mức độ quan trọng			Mức độ thực hiện		
		0	1	2	0	1	2
1	Tôi kết bạn dễ dàng	0	1	2	0	1	2
2	Tôi khen người khác khi họ làm một việc tốt	0	1	2	0	1	2
3	Khi bị đứa trẻ khác đánh hoặc trêu, tôi mách người lớn	0	1	2	0	1	2
4	Tôi là người tự tin trong các cuộc trò chuyện, làm quen	0	1	2	0	1	2
5	Tôi cố gắng hiểu bạn mình khi họ bực tức, nổi cáu hoặc buồn chán	0	1	2	0	1	2

Để dễ cho việc thiết kế các item, nội dung của trắc nghiệm cần được làm rõ, chi tiết hoá tối đa thành các thành tố/tiêu chí/chỉ số... và thứ bậc hóa theo mức độ quan trọng. Từng thành tố bộ phận phải có nội dung rất cụ thể và các thành tố trọng tâm được xác định như là những khu vực ưu tiên (có nhiều item hơn) và sự lựa chọn item, lựa chọn cách cho điểm cũng phải phản ánh thứ bậc này.

3.3. Miền đo là những cấu trúc

Một số trắc nghiệm được thiết kế đo những cấu trúc, chẳng hạn như các đặc tính nhân cách, hứng thú, thái độ... Đó là những ý nghĩ, hành vi, những thói quen, những tình cảm, khí chất, những sở thích... diễn ra trong cuộc sống hàng ngày của một cá nhân. Những item của các trắc nghiệm kiểu này được thiết kế để nhận biết những đặc tính, do đó không sử dụng các câu hỏi đúng/sai, cũng không thể đo bằng các item hỏi trực tiếp. Sự có mặt của cấu trúc phải được suy luận từ những đo đạc qua một loạt những hành vi cụ thể.

Để cụ thể hoá nội dung của những cấu trúc, người thiết kế phải lập một danh sách gồm những hành vi cụ thể, niềm tin, sở thích, hứng thú... chứng minh sự có mặt của cấu trúc đó, đồng thời liệt kê những hành vi cụ thể, niềm tin, thái độ... không phù hợp với sự có mặt của nó. Những bản liệt kê này được dùng như là những “tiêu chuẩn”, “căn cứ” đối chiếu, so sánh nhằm xác định bản chất, nội dung của các item.

Quá trình cụ thể hóa miến đo thành những cấu trúc là công việc khá phức tạp nhưng rất quan trọng. Quá trình này phức tạp vì các cấu trúc thường là những khái niệm lý thuyết trừu tượng không có những miến đo cụ thể. Do vậy người thiết kế phải có những phân tích chi tiết về cấu trúc trên mô hình lý thuyết để nảy sinh những ý tưởng về các loại hành vi sẽ được đo. Đây là công việc có ý nghĩa quan trọng vì nếu không có sự phân tích chi tiết về những gì công cụ được thiết kế để đo, sẽ rất khó đánh giá độ hiệu lực của nó.

Dưới đây là một ví dụ: phân tích cấu trúc "tính cởi mở, thân thiện" trong giao tiếp:

*** Những hành vi phù hợp với sự có mặt của tính cởi mở, thân thiện:**

- *Chủ động mở đầu các cuộc tiếp xúc, giao tiếp với người khác.*
- *Thường xuyên sử dụng thời gian với người khác.*
- *Sử dụng thời gian với người khác nhiều hơn là với mình.*
- *Chủ động bắt chuyện với người lạ.*
- *Cười, nói "tự nhiên" trong cơ quan của người khác.*
- *Có giao tiếp thường xuyên bằng mắt với người khác.*

*** Những hành vi không phù hợp với sự có mặt của tính cởi mở, thân thiện:**

- *Ngại tiếp xúc với người khác.*
- *Hiếm khi sử dụng thời gian với người khác.*
- *Chỉ tham gia trò chuyện khi người khác chủ động*

bắt chuyện.

- Có những dấu hiệu stress tiêu cực khi ở trong cơ quan của người khác (im lặng, lo ngại, không tự nhiên...).

- Hiếm khi giao tiếp bằng mắt với người khác.

Kỹ thuật thiết kế item của các công cụ có miền đo là những cấu trúc thường được phân thành hai nhóm: chủ quan (phóng chiếu) và khách quan. Những trắc nghiệm sử dụng kỹ thuật thiết kế item chủ quan có các kiểu chung nhất là: tưởng tượng nhìn như trắc nghiệm vết mực loang của Rorschach (Rorschach inkblot), câu chuyện bịa, viết tiếp câu chưa hoàn thành... Trên những item kiểu phóng chiếu này, người ta đưa ra những kích thích mơ hồ, những câu chuyện bịa hay những câu trả lời viết dở... và thông qua câu trả lời để khám phá những đặc tính tâm lý của thân chủ. Còn những trắc nghiệm sử dụng kỹ thuật thiết kế item khách quan có các kiểu điển hình nhất là tự đánh giá, xử lý tình huống, trả lời những câu hỏi cụ thể... người làm trắc nghiệm báo cáo những hành vi, tình cảm, sở thích... của mình thông qua các mức độ: rất đồng ý, đồng ý, không đồng ý hoặc đang phân vân, hoặc rất thường xuyên, thường xuyên, đôi khi, chưa bao giờ...

Tóm lại công việc thiết kế công cụ không thể tiến hành tùy tiện mà phải theo các bước của một quy trình từ khái niệm hoá đến thao tác hoá, cụ thể hoá... và phải tuân thủ các nguyên tắc kỹ thuật, phải am hiểu đối tượng (khách thể nghiên cứu), nội dung nghiên cứu,

định rõ mục tiêu, nội dung đo lường, các kiểu cho điểm, kiểu hướng dẫn... để hạn chế tối đa các nhân tố làm giảm độ tin cậy và độ hiệu lực của công cụ đo.

Chương IV

CÁC KỸ THUẬT PHÂN TÍCH ITEM

Một bộ công cụ tốt phải gồm những item có các chỉ số thống kê tốt. Nhưng làm thế nào để nhận biết được những item tốt, phù hợp với mục đích của trắc nghiệm? Các nhà nghiên cứu đã sử dụng các kỹ thuật phân tích item để đánh giá mức độ phù hợp và nhận biết những item không phù hợp phải viết lại hoặc loại bỏ.

Phân tích item liên quan tới việc sử dụng các kỹ thuật thống kê để đánh giá những đặc tính của từng item. Việc phân tích các đặc tính của item có thể phân chia thành ba nhóm chung nhất sau đây : *Độ khó (difficulty)*, *độ phân biệt (discrimination)* và *năng lực gây nhiễu (distractor power)*.

1. Phân tích độ khó của item

Độ khó của item được dùng để tạo ra sự phân biệt giữa những người làm trắc nghiệm. Mức độ khó của item được xác định dựa trên tỷ lệ người trả lời đúng item đó. Phân tích độ khó của item chỉ phù hợp cho những trắc nghiệm đánh giá thành tích hay năng khiếu, những trắc nghiệm cho phép bộc lộ tối đa năng lực của người làm trắc nghiệm. Phân tích độ khó không phù hợp cho những trắc nghiệm

đánh giá nhân cách, hứng thú, quan điểm, thái độ.

Độ khó của từng item trực tiếp ảnh hưởng đến độ tin cậy và độ hiệu lực của trắc nghiệm. Chẳng hạn, một trắc nghiệm đánh giá năng lực toán gồm 30 item, trong đó có 10 item quá khó, tất cả mọi người làm trắc nghiệm đều nhận điểm 0 trên item này. Như vậy, phân bố điểm của trắc nghiệm gồm 30 item này cũng giống như phân bố điểm của một trắc nghiệm chỉ gồm 20 item. Độ khó của các item đã thực sự làm giảm độ dài của trắc nghiệm. Về mặt lý thuyết, độ dài của trắc nghiệm giảm sẽ làm giảm độ tin cậy của trắc nghiệm và cũng làm giảm độ hiệu lực của trắc nghiệm.

Công thức tính độ khó của item :

$$P = \frac{\text{Số người trả lời đúng trên item}}{N}$$

P : Độ khó của item

N : Tổng số người trả lời item

P có giá trị từ 0,0 đến 1,0. Giá trị của P càng gần 0,0 thì độ khó của item càng tăng. Ngược lại, giá trị của P càng gần 1,0 thì độ khó càng giảm. Muốn tính P đòi hỏi phải phạm trừ hoá điểm của item như là đúng và sai. Nếu dùng thang đo có nhiều mức độ (từ 1-5 điểm), người thiết kế phải ấn định một mức điểm đúng (chẳng hạn từ 4-5 điểm) và mức điểm sai (từ 1-3 điểm).

Độ khó của item không phải chỉ có một giá trị cố định mà ngược lại mỗi khi trắc nghiệm được làm với một nhóm mẫu nào đó, ta lại có một giá trị. Độ khó của item

có thể thay đổi tùy thuộc vào trình độ, năng lực của người làm trắc nghiệm. Một item có độ khó phù hợp nhất khi P nằm trong khoảng 0,5-0,7. Tuy nhiên, để đánh giá được độ khó của item phù hợp nhất cho trắc nghiệm cần căn cứ vào mục đích của trắc nghiệm và kiểu item. Ở những trắc nghiệm đánh giá năng lực liên quan đến mẫu chuẩn (Norm) thường gồm một số item khó và sự thay đổi điểm trắc nghiệm đạt mức tốt nhất khi các item có P gần 0,5

Những trắc nghiệm liên quan đến chuẩn, được dùng để xác định liệu người làm trắc nghiệm có đáp ứng một tiêu chuẩn nào đó về kiến thức hay kỹ năng, nên gồm những item khó, đáp ứng tiêu chuẩn đã đặt ra. Chẳng hạn, một trắc nghiệm đánh giá năng lực tối thiểu nhằm xác định liệu học viên có nắm được những hiểu biết và kỹ năng ở mức tối thiểu đủ theo học một khoá huấn luyện nào đó. Trên những trắc nghiệm kiểu này, người làm trắc nghiệm nhận được điểm đỗ hoặc trượt phụ thuộc vào khả năng liệu họ có đạt chuẩn đặt ra hay không. Một trắc nghiệm đánh giá kiến thức chỉ nhằm phát hiện 20% học sinh kém nhất sẽ khác với một trắc nghiệm nhằm phát hiện 20% học sinh giỏi nhất. Các item của trắc nghiệm trước sẽ gồm nhiều item dễ để 80% học sinh có thể vượt qua, còn trắc nghiệm sau sẽ gồm nhiều những item khó, chỉ để 20% học sinh vượt qua.

Độ khó của item không chỉ nói lên mức độ không hiểu (kiến thức hay kỹ năng) của người làm trắc nghiệm

mà nhiều khi trả lời đúng là do đoán mò. Đối với những item có hai lựa chọn (đúng/sai) hay có nhiều lựa chọn, xác suất đoán trúng là 50% hoặc 25% (nếu có 4 lựa chọn). Như vậy, độ khó tối ưu (optimal item difficulty) được tính theo công thức sau:

$$\text{Điểm giữa P do đoán đúng} = \frac{1,0 - (\text{khả năng có thể đoán đúng})}{2}$$

Độ khó tối ưu cho đoán đúng = điểm giữa P + khả năng đoán đúng

Như vậy, để trắc nghiệm có các đặc tính đo lường tốt nên gồm các item có độ khó từ 0,5 - 0,7.

Độ khó của item sẽ ảnh hưởng đến điểm của trắc nghiệm. Những item quá khó ($P = 0,0$) sẽ làm giảm điểm trung bình, thu hẹp phạm vi điểm vì mọi người mất điểm trên những item này. Do phạm vi điểm của những người làm trắc nghiệm bị thu hẹp dẫn đến làm giảm độ tin cậy, giảm độ hiệu lực, tức là làm giảm năng lực đo sự khác nhau giữa những người làm trắc nghiệm.

Những item quá dễ ($P = 1,0$) sẽ làm mọi người đều có điểm trên item đó, do đó sự khác nhau giữa những người đạt điểm cao nhất (được xem là có năng lực tốt nhất) và những người có điểm thấp nhất (được xem là có năng lực tồi nhất) bị thu hẹp lại. Sự thay đổi điểm số trắc nghiệm trên những người làm trắc nghiệm bị thu hẹp dẫn đến làm giảm độ tin cậy và giảm độ hiệu lực của trắc nghiệm. Như vậy, những item có P gần các giá

trị 0,0 và 1,0 thực chất không đóng góp gì cho các trắc nghiệm năng lực.

Đối với các trắc nghiệm đánh giá nhân cách, thái độ, hứng thú, về lý thuyết, không thích hợp cho việc sử dụng kỹ thuật phân tích độ khó của item, nhưng nếu ta muốn biết các item được trả lời như thế nào với các kiểu người khác nhau, ta có thể dùng một kỹ thuật khác, tính phần trăm của sự nhất trí (percentage endorsement)

$$\% \text{ nhất trí} = \frac{\text{Số người chọn câu trả lời}}{\text{Tổng số người trả lời}}$$

Ví dụ: Thiết kế một trắc nghiệm có các item được trả lời theo kiểu: đúng/sai (true/false) nhằm đánh giá kiểu khí chất hướng nội hay hướng ngoại. Một item của trắc nghiệm này có dạng như: "Tôi thích bắt đầu câu chuyện với người lạ", câu trả lời này có hai lựa chọn: đúng/sai. Trắc nghiệm này được thử trên mẫu gồm 100 người hướng ngoại và 100 người hướng nội (được nhận biết từ một vài các phép đo chuẩn khác). Nếu tất cả 100 người hướng ngoại chọn "đúng" và 100 người hướng nội chọn "sai", lúc đó ta có chỉ số % nhất trí là $100/200 = 0,5$. Tuy nhiên, nếu 50% số người trả lời chọn "đúng" lại thuộc về cả hai nhóm hướng nội và hướng ngoại, lúc đó chỉ số % nhất trí vẫn là 0,5 nhưng không nói cho biết mỗi loại người là bao nhiêu. Do vậy, phải tính % nhất trí cho từng nhóm:

$$(1) \% \text{ người hướng ngoại chọn "đúng"} = \frac{\text{Số người hướng ngoại chọn câu trả lời "đúng"}}{100}$$

$$(2) \% \text{ người hướng nội chọn "đúng"} = \frac{\text{Số người hướng nội chọn câu trả lời "đúng"}}{100}$$

Nếu 80% số người hướng ngoại chọn "đúng" sẽ còn 20% chọn "sai". Cũng vậy, nếu 30% số người hướng nội chọn "đúng" sẽ còn 70% chọn "sai".

Khi đó chuyển sang phân tích độ phân biệt item, ta có thể xác định liệu item này có khả năng phân biệt giữa những người hướng nội và hướng ngoại (tức là liệu hai kiểu người này trả lời item trên có theo những cách khác nhau hay không).

2. Phân tích độ phân biệt của item

2.1. Mô hình lý thuyết

Phân tích độ phân biệt là chỉ ra mức độ khác biệt trong cách trả lời item ở những kiểu người khác nhau. Không giống như phân tích độ khó, phân tích độ phân biệt thích hợp cho hầu hết các kiểu trắc nghiệm.

Trên những trắc nghiệm đánh giá năng lực (còn gọi là các trắc nghiệm đánh giá kết quả bộc lộ tối đa - maximal performance tests), sự phân tích độ phân biệt của item nhằm xác định từng item có năng lực phân biệt giữa những người làm tốt trắc nghiệm và những người làm không tốt. Một item có độ phân biệt tốt là khi trả lời item đó, hầu hết những người có điểm trắc nghiệm cao trả lời đúng, đồng thời hầu hết những người có điểm trắc nghiệm thấp lại trả lời sai.

Trên những trắc nghiệm đánh giá kết quả bộc lộ đặc trưng (typical performance tests) như các trắc nghiệm về nhân cách, hứng thú, thái độ, sự phân tích độ phân biệt của item nhằm đánh giá xem liệu từng item có năng lực phân biệt giữa những người đang sở hữu một đặc tính nào đó và những người đang thiếu hụt đặc tính đó không. Ví dụ, một trắc nghiệm đánh giá trầm cảm được thiết kế để phân biệt giữa người bị trầm cảm và người không bị trầm cảm. Trắc nghiệm này nên gồm những item trên đó những người bị trầm cảm và không bị trầm cảm sẽ trả lời theo những cách khác nhau.

Về lý thuyết, phân tích độ phân biệt tiếp cận từng item như là một phép đo riêng biệt về đặc tính được đo. Trong mỗi trắc nghiệm có thể có một số item nào đó đo những đặc tính không liên quan đến mục đích của trắc nghiệm. Sự phân tích độ phân biệt item có thể được dùng để xác định xem item nào đo tốt nhất cấu trúc hoặc nội dung mà trắc nghiệm có mục đích phải đo.

Nếu một item đo cùng một đặc tính như là toàn bộ trắc nghiệm đo, lúc đó những người có kết quả điểm cao trên trắc nghiệm cũng là người có kết quả điểm tốt trên item đó.

2.2. Các kỹ thuật đánh giá độ phân biệt của item

Có nhiều thủ tục kỹ thuật đánh giá độ phân biệt của item, tuy nhiên có thể xem có hai cách thường được sử dụng nhất: đánh giá chỉ số phân biệt của item (item discrimination index) và đánh giá tương quan điểm item

với điểm trắc nghiệm (item - total correlation)

2.2.1. Đánh giá chỉ số phân biệt item

Độ phân biệt item có thể tính cho các trắc nghiệm năng lực khi các item được cho điểm như là đúng/sai, đúng/không đúng.

Độ phân biệt được xác định từ kết quả so sánh điểm trắc nghiệm của hai nhóm người có điểm số cao và thấp rút ra từ hai mẫu riêng rẽ hoặc từ một mẫu. Nếu lấy từ một mẫu hãy chọn (khoảng 1/3 hay 1/4) số người làm trắc nghiệm có điểm cao nhất và số người làm trắc nghiệm có điểm thấp nhất. Việc chọn 1/3 hay 1/4 trên tổng số người làm trắc nghiệm vào các nhóm này là quyết định của người thiết kế trắc nghiệm, miễn sao số lượng so sánh không quá ít. Sau khi đã xác định được hai nhóm: nhóm có điểm cao và nhóm có điểm thấp, hãy tính tỷ lệ % số người trả lời đúng trên từng item cho mỗi nhóm. Chỉ số phân biệt của item được tính theo công thức sau đây:

$$D = P_T \cdot P_B = \frac{\text{Số người trả lời đúng ở nhóm điểm cao}}{\text{Tổng số người trả lời ở nhóm điểm cao}} - \frac{\text{Số người trả lời đúng ở nhóm điểm thấp}}{\text{Tổng số người trả lời ở nhóm điểm thấp}}$$

D = Độ phân biệt của item

P_T = Tỷ lệ % số người trong nhóm điểm cao trả lời đúng item

P_B = Tỷ lệ % số người trong nhóm điểm thấp trả lời đúng item

Với các trắc nghiệm đánh giá năng lực, ta có thể giải thích sự thay đổi các giá trị của D như sau: Nếu item có câu trả lời khó cho những người ở nhóm điểm thấp mà dễ cho những người ở nhóm điểm cao, thì khi đó chỉ số D tiếp cận là 1,0 (item có độ phân biệt lý tưởng). Nếu item có câu trả lời khó ngang bằng cho cả hai nhóm ($P_T = P_E$), thì chỉ số D tiếp cận là 0,0 (item không có độ phân biệt hoặc có độ phân biệt thấp). Nếu item có câu trả lời dễ cho những người ở nhóm điểm thấp, nhưng lại khó cho những người ở nhóm điểm cao thì chỉ số phân biệt D tiếp cận là -1,0, khi đó item vẫn có độ phân biệt lý tưởng nhưng không giống như mong đợi. Tuy nhiên, hiếm khi D có giá trị: 0,0; 1,0; -1,0.

Không giống như độ khó có giá trị P luôn luôn dương, độ phân biệt có thể có giá trị D là số âm. Tuy nhiên, những trắc nghiệm tốt đòi hỏi các chỉ số phân biệt D phải là số dương.

Nói chung, theo các chuyên gia thiết kế trắc nghiệm, chỉ số phân biệt của từng item được coi là thích hợp khi D lớn hơn hoặc bằng 0,3. Những item có chỉ số D nhỏ hơn hoặc bằng 0,2 nên được viết lại.

Độ phân biệt không chỉ giúp phân biệt những người có điểm số khác nhau mà còn nổi lên tính hiệu lực của trắc nghiệm. Nếu trắc nghiệm và item đo cùng một đặc tính thì những ai có kết quả điểm thấp trên trắc nghiệm cũng là những người có kết quả điểm thấp trên item hoặc ngược lại. Như vậy, chỉ số D dương phù hợp tiêu chuẩn này. Nếu một trắc nghiệm chỉ gồm toàn các item

có chỉ số D dương, tức là chỉ gồm toàn những item đo cùng một đặc tính với trắc nghiệm. Ngược lại, một trắc nghiệm có những item có chỉ số D âm, tức là trắc nghiệm gồm cả những item không cùng đo một đặc tính với trắc nghiệm.

Dưới đây là một ví dụ về một thang đo đánh giá năng lực giải quyết vấn đề gồm 20 item, được thử trên mẫu gồm 30 học sinh, mỗi item có 4 phương án lựa chọn (nhưng chỉ có một phương án lựa chọn tối ưu được coi là đúng, các phương án khác bị coi là sai).

Bảng đánh giá độ khó và độ phân biệt

Item	Số người trả lời đúng	Độ khó thực	Độ khó ước lượng (*)	Số người trả lời đúng trong số 10 học sinh có điểm cao nhất	P_T	Số người trả lời đúng trong số 10 học sinh có điểm thấp nhất	P_B	Độ phân biệt (D)
1	25	0,83	0,75	10	1,0	5	0,5	0,5
2	17	0,57	0,60	8	0,8	4	0,4	0,4
3	23	0,77	0,75	9	0,9	6	0,6	0,3
4	18	0,60	0,75	7	0,7	8	0,8	-0,1
5	16	0,53	0,40	8	0,8	0	0,0	0,8
6	27	0,90	0,95	9	0,9	10	1,0	-0,1
7	15	0,50	0,45	7	0,7	2	0,2	0,5
8	20	0,67	0,60	8	0,8	4	0,4	0,4
9	21	0,70	0,65	8	0,8	5	0,5	0,3
10	22	0,73	0,70	8	0,8	6	0,6	0,2
11	16	0,53	0,65	7	0,7	6	0,6	0,1
12	22	0,73	0,65	9	0,9	4	0,4	0,5

Item	Số người trả lời đúng	Độ khó thực	Độ khó ước lượng (")	Số người trả lời đúng trong số 10 học sinh có điểm cao nhất	P_T	Số người trả lời đúng trong số 10 học sinh có điểm thấp nhất	P_B	Độ phân biệt (D)
13	24	0,80	0,75	9	0,9	6	0,6	0,3
14	18	0,60	0,50	8	0,8	2	0,2	0,6
15	21	0,70	0,75	9	0,9	6	0,6	0,3
16	15	0,50	0,50	8	0,8	2	0,2	0,6
17	18	0,60	0,80	10	1,0	6	0,6	0,4
18	19	0,63	0,60	8	0,8	4	0,4	0,4
19	24	0,80	0,75	9	0,9	6	0,6	0,3
20	26	0,87	0,80	10	1,0	6	0,6	0,4

Trong số 20 item, có 16 item (80%) có D lớn hơn hoặc bằng 0,3. Item số 6 là dễ nhất ($P = 0,90$). Item số 5 chỉ có học sinh thuộc nhóm điểm cao trả lời đúng. Có hai item có giá trị D âm và hai item có giá trị D nhỏ hơn hoặc bằng 0,2. Nói chung, những item có độ phân biệt D nhỏ hoặc âm, không phải là những item tốt, vì vậy chúng nên được viết lại.

Cách tính độ phân biệt cho các trắc nghiệm đánh giá nhân cách, hứng thú, thái độ với item có nhiều mức độ cũng có thể sử dụng kỹ thuật tính toán như các trắc nghiệm đánh giá năng lực. Tuy nhiên, sự giải thích về độ phân biệt của từng item có những điểm khác. Không giống như các trắc nghiệm đánh giá năng lực, các item tốt thường có độ phân biệt D là số dương, còn ở các trắc nghiệm đánh giá về nhân cách, hứng thú, thái độ, các

item tốt có thể có độ phân biệt D là những số âm (nếu D bằng $-1,0$, item đó vẫn có độ phân biệt "lý tưởng" vì nó vẫn tách biệt các kiểu người khác nhau, tuy nhiên quan hệ giữa item và trắc nghiệm không giống như mong đợi).

2.2.2. Đánh giá tương quan giữa item và trắc nghiệm

Ngoài cách đánh giá độ phân biệt của item dựa trên sự so sánh tỷ lệ người ở hai nhóm làm trắc nghiệm có điểm cao và thấp cùng chọn lựa một câu trả lời cụ thể như trên, các nhà thiết kế trắc nghiệm còn hay sử dụng một kỹ thuật khác: *phân tích tương quan điểm của trắc nghiệm với điểm của item* để đánh giá độ phân biệt của từng item cụ thể.

Phương pháp này rất thích hợp với các trắc nghiệm đánh giá nhân cách, hứng thú, thái độ cho điểm item theo kiểu nhiều mức độ. Bản chất của phương pháp này là đánh giá tương quan điểm của item và điểm của trắc nghiệm, để xác định mức độ biến thiên điểm của trắc nghiệm phụ thuộc như thế nào vào điểm của item nhằm chỉ ra năng lực của từng item tham gia dự đoán điểm của trắc nghiệm. Ưu thế của phương pháp này là những đánh giá đó được kiểm định bằng các phép toán thống kê có ý nghĩa. Mặt khác, việc đánh giá mối quan hệ giữa điểm của item và điểm của trắc nghiệm rõ ràng tập trung vào cái mà các item có mục đích phải đo (điều này có liên quan đến độ hiệu lực của trắc nghiệm).

Nếu trắc nghiệm và item cùng đo một đặc tính hay một cấu trúc thì điểm trả lời trên item sẽ tương quan với tổng điểm của trắc nghiệm. Cũng vậy, nếu các item cùng đo một đặc tính hay một cấu trúc thì chúng có tương quan với nhau.

Công thức tính tương quan điểm item và điểm trắc nghiệm như sau:

$$R_{XY} = \frac{(\sum XY / N) - (\bar{X})(\bar{Y})}{(\sigma_X)(\sigma_Y)}$$

X : Điểm của item được phân tích

Y : Điểm tổng của các item còn lại (điểm trắc nghiệm)

$\bar{X}$: Điểm trung bình của item được phân tích

$\bar{Y}$: Điểm trung bình của trắc nghiệm

σ_X : Độ lệch chuẩn của điểm trên item được phân tích

σ_Y : Độ lệch chuẩn của điểm trắc nghiệm

Nếu trắc nghiệm có các item được cho điểm theo kiểu nhị phân (đúng/sai; đồng ý/không đồng ý; có/không có), người ta dùng công thức sau để đánh giá tương quan:

$$R_{PB} = \left(\frac{\bar{X}_1 - \bar{X}}{\sigma_X} \right) \sqrt{\frac{P}{1-P}}$$

$\bar{X}$ = Điểm trung bình trắc nghiệm (cho tất cả mọi người)

$\bar{X}_1$ = Điểm trung bình của những người làm trắc nghiệm trả lời đúng item 1

σ_X = Độ lệch chuẩn của trắc nghiệm

P = Tỷ lệ những người trả lời đúng trên item 1

Giá trị của hệ số tương quan càng tiệm cận 1,0 càng

chúng tỏ item và trắc nghiệm có tính đồng nhất cao, tức là cùng đo một đặc tính. Với các trắc nghiệm đánh giá năng lực, giá trị tương quan càng gần 1,0 sẽ cho biết càng có nhiều người có điểm trắc nghiệm cao trả lời đúng item đó. Ngược lại giá trị tương quan càng gần - 1,0 sẽ cho biết càng có nhiều người có điểm trắc nghiệm thấp trả lời đúng item đó.

Xem lại ví dụ trước:

Item 1 có $D = 0,5 \rightarrow$ item 1 có $R = 0,36$

Item 5 có $D = 0,8 \rightarrow$ Item 5 có $R = 0,73$

Item 11 có $D = 0,1 \rightarrow$ Item 11 có $R = 0,13$

Từ số liệu này ta nhận thấy, hai cách đánh giá độ phân biệt này có sự tương đương. Item 5 có độ phân biệt tốt nhất cũng là item có tương quan cao nhất, trái lại item 11 có độ phân biệt thấp nhất cũng là item có tương quan nhỏ nhất. Theo các chuyên gia trắc nghiệm, các item có R lớn hơn hoặc bằng 0,3 được coi là phù hợp cho trắc nghiệm.

Thực tế tương quan giữa item và điểm trắc nghiệm có thể dẫn đến kết quả đánh giá thổi phồng độ phân biệt (vì điểm của item được phân tích cũng được cộng vào điểm của trắc nghiệm). Do vậy, để điều chỉnh khi phân tích độ phân biệt ta loại bỏ điểm của item được phân tích ra khỏi điểm của trắc nghiệm, tức là tính tương quan điểm của item với điểm tổng của các item còn lại thay cho điểm tổng toàn bộ các item của trắc nghiệm.

Trong khi phân tích item người ta có thể kết hợp phân tích độ phân biệt và độ khó theo cùng một công

thức tính toán, nếu ta chỉ chia số người làm trắc nghiệm thành hai nhóm, khi đó:

$$D = P = \frac{P_T + P_B}{2}$$

Độ khó tính theo công thức này gọi là độ khó ước lượng, nó có giá trị xấp xỉ độ khó thực (xem bảng đánh giá độ khó và độ phân biệt).

3. Phân tích năng lực gây nhiễu của những lựa chọn thay thế

Những trắc nghiệm đánh giá năng lực thường sử dụng các item có nhiều lựa chọn (multiple choice format), trong đó chỉ có một lựa chọn đúng, còn các lựa chọn khác là những lựa chọn thay thế (câu trả lời sai) làm nhiệm vụ gây nhiễu, đánh lạc hướng (distractor) người làm trắc nghiệm. Độ khó của item phụ thuộc vào năng lực gây nhiễu của các lựa chọn thay thế này, do vậy cần phân tích từng lựa chọn thay thế trong item để đánh giá liệu chúng có khả năng phát hiện những sai lầm thường có ở người làm trắc nghiệm hay không.

Nếu các lựa chọn thay thế (lựa chọn sai) được thiết kế tốt (có khả năng phát hiện những lỗi thường gặp của người làm trắc nghiệm) sẽ làm tăng độ hiệu lực của item. Về lý thuyết, một câu hỏi hay một item có nhiều lựa chọn được thiết kế tốt phải có đủ hai đặc tính sau:

(1) - *Những người có hiểu biết, có kỹ năng (thuộc phạm vi trắc nghiệm đo lường) khi trả lời sẽ chọn được*

câu trả lời đúng.

(2) - *Những người thiếu hiểu biết, không có kỹ năng (thuộc phạm vi trắc nghiệm đo lường) khi làm trắc nghiệm sẽ khó chọn được câu trả lời đúng, họ chỉ có thể chọn ngẫu nhiên trong số các câu trả lời. Một số chọn đúng do đoán mò (may rủi), một số chọn sai.*

Như vậy năng lực “đánh lừa” của các lựa chọn sai được mong đợi bằng tỷ lệ % của số người chọn những câu trả lời sai trên số các lựa chọn sai. Số người cụ thể chọn lựa một câu trả lời sai nói lên năng lực gây nhiễu của câu trả lời sai đó. Những câu trả lời sai không được một ai chọn hoặc chỉ có rất ít người chọn bị coi là không có năng lực gây nhiễu, sẽ bị loại bỏ hoặc viết lại. Ngược lại, những câu trả lời sai được quá nhiều người làm trắc nghiệm chọn (hơn cả số người chọn câu trả lời đúng) có thể có vấn đề về bản sắc văn hoá, ngữ nghĩa, độ đọc hiểu... nên được xem xét lại. Tùy mục đích của item và trắc nghiệm mà giữ nguyên hay phải viết lại.

Các lựa chọn sai cần phải được phân tích lý giải dưới ánh sáng của những thông tin về độ khó của item. Khi item cần tăng tối đa độ khó nên gồm những lựa chọn sai được đa số người làm trắc nghiệm chọn. Do độ khó được tạo ra từ sự có mặt của các câu trả lời sai, nên những câu trả lời sai càng có vẻ “giống” với câu trả lời đúng càng làm tăng độ khó của item.

Không có một quy định cứng nhắc về độ phù hợp của những câu trả lời sai làm nhiệm vụ gây nhiễu, do vậy người thiết kế trắc nghiệm phải căn cứ vào mục đích của

trắc nghiệm, nhiệm vụ của item để xác định những lựa chọn sai nào là phù hợp nhất.

Tóm lại, độ nhiễu của những lựa chọn thay thế, độ khó, độ phân biệt của item có ảnh hưởng lẫn nhau và chúng đều ảnh hưởng đến độ tin cậy và độ hiệu lực của trắc nghiệm. Năng lực gây nhiễu của các lựa chọn sai ảnh hưởng trực tiếp đến độ khó của item. Độ khó của item lại ảnh hưởng trực tiếp đến độ phân biệt. Cả độ khó và độ phân biệt của item ảnh hưởng trực tiếp đến phạm vi phân bố điểm trắc nghiệm, vì vậy ảnh hưởng đến độ tin cậy và độ hiệu lực của trắc nghiệm. Thông thường những item có độ khó được ưa thích nằm gần 0,5 (khoảng 0,3-0,8) và độ phân biệt được coi là thích hợp phải có R lớn hơn hoặc bằng 0,3. Những item tốt nhất cho một thang đo để một cấu trúc đồng nhất là những item có quan hệ tuyến tính với điểm tổng của thang đo đó và có độ dốc đường thẳng là những trị số dương gần 1,0.

Chương V

CÁC PHƯƠNG PHÁP CHỌN MẪU

Trong các khoa học xã hội, kết quả nghiên cứu thường phụ thuộc đáng kể vào phương pháp chọn mẫu. Tuy nhiên để có một mẫu nghiên cứu tốt, đòi hỏi người nghiên cứu phải am tường lý thuyết và các phương pháp chọn mẫu.

1. Lý thuyết chọn mẫu

1.1. Các khái niệm

Mẫu (sample) là một tập hợp hay một nhóm các yếu tố được rút ra từ một tổng thể gồm tất cả các yếu tố. Chẳng hạn 10 học sinh là một mẫu được rút ra từ một tổng thể là một lớp gồm 50 học sinh.

Tổng thể (population) là một tập hợp gồm tất cả các yếu tố (đã biết và chưa biết) để từ đó có thể chọn ra một mẫu. Chẳng hạn tất cả sinh viên của một lớp, hoặc tất cả sinh viên của một trường đại học hoặc toàn bộ sinh viên trong cả nước.

Khung mẫu (sample frame) là một "danh sách" liệt kê đầy đủ các thành phần của tổng thể để từ đó mẫu được chọn lựa. Một tổng thể nhỏ có thể được liệt kê chính xác (tất cả sinh viên của một trường đại học)

nhưng một tổng thể lớn khó có thể liệt kê đầy đủ (sinh viên của tất cả các trường đại học trong cả nước). Do vậy khung mẫu là sự cố gắng lớn nhất để liệt kê đầy đủ các thành viên của tổng thể. Nếu khung mẫu không thực sự đại diện cho tổng thể thì mẫu không thể đại diện cho tổng thể.

1.2. Những đặc tính của mẫu

Một trong những đặc tính quan trọng nhất của mẫu là tính đại diện. Tính đại diện của mẫu phụ thuộc chủ yếu vào các nhân tố sau:

- Mẫu sẽ có tính đại diện (cho tổng thể được chọn lựa), nếu tất cả các thành viên của tổng thể đều có cơ hội ngang bằng để được lựa chọn vào mẫu.

- Mẫu sẽ đại diện cho tổng thể nếu bản chất, cấu trúc của nó phản ánh bản chất, cấu trúc của tổng thể (tức là mẫu phải gồm đủ những đặc tính chứa trong tổng thể).

Tính đại diện của mẫu không đòi hỏi nó phải đại diện cho mọi khía cạnh của tổng thể mà chỉ giới hạn đến những đặc tính phù hợp cho những nhu cầu cơ bản của mục tiêu nghiên cứu.

Tính đại diện của mẫu phụ thuộc chủ yếu vào kích cỡ mẫu và tính đồng nhất của mẫu. Mẫu càng lớn, càng đồng nhất thì tính đại diện càng cao. Trong nghiên cứu, người ta thường chú trọng kích cỡ của mẫu, tìm cách tăng cỡ mẫu để nhằm đạt được tính đại diện. Tuy nhiên, thực tế cho thấy tính đại diện của mẫu lại phụ thuộc nhiều hơn vào tính đồng nhất của mẫu.

2. Các phương pháp chọn mẫu cơ bản

Trong nghiên cứu các khoa học xã hội, tùy mục đích nghiên cứu mà người ta có thể áp dụng các phương pháp chọn mẫu khác nhau. Thực tế có nhiều phương pháp chọn mẫu, nhưng có thể chia các phương pháp chọn mẫu thành hai nhóm chính: *chọn mẫu theo phương pháp xác suất* và *chọn mẫu theo phương pháp phi xác suất*.

2.1. Chọn mẫu theo phương pháp xác suất

2.1.1. Khái niệm

Chọn mẫu xác suất (probability sampling) là phương pháp thiết kế mẫu dựa trên các quy tắc của lý thuyết xác suất. Mẫu xác suất liên quan đến cơ chế chọn lựa ngẫu nhiên và theo lý thuyết nó sẽ là một mẫu mang tính đại diện cho tổng thể (do tất cả các thành viên của tổng thể đều có cơ hội ngang bằng để được lựa chọn vào mẫu). Đây là phương pháp được sử dụng rất phổ biến và rất có hiệu lực trong các khoa học xã hội. Chẳng hạn với một mẫu nghiên cứu chọn theo phương pháp xác suất, chỉ gồm 1.579 cử tri, Viện Gallup CNN (Mỹ) đã dự báo chính xác kết quả bầu cử tổng thống Mỹ tháng 11-1992: kết quả dự báo Bush đạt 36%, Clinton - 43%, Perot - 16%, so với kết quả chính thức là: Bush - 38%, Clinton - 43% , Perot - 18%.

2.1.2. Những nguyên tắc cơ bản của chọn mẫu xác suất

Mẫu xác suất chưa bao giờ là đại diện hoàn hảo của tổng thể, nhưng so với những cách lấy mẫu khác, mẫu

xác suất có nhiều ưu thế. Về lý thuyết, mẫu xác suất cho phép người nghiên cứu đánh giá được mức độ chính xác hoặc mức độ đại diện của mẫu. Mức độ chính xác được cụ thể hoá qua mức độ tin cậy và khoảng tin cậy. Mẫu xác suất cũng cho phép đánh giá sai số chuẩn của mẫu, kích cỡ thích hợp của mẫu. Chẳng hạn bao nhiêu % những đánh giá của mẫu sẽ rơi trong khoảng $\pm 1,2,3$ sai số chuẩn tính từ tham số trung bình của tổng thể.

Mẫu xác suất được chọn lựa ngẫu nhiên từ tổng thể sẽ cung cấp những tham số có thể dùng để suy luận cho tổng thể. Vì theo lý thuyết xác suất, nếu có rất nhiều mẫu ngẫu nhiên được chọn độc lập từ tổng thể thì phân bố của mẫu cũng có dạng chuẩn như phân bố của tổng thể (tức là các trung bình mẫu sẽ được phân bố xung quanh trung bình tổng thể). Tuy nhiên điều nhà nghiên cứu muốn biết là liệu chỉ từ một mẫu được chọn ta có thể đạt được độ tin tưởng bao nhiêu % để trung bình của mẫu có giá trị gần bằng giá trị của trung bình tổng thể, tức là sai số chuẩn (sai số mẫu) không quá lớn. Thường thì chúng ta không biết được giá trị trung bình của tổng thể và chúng ta cũng không thiết kế ra được vô số mẫu từ tổng thể để xác định sự phân bố mẫu. Do vậy ta thường dùng đường cong chuẩn mực (đường cong biểu thị sự phân bố tổng thể) để suy luận. Kết quả của những suy luận là ta có thể đánh giá tham số trung bình của tổng thể và mức độ sai số mong đợi (sai số chuẩn) trên cơ sở tham số trung bình mẫu.

Theo lý thuyết xác suất, khi quy mô của mẫu tăng

lên thì sai số mẫu giảm xuống (đường phân bố mẫu gần giống đường cong chuẩn mực). Sai số mẫu được đo bằng sai số chuẩn, đó chính là độ lệch giữa các trung bình mẫu so với trung bình của tổng thể (còn gọi là độ lệch chuẩn của phân bố mẫu). Sai số mẫu phản ánh khả năng khác nhau có thể có của mẫu so với tổng thể (cần phân biệt sai số mẫu với các loại sai số khác mắc phải trong quá trình nghiên cứu như người được điều tra hiểu lầm câu hỏi).

- Sai số chuẩn được tính theo công thức sau:

$$S = \frac{P \times Q}{n}$$

S: Sai số chuẩn

P: % đồng ý

Q: % không đồng ý

n: Số người trả lời

Về logic của lý thuyết xác suất, các mức độ tin cậy và khoảng tin cậy có thể cung cấp cơ sở khoa học cho việc xác định cỡ mẫu thích hợp cho một nghiên cứu. Một khi chúng ta quyết định mức độ sai số mẫu có thể chấp nhận được thì ta sẽ tính được số lượng khách thể tối thiểu cần nghiên cứu trong mẫu. Chẳng hạn ta muốn đạt độ tin cậy 95% cho các số liệu nghiên cứu với độ chính xác $\pm 5\%$ từ các tham số của tổng thể ta nên chọn một mẫu có dung lượng ít nhất là 400 khách thể.

2.1.3. Các kiểu thiết kế mẫu

- *Chọn mẫu ngẫu nhiên đơn giản* (simple random

sampling), còn gọi là mẫu xác suất không đều.

Đây là phương pháp chọn mẫu trong đó các thành viên của tổng thể được đánh số sau đó lựa chọn một cách ngẫu nhiên theo danh sách hoặc chọn theo những số ngẫu nhiên do máy tính tạo ra. Các chương trình máy tính hiện đại như phần mềm SPSS có cài sẵn các chương trình này, máy sẽ chọn ra các số ngẫu nhiên sau khi ta định rõ khoảng trong đó có các con số này. Với bất kỳ phương pháp chọn số ngẫu nhiên nào, mỗi số đều có cơ hội được chọn ngang nhau.

- *Chọn mẫu ngẫu nhiên hệ thống* (systematic sampling) còn gọi là phương pháp chọn mẫu theo xác suất đều.

Đây là phương pháp chọn mẫu xác suất, trong đó những thành viên thứ n trên một khung mẫu được lựa chọn vào mẫu, sau khi đã lựa chọn được một số đầu tiên ngẫu nhiên. Chẳng hạn ta muốn có một mẫu gồm 200 sinh viên được rút ra từ một tổng thể gồm 5.000 sinh viên của một trường đại học, trước hết ta có một khung mẫu gồm tất cả 5.000 sinh viên được đánh số thứ tự từ 1 đến 5000. Sau đó trong khoảng từ 1-25, chọn ngẫu nhiên một số nào đó, chẳng hạn số 10. Như vậy, những người có số thứ tự 10, 35, 60, 75... (cách một khoảng đều là 25) được chọn cho tới khi đủ 200 người.

Điều kiện để thiết kế mẫu ngẫu nhiên hệ thống là danh sách các thành viên của khung mẫu được xếp ngẫu nhiên (có thể theo thứ tự tên A, B, C...) không theo một trật tự thiên vị nào.

$$\text{Ví dụ:} \quad \text{Khoảng mẫu} = \frac{\text{Cỡ tổng thể}}{\text{Cỡ mẫu}} = \frac{10.000}{100} = 100$$

$$\text{Tỷ lệ mẫu} = \frac{\text{Cỡ mẫu}}{\text{Tổng thể}} = \frac{100}{10.000} = 1/100$$

• **Chọn mẫu phân tầng (stratified sampling)**

Kích cỡ của mẫu và tính đồng nhất của mẫu là hai nhân tố ảnh hưởng trực tiếp đến sai số mẫu. Mẫu càng lớn thì sai số chuẩn càng nhỏ và ngược lại, mẫu càng nhỏ thì sai số mẫu càng lớn. Mẫu không đồng nhất sẽ sinh ra sai số mẫu lớn và ngược lại mẫu đồng nhất sẽ sinh ra sai số nhỏ.

Mẫu phân tầng là một phương pháp tổ chức tổng thể thành những tầng có tính đồng nhất để chọn mẫu nhằm đạt mức đại diện tốt nhất. Do tổng thể khó có tính đồng nhất nên người ta chia tổng thể thành các nhóm nhỏ (phân tầng) để tăng tính đồng nhất trong các tiểu mẫu và sau đó mới chọn ra những khách thể từ những nhóm đồng nhất này tạo thành mẫu. Thông thường người ta chia khung mẫu thành nhiều tầng, dựa theo giới tính, dân tộc, khu vực, tuổi, trình độ học vấn, điều kiện kinh tế, kinh nghiệm nghề... vì những yếu tố này có khả năng ảnh hưởng đến những câu trả lời.

Mẫu phân tầng thường được thiết kế kết hợp với mẫu ngẫu nhiên đơn giản hoặc mẫu ngẫu nhiên hệ thống. Trước hết ta phân loại các thành phần của tổng thể vào

các nhóm cụ thể dựa trên các biến phân tầng được sử dụng (các tiểu mẫu). Sau đó trên cơ sở các nhóm với tỷ lệ tương đối đại diện cho tổng thể, ta chọn ngẫu nhiên đơn giản hay ngẫu nhiên hệ thống. Ví dụ, để chọn một mẫu sinh viên đại diện cho thành phố hay cả nước nhằm đánh giá lối sống của họ có thể phân tầng thành các nhóm sinh viên sau: công lập/dân lập, nam/nữ, khu vực (Hà Nội/Đà Nẵng/thành phố Hồ Chí Minh), sinh viên nội trú (sống trong ký túc xá)/ngoại trú... theo một tỷ lệ tương đối cân xứng trong tổng thể sinh viên. Rồi từ những nhóm này lập khung mẫu (danh sách) chọn theo phương pháp ngẫu nhiên đơn giản hay ngẫu nhiên hệ thống.

Trong mẫu phân tầng những tiểu nhóm quyết định thường được chọn mẫu nặng hơn (chọn nhiều hơn) các nhóm khác. Kỹ thuật điều chỉnh hay chọn mẫu không cân xứng này được tiến hành để tạo ra một số lượng đủ lớn ở các nhóm chính trong mẫu đáp ứng mục tiêu nghiên cứu.

- *Mẫu cụm nhiều giai đoạn* (multistage cluster sampling)

Mẫu cụm nhiều giai đoạn được thiết kế khi ta không thể thu thập được một danh sách đầy đủ các thành phần từ tổng thể để thiết kế mẫu. Chẳng hạn sinh viên của tất cả các trường đại học đóng trên địa bàn Hà Nội hoặc tất cả sinh viên các trường đại học trong cả nước.

Mẫu cụm khác với mẫu phân tầng ở chỗ: trong mẫu phân tầng, mỗi tầng là một nhóm đồng nhất (các tầng

được tạo ra do mục tiêu của sự chọn mẫu nhằm làm tăng tính đại diện). Trong khi đó, ở cách chọn mẫu cụm nhiều giai đoạn, các cụm lại liên kết các nhóm không đồng nhất.

Cách chọn mẫu cụm nhiều giai đoạn đòi hỏi trước hết phải tạo ra một danh sách các cụm không đồng nhất, sau đó chọn mẫu. Từ mẫu khởi đầu gồm các cụm đại diện cho tổng thể ta lại tạo ra một danh sách gồm các thành phần của từng cụm, sau đó chọn mẫu gồm các thành phần đại diện cho từng cụm. Như vậy, mẫu cụm nhiều giai đoạn liên quan đến sự lặp đi lặp lại hai bước cơ bản: lập danh sách và chọn mẫu ngẫu nhiên.

Điều quan trọng nhất trong mẫu cụm là danh sách của tất cả các yếu tố được yêu cầu cho các đơn vị mẫu cuối cùng không cần phải có từ đầu mà chỉ cần có trong danh sách các cụm.

Chọn mẫu cụm thường được thiết kế phối hợp với chọn mẫu phân tầng. Trước hết chọn các cụm sau đó chọn các tiểu nhóm (các tầng) trong mỗi cụm, rồi đến các đơn vị (cá nhân) trong các tầng. Ví dụ, để chọn được một mẫu học sinh trung học phổ thông đại diện cho Thành phố Hà Nội ta có thể tiến hành một số bước sau:

1. Lập danh sách tất cả các trường trung học phổ thông. Từ danh sách này ta phân thành hai cụm: trung học phổ thông công lập và trung học phổ thông dân lập. Từ hai cụm lớn này phân thành các cụm nội thành ngoại thành. Căn cứ theo tỷ lệ các trường công lập/dân lập, nội thành/ngoại thành. Sau đó ta lập danh sách các

trường cho từng cụm, rồi dùng phương pháp ngẫu nhiên hệ thống hay ngẫu nhiên đơn giản để chọn ra các trường.

2. Từ các danh sách các trường được chọn, ta lập danh sách các lớp trong từng khối (phân tầng) căn cứ theo tỷ lệ các khối lớp, chọn ngẫu nhiên ra các lớp.

3. Lập danh sách học sinh ở các lớp đã chọn lại phân tầng theo giới tính, tuổi, điều kiện kinh tế... Sau đó dùng phương pháp ngẫu nhiên đơn giản hay hệ thống chọn ra một mẫu học sinh.

Một mẫu cụm đại diện tốt nhất cho tổng thể khi có một số lượng lớn các cụm được đưa vào mẫu. Trong mỗi cụm càng có nhiều thành phần càng tăng tính đại diện. Tuy nhiên, hướng dẫn chung cho thiết kế mẫu cụm là tăng tối đa số lượng cụm (nhóm) được chọn và giảm số lượng thành phần được chọn ở mỗi cụm (những mẫu gồm một số lượng nhỏ các cụm dễ thiết lập hơn và chi phí ít tốn kém hơn).

Khi có quá nhiều giai đoạn lựa chọn, các sai số mẫu có chiều hướng gia tăng, chúng ta có thể dùng kỹ thuật điều chỉnh: *định tỷ lệ các tầng trong các cụm sao cho ở giai đoạn mẫu cuối cùng các đơn vị (cá nhân) đều có cơ hội được chọn ngang nhau*. Ví dụ, có 1.000 khu nhà tập thể trong thành phố, ta cần chọn một mẫu gồm 100 khu nhà. Trong một khu cứ 5 dãy nhà ta chọn 1 dãy nhà, mỗi dãy nhà cứ 10 hộ ta chọn 1 hộ. Như vậy cơ hội cho mỗi hộ được chọn ở mẫu cuối cùng là:

$$\frac{1}{10} \times \frac{1}{5} \times \frac{1}{10} = \frac{1}{500}$$

Tuy nhiên số nhà của mỗi khu tập thể rất khác nhau, số hộ ở mỗi nhà cũng rất khác nhau. Như vậy những khu tập thể lớn có đông hộ, sẽ có cơ hội được chọn cao hơn. Nhưng do những hộ trong một nhà có tính đồng nhất và những dãy trong cùng một khu cũng có tính đồng nhất, nên không nhất thiết chọn mẫu gồm quá nhiều hộ trong cùng một dãy nhà hoặc quá nhiều dãy nhà trong cùng một khu tập thể.

2.1.4. Những gợi ý trong thực hành chọn mẫu

Tính khái quát hoá cao chỉ có thể có từ một mẫu được thiết kế có tính đại diện. Do vậy mẫu xác suất có thể đáp ứng yêu cầu này. Những nghiên cứu khảo sát thực trạng thường phải dựa trên cơ sở các mẫu xác suất còn các nghiên cứu thực nghiệm có thể không cần. Như vậy, mục tiêu nghiên cứu sẽ quyết định loại mẫu nào là cần thiết, thích hợp. Khi không thể xây dựng được một mẫu ngẫu nhiên, có một số kỹ thuật phi xác suất có thể sử dụng. Những mẫu kiểu này thường có nhiều hạn chế vì các đối tượng chọn vào mẫu không đại diện cho tổng thể các đối tượng mà chỉ đại diện cho chính nó.

2.2. Chọn mẫu phi xác suất

Trong nhiều trường hợp việc chọn mẫu xác suất không thể thực hiện được, nhà nghiên cứu cần thiết phải chọn mẫu không theo các quy tắc chọn mẫu xác

suất. Những mẫu như vậy gọi là mẫu phi xác suất (nonprobability sampling). Mặc dù mẫu phi xác suất khó có thể là mẫu khái quát đại diện cho một tổng thể nghiên cứu lớn hơn, nhưng trong nhiều trường hợp nó tỏ ra rất có hiệu quả, đặc biệt là trong các nghiên cứu phát hiện, khám phá, tìm kiếm các ý tưởng mới.

Theo các nhà nghiên cứu những kiểu chọn mẫu phi xác suất thường hay sử dụng là *mẫu tiện lợi*, *mẫu phán đoán*, *mẫu chỉ tiêu* và *mẫu tăng nhanh*.

2.2.1. Mẫu tiện lợi (convenience sampling)

Khó có thể tiến hành một cuộc nghiên cứu về gái mại dâm theo phương pháp chọn mẫu xác suất (vì không thể có được danh sách đầy đủ những người tham gia hoạt động mại dâm). Do vậy có thể tiến hành chọn một mẫu gồm những người có nhiều khả năng đã tham gia hoạt động mại dâm, họ là những người sẵn sàng trả lời các câu hỏi nghiên cứu. Chẳng hạn, một nhóm sinh viên tham gia một câu lạc bộ được yêu cầu trả lời những câu hỏi nghiên cứu, thì mẫu này gọi là mẫu tiện lợi. Mẫu tiện lợi chỉ là một mẫu có khả năng đưa ra những số liệu đại diện cho chính nó và có khả năng làm cho những số liệu nghiên cứu hấp dẫn người nghiên cứu. Tuy vậy người nghiên cứu không thể “nhắm mắt” chọn bừa mà cần phải cân nhắc kỹ trước khi chọn mẫu để đáp ứng tốt nhất cho mục đích nghiên cứu (chẳng hạn những người được dự định khảo sát liệu có hiểu biết những vấn đề ta muốn khảo sát và liệu có sẵn sàng trả

lời hay không...).

2.2.2. *Mẫu phán đoán (judgmental sampling)*

Mẫu phán đoán là một hình thức chọn mẫu phi xác suất trong đó các đối tượng được chọn có vẻ đáp ứng những yêu cầu, mục tiêu nghiên cứu. Chẳng hạn để nghiên cứu gái mại dâm, người nghiên cứu có thể đến các trại phục hồi nhân phẩm, hoặc vào các nhà hàng, khách sạn... phỏng vấn những người có nhiều khả năng nhất là gái mại dâm.

Mẫu phán đoán cũng giống như mẫu tiện lợi khó có khả năng khái quát để đại diện cho một tổng thể lớn hơn, nhưng nó có thể đáp ứng tốt nhất những yêu cầu đặc biệt của nghiên cứu (chẳng hạn muốn phỏng vấn sâu, tìm hiểu rõ nguyên nhân...).

2.2.3. *Mẫu chỉ tiêu (quota sampling)*

Đây là hình thức chọn mẫu phi xác suất hay được sử dụng nhất và thường bị hiểu lầm là mẫu xác suất phân tầng. Bởi vì mẫu chỉ tiêu đòi hỏi chọn các tiểu mẫu với dung lượng nhất định trên cơ sở các nhóm nghiên cứu đã được xác định rõ ràng.

Sự khác nhau giữa mẫu chỉ tiêu và mẫu phân tầng là: mẫu chỉ tiêu không có các khung mẫu để chọn. Hơn nữa các nhóm đã được xác định trước, dung lượng mẫu được xác định trước và các cá nhân được chọn theo chỉ tiêu bất kể họ ở đâu.

Để chọn mẫu chỉ tiêu, trước hết phải hình thành một ma trận các đặc trưng mong muốn giới tính, tuổi, chủng

tộc, điều kiện kinh tế, ngành học, năm học... Sau đó căn cứ trên tỷ lệ phần trăm để xác định dung lượng mẫu cho từng chi tiết. Ví dụ, để nghiên cứu thái độ của sinh viên đối với các sinh hoạt ngoại khoá của một trường đại học, người nghiên cứu phải xem xét số lượng, tỷ lệ % nam/ nữ sinh viên ở từng năm, tỷ lệ % sinh viên ngoại trú / nội trú... Giả sử có 80% nam, 20% nữ, 65 % thuộc các ngành tự nhiên, 35% thuộc các ngành xã hội, 70% ngoại trú, 30% nội trú, 30% năm thứ nhất, 25% năm thứ hai, 25% năm thứ ba, 20% năm thứ tư... Bây giờ người nghiên cứu quyết định lấy mẫu 100 sinh viên theo các tỷ lệ trên: 80 nam, 20 nữ, 65 sinh viên thuộc các ngành tự nhiên, 35 sinh viên các ngành xã hội, 70 sinh viên ngoại trú, 30 sinh viên nội trú, 30 sinh viên năm thứ nhất, 25 sinh viên năm thứ hai, 25 sinh viên năm thứ ba và 20 sinh viên năm thứ tư... Do các nhóm có sự gối lên nhau, nhà nghiên cứu phải lập ma trận để tính.

Khi số lượng cụ thể cho từng chỉ tiêu đã được xác định, thì thủ tục tiếp theo đơn giản chỉ là chọn đủ người vào các chỉ tiêu.

Bằng phương pháp này Gallup (1936) đã dự báo chính xác Roosevelt trúng cử tổng thống Mỹ. Liên tiếp các năm 1940, 1944 Gallup và Viện nghiên cứu dư luận Mỹ đã dùng phương pháp chọn mẫu này dự báo chính xác kết quả bầu cử tổng thống Mỹ, nhưng năm 1948 họ đã thất bại khi đưa ra dự báo rằng Thống đốc Thomas Dewey thắng cử trong khi kết quả bầu cử hai tuần sau đó, Tổng thống Harry Truman đã thắng cử. Sự thất bại

này chủ yếu là do lỗi trong chọn mẫu (các tiểu nhóm không được chọn trên cơ sở các khung mẫu gồm tất cả các thành viên của tiểu nhóm). Trong thiết kế mẫu, mỗi thành viên trong các tiểu nhóm không có cơ hội như nhau để được chọn vào mẫu nghiên cứu, tức là sự lựa chọn các trường hợp trong các nhóm chỉ tiêu có thể bị thiên vị, điều này làm hỏng tính đại diện của mẫu. Kết quả là khó có thể xác định được giá trị khoa học của những suy luận từ mẫu chỉ tiêu cho một tổng thể rộng hơn.

2.2.4. Mẫu tăng nhanh (snowball sampling)

Mẫu này thích hợp cho những nghiên cứu giải thích, khám phá mà người nghiên cứu thực sự không có khả năng biết được những đối tượng nào nên bao gồm vào mẫu nghiên cứu. Chọn mẫu tăng nhanh bắt đầu bằng việc chọn một vài đối tượng có những phẩm chất cần tìm, phỏng vấn các đối tượng này và nhờ họ giới thiệu những đối tượng khác thích hợp cho mục tiêu nghiên cứu.

Cách tiếp cận này có thể đặc biệt thích hợp cho việc nghiên cứu các tệ nạn xã hội: mãi dâm, ma túy, cờ bạc...

Tóm lại, tuy mục đích nghiên cứu, nhà nghiên cứu lựa chọn kiểu thiết kế mẫu thích hợp. Nhà nghiên cứu cần hiểu rõ những đặc điểm lý tưởng của từng kiểu chọn mẫu để đáp ứng ở mức cao nhất các mục tiêu nghiên cứu. Quá trình thiết kế mẫu để bảo đảm tính khả thi, giảm chi phí, có thể bỏ qua một số đặc điểm nào đó của thiết kế mẫu lý tưởng. Song bất kể chọn kiểu thiết kế

mẫu nào thì nhà nghiên cứu cũng phải giải thích chính xác, cặn kẽ quá trình xây dựng mẫu, cảnh báo những vấn đề có thể có của quá trình thực hiện và khả năng điều chỉnh tối ưu để có một mẫu phù hợp nhất cho mục đích nghiên cứu.

Chương VI

MỘT SỐ PHƯƠNG PHÁP CƠ BẢN ĐÁNH GIÁ ĐỘ TIN CẬY CỦA CÔNG CỤ ĐO LƯỜNG

Trong nghiên cứu khoa học giáo dục và khoa học tâm lý, giá trị khoa học của các số liệu thu được phụ thuộc chủ yếu vào độ tin cậy và độ hiệu lực của bộ công cụ dùng để đo đạc. Tuy nhiên nhà nghiên cứu không phải lúc nào cũng sẵn có những bộ công cụ đã được chuẩn hoá (có độ tin cậy và độ hiệu lực bảo đảm) thích hợp cho việc thu thập số liệu nghiên cứu. Do vậy khi sử dụng những công cụ do mình tự thiết kế hoặc dịch những bộ công cụ của nước ngoài chưa qua quá trình thích nghi hoá, thì một đòi hỏi có tính “bắt buộc” là nhà nghiên cứu phải đánh giá độ tin cậy và độ hiệu lực của công cụ đo, trước khi sử dụng các số liệu đo lường của bộ công cụ đó nhằm rút ra những kết luận hoặc dùng số liệu đó để suy đoán.

Dưới đây là những phương pháp cơ bản thường được các chuyên gia thiết kế trắc nghiệm ưa dùng trong quá trình phân tích đánh giá độ tin cậy của trắc nghiệm.

1. Mô hình lý thuyết về độ tin cậy

Một phép đo tốt, trước hết phải có đủ độ tin cậy, đo

đúng cái nó được thiết kế để đo. Nói một cách cụ thể, hình ảnh hơn là: “nếu A trầm cảm hơn B thì A sẽ có điểm số cao hơn B trên thang đo trầm cảm”. Tuy nhiên để xác định độ tin cậy của một phép đo, người ta lại xuất phát từ một tiền đề có tính giả thiết rằng phép đo sẽ đo lường một đặc tính tương đối ổn định. Bởi vì sẽ không có một phương pháp đo đặc nào có giá trị trừ phi nó đo một cái gì đó có tính tương đối ổn định và bản thân phép đo đó phải có tính ổn định tương đối, ít hoặc không thay đổi giữa các lần đo khác nhau. Những số liệu phân tích kết quả làm trắc nghiệm chỉ ra rằng mọi người thường có kết quả làm trắc nghiệm tương đối ổn định qua các lần đo trừ phi có những thay đổi thực sự về hiểu biết, kỹ năng liên quan đến đặc tính được đo giữa các lần làm trắc nghiệm. Như vậy, một thang đo có độ tin cậy tốt được xác định như là một phép đo trên đó mọi người đạt số điểm tương tự ở những lần đo khác nhau. Điều này không có nghĩa là điểm trắc nghiệm của một cá nhân phải tuyệt đối giống nhau ở các lần đo. Điểm trắc nghiệm có thể khác nhau giữa các lần đo, nhưng nếu không có sự thay đổi thực sự của tri thức, kỹ năng của đặc tính được đo thì nếu một người có điểm trắc nghiệm cao hơn điểm trung bình một độ lệch chuẩn ở lần đo thứ nhất cũng sẽ có số điểm cao hơn điểm trung bình một độ lệch chuẩn ở lần đo sau.

Các lý thuyết về trắc nghiệm đều cho rằng điểm số của trắc nghiệm đo lường một đặc tính, phẩm chất hay năng lực nào đó ở một cá nhân bằng điểm thực của cá

nhân cộng với điểm sai số của sự đo lường.

$$X = T + E$$

X: *Điểm trắc nghiệm*

T: *Điểm thực*

E: *Điểm sai số của phép đo*

Điểm thực được xác định như là điểm trung bình của vô số những lần đo của trắc nghiệm trên một cá nhân. Còn điểm sai số của sự đo lường là sự khác biệt ở mỗi lần đo do các yếu tố cá nhân (động cơ, xúc cảm, tâm thế, stress...), do các yếu tố tình huống, hoàn cảnh ảnh hưởng (tiếng ồn, ánh sáng, lời hướng dẫn...), nhưng chúng ta chưa bao giờ đo được điểm thực ở cá nhân một cách thật chính xác. Như vậy, để phân tích độ tin cậy của trắc nghiệm người ta phải dựa vào kết quả làm trắc nghiệm của một nhóm người (điểm trắc nghiệm của mẫu) nhằm nhận diện tỷ lệ thành phần của điểm thực và điểm sai số trong kết quả điểm trắc nghiệm của cả nhóm. Nói một cách cụ thể hơn, phân tích độ tin cậy là tìm cách xác định mức độ trên đó sự khác biệt điểm số trắc nghiệm phản ánh sự khác biệt điểm thực ở nhóm người làm trắc nghiệm. Do vậy để đánh giá độ tin cậy của trắc nghiệm, người ta phải giả thiết rằng phương sai của những điểm số đo được bằng trắc nghiệm ở một nhóm nghiệm thể (mẫu) phải bằng phương sai của những điểm số thực cộng với phương sai của những điểm sai số của mỗi lần đo, như công thức sau:

$$\sigma^2_X = \sigma^2_T + \sigma^2_E$$

σ^2_X = giá trị phương sai của điểm trắc nghiệm

σ^2_T = Giá trị phương sai của điểm thực

σ^2_E = Giá trị phương sai điểm sai số của sự đo lường

Điều này được hiểu là sự khác nhau của mọi người về điểm số trắc nghiệm phản ánh sự khác nhau giữa họ về những kiến thức, hiểu biết, kỹ năng, cộng với sự khác biệt do sai số ngẫu nhiên của sự đo lường. Khi đó độ tin cậy của trắc nghiệm được xác định như là hệ số tương quan (R_{XX}) biểu thị tỷ lệ hay phần trăm giữa phương sai của điểm thực trên phương sai của điểm trắc nghiệm, theo công thức sau:

$$R_{XX} = \frac{\sigma^2_T}{\sigma^2_X} = \frac{\sigma^2_T}{\sigma^2_T + \sigma^2_E}$$

Theo công thức này, nếu $R_{XX} = .80$, điều này có nghĩa rằng 80% phương sai của điểm trắc nghiệm là phương sai của điểm thực và chỉ có 20% là phương sai của điểm do sai số ngẫu nhiên của sự đo lường, điều này nói rằng trắc nghiệm có độ tin cậy cao. Ngược lại, nếu $R_{XX} = .30$, điều này có nghĩa rằng chỉ có 30% phương sai của điểm trắc nghiệm là phương sai của điểm thực, còn 70% là phương sai của điểm sai số của sự đo lường, điều này nói lên rằng trắc nghiệm có độ tin cậy rất thấp.

Một trắc nghiệm có độ tin cậy phải có một tỷ lệ lớn hơn của phương sai điểm trắc nghiệm thuộc về sự khác biệt ở điểm thực (ngược lại, một trắc nghiệm không có độ tin cậy sẽ có một tỷ lệ lớn hơn của phương

sai điểm trắc nghiệm thuộc về sự khác biệt ở điểm sai số của sự đo lường), tức là sự khác biệt thực sự về kiến thức, kỹ năng hoặc phẩm chất, thái độ của người làm trắc nghiệm. Điều này được minh họa trong bất phương trình sau:

$$\frac{\sigma^2_T}{\sigma^2_X} > \frac{\sigma^2_E}{\sigma^2_X}$$

Phối hợp hai công thức trên ta có:

$$\frac{\sigma^2_E}{\sigma^2_X} = 1 - R_{XX}$$

Như vậy, độ tin cậy của một phép đo luôn là số dương thập phân từ mức thấp nhất là .00 đến mức cao nhất là 1.00. Khi sai số của sự đo lường càng nhỏ thì độ tin cậy càng tăng và độ tin cậy đạt mức lý tưởng 1.00 khi sai số của sự đo lường bằng 0. Do đó mục tiêu của các chiến lược đánh giá độ tin cậy là đánh giá những sai số của sự đo lường nhằm tìm cách giảm thiểu những sai số này, nhờ đó làm tăng độ tin cậy của phép đo.

2. Các phương pháp cơ bản đánh giá độ tin cậy của phép đo

Có nhiều phương pháp khác nhau để đánh giá độ tin cậy của phép đo. Theo các chuyên gia thiết kế trắc nghiệm, có bốn nhóm phương pháp cơ bản sau: *nhóm phương pháp đánh giá mức độ kiên định về điểm số giữa*

hai lần đo (*test-retest methods*); nhóm phương pháp sử dụng form thay thế tương đương (*alternate form methods*); nhóm phương pháp phân đôi số item của trắc nghiệm (*split-half methods*); và nhóm phương pháp đánh giá độ phù hợp của từng item (*internal consistency methods*).

Nhóm phương pháp thứ nhất: Để đánh giá độ tin cậy của trắc nghiệm người ta thường so sánh tương quan điểm giữa hai lần đo của cùng một trắc nghiệm trên cùng một đối tượng (*test-retest correlation*). Một trắc nghiệm có độ tin cậy phải có tính ổn định, tức là có sự tương quan đáng kể giữa hai lần đo. Trong thống kê, nếu trắc nghiệm được thiết kế theo thang định khoảng hay thang định tỷ lệ và mối quan hệ điểm số trắc nghiệm giữa hai lần đo là quan hệ tuyến tính, thì người ta dùng tương quan Pearson (*Pearson correlation coefficient*) để đánh giá hệ số tin cậy của trắc nghiệm. Còn nếu trắc nghiệm được thiết kế theo thang định hạng, hoặc mối quan hệ điểm số trắc nghiệm giữa hai lần đo không phải là quan hệ tuyến tính, thì người ta dùng tương quan Spearman (*Spearman rank correlation coefficient*) để đánh giá hệ số tin cậy của trắc nghiệm.

Vì cùng một trắc nghiệm đo hai lần trên cùng một người, nên sự khác nhau về điểm số giữa hai lần đo được xem là “hoàn toàn do sai số của sự đo lường”. Nếu có sự khác biệt nhỏ giữa hai lần đo thì người ta có thể kết luận là sai số của phép đo nhỏ và phép đo đó có độ tin cậy cao. Tuy nhiên nhóm phương pháp

này cũng có những vấn đề (hạn chế), chẳng hạn khoảng cách về thời gian giữa hai lần đo thường có ảnh hưởng đến hệ số tương quan (do thời gian làm biến đổi những đặc tính được đo) hoặc nghiệm thể có thêm kinh nghiệm, nghiệm thể có thể quen, nhớ được các câu trả lời... Do vậy hệ số tương quan điểm giữa hai lần đo thường được gọi là hệ số ổn định của trắc nghiệm hơn là hệ số tin cậy. Tuy nhiên, nhóm phương pháp này đặc biệt có lợi khi người ta quan tâm đến tính ổn định tương đối của trắc nghiệm qua các lần đo (chẳng hạn những trắc nghiệm thăm dò thái độ của cử tri). Cũng tùy mục đích của trắc nghiệm và điều kiện thực tế mà khoảng cách giữa hai lần đo có thể được điều chỉnh từ vài ngày, vài tuần đến vài tháng.

Nhóm phương pháp thứ hai: Một trắc nghiệm sử dụng những form tương đương: như form A và form B có hình thức, câu chữ khác nhau nhưng có nội dung giống nhau hoặc tương đương và có những đặc tính thống kê như nhau hoặc tương đương. Tương quan điểm số giữa form A (lần đo thứ nhất) và form B (lần đo thứ hai) được xem là hệ số tin cậy của trắc nghiệm. Về lý thuyết nhóm phương pháp này có thể khắc phục được những hạn chế ở nhóm phương pháp trước (trắc nghiệm được đo hai lần trên cùng một đối tượng) nhưng lại nảy sinh những vấn đề mới. Chẳng hạn, rất tốn kém và rất khó thiết kế một trắc nghiệm có nhiều form hoàn toàn tương đương.

Nhóm phương pháp thứ ba: Thường được dùng để đánh giá độ tin cậy của một trắc nghiệm khi ta không

có điều kiện đo trắc nghiệm hai lần trên cùng một đối tượng và cũng không thiết kế được các trắc nghiệm tương đương, là phương pháp phân đôi số item của trắc nghiệm thành hai phần (thường chia theo số chẵn và số lẻ) rồi so sánh tương quan điểm giữa hai nửa trắc nghiệm. Đây gọi là phương pháp tính độ tin cậy phân đôi trắc nghiệm (split-half reliability). Về lý thuyết, hai nửa trắc nghiệm có thể được xem là hai form tương đương của cùng một trắc nghiệm. Do vậy tương quan điểm số của hai nửa này được xem là hệ số tin cậy của trắc nghiệm.

Công thức tính như sau:

$$R_{X0XE} = \frac{(\Sigma X0XE) - (\bar{X}0)(\bar{X}E)}{(\sigma_{X0})(\sigma_{XE})}$$

$X0$ = Các item số lẻ

XE = Các item số chẵn

Tuy nhiên, khi sử dụng phương pháp này hệ số tin cậy sẽ giảm đi so với độ tin cậy thật của trắc nghiệm vì trắc nghiệm vốn có số item gấp 2 lần (mà theo lý thuyết, trắc nghiệm càng có nhiều item thì độ tin cậy càng cao). Rất may trong thống kê có công thức khác của Spearman chuyển hệ số tin cậy phân đôi thành hệ số tin cậy của toàn bộ trắc nghiệm gọi là công thức Spearman-Brown prophesy:

$$R = \frac{(N)(R_{XX})}{1 + (N-1)(R_{XX})}$$

$$N = 2$$

$R =$ Hệ số tin cậy mới (đã được điều chỉnh)

$R_{xx} =$ Hệ số tin cậy hiện tại (hệ số tin cậy phân nửa trắc nghiệm, chưa điều chỉnh)

Phương pháp này thường thích hợp cho các thang đo định danh (các phép đo có item gồm 2 lựa chọn: đúng, sai). Không thể áp dụng phương pháp này đánh giá độ tin cậy ở những trắc nghiệm có các item không tương đương (chẳng hạn các item có độ khó khác nhau).

Nhóm phương pháp thứ tư: Thường được dùng để đánh giá độ tin cậy của trắc nghiệm là các phương pháp đánh giá độ phù hợp của từng item (internal consistency methods) sử dụng mô hình hệ số tương quan alpha (Cronbach's Coefficient Alpha). Mô hình này đánh giá độ tin cậy của phép đo dựa trên sự tính toán phương sai của từng item trong toàn phép đo và tính tương quan điểm của từng item với điểm của tổng các item còn lại của phép đo. Phương pháp này thích hợp cho việc xác định độ tin cậy của các loại trắc nghiệm có các item được thiết kế theo kiểu nhiều mức độ tính theo điểm số (kiểu thang định khoảng hay thang định tỷ lệ).

Công thức tính như sau:

$$R_{xx} = \alpha = \frac{K}{K - 1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right)$$

$R_{xx} =$ Hệ số tương quan α

K = Số các item của trắc nghiệm

σ_i^2 = Giá trị phương sai của từng item cụ thể của trắc nghiệm

σ_x^2 = Giá trị phương sai của toàn bộ trắc nghiệm

$\Sigma\sigma_i^2$ = Tổng các giá trị phương sai của tất cả các item của trắc nghiệm

3. Những yếu tố ảnh hưởng đến độ tin cậy của trắc nghiệm

Ngoài lỗi của sự đo lường còn có một loạt những yếu tố khác làm ảnh hưởng đến độ tin cậy của trắc nghiệm. Có thể phân loại các yếu tố ảnh hưởng này thành các phạm trù sau:

- Các yếu tố gây ra kết quả điểm trắc nghiệm không ổn định qua các lần đo: Điểm số trắc nghiệm không ổn định có thể là do người làm trắc nghiệm có sự thay đổi thực sự về kiến thức, hiểu biết, kỹ năng giữa hai lần đo, do vậy kết quả điểm trắc nghiệm giữa hai lần đo khác nhau. Sự không ổn định cũng có thể là do những thay đổi ở phía người làm trắc nghiệm như mệt mỏi, bị stress, xúc cảm không ổn định, thiếu tập trung, hiểu nhầm câu hỏi, nhớ câu hỏi ở lần đo trước... hoặc câu hỏi trắc nghiệm quá khó, không thể trả lời bằng suy nghĩ sẵn có mà phải đoán, do vậy dẫn đến sai lệch điểm giữa các lần đo. Những điều kiện khác nhau giữa các lần trắc nghiệm như tiếng ồn xung quanh, sự không nhất quán về lời hướng dẫn, sự không thống nhất về thời gian, các thủ tục cho điểm khác nhau... nhất là với các trắc nghiệm có câu hỏi mở, trắc nghiệm phóng chiếu.

- *Các yếu tố liên quan đến item, câu hỏi của trắc nghiệm:* Các trắc nghiệm chỉ là một tập hợp có giới hạn các item nhằm đo một mẫu hành vi đại diện cho đặc tính được nghiên cứu mà không đo tất cả các biểu hiện của đặc tính đó. Do vậy, tính “đầy đủ” của mẫu hành vi được dùng để nảy sinh, phát triển các item của trắc nghiệm sẽ ảnh hưởng đáng kể đến độ tin cậy của trắc nghiệm. Một trắc nghiệm càng có nhiều item thì độ tin cậy của trắc nghiệm càng tăng, vì càng nhiều item, trắc nghiệm sẽ có nhiều cơ hội để cung cấp một mẫu rộng hơn những hiểu biết về đặc tính cần đo. Tuy nhiên trắc nghiệm không thể quá dài, điều này sẽ đe dọa đến độ tin cậy vì sự mệt mỏi, thiếu tập trung của người làm trắc nghiệm. Mặt khác các item của trắc nghiệm phải có tính đại diện, đủ “bao quát hết” miền đo. Những trắc nghiệm đáng tin cậy nhất là những trắc nghiệm gồm một mẫu các item đại diện cho miền đo đó, tức là phải gồm những item “được đánh giá là tốt nhất” (đáp ứng các tiêu chuẩn phân tích thống kê) rút ra từ một “bể” các item có thể có của miền đo đó.

- *Các yếu tố liên quan đến thống kê:* Khi phạm vi điểm trắc nghiệm càng bị thu hẹp thì tương quan càng giảm, vì vậy độ tin cậy của trắc nghiệm sẽ giảm. Ngược lại phạm vi điểm trắc nghiệm càng được mở rộng thì tương quan càng tăng, do đó độ tin cậy của trắc nghiệm sẽ tăng. Như vậy, muốn tăng độ tin cậy của trắc nghiệm người ta thường mở rộng phạm vi

điểm của trắc nghiệm bằng cách thu thập các điểm trắc nghiệm từ một mẫu gồm những người có năng lực khác nhau, rồi gộp lại để phân tích. Tuy nhiên điều này lại gây ra sự “thối phồng” độ tin cậy. Chẳng hạn, một trắc nghiệm năng lực được thiết kế dùng cho các lứa tuổi: 5, 10 và 15, thay cho việc tính độ tin cậy giữa hai lần đo cho từng nhóm tuổi riêng rẽ, ta gộp cả ba nhóm tuổi vào cùng một mẫu để phân tích độ tin cậy. Vì năng lực phát triển theo độ tuổi dẫn đến phạm vi điểm (range) theo đó được mở rộng, điều này dẫn đến phương sai của điểm thực tăng lên, và nếu phương sai của điểm sai số đo lường giữ nguyên, thì hệ số tin cậy sẽ tăng theo. Do đó tính hệ số tin cậy từ mẫu phân tích tổng hợp này thực chất là sự đánh giá thối phồng độ tin cậy của trắc nghiệm khi nó được sử dụng để so sánh trẻ em ở từng nhóm tuổi.

Tóm lại, khi thiết kế một trắc nghiệm đo lường các đặc tính tâm lý - giáo dục, người nghiên cứu cần am hiểu mô hình lý thuyết về độ tin cậy. Biết rõ điểm mạnh, điểm yếu của từng phương pháp phân tích độ tin cậy, các giả thiết đi kèm khi phân tích số liệu, để sử dụng hợp lý các phương pháp này nhằm đánh giá xác thực độ tin cậy của trắc nghiệm, hạn chế tối đa các yếu tố làm giảm thiểu hoặc tăng giả tạo độ tin cậy.

Chương VII

ĐỘ HIỆU LỰC VÀ CÁC PHƯƠNG PHÁP ĐÁNH GIÁ ĐỘ HIỆU LỰC

1. Các lý thuyết làm cơ sở cho sự phân tích độ hiệu lực

Trong khoa học hành vi, một công cụ đo lường tốt đòi hỏi không chỉ có độ tin cậy cao mà còn đòi hỏi phải có độ hiệu lực hay độ ứng nghiệm (validity). Một phép đo có độ tin cậy tốt là điều kiện cần chứ chưa phải là điều kiện đủ bảo đảm cho phép đo đó có độ hiệu lực tốt.

Về lý thuyết, một trắc nghiệm có độ tin cậy tốt có thể không có độ hiệu lực. Một trắc nghiệm như vậy có thể cung cấp một phép đo kiên định về sự khác biệt giữa những người làm trắc nghiệm, nhưng có thể đo cái gì đó không phù hợp với mục đích của trắc nghiệm. Ngược lại, tất cả các trắc nghiệm có tính hiệu lực đều có độ tin cậy.

Về mặt khái niệm, độ hiệu lực của trắc nghiệm được xác định như là mức độ chính xác mà trắc nghiệm đó đo đúng cái cấu trúc nó được thiết kế để đo. Phân tích độ tin cậy nhằm chỉ ra năng lực của trắc nghiệm sản sinh ra những điểm kiên định nhưng không nhận diện được những đặc điểm thực nào sinh ra kết quả

kiên định. Trái lại phân tích độ hiệu lực nhằm chỉ ra điểm trắc nghiệm đo những đặc tính kiên định nào và tập trung vào các biến sản sinh ra sự khác biệt điểm thực. Về lý thuyết điểm thực có hai thành phần: *những đặc tính kiên định của cá nhân phù hợp với mục đích của trắc nghiệm và những đặc tính kiên định của cá nhân không phù hợp với mục đích của trắc nghiệm*. Nếu trắc nghiệm là một công cụ đo lý tưởng, thì chỉ có những đặc tính kiên định của cá nhân phù hợp với mục đích của trắc nghiệm được đo. Nhưng trắc nghiệm chưa bao giờ là phép đo lý tưởng, do đó:

$$T = R + I$$

T: Điểm thực của người làm trắc nghiệm

R: Điểm đo những đặc tính phù hợp với mục đích của trắc nghiệm

I: Điểm đo những đặc tính không phù hợp với mục đích của trắc nghiệm (điểm sai số đo lường có tính hệ thống).

Ảnh hưởng của những đặc tính không phù hợp được gọi là sai số đo lường có tính hệ thống (gọi tắt là sai số hệ thống), vì sự có mặt của nó là không thể tránh khỏi nên điểm trắc nghiệm của X được hiểu là:

$$X = R + I + E \quad (1)$$

E: Điểm sai số đo lường ngẫu nhiên hay ảnh hưởng của các yếu tố không thể kiểm soát được.

Độ hiệu lực cũng giống như độ tin cậy được xác định bằng phân tích kết quả điểm trắc nghiệm của một nhóm người. Phân tích độ hiệu lực chính là đánh

giá ảnh hưởng của các thành tố này đối với phương sai của điểm trắc nghiệm. Do vậy công thức (1) được viết lại là:

$$\sigma^2_X = \sigma^2_R + \sigma^2_I + \sigma^2_E$$

σ^2_X = Phương sai điểm trắc nghiệm

σ^2_R = Phương sai điểm do những đặc tính kiên định phù hợp với mục đích của trắc nghiệm (gọi là phương sai điểm phù hợp)

σ^2_I = Phương sai điểm do những đặc tính kiên định không phù hợp với mục đích của trắc nghiệm (gọi là phương sai điểm sai số hệ thống)

σ^2_E = Phương sai điểm do các yếu tố ngẫu nhiên (gọi là phương sai điểm sai số ngẫu nhiên)

Do vậy nếu một trắc nghiệm tin cậy phải có $\sigma^2_R + \sigma^2_I > \sigma^2_E$ thì một trắc nghiệm có hiệu lực phải có $\sigma^2_R > \sigma^2_I$

Như vậy độ hiệu lực được định nghĩa dưới góc độ toán học là :

$$R^2_{XY} = \frac{\sigma^2_I}{\sigma^2_X}$$

2. Các kiểu phân tích độ hiệu lực

Người ta thường lầm tưởng mỗi trắc nghiệm chỉ có một độ hiệu lực duy nhất. Ngược lại, mỗi một trắc nghiệm có nhiều “kiểu hiệu lực” khác nhau. Vì vậy độ hiệu lực nên được hiểu như là những dữ liệu điều tra một cách khoa học về ý nghĩa của điểm số của trắc nghiệm khi đo lường.

Độ hiệu lực của một phép đo thường được đánh giá như là độ hiệu lực nội dung (content validity), độ hiệu lực cấu trúc (construct validity), độ hiệu lực tiêu chuẩn (criterion validity) và độ hiệu lực dự báo (predictive validity). Trong suốt nhiều năm, nhiều chuyên gia thiết kế trắc nghiệm cho rằng những kiểu hiệu lực khác nhau này được sử dụng phục vụ những mục đích khác nhau. Chẳng hạn, những trắc nghiệm có mục tiêu đánh giá chất lượng giáo dục, người ta tập trung vào độ hiệu lực nội dung, còn những trắc nghiệm về nhân cách, người ta coi trọng độ hiệu lực dự báo... Tuy nhiên gần đây các chuyên gia về trắc nghiệm đều cho rằng các kiểu hiệu lực này không phải là những kiểu hiệu lực tách biệt theo đuổi những mục đích khác nhau, mà chúng chỉ là những cách tiếp cận đại diện cho những chiến lược khác nhau cùng chung một mục đích là hiểu ý nghĩa của điểm số trắc nghiệm và đều nhằm hiệu lực hoá những kết luận, những suy đoán, suy luận được rút ra từ điểm số của trắc nghiệm. Như vậy, thực chất của các chiến lược đánh giá độ hiệu lực là làm những suy luận có cơ sở khoa học dựa trên những điểm số của trắc nghiệm.

Theo các chuyên gia thiết kế trắc nghiệm nổi tiếng như Murphy và Davidshofer (1994) thì các chiến lược đánh giá độ hiệu lực nội dung và độ hiệu lực cấu trúc thực chất là xác định xem liệu một trắc nghiệm có đưa ra được một nội dung, một cấu trúc phù hợp để đo một đặc tính cụ thể nào đó hay không và một trắc nghiệm

được coi là có độ hiệu lực nếu nó đưa ra một cách đo phù hợp, đo đúng cái nó được thiết kế để đo. Do vậy, người ta gọi đây là độ hiệu lực của sự đo lường (validity of measurement). Ngược lại, chiến lược đánh giá độ hiệu lực tiêu chuẩn hay độ hiệu lực dự báo thực chất là xem xét mức độ hiệu lực đến đâu của những dự đoán hay những quyết định dựa trên điểm số của trắc nghiệm và một trắc nghiệm được coi là có độ hiệu lực nếu nó được sử dụng để đạt đến những quyết định đúng, chính xác. Do vậy, người ta gọi đây là độ hiệu lực cho việc làm những quyết định (validity for decisions).

Để hiểu sự khác nhau giữa hai cách tiếp cận này ta hãy xem một ví dụ cụ thể. Chẳng hạn một trắc nghiệm đánh giá kỹ năng lãnh đạo nhằm chọn ra được những người có năng lực quản lý. Có hai câu hỏi liên quan đến độ hiệu lực của trắc nghiệm này. Câu hỏi thứ nhất: liệu trắc nghiệm này có thực sự cho ta biết những điều gì đó về kỹ năng lãnh đạo của một cá nhân (đây chính là câu hỏi liên quan đến độ hiệu lực của sự đo lường vì nó liên quan đến cái gì trắc nghiệm đo lường). Câu hỏi thứ hai: liệu những người đạt điểm trắc nghiệm cao có tỏ ra là người có kỹ năng quản lý lãnh đạo trong thực tế (đây chính là câu hỏi liên quan đến độ hiệu lực làm quyết định vì nó liên quan đến việc dùng điểm trắc nghiệm để làm một quyết định về một lĩnh vực mà trắc nghiệm không trực tiếp đánh giá nhưng được xem là có liên quan... đó là nhà quản lý

giỏi). Theo các chuyên gia (Guion, 1991), một trắc nghiệm có độ hiệu lực đo lường tốt không luôn luôn đảm bảo nó sẽ có độ hiệu lực tốt cho việc làm những quyết định chính xác, do vậy nhà nghiên cứu bắt buộc phải đồng thời cùng xem xét tất cả các mặt hiệu lực khi đánh giá độ hiệu lực của trắc nghiệm.

3. Các phương pháp cơ bản đánh giá độ hiệu lực

Mặc dù độ hiệu lực nội dung và độ hiệu lực cấu trúc đều nói lên mức độ hiệu lực của sự đo lường, nhưng việc đánh giá một trắc nghiệm có độ hiệu lực nội dung tốt hay không người ta lại xem xét nội dung của trắc nghiệm, điều tra quá trình thiết kế item và quá trình hiệu lực hoá các item của trắc nghiệm. Trong khi đó, để đánh giá độ hiệu lực cấu trúc người ta điều tra mối quan hệ giữa các cấu trúc thành phần của trắc nghiệm với cấu trúc của cái nó được thiết kế để đo và xem xét mối quan hệ giữa điểm của trắc nghiệm với điểm của các phép đo khác.

Như vậy, độ hiệu lực nội dung của một trắc nghiệm chính là tính hiệu lực và tính đại diện của các item (liệu mỗi item có nội dung phù hợp với nội dung cụ thể cần đo của một chỉ số cụ thể, thuộc một miền đo cụ thể hay từng item có liên quan trực tiếp với một thành phần cụ thể của miền đo, liệu các item có đại diện cho miền đo rộng lớn hơn...). Còn độ hiệu lực cấu trúc của một trắc nghiệm liên quan tới mức độ qua đó trắc nghiệm được đánh giá là một phép đo về mặt cấu trúc (phép đo phải

có cấu trúc trùng với cấu trúc của cái định đo). Trong khi đó độ hiệu lực tiêu chuẩn được kiểm định thông qua mức độ tương quan của phép đo với một phép đo chuẩn khác có cùng cấu trúc. Còn độ hiệu lực dự báo nói lên những kết quả dự báo qua điểm số của trắc nghiệm có đúng với thực tế quan sát được hay không.

Một phương pháp khác nữa thường được dùng để đánh giá độ hiệu lực là phương pháp phân tích yếu tố. Một trắc nghiệm có những miến đo nhất định. Những item được thiết kế cho một miến đo nào đó đòi hỏi phải có tính đồng nhất, do vậy có thể dùng phương pháp phân tích yếu tố để đánh giá tính đồng nhất của item (các item có cùng miến đo được mong đợi có quan hệ đáng kể với cùng một yếu tố) và đánh giá cấu trúc của phép đo. Thang đo có độ tin cậy tốt đòi hỏi các item phải có tính đồng nhất, tức là có độ chứa tương quan yếu tố (factor loading) đủ lớn (theo các chuyên gia độ chứa này phải lớn hơn 0,30 mới được coi là phù hợp).

4. Những yếu tố ảnh hưởng đến độ hiệu lực

Tất cả những phép phân tích độ hiệu lực đều nhằm cùng trả lời một câu hỏi rất cơ bản: *liệu trắc nghiệm có đo những đặc tính phù hợp với mục đích khi thiết kế (đo sự hiểu biết, những kỹ năng, hay những phẩm chất nhân cách...)?*

Như vậy có những yếu tố nào chi phối độ hiệu lực?

- Trước hết là độ tin cậy của trắc nghiệm ảnh hưởng đến độ hiệu lực, độ tin cậy thấp sẽ hạn chế độ

hiệu lực. Bất cứ yếu tố nào ảnh hưởng đến độ tin cậy đều ảnh hưởng trực tiếp hay gián tiếp đến độ hiệu lực.

- Độ khó của item, độ phân biệt của item gián tiếp ảnh hưởng đến độ hiệu lực (những item quá khó hoặc có độ phân biệt yếu không giúp gì cho độ hiệu lực của trắc nghiệm). Tính đồng nhất của item ảnh hưởng trực tiếp đến độ hiệu lực.

- Những đặc tính của mẫu (quy mô mẫu, tính đại diện của mẫu...) ảnh hưởng đến độ hiệu lực.

- Độ tin cậy và độ hiệu lực của các phép đo chuẩn (dùng đánh giá độ hiệu lực tiêu chuẩn của trắc nghiệm).

Tóm lại khi thiết kế một trắc nghiệm đo lường các đặc tính tâm lý - giáo dục, người nghiên cứu cần am hiểu các lý thuyết làm cơ sở phân tích độ hiệu lực; hiểu được các kiểu hiệu lực, các chiến lược phân tích đánh giá độ hiệu lực; biết rõ điểm mạnh, điểm yếu của từng chiến lược phân tích, các giả thiết đi kèm khi phân tích số liệu; nên sử dụng phối hợp các chiến lược phân tích (ít nhất là hai kiểu phân tích) khi đánh giá độ hiệu lực của công cụ đo.

Chương VIII

THÍCH NGHI VÀ CHUẨN HOÁ CÔNG CỤ

1. Chuẩn hoá công cụ

1.1. Khái niệm lý thuyết

Sự chuẩn hoá bất kỳ một công cụ đánh giá nào, dù đó là những trắc nghiệm tâm lý hay giáo dục, đòi hỏi không chỉ là sự chuẩn hoá các thủ tục hướng dẫn cách làm, cách cho điểm mà còn đòi hỏi sử dụng công cụ đó đo lường trên một mẫu đủ lớn được chọn lựa như là mẫu đại diện cho tổng thể các đối tượng (mà trắc nghiệm có mục tiêu phải đo) để xây dựng bảng điểm chuẩn. Mẫu dùng để chuẩn hoá thường là một nhóm mẫu (norm groups) đại diện cho tổng thể các đối tượng ở một cấp độ nào đó: quốc gia, khu vực hay địa phương.

Mục đích chủ yếu của các thủ tục chuẩn hoá công cụ là xác định phân bố điểm thô trong các nhóm mẫu chuẩn hoá. Sau đó điểm thô của các nhóm mẫu được chuyển sang dạng điểm norm (điểm phân bố của mẫu). Những điểm norm này thường gồm giới tính, tuổi tương đương, lớp tương đương, phần trăm thứ hạng và điểm chuẩn.

Hầu hết các trắc nghiệm được chuẩn hoá đều có bản hướng dẫn có chứa các bảng điểm norm trong đó lập danh sách các điểm thô và những điểm chuyển đổi tương ứng. Việc chuyển điểm thô sang một kiểu điểm khác, thường có lợi cho việc so sánh, định chuẩn, định thứ hạng phục vụ cho các mục tiêu chọn lựa, phân loại đối tượng nhằm đáp ứng một mục đích nào đó của trắc nghiệm.

Từ điểm thô do cá nhân đạt được khi làm trắc nghiệm, ta có thể tìm một bảng norm phù hợp với giới tính, tuổi, lớp... trong mẫu chuẩn hoá để đối chiếu tìm điểm chuyển đổi tương ứng qua đó suy luận được vị trí của cá nhân vừa đánh giá.

Norm không phải là những chuẩn mong đợi về kết quả làm trắc nghiệm mà đơn giản chỉ là những "mẫu dùng để chuẩn hoá điểm" chứa đựng những thông tin liên quan đến việc phân tích, giải thích điểm thô. Norm giúp ta chỉ ra vị trí tương đối của một người làm trắc nghiệm so với phân bố điểm đạt được ở những người khác có cùng độ tuổi, cùng lớp, cùng giới tính hoặc có cùng các đặc trưng dân số học khác.

Kích cỡ của norm, tính đại diện của norm thường được xác định theo những tiêu chuẩn khác nhau. Norm của những trắc nghiệm làm với nhóm, có thể dựa trên một mẫu cực lớn : 100.000 người. Trái lại, norm của những trắc nghiệm làm với cá nhân, được chuẩn hoá chỉ gồm từ 2.000-4.000 người. Sử dụng mẫu chuẩn hoá lớn không có nghĩa là tính đại diện của mẫu chuẩn hoá được

đảm bảo.

1.2. Chọn lựa mẫu chuẩn hoá

Những trắc nghiệm được chuẩn hoá có các bảng điểm norm phục vụ có hiệu quả cho việc giải thích, suy luận điểm thô của một cá nhân hay một nhóm những người làm trắc nghiệm. Tuy nhiên, việc sử dụng điểm norm để giải thích điểm thô của một cá nhân đòi hỏi phải để ý đến những đặc tính, bản chất của từng nhóm norm cụ thể (như tuổi, giới tính, dân tộc, học vấn, địa vị kinh tế, vùng địa lý...). Kết quả suy luận giải thích chỉ phù hợp khi người làm trắc nghiệm có những đặc tính, bản chất phù hợp với những đặc tính, bản chất của norm.

Khi sử dụng norm để so sánh giải thích kết quả làm trắc nghiệm của một cá nhân, ta cần phải để ý norm được thành lập khi nào. Norm của trắc nghiệm có thể bị lạc hậu do những thay đổi nhanh chóng của xã hội, của giáo dục. Chẳng hạn, những thay đổi về chương trình có thể làm cho norm của những trắc nghiệm đánh giá thành tích, năng lực trở nên lạc hậu, không thích hợp. Do vậy, những trắc nghiệm đánh giá kiến thức, kỹ năng, thành tích thường phải sửa chữa, bổ sung về nội dung và phải chuẩn hoá lại sau khoảng thời gian 5 năm.

Cách tiến hành chọn mẫu để chuẩn hoá có thể rất khác nhau, từ các chiến lược chọn mẫu đơn giản như chọn mẫu ngẫu nhiên đơn giản đến các chiến lược chọn mẫu phức tạp hơn như chọn mẫu phân tầng, mẫu cụm nhiều giai đoạn. Các mẫu chuẩn hoá đạt được từ cách chọn mẫu phân tầng, mẫu cụm nhiều giai đoạn sẽ cung

cấp một cơ sở tốt hơn cho việc phân tích, giải thích điểm trắc nghiệm.

Tuy nhiên, để tạo ra chiến lược khả thi hơn (ít tốn kém, dễ thực hiện hơn), người ta có thể thiết kế các mẫu item. Thay cho việc đo lường toàn bộ trắc nghiệm trên một mẫu đại diện lớn, người ta thiết lập các mẫu item khác nhau và đo lường các mẫu item này trên các nhóm đối tượng khác nhau. Mỗi nhóm đối tượng chỉ trả lời một nhóm các item. Phương pháp này tỏ ra có hiệu quả vì trong một khoảng thời gian giới hạn sẽ có nhiều item hơn được làm với nhiều nhóm mẫu (do đó mẫu tổng sẽ lớn hơn). Quy trình phân tích item được tiến hành sau đó và điểm norm dựa trên các nhóm mẫu đại diện được xác định có thể phản ánh một phạm vi rộng hơn của nội dung trắc nghiệm, phương pháp chọn mẫu item đã được các chuyên gia thiết kế trắc nghiệm chứng minh là có hiệu quả. Các norm của phương pháp này cũng có những đặc tính bản chất tương tự như các norm được thiết kế dựa trên các phương pháp chọn mẫu truyền thống.

1.3. Các kiểu mẫu chuẩn hoá

1.3.1. Norm quốc gia, khu vực, địa phương

Mẫu chuẩn hoá (norm) được công bố trong các bản hướng dẫn (manual) trắc nghiệm được sử dụng để so sánh kết quả làm trắc nghiệm (điểm trắc nghiệm) của một cá nhân với các cá nhân khác trong mẫu. Tuy nhiên, nhiều khi ta quan tâm hơn tới việc xác định xem

liệu một học sinh làm trắc nghiệm tốt đến mức nào khi so sánh với các học sinh khác trong trường, trong một hệ thống trường của khu vực hơn là với tất cả các trường của quốc gia.

Khi sự quan tâm của người thiết kế giới hạn điểm trắc nghiệm ở phạm vi trường, ta có thể chuyển điểm thô của trắc nghiệm sang kiểu norm địa phương.

1.3.2. Norm lớp, norm tuổi

Norm tuổi (tuổi tương đương, hay tuổi học vấn) là điểm trung vị (median score) của trắc nghiệm trên một mẫu gồm những người làm trắc nghiệm có cùng độ tuổi. Norm lớp là điểm trung vị đạt được từ một mẫu gồm những người cùng lớp, cùng làm trắc nghiệm.

Mặc dù norm lớp, norm tuổi được sử dụng nhiều, nhưng chúng thường có những sai sót nghiêm trọng. Sai sót nghiêm trọng nhất là sự phát triển về nhận thức, tâm vận động và các đặc tính nhân cách khác không đồng nhất qua toàn bộ phạm vi các độ tuổi hoặc các lớp. Vì vậy sự khác biệt của hai tháng tuổi trên một kiểu trắc nghiệm thành tích ở lớp 4 sẽ không tương đương với 2 tháng tuổi ở các lớp cao hơn. Do vậy norm tuổi, norm lớp thường không được khuyến khích sử dụng ở các lớp học sinh lớn, mà chủ yếu chỉ được sử dụng ở cấp tiểu học (nơi được giả thiết các đơn vị phát triển tương đối hằng định qua thời gian).

Trong các trắc nghiệm đánh giá trí tuệ, người ta còn sử dụng norm tuổi tâm thần (khái niệm này do Alfred Binet

đề xuất). Tuổi tâm thần của một đứa trẻ làm trắc nghiệm được suy luận tương đương với tuổi thực của một nhóm trẻ em trong nhóm mẫu chuẩn hoá khi điểm trắc nghiệm của đứa trẻ ngang bằng điểm trung vị của nhóm trẻ trong mẫu chuẩn hoá. Norm tuổi của nghiệm thể có thể được chuyển thành thương số, chẳng hạn thương số trí tuệ được tính theo công thức sau:

$$IQ = 100 \frac{MA}{CA}$$

MA : Tuổi tâm thần

CA : Tuổi thực (tính theo tháng)

1.3.3. Norm phần trăm (percentile norm)

Đây là kiểu chuyển đổi phi tuyến tính (nonlinear transformation), thường sử dụng cho các kiểu thang đo định hạng. Điểm thô của nghiệm thể được chuyển thành thứ hạng phần trăm (percentile rank) nhằm xác định có bao nhiêu phần trăm các điểm khác đứng dưới điểm thô đó. Chẳng hạn, điểm 7 ở ví dụ dưới đây có thứ hạng là 25% (tức là ai đó đạt điểm 7 về kết quả thì kết quả này nói rằng người đó có kết quả điểm tốt hơn 25% số học sinh cùng làm trắc nghiệm).

Norm phần trăm được tính theo công thức sau :

$$PR_{mp} = \frac{cf - mp}{N} \times 100 = \frac{cf_B(X) + [f(X)]}{N} \times 100$$

PR_{mp} : Thứ hạng phần trăm (sử dụng điểm giữa của tần suất tập hợp)

cf - mp : Điểm giữa của tần suất tập hợp

$cf_B(X)$: Tần suất tập hợp dưới (bằng tổng số người có điểm dưới điểm X)

$f(X)$: Tần suất của điểm thô (bằng tổng số người có cùng điểm số X)

N : Tổng số người làm trắc nghiệm.

Ví dụ : Bạn có một nhóm 16 học sinh đạt điểm thô từ 5-10 điểm trên một trắc nghiệm đọc hiểu, bạn muốn biết thứ hạng của từng điểm sẽ là bao nhiêu phần trăm? Dựa theo công thức trên bạn tính được bảng thứ hạng phần trăm (%) sau đây:

Điểm thô	$f(X)$	$cf_B(X)$	$cf - mp$	PR_{mp}
10	1	15	15,5	96,875
9	3	12	13,5	84,375
8	6	6	9,0	56,25
7	4	2	4,0	25,0
5	2	0	1	6,25

Điểm ưu của thứ hạng phần trăm là nó rất linh hoạt, có thể được dùng với bất kỳ loại phân bố điểm nào (dù nó là phân bố chuẩn hay không chuẩn). Nhưng điểm yếu của nó là chỉ nói được một điểm có thứ hạng cao hay thấp hơn một điểm khác mà không nói rõ cao hơn, thấp hơn bao nhiêu. Do chỉ tập trung vào tần suất nên quan hệ giữa các điểm của trắc nghiệm sau khi chuyển sang dạng thứ hạng phần trăm đã có sự thay đổi. Các đơn vị xếp hạng phần trăm không đều nhau trên tất cả các phần của thang đo (chúng có khuynh hướng bó lại ở

giữa và mở rộng ra ở hai cực của thang đo), do đó gây khó khăn cho sự phân tích, giải thích những thay đổi. Chẳng hạn, điểm thô giữa 8 và 9 chỉ khác nhau 1 điểm nhưng phần trăm thứ hạng khác nhau khá xa ($PR = 84,375 - 56,25 = 28,125$) - trong khi đó giữa điểm 5 và 7 cách nhau 2 điểm nhưng phần trăm thứ hạng khác nhau chỉ là 18,75.

1.3.4. Norm điểm chuẩn

Đây là kiểu chuyển đổi tuyến tính (linear transformation), điểm chuẩn đại diện cho điểm thô trên thang đo có điểm trung bình: $\bar{X} = 0$; độ lệch chuẩn: $S = 1$ thường sử dụng cho các kiểu thang đo định khoảng. Kiểu chuyển đổi này không thay đổi bản chất thang đo, chỉ chuyển đổi đơn vị của sự đo lường (các giá trị trung bình, độ lệch chuẩn được chuyển đổi) nhưng quan hệ giữa các điểm thô, hình dạng phân bố điểm thô không thay đổi.

Từ điểm thô có thể chuyển sang các kiểu điểm chuẩn khác nhau: điểm z (z - scores); điểm CEEB (CEEB-scores); điểm IQ (IQ - scores); điểm T (T - scores) và điểm Stanine (Stanine - scores).

Điểm z là một phép đo nhằm đánh giá một điểm thô cách điểm trung bình của phân bố bao nhiêu độ lệch chuẩn. Một điểm z cũng được gọi là một đơn vị chuẩn hoặc điểm chuẩn. Điểm z được xác định theo công thức sau :

$$z - Score = \frac{X - \bar{X}}{S}$$

Phân bố của điểm thô chuyển sang điểm z không

thay đổi hình dạng, nhưng có giá trị trung bình và độ lệch chuẩn thay đổi ($\bar{X} = 0$; $S = 1$). Tuy nhiên, điểm z có bất lợi là có các giá trị thập phân và giá trị âm, do vậy có thể chuyển điểm z thành Z theo công thức :

$$Z = (z \times 10) + 50$$

Phân bố điểm Z cũng có hình dạng giống phân bố điểm thô nhưng có giá trị trung bình và độ lệch chuẩn thay đổi ($\bar{X} = 50$; $S = 10$).

Một số trắc nghiệm (chẳng hạn GRE) sử dụng $\bar{X} = 500$; $S = 100$ (CEEB - scores), một số trắc nghiệm khác sử dụng $\bar{X} = 100$; $S = 20$ (AGCT - scores). Trắc nghiệm trí tuệ của Wechsler sử dụng $\bar{X} = 10$; $S = 3$ cho các tiểu trắc nghiệm và sử dụng $\bar{X} = 100$; $S = 15$ cho toàn bộ trắc nghiệm (IQ - scores).

1.3.5. Điểm chuẩn có phân bố chuẩn (normalized standard scores)

Để làm cho điểm số từ những trắc nghiệm khác nhau có thể so sánh trực tiếp với nhau, người ta sử dụng một thủ tục chuyển các điểm thô sang dạng phân bố chuẩn, thủ tục chuyển này ảnh hưởng không chỉ đến $\bar{X}$, S mà còn thay đổi hình dạng phân bố điểm thô thành dạng phân bố chuẩn.

Để chuyển sang dạng phân bố chuẩn, trước hết tính phần trăm thứ hạng của điểm thô, rồi chuyển sang dạng tỷ lệ. Sau đó sử dụng bảng các khu vực dưới đường cong chuẩn để tìm điểm z tương ứng với từng thứ hạng. Để loại bỏ điểm thập phân và điểm âm, điểm z (z_n) được chuyển sang điểm T theo công thức sau:

$$T - \text{score} = 10z_n + 50$$

Phân bố mới của điểm T có $\bar{X} = 50$ và $S = 10$.

Dưới đây là một ví dụ về điểm của 250 sinh viên trên một trắc nghiệm đánh giá thái độ. Điểm thô của các sinh viên được phân khoảng rồi chuyển thành điểm thứ hạng phân trăm và điểm chuẩn tương ứng với các điểm giữa của phân bố tần suất.

Bảng điểm thứ hạng phân trăm và điểm chuẩn
(tương ứng với các điểm giữa của phân bố tần suất)

Khoảng điểm (1)	Điểm giữa (2)	Tần suất (3)	Tần suất tập hợp dưới điểm giữa (4)	Thứ hạng phần trăm (5)	z (6)	Z (7)	z _n (8)	T (9)
750 - 799	774,5	3	248,5	99,4 (99)	2,59	76	2,51	75
700 - 749	724,5	11	241,5	96,6 (97)	2,03	70	1,83	68
650 - 699	674,5	18	227,0	90,8 (91)	1,48	65	1,33	63
600 - 649	624,5	27	204,5	81,8 (82)	0,92	54	0,91	59
550 - 599	574,5	49	166,5	66,6 (67)	0,37	54	0,43	54
500 - 549	524,5	65	109,5	43,8 (44)	- 0,19	48	- 0,16	48
450 - 499	474,5	38	58,0	23,2 (23)	- 0,74	43	- 0,73	43
400 - 449	424,5	25	26,5	10,6 (11)	- 1,30	37	- 1,25	37
350 - 399	374,5	13	7,5	3,0 (3)	- 1,86	31	- 1,88	31
300 - 349	324,5	1	0,5	0,2 (0)	- 2,41	26	- 2,88	21

Khi điểm thô của trắc nghiệm có dạng phân bố chuẩn thì có thể sử dụng bảng phân bố chuẩn của điểm

chuẩn (z-score) để tính phần trăm thứ hạng theo cách sau:

1. Chuyển điểm thô sang điểm z sử dụng giá trị trung bình và độ lệch chuẩn của phân bố điểm thô.
2. Tìm thứ hạng phần trăm tương ứng với điểm từng điểm z.

Chú ý khi điểm thô của trắc nghiệm *không có dạng phân bố chuẩn* thì không được sử dụng bảng điểm chuẩn Z để suy luận, thay vào đó phải trực tiếp tính phần trăm thứ hạng.

2. Thích nghi hoá công cụ

Trong nhiều công trình nghiên cứu, ta có thể sử dụng các phép đo đã được chuẩn hoá của nước ngoài. Tuy nhiên, điều bắt buộc là các thang đo này phải được thích nghi hoá.

Thích nghi hoá không nên hiểu một cách đơn giản chỉ là dịch trắc nghiệm, dịch bản hướng dẫn mà đó chính là quá trình chuẩn hoá lại. Quá trình này thường trải qua những bước cơ bản sau đây:

- *Bước 1:* Hiểu rõ những đặc tính bản chất của trắc nghiệm. Hiểu rõ xuất xứ của trắc nghiệm (các khái niệm lý thuyết), những đặc tính thiết kế của trắc nghiệm (trắc nghiệm có mục đích đo cái gì, được dùng cho đối tượng nào, các thủ tục cho điểm), cấu trúc của trắc nghiệm (trắc nghiệm gồm những miền đo nào, những chỉ báo nào), nội dung mục tiêu cụ thể của từng item (từng item đo những nội dung,

biểu hiện cụ thể nào, thuộc miền đo nào)...

- *Bước 2:* Trên cơ sở hiểu rõ những đặc tính vừa nêu, nội dung của trắc nghiệm được chuyển ngữ tương đương. Tức là các item được dịch ra vẫn không bị thay đổi nội dung, không đi chệch khỏi mục tiêu cụ thể (đo một biểu hiện nào đó mà có nhiệm vụ phải đo), không chệch khỏi cấu trúc của trắc nghiệm. Các item phải bảo đảm dễ hiểu, đơn nghĩa. Trong quá trình thích nghi hoá có một số item có nội dung, hình thức không thích hợp phải được sửa chữa, thay thế.
- *Bước 3:* Xác định mẫu chuẩn hoá mới (new noms). Các nhóm mẫu chuẩn hoá được chọn phải đáp ứng mục đích của trắc nghiệm.
- *Bước 4:* Đánh giá lại độ tin cậy và độ hiệu lực trên các nhóm mẫu mới. Trên cơ sở các nhóm mẫu chuẩn hoá xây dựng lại các điểm chuẩn (norms). Bước này cũng bao hàm những gợi ý, những cảnh báo về việc nên sử dụng trắc nghiệm như thế nào trong tương lai.

Như vậy thích nghi hoá một trắc nghiệm chính là quá trình “thiết kế lại”, “chuẩn hoá lại” một trắc nghiệm làm cho nó phù hợp với những mẫu chuẩn hoá mới (mẫu này gồm các đối tượng có sự khác biệt về chủng tộc, ngôn ngữ, văn hoá, tuổi...). Trên cơ sở “trung thành” với những đặc tính bản chất vốn có (đặc tính thiết kế) của trắc nghiệm.

Chương IX

THỰC HÀNH CÁC KỸ NĂNG THIẾT KẾ CÔNG CỤ ĐO

Dưới đây là một ví dụ *thực hành kỹ năng thiết kế công cụ đo* nhằm đáp ứng mục tiêu đo lường của đề tài: “Thực trạng đạo đức, tư tưởng chính trị và lối sống của học sinh, sinh viên...”. Đề tài này sử dụng điều tra như là phương pháp nghiên cứu cơ bản nhằm phát hiện thực trạng.

1. Mục tiêu, nội dung, đối tượng điều tra và phương pháp chọn mẫu

1.1. Mục tiêu điều tra

Phát hiện, mô tả thực trạng (chất lượng) đạo đức, tư tưởng - chính trị và lối sống của thanh niên học sinh, sinh viên: *thực trạng chất lượng đạo đức, tư tưởng -*

* Nhóm chuyên gia thiết kế bộ công cụ này gồm: PGS, TS. Lê Đức Phúc; PGS. Trần Trọng Thủy; PGS, TS. Hà Nhật Thăng; PGS, TS. Vũ Trọng Rỹ; TS. Lưu Thị Thu Thủy; TS. Nguyễn Công Khanh.

chính trị và lối sống của học sinh, sinh viên hiện nay biểu hiện như thế nào? có những thiếu hụt gì?. Từ thực trạng dự báo xu hướng, làm cơ sở cho những giải pháp giáo dục đạo đức, tư tưởng chính trị, lối sống của học sinh, sinh viên trong chiến lược phát triển toàn diện con người Việt Nam trong thời kỳ công nghiệp hoá, hiện đại hoá.

1.2. Nội dung điều tra

Đánh giá ba mặt: *đạo đức, chính trị tư tưởng và lối sống* của học sinh, sinh viên. Trên từng mặt đánh giá ở ba bình diện: *nhận thức, thái độ và hành vi*. Sau đó đánh giá tổng hợp cả ba mặt đạo đức, tư tưởng - chính trị và lối sống (đồng thời đánh giá tổng hợp cả ba bình diện nhận thức, thái độ và hành vi) ở từng khách thể điều tra.

1.3. Đối tượng điều tra

Là học sinh trung học phổ thông (lớp 12), sinh viên, cán bộ, giáo viên:

- Về phía học sinh, sinh viên: Có 2.271 học sinh, sinh viên tham gia cuộc khảo sát này, trong đó có 686 học sinh lớp 12 của 13 trường trung học phổ thông và 1.585 sinh viên năm thứ hai và thứ tư/thứ năm của 13 trường đại học thuộc ba vùng Hà Nội, Thành phố Hồ Chí Minh và Đà Nẵng.

- Về phía cán bộ, giáo viên: Có 176 người tham gia cuộc khảo sát này, gồm các thầy cô giáo, cán bộ đoàn, cán bộ hội sinh viên, ban giám hiệu, phòng đào tạo, đảng uỷ, cán bộ quản lý ký túc xá ở các trường trong

mẫu điều tra.

1.4. Phương pháp chọn mẫu

Để tài thực hiện phương pháp chọn mẫu cụm, nhiều giai đoạn kết hợp với mẫu ngẫu nhiên (các phương pháp này đã được cải tiến cho phù hợp với điều kiện của Việt Nam.

Cụ thể: mẫu điều tra sinh viên chia theo ba vùng: Hà Nội, Thành phố Hồ Chí Minh, Đà Nẵng (đây là ba trung tâm lớn tập trung phần lớn sinh viên các trường đại học). Mỗi vùng, mẫu sinh viên được phân thành 5 tầng theo ngành học: sư phạm, văn hoá, công nghệ thông tin, y - dược, kinh tế - quản trị kinh doanh. Các trường được chọn ngẫu nhiên, mỗi trường chọn ngẫu nhiên 2 khoa, mỗi khoa chọn ngẫu nhiên 2 lớp năm thứ hai và năm thứ tư hoặc thứ năm (nếu ngành đó học 5 năm). Mỗi lớp chọn ngẫu nhiên 30 sinh viên.

Đối với mẫu học sinh, mỗi khu vực chọn ngẫu nhiên 4 trường, gồm 2 trường nội thành, 2 trường ở khu vực nông thôn. Mỗi trường chọn ngẫu nhiên 2 lớp, mỗi lớp chọn ngẫu nhiên 25 học sinh (có sự cân bằng nam nữ).

2. Các khái niệm công cụ, các cấu trúc và các chỉ số

2.1. Đạo đức

2.1.1. Khái niệm đạo đức

Đạo đức là những phẩm chất tốt đẹp của con người do tu dưỡng theo những tiêu chuẩn, nguyên tắc đạo đức mà có.

Những tiêu chuẩn, nguyên tắc đạo đức được dư luận xã hội thừa nhận, quy định hành vi, quan hệ của con người đối với nhau, đối với công việc, môi trường tự nhiên và bản thân.

2.1.2. Các cấu trúc (nội dung) cần đánh giá

Nhận thức đạo đức: nhận biết, nhận dạng cái thiện, hiểu và lý giải đúng các quan niệm, các quy ước, các chuẩn mực, các giá trị đạo đức.

Thái độ đạo đức (mang tính xúc cảm): thể hiện niềm tin, sự mong muốn, bản lĩnh trước các sự kiện, các hiện tượng của xã hội... hành vi của người khác, của bản thân theo các chuẩn mực đạo đức xã hội.

Hành vi đạo đức: biểu hiện qua những hành động phù hợp với các giá trị, quy ước và chuẩn mực đạo đức.

2.1.3. Các chỉ số cơ bản cần đánh giá

- (1) Kính trọng, biết ơn;
- (2) Tôn trọng nội quy, pháp luật;
- (3) Trung thực, thẳng thắn, trọng lễ phải;
- (4) Có tinh thần trách nhiệm;
- (5) Nhân ái, tình nghĩa;
- (6) Giữ chữ tín;
- (7) Yêu lao động;
- (8) Bảo vệ môi sinh.

2.2. Tư tưởng - chính trị

2.2.1. Khái niệm tư tưởng - chính trị

Quan điểm và ý nghĩ chung của con người đối với

hiện thực khách quan của xã hội.

Quan điểm và thái độ thể hiện ý thức tư tưởng - chính trị liên quan đến chủ trương, đường lối của Đảng, chính sách, pháp luật của Nhà nước.

Sự tham gia của cá nhân vào các hoạt động chính trị - xã hội và thực hiện nhiệm vụ chính trị của mình.

2.2.2. Các cấu trúc (nội dung) cần đánh giá

Sự hiểu biết về thể chế, đường lối, chủ trương, chính sách hiện hành trong thời kỳ công nghiệp hoá, hiện đại hoá.

Sự quan tâm tới các sự kiện chính trị, thời sự ở trong và ngoài nước;

Thái độ đối với đường lối, chủ trương, chính sách của Đảng và Nhà nước.

Thực hiện nhiệm vụ chính trị của học sinh, sinh viên là học tập và tu dưỡng bản thân.

Sự tham gia của cá nhân vào hoạt động chính trị - xã hội với tư cách là học sinh, sinh viên.

2.2.3. Các chỉ số cơ bản

- (1) Lý tưởng phấn đấu (sống có lý tưởng);*
- (2) Tính tích cực xã hội;*
- (3) Quan tâm tới chính trị xã hội;*
- (4) Yêu nước, tự hào dân tộc.*

2.3. Lối sống

2.3.1. Khái niệm lối sống

Những cách thức suy nghĩ, kỹ năng ứng xử (cách

nghe, nếp sinh hoạt, thói quen, phong cách học tập, làm việc, xử sự...) tạo nên cái riêng của mỗi cá nhân hay một nhóm người nào đó.

2.3.2. Các cấu trúc (nội dung) cần đánh giá

Toàn bộ các hình thức biểu hiện lối sống của con người, được chia thành hai loại:

a) Lối sống đúng đắn, lành mạnh, có ích, được chấp nhận;

b) Lối sống không phù hợp với những yêu cầu, chuẩn mực khách quan, gây hậu quả xấu, không được xã hội chấp nhận.

Sự tự đánh giá của chủ thể về những biểu hiện trong lối sống của bản thân (nhận thức, thái độ, hành vi).

Chủ thể đánh giá các biểu hiện lối sống (lành mạnh hay tiêu cực) của các thành viên khác trong lớp, trong trường.

2.3.3. Các chỉ số cơ bản:

(1) Tích cực, chủ động, linh hoạt;

(2) Giản dị, tiết kiệm, lành mạnh;

(3) Tính có kế hoạch, có mục đích;

(4) Sự đồng cảm, chia sẻ;

(5) Khả năng thích ứng;

(6) Tinh thần kỷ luật;

(7) Tinh thần hợp tác;

(8) Tín ngưỡng;

(9) Tình yêu, hôn nhân.

3. Các nguyên tắc thiết kế công cụ

Công cụ điều tra quyết định chất lượng cuộc điều tra. Một bộ công cụ điều tra tốt phải bảo đảm có các đặc tính thiết kế tốt (đo cái gì, dùng cho đối tượng nào...) và có các đặc tính đo lường (độ tin cậy và độ hiệu lực) tốt. Dưới đây là những nguyên tắc chỉ đạo khi thiết kế công cụ điều tra:

1. Công cụ được thiết kế trên cơ sở xác định rõ mô hình lý thuyết về cái định đo. Tức là phải xác định rõ bản chất của các khái niệm chỉ đối tượng nghiên cứu sẽ được đo lường (đó là các khái niệm: đạo đức, tư tưởng-chính trị và lối sống ở học sinh, sinh viên).

2. Cấu trúc của phép đo phải phù hợp với mô hình lý thuyết định hướng về cái định đo. Tức là mô hình lý thuyết định hướng về cái định đo gồm ba mặt: đạo đức, tư tưởng- chính trị và lối sống, mỗi mặt trên ba bình diện: nhận thức, thái độ hành vi, vậy cấu trúc của phép đo (gồm các thang đo, các tiểu thang đo...) phải phản ánh được cấu trúc lý thuyết trên.

3. Việc thiết kế phép đo phải tuân thủ theo các bước của quy trình thiết kế. Ở từng bước phải luôn đối chiếu với mục tiêu nghiên cứu, nếu không phù hợp phải quay lại những bước trước đó.

Quá trình này được triển khai qua các bước hay công đoạn như: xác định khái niệm, định ra được cấu trúc phép đo, các miền đo, các thang bậc, các chỉ số, các biểu hiệu cụ thể, mức độ đánh giá, viết các item, hiệu lực hoá các item (xem quy trình thiết kế).

4. Đảm bảo sự phù hợp với đặc điểm của nhóm khách thể và hoàn cảnh, điều kiện thực tế. Đồng thời hiện thực hoá những yêu cầu xã hội và giáo dục (qua các tiêu chuẩn thể hiện trong nội dung đo) trong thời kỳ công nghiệp hoá, hiện đại hoá

5. Bộ công cụ phải được đo trên mẫu thử để phát hiện các lỗi thiết kế (phát hiện các câu hỏi không phù hợp, ngôn ngữ khó hiểu, tối nghĩa, không muốn trả lời...) và để kiểm tra những đặc tính đo lường (độ tin cậy, độ hiệu lực, độ phù hợp của item...). Những sai sót được chỉnh sửa trước khi đo trên mẫu lớn.

6. Phải thống nhất cách hướng dẫn điều tra (tất cả điều tra viên được tập huấn thống nhất cách điều tra, quy định các điều kiện đáp ứng tiêu chuẩn chọn mẫu, dự kiến xử lý các tình huống có thể có).

4. Quy trình thiết kế công cụ

Bộ công cụ dùng cho cuộc điều tra này được thiết kế theo mô hình *đa diện, đa tầng* (xem mô hình đa diện của phép đo), có độ tin cậy và độ hiệu lực bảo đảm, có cấu trúc phù hợp với những nội dung điều tra, đáp ứng mục tiêu điều tra nhằm định lượng một cách khách quan ba mặt: *đạo đức, chính trị - tư tưởng và lối sống* của học sinh, sinh viên (đánh giá ở cấp độ cá nhân). Mỗi mặt đánh giá ở ba bình diện: *nhận thức, thái độ và hành vi*.

Thực tế ở mỗi người, đạo đức, tư tưởng chính trị, lối sống là những phạm trù có sự giao thoa khó tách rời, do

đó, nhận thức, thái độ, hành vi có sự liên quan chặt chẽ với nhau. Như thế trong mô hình cấu trúc của phép đo có sự đánh giá tổng hợp (gộp) cả ba mặt đạo đức, tư tưởng-chính trị và lối sống cũng như đánh giá gộp cả ba bình diện nhận thức, thái độ và hành vi ở từng khách thể điều tra để phân loại học sinh, sinh viên theo những tiêu chuẩn khách quan.

Để thiết kế bộ công cụ này, nhóm chuyên gia thiết kế công cụ đã triển khai một quy trình lý thuyết xây dựng trắc nghiệm (xem sơ đồ quy trình thiết kế) đi qua các bước cơ bản sau đây:

- Xây dựng mô hình lý thuyết định hướng về những mặt cần đo (đạo đức, tư tưởng-chính trị và lối sống).

- Xác định rõ các khái niệm theo mô hình lý thuyết

- Xác định cấu trúc của khái niệm, cấu trúc của phép đo, xác định các miền đo cụ thể.

- Thao tác hoá các khái niệm thành các chỉ số cụ thể có thể đo đạc được.

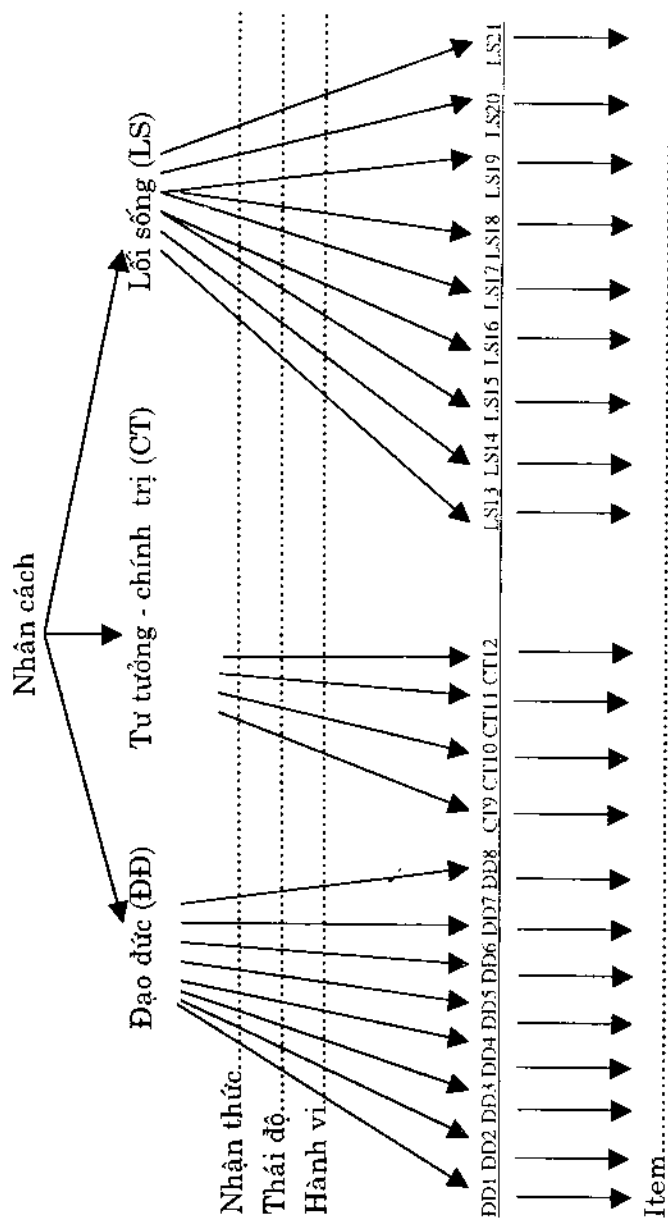
- Thống nhất nội hàm, cái cần đo ở từng chỉ số.

- Phát triển, thiết kế được nhiều nhất các item cho từng chỉ số đo đạc, chọn lựa các hình thức biểu đạt phù hợp (xây dựng “ngân hàng item” hay “bể item”).

- Hiệu lực hoá các item (chắt lọc, để mỗi item nhằm đo một biểu hiện nhất định, đo đúng cái cần đo ở từng chỉ số, có hình thức ngắn gọn đơn giản, dễ hiểu).

- Sắp xếp các item vào phiếu hỏi, viết lời hướng dẫn.

MÔ HÌNH ĐA DIỆN CỦA PHÉP ĐO



Khởi đầu, các chuyên gia trong nhóm thiết kế xác định rõ khái niệm, cấu trúc phép đo, các chỉ số đo lường cụ thể, nội dung đo lường ở từng chỉ số, đồng thời thống nhất lựa chọn các kiểu item thích hợp. Sau đó từng chuyên gia trong nhóm thiết kế độc lập đưa ra các item. Một ngân hàng gồm trên 300 item bao quát khắp các chỉ số đã được tập hợp để xem xét chọn lọc. Những item nào được cả nhóm chuyên gia đồng ý mới được giữ lại, những item có nội dung tốt nhưng hình thức chưa thích hợp được viết lại. Đối chiếu với mục tiêu đo lường, nội dung đo lường để xem xét tính đại diện, tính bao quát, số lượng item cho từng chỉ số bảo đảm những nội dung trọng tâm phải có số lượng item nhiều hơn), nếu chỉ số nào có số lượng item chưa đủ chưa bao quát hết nội dung đo lường phải bổ sung thêm item. Sau nhiều lần chỉnh sửa theo các tiêu chuẩn hiệu lực hoá item như vừa nêu, chúng tôi chỉ giữ lại 173 item tốt nhất đáp ứng các tiêu chuẩn thiết kế. Đây là những item có tính đại diện, đo đúng cái nó có nhiệm vụ phải đo, có độ phân biệt bảo đảm. Những item này được sắp xếp thành 127 câu, trong đó chỉ có 117 item được cấu trúc thành 3 thang đo dùng để đánh giá một cách có hệ thống từng cá nhân theo 21 chỉ số (tất cả các item này được điểm hoá theo một tiêu chuẩn thống nhất). Các item còn lại là những nhận định của học sinh/ sinh viên về sự biểu hiện của những hành vi đạo đức, lối sống ở tập thể lớp, nơi họ đang học tập, sinh hoạt, không nhằm đánh giá cá nhân mà chỉ dùng để cung cấp những thông tin chung.

5. Mô tả bộ công cụ điều tra

Bộ công cụ điều tra dùng cho học sinh/ sinh viên, gồm: 127 câu hỏi dài 13 trang được chia thành ba thang đo đánh giá ba mặt: *đạo đức, chính trị - tư tưởng và lối sống* của học sinh/sinh viên (xem phụ lục 2). Trên từng mặt đánh giá ở ba bình diện: nhận thức, thái độ và hành vi (xem mô hình đa diện của phép đo). Đánh giá tổng hợp (gộp) cả ba mặt đạo đức, tư tưởng - chính trị và lối sống hoặc đánh giá gộp cả ba bình diện nhận thức, thái độ và hành vi gọi là phép đo "Nhân cách" (phép đo tổng)*. Bộ công cụ điều tra được thiết kế rất công phu, theo đúng quy trình khoa học (xem quy trình thiết kế, mô hình đa diện của phép đo).

- Thang đo Đạo đức (ĐĐ) đánh giá nhận thức, thái độ và hành vi đạo đức, gồm 47 item đánh giá 8 chỉ số cơ bản sau đây:

1) Kính trọng, biết ơn (ĐĐ1) (tổ tiên, ông bà, cha mẹ, thầy cô giáo), gồm 6 item: 1, 2, 7, 45, 79, 121.

2) Tôn trọng luật pháp (ĐĐ2), gồm 6 item: 3, 4, 42, 44, 80, 81.

3) Trung thực, thẳng thắn, trọng lẽ phải (ĐĐ3), gồm 7 item: 5, 6, 46, 48, 82, 87, 106.

4) Tình thần trách nhiệm (ĐĐ4), gồm 6 item: 8, 43, 49, 50, 86, 88.

5) Nhân ái, tình nghĩa (ĐĐ5), gồm 5 item: 10, 51, 52, 83, 84.

* Nhân cách chỉ là tên gọi của phép đo tổng, không phải là cấu trúc nhân cách trong tâm lý học (TG).

6) Giữ chữ tín (ĐĐ6), gồm 5 item: 11, 12, 47, 53, 107.
7) Yêu lao động (ĐĐ7), gồm 7 item: 14, 54, 55, 85, 90, 108, 109.

8) Bảo vệ môi sinh (ĐĐ8), gồm 5 item: 18, 93, 96, 111, 126.

- Thang đo tư tưởng - chính trị (CT) đánh giá nhận thức, thái độ, niềm tin và hành vi ở góc độ tư tưởng - chính trị gồm 23 item đánh giá 4 chỉ số cơ bản sau đây:

9) Lý tưởng phấn đấu (CT9), gồm 6 item: 19, 20, 58, 59, 97, 98.

10) Tính tích cực xã hội (CT10), gồm 6 item: 21, 22, 60, 126a, 126c, 126d.

11) Quan tâm đến chính trị và thời sự (CT11): quan tâm và có thái độ đúng đắn đối với tình hình chính trị - xã hội, thời sự ở trong và ngoài nước, gồm 6 item: 23, 24, 61, 123a, 124a.

12) Yêu nước, tự hào về truyền thống dân tộc (CT12), gồm 5 item: 25, 26, 63, 64, 89.

- Thang đo Lối sống (LS) đánh giá lối sống (nhận thức, thái độ, hành vi): gồm 47 item đánh giá 9 chỉ số cơ bản sau đây:

13) Tích cực chủ động linh hoạt (LS13), gồm 6 item: 27, 29, 65, 66, 99, 100.

14) Giản dị, tiết kiệm, lành mạnh (LS14), gồm 5 item: 28, 30, 67, 68, 95.

15) Tính có kế hoạch, có mục đích (LS15), gồm 6 item: 17, 31, 69, 70, 94, 101.

16) Đồng cảm, biết chia sẻ (LS16), gồm 5 item: 33,

34, 71, 72, 102.

17) Khả năng thích ứng (LS17), gồm 5 item: 35, 36, 73, 74, 103

18) Tinh thần kỷ luật (LS18), gồm 5 item: 37, 38, 75, 76, 105

19) Tinh thần hợp tác (LS19), gồm 7 item: 15, 16, 56, 57, 91, 92, 110

20) Tín ngưỡng (LS20), gồm 4 item: 39, 40, 77, 78

21) Tình yêu, hôn nhân (LS21), gồm 4 item: 41, 125a, 125b, 125c.

6. Nguyên tắc và các phương pháp xử lý số liệu

Toàn bộ số liệu điều tra khảo sát được xử lý theo chương trình SPSS (SPSS 9.0 là phần mềm chuyên dụng, xử lý phân tích số liệu thống kê dành cho các khoa học xã hội). Số liệu được xử lý phân tích theo các nguyên tắc và phương pháp sau:

6.1. Các nguyên tắc

- Trung thực, khách quan tối đa khi xử lý phân tích số liệu.

- Sử dụng hợp lý các phép toán kiểm định thống kê. Thoả mãn các yêu cầu giả thiết trước khi áp dụng một phương pháp thống kê cụ thể.

- Sử dụng ít nhất hai phương pháp (cách tiếp cận) khác nhau khi phân tích đánh giá độ tin cậy, độ hiệu lực.

- Các kết luận chỉ căn cứ trên số liệu thu được sau khi xử lý phân tích một cách khách quan, có hệ thống.

- Loại bỏ các phiếu trả lời nếu có trên 10 item không trả lời. Làm sạch số liệu thô (loại bỏ các giá trị cực trị/ ngoại lai) trước khi xử lý tính, phân tích.

6.2. Các phương pháp xử lý, phân tích số liệu

Dùng thống kê mô tả và thống kê suy luận để giải thích số liệu, cụ thể là:

- Dùng mô hình Cronbach's alpha đánh giá độ tin cậy của từng thang đo, toàn bộ phép đo, đánh giá độ phù hợp của từng item.

- Dùng phương pháp phân tích yếu tố để đánh giá độ hiệu lực cấu trúc của phép đo, tính đồng nhất của các item ở từng thang đo, từng chỉ số.

- Dùng thống kê mô tả đánh giá tính chuẩn của phân bố điểm của hai nhóm mẫu điều tra học sinh và sinh viên.

- Dùng kiểm định t-Student, kiểm định F (ANOVA) để so sánh điểm trung bình của các nhóm mẫu.

- Dùng phương pháp hồi quy hợp biến để đánh giá tương quan, dự báo xu hướng, phát hiện các nhân tố chi phối nhiều nhất thực trạng đạo đức tư tưởng - chính trị, lối sống của học sinh/ sinh viên.

7. Điều tra thử, kiểm tra các đặc tính thiết kế và các đặc tính đo lường của bộ công cụ

Để bảo đảm tính khoa học, tính chuẩn xác của công cụ đo đạc, bộ công cụ khảo sát học sinh/sinh viên sau khi được các chuyên gia thiết kế hoàn chỉnh đã được đo thử trên nhóm mẫu gồm 215 học sinh và sinh viên

của 8 lớp, gồm hai trường trung học phổ thông (trung học phổ thông Nguyễn Trãi, trung học phổ thông Trần Hưng đạo, Hà nội) và hai trường đại học (Đại học Sư phạm Hà Nội I và Đại học Bách khoa). Tất cả các điều tra viên được tập huấn, quán triệt mục đích, thống nhất lời hướng dẫn, thống nhất cách tiến hành đo đạc, giám sát, kiểm tra từng phiếu. Đồng thời các điều tra viên được yêu cầu thu thập tất cả những thông tin phản hồi, phát hiện những vấn đề liên quan đến phép đo như: thời gian tối đa/thời gian tối thiểu cho một học sinh/sinh viên trả lời xong bộ câu hỏi, những item không rõ nghĩa (đa nghĩa hoặc tối nghĩa) khó trả lời, những item có thể làm cho cá nhân không muốn trả lời hoặc khó trả lời trung thực... Tất cả các thông tin này được phản ánh cho nhóm chuyên gia thiết kế xử lý nhằm hoàn chỉnh bộ công cụ ở mức *tối đa* trước khi điều tra trên diện rộng.

Về mặt lý thuyết, một bộ công cụ tốt phải có đủ độ tin cậy và độ hiệu lực, Độ tin cậy của một phép đo thể hiện ở khả năng nó đo đúng cái nó được thiết kế để đo. Những đặc tính quan trọng nhất của một phép đo tin cậy là độ ổn định (stability) và độ phù hợp (internal consistency).

Để thử độ tin cậy của bộ công cụ, chúng tôi đã dùng các phương pháp đánh giá độ phù hợp của từng item (internal consistency methods) sử dụng mô hình tương quan alpha của Cronbach (Cronbach's coefficient alpha): mô hình này đánh giá độ tin cậy của một phép đo dựa trên sự tính toán phương sai của từng item trong toàn

phép đo và tính tương quan điểm của từng item với điểm của các item còn lại của phép đo (mô hình này rất được các chuyên gia thiết kế trắc nghiệm tin nhiệm, dùng để đánh giá độ tin cậy).

Số liệu thu được qua mẫu thử học sinh/sinh viên được xử lý trên phần mềm chuyên dụng SPSS. Kết quả tính toán được trình bày trong bảng 1.

Phân tích kết quả cho thấy hệ số tin cậy Cronbach alpha của từng thang đo ĐĐ, CT, LS lần lượt là 0,81, 0,73, 0,77 và của toàn bộ phép đo (Nhân cách) là $\alpha = 0,90$. Theo lý thuyết về độ tin cậy, những hệ số α này đủ cao, đáp ứng được yêu cầu về độ tin cậy của một trắc nghiệm đánh giá các thuộc tính của nhân cách.

Bảng 1: Độ tin cậy của các phép đo trên mẫu thử
(Cronbach's coefficient alpha)

Các thang đo	Hệ số tin cậy alpha (Mẫu thử học sinh, sinh viên, N (số lượng) = 215)
Đạo đức	0,81
Tư tưởng-chính trị	0,73
Lối sống	0,77
Nhân cách (Tổng)	0,90

Để kiểm tra độ tin cậy của từng chỉ số của trắc nghiệm, chúng tôi đánh giá độ tin cậy của từng chỉ số bằng cách tính hệ số tương quan (r) giữa điểm số của từng chỉ số với điểm số của các chỉ số còn lại của toàn

phép đo. Những chỉ số nào có hệ số tương quan nhỏ hơn 0,20 được coi là những "item" có độ tin cậy thấp (không phù hợp) cần phải sửa chữa (thiết kế lại hoặc loại bỏ), những chỉ số nào có hệ số tương quan lớn hơn 0,30 được coi là có độ tin cậy bảo đảm (phù hợp).

Kết quả tính toán trình bày trong bảng 2 cho thấy đa số các item đều có r lớn hơn 0,30 (chỉ có 2 item có r lớn hơn 25), ngoại trừ item: CT 11 có tương quan yếu nhỏ hơn 20 (tương quan khá nhỏ: .1304) tức là chỉ số này có độ tin cậy thấp, cần phải điều chỉnh.

Bảng 2: Hệ số tin cậy của từng item

(Phân tích độ tin cậy của item)

Các item	Điểm trung bình của thang đo nếu item bị xoá	Phương sai của thang đo nếu item bị xoá	Hệ số tương quan của item với các item còn lại	Hệ số alpha nếu item bị xoá
ĐD1	157.6009	352.1938	.4667	.8514
ĐD2	159.9061	335.5383	.5773	.8463
ĐD3	158.8451	326.5183	.5516	.8474
ĐD4	158.9343	343.8919	.5357	.8485
ĐD5	160.1502	356.7226	.4163	.8531
ĐD6	160.6338	365.0822	.3517	.8553
ĐD7	156.8357	341.5531	.5345	.8484
ĐD8	160.6385	332.2036	.6123	.8448
CT9	158.6714	340.9481	.5812	.8468
CT10	159.4319	334.1522	.5935	.8456
CT11	158.6291	359.8476	.1304	.8720

CT12	159.1690	342.6600	.5494	.8480
LS13	158.9249	338.8245	.5568	.8474
LS14	159.2441	365.0156	.3153	.8561
LS15	158.4085	355.5729	.4294	.8527
LS16	160.7183	351.5146	.4164	.8529
LS17	161.2394	362.3245	.3248	.8558
LS18	161.0469	352.2902	.3811	.8542
LS19	157.4507	347.2770	.4321	.8523
LS20	161.5164	362.6660	.2825	.8572
LS21	161.9577	358.7482	.3325	.8557

Bảng 3: Ma trận tương quan (Correlation Matrix)
(Mẫu thử học sinh và sinh viên, N=215)

	Đ01	Đ02	Đ03	Đ04	Đ05	Đ06	Đ07	Đ08	CT9	CT10	CT11	CT12	LS13	LS14	LS15	LS16	LS17	LS18	LS19	LS20	LS21
Đ01																					
Đ02	.271																				
Đ03	.338	.532																			
Đ04	.284	.271	.340																		
Đ05	.094	.215	.302	.248																	
Đ06	.255	.137	.133	.311	.173																
Đ07	.135	.364	.297	.411	.202	.252															
Đ08	.305	.442	.356	.305	.176	.114	.267														
CT9	.377	.440	.374	.306	.092	.079	.331	.321													
CT10	.366	.430	.376	.341	.125	.271	.252	.388	.368												
CT11	.137	.019	.000	.087	-.078	.008	.097	.171	.148	.162											
CT12	.316	.379	.281	.328	.252	.174	.353	.325	.441	.339	.218										
LS13	.231	.319	.281	.462	.134	.340	.396	.280	.336	.418	.021	.322									
LS14	.128	.342	.259	.169	.109	.082	.237	.230	.246	.169	-.024	.141	.175								
LS15	.124	.289	.225	.336	.114	.230	.453	.137	.299	.212	-.051	.216	.362	.253							
LS16	.270	.203	.298	.237	.192	.304	.082	.178	.260	.318	.053	.206	.344	.121	.231						
LS17	.136	.125	.080	.220	.076	.253	.339	.081	.190	.149	.194	.170	.380	.058	.217	.185					
LS18	.195	.310	.413	.218	.099	.005	.183	.243	.373	.240	-.084	.169	.241	.255	.274	.197	.002				
LS19	.290	.248	.269	.263	.323	.280	.251	.257	.258	.303	.055	.336	.231	.088	.183	.357	.113	.153			
LS20	.188	.191	.116	.147	.082	.143	.203	.115	.158	.061	.163	.301	.207	.084	.112	.092	.260	.084	.111		
LS21	.195	.211	.217	.144	.110	.130	.256	.255	.204	.223	.051	.210	.174	.112	.208	.047	.109	.264	.101	.136	

Bảng 4: Phân tích yếu tố: chiết xuất yếu tố

	Hệ số chung	Yếu tố	Hệ số riêng	Phần trăm	Phần trăm gộp
ĐĐ1	.47548	1	5,95158	28,3%	28,3%
ĐĐ2	.56264	2	1,61219	7,7	36,0
ĐĐ3	.58244	3	1,33584	6,4	42,4
ĐĐ4	.44686	4	1,31697	6,3	48,7
ĐĐ5	.66247	5	1,04590	5,0	53,6
ĐĐ6	.54969				
ĐĐ7	.59523				
ĐĐ8	.56594				
CT9	.55513				
CT10	.64544				
CT11	.64838				
CT12	.55206				
LS13	.60635				
LS14	.34968				
LS15	.56895				
LS16	.49867				
LS17	.55434				
LS18	.50210				
LS19	.50734				
LS20	.55654				
LS21	.27675				

Bảng 5: Phân tích yếu tố: đánh giá hệ số chứa của các yếu tố

Item	Yếu tố 1	Yếu tố 2	Yếu tố 3	Yếu tố 4	Yếu tố 5
CT10	.67317				
ĐĐ 8	.67302				
ĐĐ2	.66043				
CT9	.64814				
LS13	.64442				
ĐĐ3	.64000				
ĐĐ4	.61614				
CT12	.60757				
ĐĐ7	.60597			.41664	
ĐĐ1	.52867				
LS15	.51719			.51650	
LS19	.50994				
LS16	.48961				
LS21	.39289				
LS14	.38453				
LS17	.36818	.57504			
LS18	.46552				
ĐĐ6	.42175	.43658			
CT11	.17085	.35434	.61546		
ĐĐ5	.49128				.55548
LS20	.32111	32353	.34456		.47787

Để đánh giá độ hiệu lực của bộ công cụ, chúng tôi áp dụng phương pháp phân tích yếu tố (Factor Analys): sử dụng mô hình phân tích thành tố cơ bản PC (Principal Component) để kiểm tra mức độ phù hợp của cấu trúc phép đo, nhằm đánh giá độ hiệu lực về mặt cấu trúc, nội dung và độ phù hợp của từng item. Điều kiện để sử dụng phương pháp này là: ma trận phải có một số tương quan lớn hơn 30, hệ số của phép thử KMO (Kaiser-Meyer-Olkin Measure of Sampling Adequacy) phải lớn hơn 60 và phép thử Bartlett (Bartlett Test of Sphericity) phải ở mức có ý nghĩa (đây là hai phép thử cho biết khả năng chiết xuất yếu tố của ma trận tương quan). Kết quả của phương pháp phân tích ma trận tương quan các chỉ số (xem bảng 3) cho biết, trong ma trận này có một số đáng kể các tương quan lớn hơn 0,30, hệ số của phép thử KMO = 0,88, phép thử Bartlett ở mức có ý nghĩa (p nhỏ hơn 0,0001), do vậy thích hợp cho việc dùng phương pháp phân tích yếu tố đánh giá độ hiệu lực của các item.

Bảng 4 và 5 trình bày kết quả của phép phân tích yếu tố, cho biết: các item của cả 3 phép đo: đạo đức, chính trị, tư tưởng và lối sống thuộc về 5 yếu tố nhưng chỉ có yếu tố 1 có giá trị riêng khá lớn (Eigenvalue = 5,95), giải thích cho 28,3% độ biến thiên của toàn phép đo. Các yếu tố còn lại có giá trị riêng khá nhỏ từ 1,0 đến 1,61 giải thích cho từ 5-7,7 % độ biến thiên của phép đo. Tất cả 20 chỉ báo của toàn phép đo đều có hệ số tương quan đáng kể với yếu tố 1. Hệ số chứa (factor loading) của 20 chỉ báo trên yếu tố 1 đều nhỏ hơn 0,30 và được

coi là phù hợp với cấu trúc của phép đo, ngoại trừ CT11 có hệ số chứa thấp (0,17) và được coi là không phù hợp vì nó không có tính đồng nhất cao, do vậy nó đóng góp không đáng kể cho yếu tố chính này.

Độ hiệu lực của bộ công cụ còn được kiểm tra qua việc tính hệ số tương quan của các phép đo với nhau. Đạo đức, tư tưởng - chính trị và lối sống của một cá nhân là ba mặt của một nhân cách, chúng phải có mối quan hệ khăng khít. Do vậy một phép đo, gồm ba thang đo, đánh giá ba mặt này phải cung cấp ba loạt điểm có tương quan đáng kể với nhau (vì theo lý thuyết, một phép đo có hiệu lực phải là phép đo có cùng cấu trúc với cái nó được thiết kế để đo). Kết quả tính toán hệ số tương quan điểm của các phép đo được trình bày trong bảng 6.

Bảng 6: Tương quan điểm của các phép đo trên mẫu thử học sinh và sinh viên (N = 215)

Các thang đo	Đạo đức	Tư tưởng- Chính trị	Lối sống	Nhân cách
Đạo đức	1,00			
Tư tưởng - chính trị	0,64**	1,00		
Lối sống	0,71**	0,54**	1,00	
Nhân cách (Tổng)	0,93**	0,79**	0,88**	1,00

** $P < 0,0001$

Qua phân tích tương quan điểm ở bảng 6 cho thấy cả ba thang đo: Đạo đức, Tư tưởng - chính trị và Lối sống đều có tương quan đáng kể với nhau (r lớn hơn 0,54), kết

quả này thoả mãn giả thiết đã nêu rằng phép đo có cùng cấu trúc với cái nó được thiết kế để đo. Điều này cho thấy phép đo có độ hiệu lực về nội dung và cấu trúc.

Tóm lại kết quả phân tích mẫu thử ($N = 215$) cho thấy bộ công cụ dùng để điều tra khảo sát được xem là có độ tin cậy và độ hiệu lực phù hợp. Ngoại trừ một chỉ số (CT11) có độ tin cậy thấp và có độ hiệu lực chưa phù hợp, cần phải điều chỉnh (cũng có một vài item có hình thức thể hiện và ngôn từ chưa thích hợp như item số 9, 13, 32, 121, 124b, 122e, những item này có thể làm giảm độ tin cậy của các chỉ số đo, do đó chúng cũng được sửa hoặc bị loại ra) còn tất cả các item còn lại của phép đo đều có độ tin cậy và hiệu lực bảo đảm theo các yêu cầu thiết kế, đo đúng cái cần đo.

8. Đánh giá độ tin cậy và độ hiệu lực của công cụ

8.1. Đánh giá độ tin cậy của bộ công cụ điều tra học sinh / sinh viên

Bộ công cụ sau khi đã hiệu chỉnh, được điều tra trên 2.271 đối tượng, trong đó có 686 học sinh và 1585 sinh viên. Cả hai mẫu này đều được sử dụng để phân tích đánh giá độ tin cậy của trắc nghiệm.

Theo mô hình lý thuyết một bộ công cụ chỉ đáng tin cậy khi nó đo đúng cái nó được thiết kế để đo. Những đặc tính quan trọng nhất của một phép đo đáng tin cậy là độ ổn định (stability) và độ phù hợp (consistency). Tuy nhiên có nhiều cách để đánh giá độ tin cậy của một phép đo.

Để đánh giá độ tin cậy của bộ công cụ này, chúng tôi dùng các phương pháp đánh giá độ phù hợp của từng item (internal consistency methods), sử dụng mô hình tương quan alpha của Cronbach (Cronbach's Coefficient alpha): mô hình này đánh giá độ tin cậy của phép đo dựa trên sự tính toán phương sai của từng item trong toàn phép đo và tính tương quan điểm của từng item với điểm của toàn bộ các item còn lại của phép đo. Phương pháp này, theo ý kiến của các chuyên gia về test (Murphy & Davishofer, 1995), rất thích hợp cho việc xác định độ tin cậy của các loại phép đo (trắc nghiệm) có các item nhiều mức độ tính theo điểm số.

$$\text{Công thức tính: } R_{xx} = \alpha = \frac{k}{k-1} \left[1 - \frac{\sum \sigma^2 i}{\sigma^2 x} \right]$$

k: Tổng số item của phép đo

$\sigma^2 x$ = Phương sai của cả phép đo

$\sigma^2 i$ = Phương sai của từng item của phép đo

$\sum \sigma^2 i$ = Tổng phương sai của từng item của phép đo

Kết quả tính toán cho thấy độ tin cậy của từng phép đo và toàn bộ thang đo được trình bày trong bảng 7 dưới đây:

Bảng 7: Hệ số tương quan alpha của các phép đo

Các thang đo	Hệ số tin cậy α (Mẫu điều tra học sinh) N = 686	Hệ số tin cậy α (Mẫu điều tra sinh viên) N = 1.581
Đạo đức	0,81	0,78
Tư tưởng-chính trị	0,71	0,68
Lối sống	0,71	0,72
Nhân cách (Tổng)	0,89	0,88

Nhìn vào bảng 7 ta thấy hệ số tương quan alpha của từng thang đo và của toàn phép đo trên cả hai mẫu học sinh và sinh viên đều ở mức cao vừa phải, chấp nhận được cho một phép đo về nhân cách.

Độ tin cậy của trắc nghiệm còn được đánh giá bằng phương pháp kiểm tra độ tin cậy (phù hợp) của từng item. Để đánh giá độ phù hợp này, chúng tôi tính hệ số tương quan giữa điểm số của từng chỉ số với điểm số của tất cả các item còn lại của phép đo. Kết quả tính toán tương quan điểm của từng chỉ số đo với điểm của các item còn lại được trình bày trong bảng 8 và 9 dưới đây.

Nhìn vào bảng 8 ta thấy tương quan của từng chỉ số đo với toàn phép đo trên mẫu học sinh cho thấy hầu hết các tương quan này đều có r lớn hơn 0,30, ngoại trừ 4 chỉ số đo (CT11, LS17, LS20, LS21) có r nhỏ hơn 0,30 nhưng các chỉ số này đều có r lớn hơn 0,23.

Trên nhóm mẫu sinh viên (xem bảng 9) ta thấy hầu hết các tương quan của từng chỉ số đo với toàn phép đo đều có r lớn hơn 30, ngoại trừ 4 chỉ số đo (CT11, LS17, LS20, LS21) có r nhỏ hơn 30 nhưng cả 4 chỉ số này đều có r lớn hơn 26.

Kết quả này nói lên rằng điểm số của các chỉ số đo đều có tương quan đáng kể với tổng điểm của các item còn lại của phép đo. Như vậy các chỉ số đo có cùng miền đo, các item của toàn phép đo có tính đồng nhất (tức là đo đúng cái cần đo) và điều này có nghĩa là tất cả các item đều đóng góp đáng kể cho độ tin cậy của toàn phép đo.

Bảng 8: Hệ số tin cậy của từng item trên mẫu học sinh (N=685)

Các item	Điểm trung bình của thang đo nếu item bị xóa (Seale Mean if Item Deleted)	Phương sai của thang đo nếu item bị xóa (Seale Variance if Item Deleted)	Hệ số tương quan của item với các item còn lại (Corrected Item Total Correlation)	Hệ số alpha nếu item bị xóa (Alpha if Item Deleted)
ĐĐ1	152.7985	370.6933	.4241	.8640
ĐĐ2	154.8350	349.3923	.5971	.8575
ĐĐ3	154.4905	345.5017	.5503	.8594
ĐĐ4	154.5562	361.4197	.5312	.8605
ĐĐ5	155.9474	369.9914	.4449	.8634
ĐĐ6	156.3445	372.7642	.4443	.8637
ĐĐ7	152.3197	352.8611	.5850	.8582
ĐĐ8	155.6905	350.4099	.5636	.8588
CT9	154.0000	356.4503	.5966	.8582
CT10	154.9431	349.5801	.5914	.8577
CT11	155.3927	369.3324	.2872	.8704
CT12	154.4978	361.9287	.4861	.8619
LS13	154.4949	361.3848	.5004	.8614
LS14	154.5723	372.9147	.4183	.8643
LS15	153.7606	366.6473	.4753	.8624
LS16	156.4715	367.7846	.4027	.8647
LS17	157.1139	380.6975	.2359	.8698
LS18	156.2818	363.2904	.4511	.8631
LS19	152.6146	361.8074	.4487	.8632
LS20	156.7737	379.1315	.2837	.8681
LS21	157.2248	381.2915	.2410	.8694

**Bảng 9: Hệ số tin cậy của từng item trên mẫu
sinh viên (N=1.581)**

Phân tích độ tin cậy

Các item	Điểm trung bình của thang đo nếu item bị xoá	Phương sai của thang đo nếu item bị xoá	Hệ số tương quan của item với các item còn lại	Hệ số alpha nếu item bị xoá
ĐĐ1	162.1347	303.8875	.3931	.8388
ĐĐ2	164.3479	290.3308	.5183	.8332
ĐĐ3	163.2334	276.6600	.5419	.8319
ĐĐ4	163.3865	299.1436	.4472	.8366
ĐĐ5	165.0240	300.8589	.4279	.8374
ĐĐ6	165.5054	306.4033	.4127	.8385
ĐĐ7	161.2524	291.6850	.4926	.8343
ĐĐ8	165.0980	287.5860	.5460	.8318
CT9	163.1765	297.7011	.4513	.8363
CT10	163.9804	286.4825	.5415	.8319
CT11	164.2612	301.7842	.2645	.8461
CT12	163.4839	304.1246	.4059	.8384
LS13	163.4314	297.2189	.4307	.8371
LS14	163.8172	306.1026	.3827	.8392
LS15	162.9096	301.6722	.4230	.8376
LS16	165.3814	296.8943	.4292	.8372
LS17	166.1474	310.0624	.2682	.8432
LS18	165.7989	296.6342	.4166	.8378
LS19	162.0677	296.5910	.3954	.8388
LS20	166.3555	307.5027	.2988	.8422
LS21	166.3416	309.4276	.2610	.8436

8.2. Đánh giá độ hiệu lực (ứng nghiệm) của bộ công cụ điều tra học sinh/sinh viên

Độ hiệu lực của bộ công cụ được đánh giá trên cơ sở phân tích điểm số của hai mẫu điều tra khảo sát gồm 686 học sinh và 1.585 sinh viên.

Độ hiệu lực của một trắc nghiệm được xác định như là mức độ chính xác mà trắc nghiệm đó đo đúng cái cấu trúc nó được thiết kế để đo. Một trắc nghiệm có độ tin cậy cao mới là điều kiện cần chứ chưa phải là điều kiện đủ bảo đảm cho nó có độ hiệu lực tốt.

Độ hiệu lực của một phép đo thường được đánh giá như là độ hiệu lực nội dung (content validity), độ hiệu lực cấu trúc (construct validity) và độ hiệu lực tiêu chuẩn (criterion validity) hay độ hiệu lực dự báo (predictive validity). Các yếu tố ảnh hưởng đến độ hiệu lực nội dung chính là độ hiệu lực của từng item. Nó bao gồm quá trình hiệu lực hoá các item của các chuyên gia thiết kế và kết quả đánh giá mức độ phù hợp của từng item trên toàn phép đo.

Các chiến lược đánh giá độ hiệu lực nội dung và độ hiệu lực cấu trúc thực chất là xác định xem liệu một trắc nghiệm có đưa ra được một cách đo phù hợp để đo một đặc tính cụ thể nào đó hay không (bằng cách điều tra ý nghĩa của điểm số của trắc nghiệm khi đo lường).

Để thực hiện nhiệm vụ này, chúng tôi dùng phương pháp phân tích yếu tố để kiểm tra tính tương đồng giữa cấu trúc của phép đo và cấu trúc của đặc tính cần đo.

Phương pháp này cũng cho phép xác định độ phù hợp cho từng chỉ số (độ hiệu lực của từng item). Điều kiện để sử dụng phương pháp này là: ma trận tương quan điểm của phép đo phải có một số đáng kể các tương quan lớn hơn 30, hệ số của phép thử KMO phải lớn hơn 60 và phép thử Bartlett phải ở mức có ý nghĩa.

Điểm số của 686 học sinh được đưa vào phân tích. Kết quả tính toán ma trận tương quan điểm của các chỉ số (xem bảng 10) cho biết, trong ma trận này có một số đáng kể các tương quan lớn hơn 0,30, hệ số của phép đo $KMO = 0,92$, phép thử Bartlett ở mức có ý nghĩa (p nhỏ hơn 0,0001) (xem bảng 11), do vậy thích hợp cho việc dùng phương pháp phân tích yếu tố đánh giá độ hiệu lực của các item.

Bảng 10: Ma trận tương quan (Correlation Matrix)
(Mẫu học sinh, N = 685)

	ĐD1	ĐD2	ĐD3	ĐD4	ĐD5	ĐD6	ĐD7	ĐD8B	CT9	CT10B	CT11B	CT12	LS13	LS14	LS15	LS16	LS17B	LS18	LS19	LS20B	LS21
ĐD1	1,000	,300	,225	,241	,185	,162	,228	,313	,314	,357	,182	,276	,222	,211	,138	,246	,068	,195	,287	,130	,146
ĐD2	,300	1,000	,472	,348	,299	,273	,360	,493	,393	,450	,333	,315	,307	,296	,248	,061	,334	,334	,334	,176	,193
ĐD3	,225	,472	1,000	,325	,337	,267	,380	,433	,373	,355	,117	,251	,278	,325	,263	,249	,026	,372	,332	,132	,199
ĐD4	,241	,348	,325	1,000	,279	,328	,368	,322	,366	,370	,182	,261	,366	,246	,393	,225	,203	,236	,263	,152	,087
ĐD5	,185	,299	,337	,279	1,000	,279	,315	,204	,249	,260	,125	,242	,208	,213	,230	,303	,136	,204	,287	,169	,168
ĐD6	,162	,273	,267	,328	,279	1,000	,268	,205	,229	,349	,136	,221	,335	,153	,237	,285	,196	,215	,237	,161	,153
ĐD7	,228	,360	,380	,368	,315	,268	1,000	,348	,413	,339	,171	,375	,335	,325	,340	,204	,208	,339	,325	,261	,201
ĐD8	,313	,493	,433	,322	,204	,205	,348	1,000	,455	,515	,201	,299	,219	,299	,274	,300	,040	,320	,313	,095	,104
CT9	,314	,393	,373	,366	,249	,229	,413	,455	1,000	,459	,237	,440	,293	,269	,350	,243	,152	,291	,305	,207	,089
CT10	,357	,450	,355	,370	,260	,349	,339	,515	,459	1,000	,238	,268	,313	,257	,283	,339	,110	,307	,308	,086	,128
CT11	,182	,150	,117	,182	,125	,136	,171	,201	,237	,238	1,000	,169	,262	,028	,150	,123	,216	,130	,053	,154	,058
CT12	,276	,333	,251	,261	,242	,221	,375	,299	,440	,268	,169	1,000	,275	,221	,225	,178	,140	,233	,276	,261	,178
LS13	,222	,315	,278	,366	,208	,335	,335	,219	,293	,313	,262	,275	1,000	,234	,430	,213	,259	,240	,204	,169	,112
LS14	,211	,307	,325	,246	,213	,153	,325	,299	,269	,257	,028	,221	,234	1,000	,275	,147	,038	,328	,190	,195	,141
LS15	,138	,296	,263	,393	,230	,327	,340	,274	,350	,283	,150	,225	,430	,275	1,000	,210	,203	,246	,207	,176	,085
LS16	,246	,248	,249	,225	,303	,285	,204	,300	,243	,339	,123	,178	,213	,147	,210	1,000	,167	,147	,276	,052	,107
LS17	,068	,061	,025	,203	,136	,196	,209	,040	,152	,110	,216	,140	,259	,038	,203	,167	1,000	,063	,079	,198	,027
LS18	,195	,334	,372	,236	,204	,215	,339	,320	,291	,307	,130	,233	,240	,328	,246	,147	,060	1,000	,212	,068	,232
LS19	,287	,334	,332	,263	,287	,237	,325	,313	,305	,308	,053	,278	,204	,190	,207	,276	,079	,212	1,000	,141	,081
LS20	,130	,176	,132	,152	,169	,161	,261	,095	,207	,096	,154	,261	,169	,195	,176	,052	,198	,068	,141	1,000	,102
LS21	,146	,193	,199	,087	,168	,153	,201	,104	,068	,128	,058	,178	,112	,141	,085	,107	,027	,232	,081	,102	1,00

Bảng 11: Phân tích yếu tố (Mẫu học sinh, N = 685

Chiết xuất yếu tố: tổng biến thiên được giải tích bởi các yếu tố

Item	Hệ số chung	Yếu tố	Trị số riêng	Phần trăm	Phần trăm gộp
ĐĐ1	.47281	1	6,02932	28,7%	28,7%
ĐĐ2	.51671	2	1,44441	6,9	35,6
ĐĐ3	.52353	3	1,17693	5,6	41,2
ĐĐ4	.47514	4	1,08695	5,2	46,4
ĐĐ5	.48821	5	1,05210	5,0	51,4
ĐĐ6	.49308				
ĐĐ7	.50186				
ĐĐ8	.62774				
CT9	.56689				
CT10	.59135				
CT11	.45656				
CT12	.50938				
LS13	.53220				
LS14	.47711				
LS15	.57231				
LS16	.57469				
LS17	.54252				
LS18	.43977				
LS19	.39111				
LS20	.57913				
LS21	.45763				

Phân tích yếu tố: đánh giá hệ số chứa của từng yếu tố (Mẫu học sinh $N_s = 685$)

	Yếu tố 1	Yếu tố 2	Yếu tố 3	Yếu tố 4	Yếu tố 5
ĐĐ1	.48833				.35846
ĐĐ2	.67322				
ĐĐ3	.63213				
ĐĐ4	.60686				
ĐĐ5	.50876			.42151	
ĐĐ6	.50716	.43658		.43006	
ĐĐ7	.65341				
ĐĐ8	.64373				
CT9	.66881				
CT10	.66799				
CT11	.33424	.38122	.61546		
CT12	.55523				.31273
LS13	.56482	.36436			
LS14	.49259		.37192		
LS15	.54851				
LS16	.46893			.39809	
LS17	.27295	.67564			
LS18	.52304				
LS19	.52708				
LS20	.32913	.35996	.42832		.33541
LS21	.28605		.44316		.30951

Bảng 12: Ma trận tương quan (Correlation Matrix)
(Mẫu sinh viên, N = 1581)

DD1	DD2	DD3	DD4	DD5	DD6	DD7	DD8	CT9	CT10	CT11	CT12	LS13	LS14	LS15	LS16	LS17	LS18	LS19	LS20	S21
DD2	.201																			
DD3	.198	.446																		
DD4	.146	.188	.267																	
DD5	.210	.276	.339	.192																
DD6	.162	.225	.260	.327	.214															
DD7	.210	.295	.341	.294	.220	.245														
DD8	.256	.409	.365	.227	.234	.208	.305													
CT9	.269	.299	.227	.195	.155	.196	.257	.364												
CT10	.344	.347	.289	.315	.239	.229	.226	.508	.411											
CT11	.143	.154	.129	.173	.104	.107	.161	.195	.168	.188										
CT12	.202	.234	.229	.204	.218	.160	.308	.262	.316	.219	.143									
LS13	.202	.207	.242	.394	.186	.303	.222	.243	.195	.322	.111	.211								
LS14	.173	.214	.335	.194	.219	.172	.250	.249	.165	.204	.077	.168	.154							
LS15	.182	.219	.245	.297	.221	.217	.291	.218	.189	.214	.145	.184	.276	.209						
LS16	.223	.242	.267	.204	.285	.190	.190	.287	.230	.346	.131	.152	.221	.152	.215					
LS17	.123	.098	.067	.265	.118	.220	.191	.081	.137	.146	.143	.138	.282	.042	.173	.165				
LS18	.173	.303	.360	.174	.235	.184	.248	.278	.228	.234	.097	.131	.174	.321	.234	.195	.043			
LS19	.192	.276	.262	.193	.265	.220	.248	.251	.160	.243	.091	.188	.141	.194	.170	.279	.098	.190		
LS20	.149	.156	.174	.171	.121	.191	.19	.141	.174	.135	.125	.241	.188	.099	.185	.115	.208	.064	.148	
LS21	.143	.197	.233	.032	.167	.080	.178	.148	.109	.120	.088	.131	.065	.168	.135	.132	.030	.230	.123	.672

Bảng 13: Phân tích yếu tố (Mẫu sinh viên, N = 1.581)

Chiết xuất yếu tố: tổng biến thiên được giải thích bởi các yếu tố

Item	Hệ số chung	Yếu tố	Hệ số riêng	Phần trăm	Phần trăm gộp
ĐĐ1	.31622	1	5,27095	25,1%	25,1%
ĐĐ2	.43052	2	1,47555	7,0	32,1
ĐĐ3	.52655	3	1,18657	5,7	37,8
ĐĐ4	.52466	4	1,18657	5,1	42,8
ĐĐ5	.32318				
ĐĐ6	.40096				
ĐĐ7	.43458				
ĐĐ8	.55364				
CT9	.51218				
CT10	.66452				
CT11	.24325				
CT12	.48262				
LS13	.49425				
LS14	.38241				
LS15	.34025				
LS16	.37785				
LS17	.45755				
LS18	.43917				
LS19	.25483				
LS20	.46297				
LS21	.37275				

Phân tích yếu tố: đánh giá hệ số chứa của từng yếu tố (Mẫu sinh viên $N=1.581$)

	Yếu tố 1	Yếu tố 2	Yếu tố 3	Yếu tố 4
ĐĐ8	.63148			
CT10	.63000			
ĐĐ3	.62933			
ĐĐ2	.59894			
ĐĐ7	.57245			
CT9	.53472			
ĐĐ4	.52695	.40720		
LS13	.51450	.40561		
LS16	.50656			
ĐĐ5	.50442			
LS15	.49825			
LS18	.49652			
ĐĐ6	.48865			
CT12	.48004			4.728
LS19	.47447			
ĐĐ1	.46701			
LS14	.45725			
CT11	.32100			
LS17	.32533	.58445		
LS21	.31490			.32751
LS20	.35929	.32379		.47709

Kết quả phân tích yếu tố trên mẫu học sinh (xem bảng 11) cho thấy có 5 yếu tố nhưng chỉ có yếu tố 1 có giá trị riêng lớn vượt trội hẳn lên (hệ số riêng = 6,03) yếu tố này giải thích cho 28,7% độ biến thiên của toàn phép đo (ma trận tương quan điểm của tất cả các chỉ số (item) của toàn phép đo chỉ chiết xuất được một yếu tố trội, điều này phù hợp với thực tế vì các miễn đo: đạo đức, tư tưởng chính trị và lối sống có cấu trúc trùng nhau ở nhiều điểm). Các yếu tố còn lại có giá trị riêng khá nhỏ từ 1, đến 1,4 giải thích cho từ 5-6,9% độ biến thiên của phép đo. Đánh giá hệ số chứa yếu tố (loading factor) của yếu tố 1 cho thấy hầu như tất cả các chỉ báo của toàn thang đo đều có hệ số chứa lớn hơn 33 và như vậy được coi là phù hợp với cấu trúc của phép đo, riêng hai chỉ báo (LS17, LS20) có r nhỏ hơn 0,30, nhưng lớn hơn 0,27 vì vậy chúng cũng được coi là phù hợp.

Điểm số của 1.586 sinh viên được đưa vào phân tích. Kết quả tính toán ma trận tương quan giữa các chỉ số (xem bảng 12) cho thấy các chỉ số trong bảng ma trận này có nhiều tương quan lớn hơn 0,30, điều này thích hợp cho việc dùng phương pháp phân tích này. Kết quả tính toán cũng cho biết hệ số của phép thử KMO = 0,91, phép thử Barlett ở mức có ý nghĩa (p nhỏ hơn 0,0001) (xem bảng 13), điều này thích hợp cho việc chiết xuất yếu tố.

Kết quả phân tích yếu tố trên mẫu điều tra sinh viên này cho thấy có 4 yếu tố nhưng chỉ có yếu tố 1 có giá trị riêng lớn vượt trội hẳn lên (hệ số riêng = 5,27), yếu tố

này giải thích cho 25% độ biến thiên của toàn phép đo. Các yếu tố còn lại có giá trị riêng khá nhỏ từ 1,1 đến 1,48 giải thích cho từ 5-7% độ biến thiên của phép đo. Như vậy, phép đo có cùng cấu trúc với cái được thiết kế để đo (tức là phép đo có độ hiệu lực về cấu trúc).

Kết quả tính toán hệ số chứa yếu tố (loading factor) cho thấy tất cả 21 chỉ báo của toàn thang đo đều có hệ số chứa lớn hơn 0,31 trên yếu tố 1 và như vậy tất cả 21 chỉ báo đều được coi là phù hợp với cấu trúc của phép đo.

Tóm lại phân tích điểm số thu được trên 2 nhóm mẫu khác nhau (học sinh và sinh viên) đều cho cùng một kết quả: chỉ có một yếu tố trội, tất cả các item đều có hệ số chứa yếu tố phù hợp (lớn hơn 0,30), do đó có cơ sở khoa học để có thể khẳng định rằng đây là phép đo có độ hiệu lực tốt, những số liệu mà nó cung cấp rất đáng tin cậy.

Độ hiệu lực của bộ công cụ còn được kiểm tra qua việc tính hệ số tương quan của các phép đo với nhau. Đạo đức, tư tưởng chính trị và lối sống của một cá nhân là ba mặt của một chỉnh thể nhân cách (đây là những phạm trù giao nhau, làm cơ sở cho nhau), chúng phải có mối quan hệ khăng khít có tính nhân quả, ảnh hưởng chi phối lẫn nhau). Do vậy một phép đo có độ hiệu lực, gồm 3 thang đo điều tra, đánh giá 3 mặt này phải cung cấp 3 loạt điểm có tương quan đáng kể với nhau (vì theo lý thuyết, một phép đo có hiệu lực phải là phép đo có cùng cấu trúc với cái nó được thiết kế để đo). Kết quả tính toán hệ số tương quan của các phép đo trên cả hai mẫu học sinh/sinh viên được trình bày trong bảng 14 và 15.

Bảng 14: Tương quan của các phép đo trên mẫu học sinh (N = 686)

Các thang đo	Đạo đức	Tư tưởng - chính trị	Lối sống	Nhân cách
Đạo đức	1,00			
Tư tưởng-chính trị	0,66**	1,00		
Lối sống	0,72**	0,58**	1,00	
Nhân cách	0,93**	0,81**	0,89**	1,00

** $p < 0,0001$

Bảng 15: Tương quan của các phép đo trên mẫu sinh viên (N = 1.581)

Các thang đo	Đạo đức	Tư tưởng - chính trị	Lối sống	Nhân cách
Đạo đức	1,00			
Tư tưởng-chính trị	0,59**	1,00		
Lối sống	0,70**	0,51**	1,00	
Nhân cách	0,92**	0,75**	0,88**	1,00

** $p < 0,0001$

Kết quả phân tích tương quan điểm ở bảng 14 và 15 cho thấy cả ba thang đo: Đạo đức, tư tưởng - chính trị và lối sống đều có tương quan đáng kể với nhau ($r > 0,51$), kết quả này lại một lần nữa thoả mãn giả thuyết về một phép đo có độ tin cậy và độ hiệu lực tốt đã nêu ở trên rằng phép đo đó phải có cùng cấu trúc với cái nó được thiết kế để đo.

Tóm lại kết quả phân tích trên hai mẫu học sinh và sinh viên cho thấy bộ công cụ dùng để điều tra khảo sát được xem là có độ tin cậy và độ hiệu lực phù hợp. Cả ba thang đo đạo đức, tư tưởng - chính trị, lối sống và tất cả các item của toàn phép đo đều có độ tin cậy và độ hiệu lực bảo đảm theo các yêu cầu thiết kế, đo đúng cái cần đo.

Chương X

THỰC HÀNH CÁC KỸ NĂNG THÍCH NGHI HOÁ CÔNG CỤ ĐO

Dưới đây là một ví dụ *thực hành kỹ năng thích nghi hoá công cụ đo* nhằm phục vụ đề tài: "Bước đầu thử thích nghi hoá trắc nghiệm đánh giá lo âu (stai) của Spielberger"*

1. Các khái niệm công cụ

Trong cuộc sống mọi người đều ít nhiều trải nghiệm lo âu với những triệu chứng thường gặp như cảm giác nhâm lẫn, không vui, bất an, trạng thái mơ hồ về sự hiểu biết... kèm theo những biểu hiện khác như đau đầu, cảm giác tức ngực, khó thở, trạng thái khó tiêu đầy bụng hay cảm giác đứng ngồi không yên... Tuy nhiên, những biểu hiện này có thể là triệu chứng của sự trải nghiệm lo âu ở mức bình thường (lo âu không ở mức bệnh lý), đó là những phản ứng tích cực, có lợi nhằm đáp lại một tình huống không quen thuộc, hoặc tình

* Đề tài này liên quan đến khoá luận của Cử nhân tâm lý Trần Mạnh Hà và một nhóm sinh viên Khoa Tâm lý, Trường Đại học Khoa học xã hội và nhân văn thuộc Đại học Quốc gia Hà Nội.

hướng đe dọa, chẳng hạn như trẻ em bị đe dọa tách khỏi bố mẹ, những đứa trẻ buổi đầu tiên đến trường, một ai đó lần đầu phải nói trước đám đông...

Người ta thường trải nghiệm sự lo âu khi phải đối mặt với những thay đổi, những môi trường không quen thuộc, những trải nghiệm mới và những lo âu này là một phần không thể thiếu trong suốt cuộc đời của mỗi cá nhân.

Tiêu chuẩn để đánh giá lo âu ở mức độ bệnh lý là cường độ, mức độ khó kiểm soát và thời gian tồn tại.

Theo tiêu chuẩn phân loại lo âu của hội Tâm thần học Mỹ DSM-IV, sự khác biệt giữa lo âu không cụ thể (generalized anxiety) ở mức bệnh lý và sự lo âu không ở mức bệnh lý là trạng thái lo lắng thái quá và thân chủ có khó khăn hoặc mất sự kiểm soát những lo âu này.

Theo DSM-IV, rối loạn lo âu không cụ thể là những lo lắng hoặc lo sợ thái quá về một số sự kiện hoặc các hành vi kéo dài trong nhiều ngày, thường xảy ra lặp đi lặp lại. Thân chủ thường có khó khăn trong sự kiểm soát những lo lắng và thường có những dấu hiệu thực thể chẳng hạn như là sự căng cơ, cáu bẳn, khó ngủ và cảm giác bất an... Những lo lắng không cụ thể cũng có thể gây ra từ những nguyên nhân kích thích hoặc những điều kiện y học nói chung. Những người trải nghiệm sự lo âu thường khó kiểm soát những xúc cảm của bản thân, luôn có những xúc cảm tiêu cực, họ thường cảm nhận có sự đảo lộn hay sự mất năng lực trong phạm vi hoạt động của bản thân.

Những biểu hiện chủ yếu của rối nhiễu lo âu không cụ thể là:

- + Sự lo lắng thái quá.
- + Sự căng cứng các cơ: thường được biểu hiện ra bên ngoài như đau đầu, bồn chồn, bất an, sự run chân tay.
- + Sự hoạt động thái quá của thần kinh tự chủ biểu hiện như: hơi thở nhanh, tiết mồ hôi thái quá, tim đập nhanh, các triệu chứng khó tiêu đầy bụng.
- + Sự cảnh giác thái quá của nhận thức biểu hiện dễ bực dọc, cáu gắt, bồn chồn, bất an...

2. Xuất xứ trắc nghiệm

Trắc nghiệm tự đánh giá trạng thái - nét lo âu của Spielberger viết tắt là STAI (State-Trait Anxiety Inventory) đã được sử dụng rộng rãi trong nghiên cứu và thực hành lâm sàng. Trắc nghiệm này bao gồm hai phép đo tách biệt nhau như là hai tiểu trắc nghiệm, đánh giá trạng thái lo âu (tâm trạng lo ngại tức thời xảy ra tại thời điểm đánh giá) và nét lo âu (cảm giác lo lắng, căng thẳng bất an ở mức chung nhất mà cá nhân có).

Khái niệm trạng thái và những nét lo âu (những dấu vết lo âu) lần đầu tiên được Cattell giới thiệu năm 1966, sau đó được Spielberger phát triển và hoàn thiện năm 1966, 1972, 1979, nó được xem như là các trạng thái của nhân cách hoặc những phản ứng xúc cảm tồn tại ở những thời điểm nào đó và ở một mức độ căng thẳng nào đó. Các trạng thái lo âu được đặc trưng bởi cảm giác của chủ thể về *sự căng thẳng*, về *sự lo sợ* và *sự hoạt động*

thái quá của hệ thống thần kinh tự chủ. Những nét lo âu thường liên quan đến những khuynh hướng hay xu hướng của một cá nhân (khuynh hướng này có tính ổn định tương đối), khi nhận thức hoặc xem một tình huống gây stress như là một tình huống nguy hiểm đe dọa và đáp ứng lại những tình huống gây stress này thường có một sự căng thẳng đáng kể ở góc độ tâm sinh lý.

Những người có trạng thái lo âu thường liên quan đến phản ứng hoặc một quá trình xảy ra tại một thời điểm nào đó và với mức độ đáng kể của sự căng thẳng mà cá nhân khó kiểm soát.

Những người với những nét lo âu ở mức cao thường bộc lộ tính dễ bị kích thích, dễ sợ hãi, họ có những khuynh hướng xem những tình huống gây stress như là những tình huống đe dọa và gây nguy hiểm.

Phép đo trạng thái lo âu SA (state anxiety) form Y1 bao gồm 20 item đánh giá những trạng thái xúc cảm tình cảm đặc trưng của trạng thái lo âu như sự lo ngại, sự căng thẳng, sự bất an và lo lắng mà một cá nhân đang trải nghiệm tại thời điểm đánh giá hoặc trong những tình huống cụ thể vừa diễn ra trong quá khứ hoặc sẽ diễn ra trong tương lai hay trong các tình huống giả định.

Theo Spielberger, tác giả của trắc nghiệm, tiểu trắc nghiệm SA tỏ ra khá nhạy cảm với những trải nghiệm lo lắng của thân chủ trong các tình huống gây stress, tư vấn, trị liệu tâm lý và các chương trình can thiệp điều

chính hành vi. Điểm số của phép đo này tăng đáng kể khi thân chủ bị stress chẳng hạn đối mặt với các tình huống nguy hiểm, các tình huống gây căng thẳng như trước các cuộc phẫu thuật, trước các kỳ thi, các cuộc phỏng vấn tuyển người vào làm việc của các công ty... Điểm số của phép đo này giảm đáng kể khi thân chủ vừa trải qua một bài tập thư giãn.

Phép đo những nét lo âu TA (trait anxiety) form Y2 gồm 20 item đánh giá mức độ lo âu chung nhất mà mọi người thường có như là hậu quả của những trải nghiệm của người có rối nhiễu lo âu. Những bệnh nhân tâm thần, những người bị trầm nhược, những người bị rối nhiễu tâm thể... thường có điểm số cao khi đánh giá bằng phép đo này.

Trắc nghiệm tự đánh giá trạng thái lo âu - nét lo âu được Spielberger xây dựng năm 1973. Thoạt đầu nó được dùng để đánh giá mức độ lo âu ở trẻ em các lớp 4, 5, 6 và năm 1976 được Spielberger chuẩn hoá cho những lứa tuổi lớn hơn: học sinh phổ thông trung học, sinh viên, công nhân, tân binh và các đối tượng là người lớn khác từ 19-60 tuổi. Trên cơ sở các kết quả nghiên cứu gần một thập kỷ, năm 1979 tác giả đã sửa chữa lại. Theo tác giả, lần sửa chữa này nhằm các mục tiêu: phát triển một phép đo *thuần nhất về lo âu, có độ nhạy cảm cao về rối nhiễu lo âu* (khả dĩ có thể tách biệt được những xúc cảm tình cảm tiêu cực biểu hiện rối nhiễu lo âu với những xúc cảm tình cảm tiêu cực biểu hiện rối nhiễu trầm cảm); một số các item có độ tin cậy thấp, ít

nhạy cảm đã được thay thế; và cấu trúc của phép đo cũng được cải tiến để có sự cân bằng về số lượng giữa các item đánh giá trạng thái lo âu và nét lo âu.

Tính đến đầu những năm 1990, trắc nghiệm này đã được thích nghi hoá và sử dụng ở rất nhiều nước như Hà Lan, Đức, Hy Lạp, Đan Mạch, Bồ Đào Nha, Tây Ban Nha, Thổ Nhĩ Kỳ, Nga, Ôxtrâylia, Canada, Niu Dilân... Đến thời điểm đó đã có trên 2.000 công trình nghiên cứu thuộc các lĩnh vực y học, tâm thần học, tâm lý học, giáo dục học, quân sự... sử dụng trắc nghiệm này và bộ công cụ này đã được đánh giá là có độ tin cậy cao, độ hiệu lực bảo đảm.

3. Các thông số chuẩn hoá của trắc nghiệm

Trắc nghiệm STAI được dùng để Việt hoá và thích nghi cho lứa tuổi học sinh phổ thông trung học là trắc nghiệm công bố năm 1983 đã qua sửa chữa lần cuối cùng (và hiện đang được sử dụng rộng rãi ở Anh, Mỹ, Ôxtrâylia...

Theo tác giả trắc nghiệm, có nhiều nghiên cứu nhằm chuẩn hoá trắc nghiệm. Những mẫu được công bố dùng để chuẩn hoá trắc nghiệm này được mô tả trong bản hướng dẫn trắc nghiệm là các nhóm mẫu sau đây:

- Người lớn đang làm việc (working adult): gồm 1.838 người từ 19-69 tuổi, trong đó có 1.387 nam và 451 nữ.

- Sinh viên: gồm 855 sinh viên đăng ký vào học các khoá "Nhập môn tâm lý" ở Trường Đại học South Florida, trong đó có 324 nam và 531 nữ.

- Tân binh: nhóm này gồm 1.701 thanh niên nam mới tuyển vào phục vụ trong lực lượng không quân và 263 lính mới tuyển vào lực lượng Hải quân (trong đó có 192 nam và 71 nữ).

Đặc điểm của các mẫu này là không chọn lựa mẫu theo phương pháp ngẫu nhiên, cũng không chọn lựa mẫu theo phương pháp đại diện hay phương pháp phân tầng. Như vậy các nhóm mẫu được tác giả trắc nghiệm dùng để đánh giá các thông số kỹ thuật đều là mẫu tiện lợi, tức là các mẫu này chỉ đại diện cho chính nó. Tuy nhiên theo tác giả trắc nghiệm, các mẫu này đều có thể dùng để so sánh với các nhóm mẫu tương đương khác.

Kết quả phân tích các thông số định tâm, độ tin cậy được trình bày trong bảng 1 dưới đây:

Bảng 1: Các tham số định tâm của trắc nghiệm STAI qua các nhóm mẫu chuẩn hoá

Các thang đo	Người lớn		Học sinh THPT		Sĩ sinh viên		Tân binh	
	Nam (1.387)	Nữ (451)	Nam (202)	Nữ (351)	Nam (324)	Nữ (531)	Nam (1.893)	Nữ (71)
SA								
Giá trị trung bình	35,72	35,20	39,45	40,54	36,47	38,76	44,05	47,01
Độ lệch chuẩn	10,40	10,61	9,74	12,86	10,02	11,95	12,18	14,42
Hệ số tin cậy	0,93	0,93	0,86	0,94	0,91	0,93	0,93	0,95
TA								
Giá trị trung bình	34,89	34,79	40,17	40,97	38,30	40,40	37,64	40,03
Độ lệch chuẩn	9,19	9,22	10,53	10,63	9,18	10,15	9,51	9,90
Hệ số tin cậy	0,91	0,91	0,90	0,90	0,90	0,91	0,89	0,90

Độ tin cậy của trắc nghiệm được đánh giá theo phương pháp tính tương quan giữa hai lần đo lặp lại trên cùng một đối tượng (test - retest reliability) trình bày trong bảng 2 dưới đây:

Bảng 2: Độ tin cậy test-retest của trắc nghiệm STAI

Các phép đo	Học sinh trung học phổ thông				Sinh viên					
	30 ngày		60 ngày		1 giờ		20 ngày		104 ngày	
SA (trạng thái lo âu)	N	R	N	R	N	R	N	R	N	R
Nam	178	,62	177	,51	88	,33	38	,54	25	,33
Nữ	179	0,34	205	0,36	109	0,16	75	0,27	22	0,31
TA (nét lo âu)										
Nam	173	,71	174	,68	88	,84	38	,86	25	,73
Nữ	178	0,75	201	0,65	109	0,76	75	0,76	22	0,77

Do trạng thái lo âu phụ thuộc nhiều vào tình huống, hoàn cảnh nên độ ổn định của trắc nghiệm qua các lần đo ở mức trung bình (trên một vài nhóm mẫu, hệ số này khá thấp). Trong đó tiểu trắc nghiệm SA được thiết kế để đo trạng thái lo âu, có độ nhạy cao với những tình huống gây stress nên có độ tương quan test - retest thấp (từ 0,16 đến 0,62). Còn tiểu trắc nghiệm TA được thiết kế đo những nét lo âu, do vậy có tính ổn định tốt hơn, tương quan giữa hai lần đo test - retest qua các khoảng cách giữa các lần đo từ 1 giờ đến 104 ngày đều tương đối cao (có r từ 0,65 đến 0,86).

Độ tin cậy của trắc nghiệm còn được đánh giá theo mô hình alpha của Cronbach. Nhìn chung trắc nghiệm

lo âu - STAI có độ tin cậy alpha cao (Cronbach's coefficient alpha) từ 0,86 đến 0,93.

Độ hiệu lực của trắc nghiệm được tác giả Spielberger chú ý ngay từ khâu xây dựng cấu trúc của trắc nghiệm và các bước hiệu lực hóa từng item (độ hiệu lực cấu trúc và độ hiệu lực nội dung). Theo ông, từng item phải đáp ứng các tiêu chuẩn ở từng giai đoạn của quá trình phát triển trắc nghiệm.

Bảng 3: Tương quan giữa SA và TA qua các nhóm mẫu

	Người lớn	Hệ số tương quan	Học sinh THPT	Hệ số tương quan	Sinh viên	Hệ số tương quan	Tân bình	Hệ số tương quan
	N	r	N	r	N	r	N	r
Nam	1.387	0,75	202	0,65	324	0,72	1.893	0,59
Nữ	451	0,70	222	0,59	531	0,64	1.893	0,59

4. Quá trình thiết kế và xây dựng trắc nghiệm

Quá trình thiết kế và xây dựng trắc nghiệm, theo tác giả Spielberger gồm các bước cơ bản sau đây:

Bước 1: Việc xây dựng trắc nghiệm bắt đầu bằng xác định rõ mô hình lý thuyết và thành lập "bể item" (establish a pool of items). Ba trắc nghiệm đánh giá lo âu của Taylor, Welsh, Cattell và Schier cùng được 288 sinh viên năm thứ nhất Khoa Tâm lý của Đại học Vanderbilt (Mỹ) tự đánh giá. Kết quả trả lời của từng

item được phân tích, sau đó người ta chọn ra 177 item có hệ số tương quan lớn hơn hoặc bằng 0,25 với các item còn lại của 3 trắc nghiệm này. Tất cả 177 item chọn ra được viết lại (nội dung tâm lý của từng item được giữ nhưng hình thức, ngôn từ được thay đổi) để mỗi item có thể dùng để đánh giá trạng thái lo âu hoặc nét lo âu (trait anxiety).

Bước 2: Chọn lọc item. Một nhóm sinh viên khoa tâm lý (được trang bị kiến thức về trắc nghiệm, về khái niệm lo âu) được đề nghị xem xét lại 177 item vừa được viết lại nhận xét về hình thức thể hiện, về nội dung (xem có đúng item đó đánh giá lo âu hay không, hình thức thể hiện có phù hợp không). Dựa trên những nhận xét của nhóm sinh viên này, những item nào mơ hồ về nội dung, không rõ nghĩa, hình thức thể hiện không phù hợp đều bị loại bỏ. Bước hiệu lực hoá sơ bộ này đã loại bỏ 53 item, chỉ giữ lại 124 item (được đánh giá là có tiềm năng cho việc đo trạng thái lo âu và nét lo âu) cho quá trình hiệu lực hoá tiếp theo.

Bước 3: Hiệu lực hóa item. Một nhóm khác gồm 54 sinh viên năm thứ ba Khoa Tâm lý của Đại học Vanderbilt được yêu cầu đánh giá lại 124 item này xem từng item phát hiện cảm giác chung nhất về lo âu như thế nào theo bốn mức độ: *hầu như không, đôi khi, thường xuyên, hầu như mọi lúc*. Sau đó tất cả các item này lại được những sinh viên này đánh giá lại xem từng item có độ nhạy như thế nào khi đánh giá trạng thái lo

âu ở thời điểm hiện tại với các mức độ "Có", "Không" hoặc "Không rõ". Kết quả phân tích xử lý đã chọn ra được 66 item. Đây là những item có hệ số tương quan cao với các item còn lại của cả phép đo ($r \geq 0,35$).

Bước 4: Hiệu lực item tiếp theo. Chỉ có 66 item đáp ứng các tiêu chuẩn chọn lọc ở bước 3 được giữ lại. Những item này được đưa cho 265 sinh viên của Đại học Vanderbilt đánh giá trong hai tình huống: lúc thư giãn (không có stress) và lúc chuẩn bị thi (có stress).

Thoạt đầu các sinh viên được hướng dẫn trả lời từng item với cảm giác chung nhất (nét lo âu). Sau đó nhóm sinh viên này được hướng dẫn hãy trả lời trên từng item trong tình huống thư giãn và tình huống có stress (trước lúc thi). Hầu hết các sinh viên này trước đó 10 tuần đã được đánh giá bằng hai trắc nghiệm lo âu của Taylor và của Cattell & Schier. Những item bị loại bỏ là những item có hệ số tương quan yếu với điểm số kết hợp của hai trắc nghiệm lo âu vừa nêu trên (chỉ giữ lại những item có r lớn hơn hoặc bằng 0,20).

Một item cũng bị loại bỏ nếu có hệ số tương quan thấp với các item còn lại (item - remainder correlation) ($r < 0,30$). Lúc này chỉ có 44 item thoả mãn cả hai điều kiện trên được giữ lại cho quá trình hiệu lực hoá tiếp theo.

Bước 5: Phân loại các item. Cả 44 item còn lại được thử trên mẫu lớn gồm 810 sinh viên của Đại học Vanderbilt (trong đó có 561 nam và 249 nữ). Thoạt đầu

người ta hướng dẫn các sinh viên trả lời từng item với cảm giác chung nhất (nét lo âu). Sau đó các sinh viên được hướng dẫn hãy trả lời từng item với tình huống thư giãn và tình huống sắp có cuộc thi (stress). Để tìm những item có độ hiệu lực tốt nhất theo cách hướng dẫn đầu tiên (đánh giá nét lo âu chung nhất), người ta tính hệ số tương quan của từng item với các item còn lại. Để tìm những item có độ hiệu lực tốt nhất theo cách hướng dẫn sau trả lời với hai tình huống thư giãn và có stress (đánh giá trạng thái lo âu) bằng cách tính độ phân biệt của từng item giữa hai tình huống thư giãn và stress. Chỉ những item nào có hệ số tương quan với các item còn lại lớn hơn hoặc bằng 0,24 và có độ phân biệt rõ giữa hai lần đo trong tình huống thư giãn và tình huống có stress được giữ lại. Tất cả có 32 item thỏa mãn các đặc tính đo lường vừa nêu được giữ lại. Cả 32 item này đều là những item đáp ứng tiêu chuẩn hiệu lực có thể đo cả nét lo âu và đồng thời đo cả trạng thái lo âu.

Bước 6: Mục tiêu chính của bước này là tìm những item thích hợp nhất cho phép đo trạng thái lo âu. Cả 32 item giữ lại, một lần nữa được thử trên 400 sinh viên năm thứ nhất Khoa Tâm Lý ở hai thời điểm khác nhau. Thời điểm học giữa kỳ 1 không có thi cử và 2 tháng sau cũng những học sinh đó bắt đầu vào kỳ thi hết học kỳ. Người ta đánh giá sự thay đổi điểm số trên từng item bằng cách tính tương quan điểm giữa hai lần đo.

Sau đó người ta chọn 20 item đáp ứng tốt nhất các tiêu chuẩn vừa nêu (tức là những item có số điểm thay đổi đáng kể giữa lần đo trong tình huống thư giãn và tình huống chuẩn bị thi hết học kỳ) vào form A như là một phép đo có độ hiệu lực tốt để đánh giá trạng thái lo âu. Thực tế các item của form A đã đáp ứng các đặc tính đo lường kiểm tra ở bước 4 do vậy nó cũng là phép đo có hiệu lực tốt để đánh giá nét lo âu (nếu thay đổi lời hướng dẫn).

Bước 7: Mục tiêu chính của bước này là tìm những item thích hợp nhất cho phép đo nét lo âu. Người ta chọn 20 item đáp ứng tốt nhất các đặc tính đo lường được chọn ở bước 4 và bước 5 nhưng không được chọn vào form A để hình thành form B. Các item được chọn vào form B là thích hợp để đo lường nét lo âu.

Tuy nhiên, nhiều nghiên cứu sau đó cho thấy ở cả form A và form B đều có những item không đủ độ nhạy để đánh giá trạng thái lo âu (điểm không thay đổi đáng kể khi chuyển từ tình huống thư giãn sang tình huống có stress) hoặc không đủ độ ổn định để đánh giá nét lo âu. Do vậy cả hai form A và B tiếp tục được sửa chữa để hình thành form X. Form X gồm những item đã qua quá trình hiệu lực hoá chứng tỏ có những đặc tính đo lường tốt nhất (hoặc là đo lường trạng thái lo âu hoặc là đo lường nét lo âu).

Quá trình chuẩn hoá form X có hơn 6.000 học sinh

phổ thông trung học và sinh viên, khoảng 600 bệnh nhân và 200 tù nhân được yêu cầu tham gia vào các nhóm mẫu khác nhau. Form X qua gần một thập kỷ tồn tại, nó được sử dụng nhiều hơn bất cứ các phép đo lo âu nào khác cùng thời. Tuy nhiên người ta phát hiện thấy form X không có sự cân đối giữa hai yếu tố đánh giá tình trạng có mặt và vắng mặt của những triệu chứng lo lắng. Hơn nữa có một vài item có độ tin cậy và hiệu lực không ổn định qua một số nhóm mẫu. Form X một lần nữa được sửa chữa lại thành form Y. Mục tiêu của sự sửa chữa lần này là tạo ra một thang đo lo âu gồm 2 form: Y1 và Y2 có sự cân đối về số lượng item và tương đối tách biệt khi đánh giá trạng thái lo âu và nét lo âu. Quá trình sửa chữa và chuẩn hoá lại này có trên 5.000 đối tượng thuộc các nhóm mẫu khác nhau tham gia. Đặc biệt là người ta đã sử dụng phương pháp phân tích yếu tố để đánh giá độ chứa yếu tố của form Y1 và Y2. Kết quả phân tích trên 1.728 lính tân binh (xem bảng 4) cho thấy các item từ 1 đến 20 đều có độ chứa lớn hơn 0,40 ở yếu tố I có tên là form Y1 do trạng thái lo âu. Các item từ 21 đến 40 có độ chứa lớn hơn 0,32 ở yếu tố II (form Y2) do nét lo âu. Kết quả phân tích yếu tố này xác nhận form Y có độ hiệu lực về cấu trúc phù hợp. Mặt khác độ hiệu lực của thang đo lo âu này còn được khẳng định qua hệ số tương quan đáng kể giữa trạng thái lo âu và nét lo âu trình bày ở bảng 3.

Bảng 4: Cấu trúc yếu tố của form Y (N = 1.728)

<i>Item của form Y</i>	<i>Độ chứa yếu tố I</i>	<i>Độ chứa yếu tố II</i>
Y1		
1	0,64	
2	0,59	
3	0,68	
4	0,54	
5	0,66	
6	0,59	
7	0,49	
8	0,54	
9	0,58	
10	0,67	
11	0,42	
12	0,66	
13	0,59	
14	0,46	
15	0,68	
16	0,48	
17	0,61	
18	0,57	
19	0,58	
20	0,60	
Y2		

<i>Item của form Y</i>	<i>Độ chứa yếu tố I</i>	<i>Độ chứa yếu tố II</i>
21		0,60
22		0,51
23		0,62
24		0,36
25		0,50
26		0,56
27		0,63
28		0,48
29		0,43
30		0,72
31		0,46
32		0,46
33		0,70
34		0,47
35		0,46
36		0,60
37		0,37
38		0,32
39		0,54
40		0,38
Tên yếu tố	Trạng thái lo âu	Nét lo âu

Tóm lại thang đo lo âu của Spielberg (form Y) đã có một lịch sử phát triển, sửa đổi và chuẩn hoá rất công phu bài bản. Độ tin cậy và độ hiệu lực của phép đo đã

được chứng minh là một tiêu chuẩn (đủ cao) chẩn đoán trong đánh giá lâm sàng.

5. Quy trình thích nghi hoá trắc nghiệm

Công việc thích nghi hoá trắc nghiệm không phải chỉ là dịch các trắc nghiệm từ tiếng nước ngoài sang tiếng Việt, dịch bản hướng dẫn sử dụng mà cái chính là phải biết xuất xứ trắc nghiệm được thiết kế để đo cái gì, dựa trên mô hình lý thuyết nào, gồm những miền đo nào, dựa trên những nguyên tắc hay phương pháp nào để thiết kế và xây dựng các item... áp dụng những kỹ thuật nào để xác định các thông số về độ tin cậy và độ hiệu lực. Bởi vì tất cả những trắc nghiệm bằng lời đều có bản chất văn hoá và sẽ không thích hợp nếu dùng cho các đối tượng trong các nền văn hoá khác mà không qua thích nghi và chuẩn hoá lại.

Quy trình được sử dụng để thích nghi hoá trắc nghiệm tự đánh giá trạng thái - nét lo âu của Spielberger gồm những bước sau đây:

Bước 1: Trắc nghiệm được chuyển ngữ tương đương từ tiếng Anh sang tiếng Việt. Tất cả các item của trắc nghiệm khi dịch sang tiếng Việt đều được kiểm tra lại xem từng item có bị thay đổi nghĩa, thay đổi mục tiêu đo lường, có đi chệch ra khỏi từng mục tiêu cụ thể: đo cái nó có nhiệm vụ phải đo không? Các item sau khi chuyển sang tiếng Việt phải bảo đảm dễ hiểu, đơn nghĩa và vẫn đo đúng trạng thái lo âu (công việc này do một chuyên gia tâm lý chuyên sâu về thiết kế, thích nghi hoá trắc

nghiệm và một nhóm sinh viên tâm lý thực hiện).

Bước 2: Trắc nghiệm được làm thử trên mẫu nhỏ để kiểm tra những khả năng đọc hiểu trắc nghiệm của đối tượng. Một nhóm gồm 40 học sinh lớp 9 được làm thử cho thấy các em không có khó khăn gì khi đọc và hiểu để trả lời các item của trắc nghiệm.

Bước 3: Trắc nghiệm được thử trên nhóm mẫu đủ lớn để xác định lại các thông số về độ tin cậy và độ hiệu lực (ứng nghiệm) của trắc nghiệm.

Do điều kiện chưa thể có được một mẫu đại diện cho các lứa tuổi, ba nhóm mẫu khác nhau sau đây được phân tích:

Nhóm 1: Gồm 439 học sinh trung học phổ thông từ lớp 10 - 12 của 3 trường phổ thông trung học: Trường Trung học phổ thông Phạm Hồng Thái (khu vực nội thành Hà Nội), Trường Trung học phổ thông Xuân Đình (khu vực ngoại thành Hà Nội), Trường Trung học phổ thông chuyên ban Lê Quý Đôn (khu vực thị xã Hà Đông).

Nhóm 2: Gồm 130 sinh viên năm thứ hai Khoa Tâm lý và Khoa Du lịch Trường Đại học Khoa học xã hội và nhân văn thuộc Đại học Quốc gia Hà nội.

Nhóm 3: Gồm 126 học sinh lớp 9 của ba trường trung học cơ sở ở Hà Nội (Trường Trung học cơ sở Đại Kim, Trường Trung học cơ sở Thanh Trì...).

Bước 4: Xác định lại các tham số định tâm của trắc nghiệm trên nhóm học sinh được đo gồm: giá trị trung bình, độ lệch chuẩn và sai số chuẩn của phép đo.

Bước 5: Đánh giá lại độ tin cậy của trắc nghiệm chúng tôi sử dụng mô hình alpha của Cronbach để kiểm tra mối tương quan bên trong giữa các item và kiểm tra tương quan tuyến tính của từng item với toàn bộ phép đo. Chúng tôi cũng sử dụng phương pháp phân tích yếu tố để đánh giá độ tin cậy của trắc nghiệm.

Bước 6: Đánh giá lại độ hiệu lực của trắc nghiệm. Để đánh giá độ hiệu lực của trắc nghiệm chúng tôi đã so sánh điểm số của học sinh đo bằng trắc nghiệm lo âu của Spielbergger với điểm số đo bằng trắc nghiệm lo âu của Beck trên cùng một nhóm học sinh (đây cũng là một trắc nghiệm đánh giá lo âu đã được chuẩn hoá ở Mỹ, Ôxtrâyliá). Sự tương quan đáng kể giữa điểm số của hai trắc nghiệm là bằng chứng về độ hiệu lực tiêu chuẩn của trắc nghiệm. Phương pháp phân tích yếu tố cũng được dùng để kiểm tra độ hiệu lực của phép đo.

Bước 7: Sửa chữa, điều chỉnh các item có độ tin cậy yếu, đưa ra các khuyến nghị định hướng cách sử dụng trắc nghiệm.

6. Đánh giá lại các tham số định tâm của trắc nghiệm

Trắc nghiệm lo âu của Spielbergger, gồm hai tiểu trắc nghiệm: đánh giá trạng thái lo âu (form Y1) và đánh giá nét lo âu (form Y2), được kiểm tra trên 5 nhóm mẫu: Khối lớp 9, 10, 11, 12 và sinh viên để xác định lại các tham số định tâm. Kết quả tính toán được trình bày trong bảng 5.

Bảng 5: Các tham số định tâm của trắc nghiệm

Thang đo	Trạng thái lo âu (form Y 1)				Nét lo âu (form Y 2)			
Nhóm mẫu (lớp)	Cỡ mẫu (N)	Giá trị trung bình	Độ lệch chuẩn (SD)	Sai số phép đo (SE)	Cỡ mẫu (N)	Giá trị trung bình (SD)	Độ lệch chuẩn (SD)	Sai số phép đo (SE)
Khối 9	-	-	-	-	136	46,34	8,19	0,75
Khối 10	136	42,65	10,62	0,91	138	47,78	8,54	0,73
Khối 11	135	41,66	9,57	0,82	137	47,27	8,46	0,72
Khối 12	146	43,16	10,34	0,86	145	48,08	7,87	0,65
Sinh viên	119	44,79	10,83	0,99	119	48,72	8,96	0,82

Từ kết quả nghiên cứu được trình bày trong bảng 5, chúng tôi thấy rằng so với số liệu phân tích trên bảng 1 (mẫu chuẩn hoá của Mỹ) giá trị của các tham số định tâm đều bị thay đổi.

Trị số trung bình của nhóm mẫu học sinh lớp 10 ở Việt Nam cao hơn so với trị số trung bình của nhóm mẫu học sinh lớp 10 ở Mỹ. Có thể có nhiều lý do dẫn đến học sinh lớp 10 ở Việt Nam có điểm số trung bình lo âu cao hơn: do yếu tố văn hoá, yếu tố ngôn ngữ và do kích cỡ mẫu nhỏ (không phải là mẫu đại diện)... Điểm trung bình của học sinh Việt Nam cao hơn không có nghĩa là học sinh Việt Nam có nhiều lo âu hơn học sinh lớp 10 ở Mỹ, cũng không có nghĩa là trong nhóm học sinh lớp 10 ở Việt Nam số người lo âu nhiều hơn.

Việc tính toán lại giá trị trung bình, độ lệch chuẩn

trên các nhóm mẫu Việt Nam khác nhau sẽ mang lại lợi ích cho những đánh giá chuẩn đoán lâm sàng. Những người có điểm số lo âu cao hơn điểm trung bình một độ lệch chuẩn (tức là lớn hơn hoặc bằng 53 điểm trên thang đo SA và lớn hơn hoặc bằng 56 điểm trên thang đo TA) được coi là những người có biểu hiện lo âu (kém thích nghi). Nếu theo tiêu chuẩn phân loại này, thì hiện có từ 17,74% đến 18,81% học sinh trung học phổ thông (lớp 10, 11, 12) có biểu hiện lo âu. Cũng mẫu học sinh trung học phổ thông này, chúng tôi đánh giá bằng thang lo âu của Beck thì có 17,02% học sinh có biểu hiện lo âu (cao hơn điểm trung bình 1 SD). Kết quả được trình bày trong bảng 6.

Bảng 6: Mức độ lo âu theo các thang đo

Các thang đo	Học sinh trung học phổ thông được điều tra	Điểm trung bình	Độ lệch chuẩn (SD)	Số học sinh có điểm cao hơn điểm trung bình 1SD
SA	417	42,48	10,18	17,74% (≥ 53)
TA	420	47,71	8,27	18,81 (≥ 56)
Form Y (SA+TA)	413	90,28	16,69	17,97% (≥ 107)
BA	423	13,42	9,90	17,02% (≥ 23)

SA: Thang đo trạng thái lo âu (Form Y1)

TA: Thang đo nét lo âu (Form Y2)

Form Y (SA+TA): Toàn bộ thang đo của Spielberger

BA: Thang đo lo âu của Beck

7. Đánh giá lại độ tin cậy lực của trắc nghiệm

Một trắc nghiệm tốt trước hết phải là một phép đo tin cậy, đo đúng cái cần đo. Có nhiều phương pháp khác

nhau để đánh giá độ tin cậy của trắc nghiệm (reliability), nhưng trong phạm vi nghiên cứu này độ tin cậy của trắc nghiệm được kiểm tra bằng phương pháp đánh giá độ phù hợp của từng item (internal consistency methods) sử dụng mô hình tương quan alpha của Cronbach (Cronbach's coefficient alpha). Mô hình này đánh giá độ tin cậy của phép đo dựa trên sự tính toán phương sai của từng item trong toàn phép đo và tính tương quan điểm của từng item với điểm của tổng các item còn lại trong phép đo. Phương pháp này thích hợp cho việc xác định độ tin cậy của các loại phép đo (trắc nghiệm) có các item được thiết kế theo kiểu nhiều mức độ tính theo điểm số.

Kết quả được trình bày ở bảng 7 dưới đây:

Bảng 7: Hệ số tin cậy alpha của trắc nghiệm

Các thang đo	Khối lớp 10 (N = 136)	Khối lớp 11 (N = 135)	Khối lớp 12 (N = 146)	Sinh viên (N = 119)
SA	0,89	0,87	0,89	0,90
TA	0,85	0,84	0,84	0,84
Form Y	0,92	0,90	0,92	0,93

Từ kết quả nghiên cứu được trình bày trong bảng 7 ta nhận thấy hệ số tin cậy alpha của trắc nghiệm có giảm đi một chút so với mẫu chuẩn hoá, nhưng các hệ số này vẫn đủ cao, bảo đảm độ tin cậy cho một phép đo lâm sàng. Trong đó hệ số tin cậy alpha của thang đo trạng

thái lo âu (SA) tốt hơn (từ 0,87 đến 0,90) còn hệ số tin cậy của thông số nét lo âu (TA) thấp hơn một chút (từ 0,84 đến 0,85). Độ tin cậy của toàn phép đo từ 0,90 đến 0,93.

Những item được thiết kế cho một miền đo nào đó đòi hỏi phải có tính đồng nhất. Do vậy độ tin cậy của trắc nghiệm này còn được kiểm tra lại bằng phương pháp tính hệ số tương quan giữa điểm của từng item với điểm của các item còn lại (item - total correlation) để đánh giá mức độ tương quan phù hợp của từng item. Những item có hệ số tương quan lớn hơn 0,30 với các item còn lại của trắc nghiệm (cùng miền đo) được coi là có độ tin cậy phù hợp với phép đo. Kết quả đánh giá độ phù hợp của từng item của thang đo SA được trình bày trong bảng 8.

Kết quả này cho thấy tất cả các item của thang đo SA đều có hệ số tương quan cao hơn so với các item còn lại (r lớn hơn 0,36), do vậy cả 20 item của SA đều có độ tin cậy bảo đảm, đều đóng góp cho độ tin cậy của thang đo.

Bảng 8: Mức độ tin cậy của từng item của thang đo trạng thái lo âu (SA)

Các item của SA (từ 1-20)	Điểm trung bình của thang đo nếu item bị loại	Phương sai của thang đo nếu item bị loại	Tương quan giữa item và các item còn lại	Hệ số α nếu item bị loại
T1	40,3501	96,5262	0,4025	0,8823
T2	40,6283	95,0899	0,4223	0,8820
T3	40,6403	93,3078	0,5502	0,8779

Các item của SA (từ 1-20)	Điểm trung bình của thang đo nếu item bị loại	Phương sai của thang đo nếu item bị loại	Tương quan giữa item và các item còn lại	Hệ số α nếu item bị loại
T4	40,4005	92,1061	0,5586	0,8775
T5	39,7962	92,4031	0,5506	0,8778
T6	41,0048	97,3221	0,3699	0,8832
T7	40,7434	93,4220	0,5052	0,8793
T8	39,6978	93,8700	0,4766	0,8803
T9	41,0288	96,1049	0,4733	0,8805
T10	39,9880	91,5792	0,6042	0,8760
T11	40,1559	93,9492	0,4978	0,8796
T12	40,8106	94,2933	0,5158	0,8791
T13	40,7770	94,8900	0,4759	0,8802
T14	40,6355	95,4245	0,4145	0,8822
T15	39,8177	91,7792	0,6150	0,8757
T16	39,4724	94,3893	0,4539	0,8810
T17	40,5899	91,9877	0,5743	0,8770
T18	40,6882	94,9507	0,4376	0,8815
T19	40,0743	93,2997	0,5393	0,8782
T20	39,9041	91,2792	0,4438	0,8814

Kết quả này cho thấy tất cả các item của thang đo SA đều có hệ số tương quan cao với các item còn lại (r lớn hơn 0,36), do vậy cả 20 item của SA đều có độ tin cậy bảo đảm, đều đóng góp cho độ tin cậy của toàn thang đo.

Kết quả đánh giá độ phù hợp của từng item của thang đo TA được trình bày trong bảng 9.

Bảng 9: Mức độ tin cậy của từng item của thang đo nét lo âu (TA)

Các item của TA (từ 21-40)	Điểm trung bình của thang đo nếu item bị loại	Phương sai của thang đo nếu item bị loại	Tương quan giữa item và các item còn lại	Hệ số α nếu item bị loại
T21	45,0000	63,7566	0,3645	0,8358
T22	45,8333	63,9292	0,3872	0,8351
T23	44,6190	61,6397	0,4620	0,8315
T24	44,4905	65,2720	0,1803	0,8442
T25	45,7214	61,3327	0,5177	0,8292
T26	45,0071	61,6198	0,5235	0,8290
T27	45,1405	61,0041	0,5409	0,8282
T28	45,8405	62,4494	0,4336	0,8329
T29	45,8119	61,1030	0,5277	0,8287
T30	45,1143	60,3735	0,5454	0,8275
T31	45,4119	62,4385	0,4166	0,8336
T32	45,8095	62,9517	0,3688	0,8357
T33	45,4976	61,7971	0,4225	0,8334
T34	45,0262	63,8394	0,2831	0,8397
T35	45,8905	63,4772	0,3299	0,8374
T36	44,7643	62,8170	0,3802	0,8352
T37	45,6262	63,5330	0,3358	0,8371
T38	45,5786	60,7551	0,4768	0,8307
T39	45,0048	61,4940	0,4376	0,8327
T40	45,3833	61,0102	0,4413	0,8326

Kết quả này cho thấy tất cả các item của thang đo TA đều có hệ số tương quan lớn hơn hoặc bằng 0,33 với

các item còn lại, ngoại trừ item T24 có hệ số tương quan thấp ($r = 0,1803$). Như vậy, cả 19 item của thang đo nét lo âu đều có độ tin cậy bảo đảm, đều phù hợp cho phép đo và đều đóng góp cho độ tin cậy của cả thang đo, riêng item T24 có độ tương quan thấp, không đóng góp nhiều cho độ tin cậy của phép đo, do vậy cần được điều chỉnh, viết lại.

8. Đánh giá lại độ hiệu lực của trắc nghiệm

Độ hiệu lực của trắc nghiệm được xác định như là mức độ chính xác mà trắc nghiệm đó đo chính cái cấu trúc nó được thiết kế để đo. Thực chất của các chiến lược đánh giá độ hiệu lực của một trắc nghiệm là điều tra về ý nghĩa điểm số của trắc nghiệm khi đo lường và làm những suy luận có cơ sở khoa học trên cơ sở những điểm số của trắc nghiệm.

Các chiến lược đánh giá lại độ hiệu lực nội dung của phép đo lo âu trong nghiên cứu này là điều tra lại từng mục tiêu đo lường cụ thể của từng item, mức độ đọc hiểu của từng item có phù hợp với nghiệm thể, mức độ mong muốn trả lời của từng nghiệm thể trên từng item. Kết quả cho thấy không có item nào đi chệch khỏi mục tiêu đo lường cụ thể của chính nó, đồng thời cũng không có item nào không phù hợp với mục đích đo lường của trắc nghiệm.

Các chiến lược đánh giá lại độ hiệu lực cấu trúc của phép đo lo âu trong nghiên cứu này là điều tra tương quan giữa điểm số của thang đo trạng thái lo âu và

điểm số của thang đo nét lo âu. Mục đích của việc điều tra tương quan này là tìm hiểu xem liệu cấu trúc của trắc nghiệm có phù hợp với cấu trúc của trạng thái lo âu và nét lo âu (cái nó được thiết kế để đo). Trạng thái lo âu và nét lo âu ở một cá nhân là những mặt có liên quan chặt chẽ với nhau thuộc cùng một cấu trúc (lo âu). Do vậy, một trắc nghiệm đo lo âu có độ hiệu lực tốt gồm hai thang đo cụ thể đánh giá trạng thái lo âu và nét lo âu phải cho hai loạt điểm số có tương quan đáng kể. Kết quả điều tra tương quan giữa hai loạt điểm số của hai thang đo trạng thái lo âu và nét lo âu cho thấy: $r = 0,64$ (mẫu học sinh trung học phổ thông, $N = 417$) và $r = 0,76$ (mẫu sinh viên, $N = 119$) (Xem bảng 10 và bảng 11).

Kết quả này cho thấy phép đo lo âu của Spielberger có cấu trúc tương đồng với đặc tính nó được thiết kế để đo (lo âu). Điều này chứng tỏ rằng phép đo vẫn bảo đảm có độ hiệu lực tốt khi Việt hoá.

Bảng 10: Hệ số tương quan giữa các thang đo
(Mẫu học sinh trung học phổ thông $N = 417$)

Các thang đo	SA	TA
SA	1,00	
TA	0,64**	1,00
Form Y	0,924**	0,88**

** $P < 0,0001$

Bảng 11: Hệ số tương quan giữa các thang đo
(Mẫu sinh viên, N = 119)

Các thang đo	SA	TA
SA	1,00	
TA	0,76**	1,00
Form Y	0,95**	0,93**

**P< 0,0001

Độ hiệu lực cấu trúc của phép đo còn được đánh giá bằng phương pháp phân tích yếu tố nhằm kiểm tra độ hiệu lực của từng item xem chúng có phù hợp với cấu trúc mà nó được thiết kế để đo hay không. Các item cùng đo một cấu trúc đòi hỏi chúng phải có tính đồng nhất, tức là chúng được mong đợi có quan hệ đáng kể với một yếu tố trội khi dùng phương pháp phân tích yếu tố để chiết xuất yếu tố.

Điều kiện để chiết xuất yếu tố là điểm số của item phải có quan hệ đáng kể trong ma trận tương quan (có một số lượng đáng kể các tương quan lớn hơn 0,30); hệ số của phép đo KMO (Kaiser - Meyer - Olkin Measure) lớn hơn 0,60; và phép thử Bartlett phải ở mức có ý nghĩa (đây là hai phép thử cho biết khả năng chiết xuất yếu tố của ma trận tương quan).

Kết quả phân tích điểm số các item của thang đo trạng thái lo âu được trình bày trong bảng 12 và bảng 13. Bảng 12 cho biết, trong ma trận này có một số lượng

đáng kể các item có quan hệ ở mức lớn hơn 0,30, do vậy thích hợp cho việc sử dụng ma trận để phân tích yếu tố. Bảng 13 cho biết, hệ số KMO = 0,91, phép thử Bartlett ở mức có ý nghĩa ($P < 0,0001$), do vậy ma trận này có khả năng chiết xuất thành các yếu tố. Kết quả sau khi chiết xuất yếu tố cho biết, các item của thang đo trạng thái lo âu được phân tích thành bốn yếu tố, nhưng chỉ có yếu tố 1 nổi trội với giá trị riêng khá lớn (Eigenvalue = 6,34). Giải thích cho 31,7% độ biến thiên của toàn phép đo (các yếu tố còn lại có giá trị riêng nhỏ hơn 2). Tất cả 20 item đều có quan hệ đáng kể với yếu tố. Hệ số chứa yếu tố (factor loading) trên yếu tố này của tất cả 20 item đều lớn hơn hoặc bằng 0,40, điều này có nghĩa là các item có tính đồng nhất và đều phù hợp với cấu trúc của phép đo.

Kết quả phân tích điểm số của item thang đo nét lo âu được trình bày trong bảng 14 và bảng 15. Bảng 14 và bảng 15 cho biết ma trận tương quan này có một số các item có quan hệ ở mức lớn hơn 0,30, hệ số KMO = 0,86 và phép thử Bartlett cho biết mức ý nghĩa $P < 0,0001$. Do vậy ma trận này thích hợp cho phép phân tích yếu tố và có khả năng chiết xuất thành các yếu tố.

Kết quả sau khi chiết xuất cho biết có tất cả 5 yếu tố nhưng cũng chỉ có 1 yếu tố nổi trội với giá trị riêng bằng 5,16 giải thích cho 25,8% mức biến thiên của toàn phép đo (các item còn lại cùng có giá trị riêng nhỏ hơn 2).

Bảng 13: Phân tích yếu tố (thang SA)

<i>Item</i>	<i>Hệ số chung</i>	<i>Yếu tố</i>	<i>Hệ số riêng</i>	<i>Phần trăm</i>	<i>Phần trăm cộng dồn</i>
T1	.27199 *	1	6.34424	31,7	31,7
T2	.28030 *	2	1.85995	9,3	41,0
T3	.38473 *	3	1.22606	6,1	47,2
T4	.42725 *	4	1.02191	5,1	52,3
T5	.42780 *	5	.86390	4,3	56,6
T6	.22306 *	6	.85495	4,3	60,9
T7	.33258 *	7	.75703	3,8	64,6
T8	.36985 *	8	.72544	3,6	68,3
T9	.32242 *	9	.71518	3,6	71,8
T10	.50702 *	10	.68979	3,4	75,3
T11	.35732 *	11	.66043	3,3	78,6
T12	.38108 *	12	.61700	3,1	81,7
T13	.32209 *	13	.57157	2,9	84,5
T14	.25239 *	14	.54097	2,7	87,2
T15	.47288 *	15	.49123	2,5	89,7
T16	.31153 *	16	.46242	2,3	92,0
T17	.41689 *	17	.44464	2,2	94,4
T18	.26614 *	18	.43559	2,2	96,4
T19	.38874 *	19	.37075	1,9	98,3
T20	.38016 *	20	.34695	1,7	100,0

Ma trận yếu tố

Item	Yếu tố 1	Yếu tố 2	Yếu tố 3	Yếu tố 4
T15	.66279	-.20634		
T10	.65879	-.33835		
T17	.61825	.27036		
T5	.60221	-.22850		-.32932
T4	.60074	.24979		
T3	.58890			
T19	.57085	-.21007		
T12	.55862	.27656	-.23534	
T11	.53815		.34906	.20263
T7	.53784	.26954		
T8	.51641	-.36824		
T13	.51285	.29386		
T9	.50387	.24132		
T20	.48861	-.37082		
T16	.48587	-.31311		
T18	.46656			.21361
T2	.45283		.34050	
T14	.44066	.25001		
T1	.43868		.38865	-.23259
T6	.39731	.25644		

Bảng 15: phân tích yếu tố (thang TA)

<i>Item</i>	<i>Hệ số chung</i>	<i>Yếu tố</i>	<i>Hệ số riêng</i>	<i>Phần trăm</i>	<i>Phần trăm cộng dồn</i>
T21	.31060 *	1	5.15707	25,8	25,8
T22	.18238 *	2	1.86337	9,3	35,1
T23	.37635 *	3	1.13301	5,7	40,8
T24	.08314 *	4	1.04929	5,2	46,0
T25	.34341 *	5	1.04293	5,2	51,2
T26	.41191 *	6	.99542	5,0	56,2
T27	.36410 *	7	.87440	4,4	60,6
T28	.27957 *	8	.83348	4,2	64,7
T29	.36098 *	9	.78992	3,9	68,7
T30	.42824 *	10	.76549	3,8	72,5
T31	.29093 *	11	.70546	3,5	76,0
T32	.22985 *	12	.69163	3,5	79,5
T33	.24001 *	13	.63976	3,2	82,7
T34	.18597 *	14	.62282	3,1	85,8
T35	.20838 *	15	.60440	3,0	88,8
T36	.34065 *	16	.51732	2,6	91,4
T37	.22234 *	17	.49704	2,5	93,9
T38	.36673 *	18	.45854	2,3	96,2
T39	.30357 *	19	.40677	2,0	98,2
T40	.26393 *	20	.35187	1,8	100,0

Ma trận yếu tố

Item	Yếu tố 1	Yếu tố 2	Yếu tố 3	Yếu tố 4	Yếu tố 5
T30	.62695	-.24501	-.20436		
T26	.60196	-.28422	-.28532		
T27	.59399				
T25	.56778				.20545
T29	.56767	.24095			
T23	.56348	-.35297	.46316	-.25775	
T38	.51440	.36603			
T39	.50840			.47697	
T40	.48022				
T28	.47331				
T33	.46606				
T31	.46106	.32110			-.22986
T36	.43874	-.39078			
T21	.42122	-.39174			
T22	.40875				
T32	.39870	.28452			.22626
T37	.36449	.26138			-.26699
T35	.36247	.25131	.26517		.28311
T34	.30119			.21873	
T24	.18378				

Tất cả 20 item của thang đo nét lo âu đều có quan hệ đủ lớn với yếu tố 1. Hệ số chứa của tất cả các item này đều lớn hơn 0,30, ngoại trừ item T24 có quan hệ yếu ($r < 0,20$).

Như vậy, có 19 item có quan hệ tuyến tính, đồng nhất với phép đo, do vậy chúng có độ hiệu lực phù hợp với cấu trúc của phép đo, riêng item T24 không phù hợp cần phải chỉnh sửa hoặc thay thế.

Để đánh giá độ hiệu lực tiêu chuẩn, nghiên cứu này so sánh tương quan điểm của các thang đo trạng thái lo âu, nét lo âu và toàn phép đo (Form Y) với thang đo lo âu của Beck (BA). Đây là những phép đo được thiết kế khác nhau nhưng cùng đo một cấu trúc: lo âu, do vậy tương quan mong đợi là các loại điểm của các thang đo này sẽ tương quan đáng kể với nhau.

Kết quả phân tích tương quan trình bày trong bảng 16 và bảng 17. Kết quả này cho thấy cả SA và TA đều có quan hệ đáng kể với BA ở mức từ 0,53 đến 0,56 (mức ý nghĩa $P < 0,0001$) tương quan điểm của toàn phép đo (Form Y) với BA từ 0,58, đến 0,60 (mức ý nghĩa $P < 0,0001$)

Bảng 16: Tương quan điểm giữa các thang đo lo âu
(Mẫu học sinh trung học phổ thông, N = 417)

Các thang đo	SA	TA	FormY	BA
SA	1,00			
TA	0,64**	1,00		
Form Y	0,924**	0,88**	1,00	
BA	0,54**	0,56**	0,60**	1,00

** $P < 0,0001$

Bảng 17: Tương quan điểm giữa các thang đo lo âu
(Mẫu sinh viên, N = 119)

Các thang đo	SA	TA	FormY	BA
SA	1,00			
TA	0,76**	1,00		
Form Y	0,95**	0,93**	1,00	
BA	0,55**	0,53**	0,58**	1,00

** $P < 0,0001$

Tóm lại, kết quả nghiên cứu cho thấy sau khi Việt hoá thang đo lo âu STAI của Spielberger vẫn giữ được độ tin cậy và độ hiệu lực đủ cao theo tiêu chuẩn của một trắc nghiệm đánh giá lâm sàng (những trắc nghiệm dùng chẩn đoán lâm sàng yêu cầu phải có độ tin cậy lớn hơn 0,85). Tuy nhiên độ tin cậy của thang đo nét lo âu có giảm một chút so với mẫu chuẩn. Có ít nhất 1 item T24 sau khi Việt hoá không còn giữ được độ tin cậy và độ phù hợp do vậy cần phải chỉnh sửa. Điều này cần phải tiếp tục kiểm tra lại trên mẫu lớn hơn và mẫu đại diện hơn.

Chương XI

THỰC HÀNH CÁC KỸ NĂNG XÂY DỰNG TRẮC NGHIỆM ĐÁNH GIÁ KIẾN THỨC KỸ NĂNG

Dưới đây là ví dụ xây dựng một trắc nghiệm nhằm đánh giá kết quả học môn Tâm lý trị liệu (4 đơn vị học trình = 60 tiết) dành cho sinh viên năm thứ ba Khoa Tâm lý, Trường Đại học Khoa học xã hội và nhân văn thuộc Đại học Quốc gia Hà Nội (kết quả đánh giá bằng trắc nghiệm này được kết hợp với các hình thức đánh giá khác như đánh giá qua xêmina, thi viết, thi vấn đáp).

1. Mục tiêu của trắc nghiệm

- Đánh giá sự hiểu biết, nắm vững và làm chủ khái niệm;
- Khả năng vận dụng khái niệm trong những tình huống lâm sàng.

2. Nội dung đánh giá

Môn học này gồm 7 chương, nội dung của trắc nghiệm bao quát kiến thức, kỹ năng của 6 chương (một chương không nằm trong phạm vi đánh giá của trắc nghiệm) gồm những đơn vị kiến thức cơ bản nằm trong

các câu hỏi sau:

- Rối nhiễu tâm lý là gì?
- Bản chất của tâm bệnh? các dấu hiệu cơ bản của tâm bệnh?
- Khái niệm trị liệu tâm lý?
- Mục tiêu, nhiệm vụ của trị liệu tâm lý? Sự khác biệt cơ bản giữa nhà tâm lý lâm sàng / bác sĩ tâm thần?
- Những cách tiếp cận chính trong trị liệu tâm lý?
- Trị liệu tâm lý theo cách tiếp cận phân tâm?
- Bản chất của rối nhiễu theo cách hiểu của trường phái Phân tâm, nhiệm vụ của nhà trị liệu phân tâm?
- Trị liệu tâm lý theo cách tiếp cận nhân văn - hiện sinh.
- Bản chất của rối nhiễu theo cách hiểu của các nhà trị liệu nhân văn - hiện sinh, nhiệm vụ của nhà trị liệu nhân văn - hiện sinh?
- Trị liệu tâm lý theo cách tiếp cận hành vi - nhận thức.
- Bản chất của rối nhiễu theo cách hiểu của các nhà trị liệu hành vi - nhận thức, nhiệm vụ của nhà trị liệu hành vi - nhận thức?
- Trước một ca rối nhiễu lâm sàng (chẳng hạn, trầm cảm hoặc có hành vi mất tay), các cách tiếp cận trị liệu khác nhau giải thích thế nào về bản chất của rối nhiễu, những định hướng ưu tiên của từng cách tiếp cận trong chiến lược can thiệp bằng trị liệu tâm lý?

3. Các bước của quy trình thiết kế trắc nghiệm.

Bước thứ nhất: Sau khi đã xác định rõ mục tiêu, nội dung đo lường của trắc nghiệm, người thiết kế lập một danh sách các đơn vị kiến thức, kỹ năng cơ bản cần được đánh giá dưới dạng các câu hỏi (như đã liệt kê ở trên).

Bước thứ hai: Lập một đề cương gồm các kiến thức, kỹ năng cơ bản dùng để trả lời các câu hỏi này dưới dạng những câu trả lời đúng nhất/tốt nhất, gần đúng.

Bước thứ ba: Lựa chọn kiểu item cho trắc nghiệm. Những hình thức item thích hợp cho trắc nghiệm đánh giá kiến thức thường dùng là: câu hỏi đúng/sai, câu hỏi có nhiều lựa chọn, câu hỏi điền thế/điền chỗ trống, câu hỏi trả lời ngắn, câu hỏi tự luận, tình huống/các phương án giải quyết tình huống...

Bước thứ tư: Viết các item dựa theo nội dung các câu hỏi và câu trả lời, những phần trọng tâm nên có nhiều item hơn. Những câu hỏi có dạng đúng/sai nên hạn chế vì xác suất ngẫu nhiên chọn đúng khá cao (50%). Những câu hỏi mở (tự luận) cũng nên hạn chế, nên kết hợp vài kiểu item trong một trắc nghiệm để đa dạng hoá các hình thức của item. Các item trong trắc nghiệm nên sắp xếp theo trật tự từ dễ đến khó, các câu hỏi khó, câu hỏi mở nên để ở cuối.

Bước thứ năm: Chọn lựa những item tốt nhất/ quan trọng nhất vào trắc nghiệm, sao cho số lượng các item đủ bao quát khắp các đơn vị kiến thức cần đánh giá hoặc chỉ ít cũng là những kiến thức cơ bản nhất, những phần trọng tâm nhất. Xác định thời gian tối đa và tối

thiếu cho từng item và cho cả trắc nghiệm (bằng cách cho làm thử hoặc đặt mình vào vị trí của nghiệm thể để làm thử). Một trắc nghiệm đánh giá kiến thức, kỹ năng nên trong khoảng thời gian từ 40 - 80 phút là thích hợp, tùy theo mục đích kiểm tra đánh giá mà xác định số lượng câu hỏi, thời gian thích hợp, tuy nhiên không nên quá ngắn, làm dưới 20 phút và cũng không nên quá dài, làm trên 90 phút.

Bước thứ sáu: Xây dựng hệ thống điểm cho từng item và cho toàn bộ trắc nghiệm, nên sử dụng trọng số khi cho điểm, tức là những item quan trọng, item khó (những item đánh giá khả năng vận dụng, sáng tạo lại khái niệm) nên có số điểm cao hơn. Trong từng item nên sử dụng kiểu cho điểm tối đa (nếu trả lời đúng nhất/ tốt nhất) và điểm tối thiểu (nếu trả lời ngắn đúng nhất/ chấp nhận được). Điểm số của trắc nghiệm ở các phần nên có sự cân xứng với đơn vị kiến thức, kỹ năng (tương ứng với số tiết/ thời gian quy định trong chương trình cho từng chương).

Bước thứ bảy: Cần thống nhất cách hướng dẫn làm trắc nghiệm, dự kiến cách giải thích thống nhất các khái niệm...

Dưới đây là trắc nghiệm đã thiết kế theo quy trình trên:

FORM A

Trắc nghiệm kiểm tra kiến thức môn Tâm lý trị liệu
Họ và tên..... nam/nữ..... năm sinh.... lớp....

Hãy đọc kỹ các câu hỏi dưới đây, chọn lựa câu trả lời **đúng nhất** (bằng cách khoanh tròn **một** chữ cái thích hợp hoặc ghi **một** chữ cái vào ô trống của câu hỏi hoặc viết câu trả lời ngắn gọn vào khoảng trống)

1/ Rối nhiễu tâm lý là:

- a/ Những triệu chứng bất thường về tâm lý có ý nghĩa về mặt lâm sàng
- b/ Những bất thường về tâm lý làm mất năng lực vốn có của cá nhân
- c/ Những triệu chứng bất thường về tâm lý biểu hiện sự suy thoái chức năng sinh - tâm lý của cá nhân
- d/ Ý kiến khác (ghi cụ thể)

2/ Mục tiêu của trị liệu tâm lý là:

- a/ Làm giảm hoặc mất triệu chứng rối nhiễu ở thân chủ
- b/ Điều chỉnh lại những mối quan hệ nhân cách bị rối nhiễu
- c/ Phát triển các kỹ năng ứng phó, giải quyết vấn đề, tạo khả năng thích nghi tốt hơn ở thân chủ
- d/ Ý kiến khác (ghi cụ thể)

3/ Cách tiếp cận trị liệu nào sau đây:

- a/ Phân tâm
- b/ Nhân văn - hiện sinh
- c/ Nhận thức hành vi

..... cho rằng nguyên nhân của rối nhiễu tâm lý là do thiếu vắng những mối quan hệ tình người có ý nghĩa,

hoặc thiếu vắng những ý tưởng, mục tiêu quan trọng để phấn đấu.

..... cho rằng nguyên nhân của rối nhiễu tâm lý là do những xung đột, mâu thuẫn giữa những yêu cầu, ham muốn của cuộc sống bản năng với sự chống cự xuất hiện bên trong con người chống lại những yêu cầu đó.

..... cho rằng nguyên nhân của rối nhiễu tâm lý là do môi trường sống không thuận lợi và do những thiếu hụt bên trong cá nhân.

..... cho rằng rối nhiễu tâm lý nào đó tồn tại là do có những điều kiện hiện tại đang duy trì sự rối nhiễu này.

..... cho rằng rối nhiễu tâm lý nào đó xảy ra là do có những xung đột xảy ra trước đó vẫn không được giải quyết.

..... cho rằng rối nhiễu tâm lý nào đó xảy ra là do thiếu sự nhận thức đầy đủ về bản thân, thiếu lòng tự tin và không chấp nhận chính mình.

4/ Mục tiêu chủ yếu của cách tiếp cận trị liệu nào sau đây:

a/ Phân tâm

b/ Nhân văn - hiện sinh

c/ Nhận thức - hành vi

..... không phải là chữa trị cho thân thủ hoặc tìm kiếm nguyên nhân từ quá khứ... mà cái chính là khuyến khích thân chủ tự giúp mình, tự hiện thực hoá những tiềm năng của bản thân.

..... là can thiệp tích cực để làm giảm hoặc bỏ rối

hiều bằng cách thay đổi các điều kiện đang duy trì rồi nhiều.

..... là bộc lộ vô thức, giúp thân chủ được giải thoát khỏi những dồn nén (xung đột) được thiết lập từ tuổi ấu thơ.

..... là giác ngộ vô thức bằng cách hướng thân chủ tới sự hiểu rõ mối liên hệ giữa triệu chứng hiện tại và nguồn gốc quá khứ.

..... là tự biết mình, phát triển kiểu trị liệu tự giúp mình, tự khuyến khích, và tự hiện thực hoá những tiềm năng của bản thân

..... là nhằm vào môi trường, nhằm vào các yếu tố hiện tại, nhằm trực tiếp vào những hành vi kém thích nghi để tìm cách biến đổi chúng.

5/ Nhiệm vụ trọng tâm của nhà trị liệu theo trường phái nào sau đây:

a/ Phân tâm

b/ Nhân văn - hiện sinh

c/ Nhận thức - hành vi

..... là nhận diện những tác nhân, điều kiện đang duy trì rồi nhiều và tìm cách loại bỏ hoặc thay đổi chúng.

..... là tạo môi trường thuận lợi để thân chủ học cách bày tỏ những tình cảm khó xử, là tập trung vào những trải nghiệm thực tại, nhận biết những xúc cảm, thái độ, hành vi của mình xuất hiện như thế nào để giúp thân chủ tự chấp nhận mình.

..... là kiên trì lắng nghe tất cả những gì thân

chủ bộc lộ, rồi sử dụng sự liên tưởng này để hiểu bằng cách nào người bệnh đã sử dụng quá trình dồn nén để chế ngự, giải quyết xung đột.

..... là giúp thân chủ học các kỹ năng tự điều chỉnh, kỹ năng ứng phó giải quyết vấn đề, phát triển các mẫu ứng xử mới thích nghi hơn.

..... là khuyến khích thân chủ bộc lộ những cảm giác mạnh, khích lệ thân chủ trò chuyện cởi mở về những xúc cảm bị dồn nén để tái thiết lại cảm xúc lành mạnh.

..... là giúp thân chủ dỡ bỏ những "rào cản về tâm lý" đang hạn chế sự tích cực vốn có, giúp họ hiểu rõ bản thân và chấp nhận tình cảm riêng của mình.

..... là giúp thân chủ tái thiết lại ký ức bị dồn nén lâu ngày, từ từ trải nghiệm lại những cảm giác đau khổ, căng thẳng để hướng tới sự giác ngộ.

..... là giúp đỡ nhiệt tình vô điều kiện, không áp đặt, không đánh giá năng lực và tạo điều kiện để thân chủ phát hiện ra cái bản sắc cá nhân của chính họ.

..... là tìm cách nhận diện những ý nghĩ, niềm tin không hợp lý, thay thế chúng bằng những ý nghĩ, niềm tin hợp lý hơn, là tìm cách thay đổi hậu quả của hành vi theo hướng củng cố những hành vi thích nghi để giảm những hành vi kém thích nghi.

6/ Hãy đặt tên phù hợp cho từng cách tiếp cận trị liệu sau đây (vẽ mũi tên chỉ sự thích hợp)

a/ Phân tâm..... còn gọi là cách trị liệu tự quan sát nội tâm

b/ Nhân văn - hiện sinh..... còn gọi là cách trị liệu bằng hành động

c/ Nhận thức - hành vi..... còn gọi là cách trị liệu bằng trò chuyện

7/ Bản chất của tâm bệnh lý là

a/ Những rối nhiễu tâm lý hay rối loạn tâm thần dai dẳng làm thân chủ khó kiểm soát.

b/ Những rối nhiễu về xúc cảm, nhận thức hay hành vi làm thân chủ mất khả năng đạt được những mục tiêu quan trọng trong cuộc sống của bản thân.

c/ Là một "liên thể" những rối nhiễu tâm lý hay rối loạn tâm thần từ nhẹ đến nặng.

d/ Ý kiến khác (ghi cụ thể).....

8/ Những dấu hiệu bản chất để nhận biết những hành vi bất thường ở mức tâm bệnh là gì?

.....

.....

.....

.....

.....

.....

9/ Mối quan hệ giữa nhà trị liệu và người có rối nhiễu là mối quan hệ lâm sàng trên chủ thể, tức là

.....

10/ Điểm khác nhau cơ bản giữa nhà trị liệu tâm lý với các bác sĩ tâm thần:

a/ Không kê đơn thuốc

- b/ Thái độ lâm sàng và cách tiếp cận lâm sàng
- c/ Hiểu ý nghĩa (cấu trúc, cơ chế tâm lý) đằng sau các triệu chứng
- d/ Ý kiến khác (ghì cụ thể).....

11/ Cách tiếp cận nhận thức - hành vi xem rối nhiễu tâm lý ở một cá nhân là do những tác nhân nào sau đây đang cản trở (làm thay đổi các chức năng hoạt động bình thường của cá nhân đó).

- a/ Điều kiện môi trường hiện tại không thuận lợi
- b/ Những nhận thức không phù hợp của cá nhân
- c/ Tập nhiễm những mẫu ứng xử kém thích nghi
- d/ Cả ba yếu tố trên

12/ Nhiệm vụ trọng tâm của trị liệu..... là nhận diện bản chất của rối nhiễu, những tác nhân phát sinh và duy trì rối nhiễu và tìm cách loại bỏ chúng:

- a/ Phân tâm
- b/ Nhận thức - hành vi
- c/ Nhân văn - hiện sinh
- d/ Cả ba cách tiếp cận trên

13/ Những dấu hiệu nào sau đây chỉ bản chất của trị liệu hành vi - nhận thức:

- a/ Tập trung vào hiện tại
- b/ Tiếp cận bằng hành động nhiều hơn là lời nói
- c/ Thân chủ tham gia tích cực, học các kỹ năng để biến trị liệu thành tự trị liệu
- d/ Cả ba dấu hiệu trên

14/ Một đứa trẻ mút tay, nhà trị liệu theo trường phái:

- a/ Phân tâm quan tâm đến.....

.....
b/ Nhân văn - hiện sinh quan tâm đến

.....
c/ Hành vi (hiện đại) quan tâm đến

.....
d/ Nhận thức quan tâm đến

.....
15/ Để lý giải tại sao một đứa trẻ bị trầm cảm, nhà trị liệu tâm lý theo trường phái:

a/ Phân tâm cho rằng

.....
a/ Nhân văn - hiện sinh cho rằng

.....
c/ Hành vi (hiện đại) cho rằng

.....
d/ Nhận thức cho rằng

.....
16/ Để định hướng ưu tiên trị liệu rồi nhiều trầm cảm, nhiệm vụ trọng tâm của nhà trị liệu tâm lý theo trường phái

a/ Phân tâm là

.....
b/ Nhân văn - hiện sinh là

.....
c/ Hành vi (hiện đại) là

.....
d/ Nhận thức là

Cách cho điểm:

- Câu 1: trả lời đúng nhất được 4 điểm (câu trả lời gần đúng nhất 1 điểm)
- Câu 2: trả lời đúng nhất được 4 điểm (câu trả lời gần đúng nhất 1 điểm)
- Câu 3: 6 điểm, mỗi ý trả lời đúng nhất 1 điểm
- Câu 4: 6 điểm, mỗi ý trả lời đúng nhất 1 điểm
- Câu 5: 9 điểm, mỗi ý trả lời đúng nhất 1 điểm
- Câu 6: 3 điểm, mỗi ý trả lời đúng nhất 1 điểm
- Câu 7: trả lời đúng nhất được 4 điểm, (câu trả lời gần đúng nhất 1 điểm)
- Câu 8: 6 điểm, trả lời đúng nhất mỗi ý được 1 điểm
- Câu 9: 3 điểm, trả lời đúng nhất mỗi ý được 1 điểm
- Câu 10 trả lời đúng nhất được 4 điểm (câu trả lời gần đúng nhất 1 điểm)
- Câu 11: trả lời đúng nhất được 3 điểm (câu trả lời gần đúng nhất 1 điểm)
- Câu 12: trả lời đúng nhất được 3 điểm (câu trả lời gần đúng nhất 1 điểm)
- Câu 13: trả lời đúng nhất được 3 điểm (câu trả lời gần đúng nhất 1 điểm)
- Câu 14: 8 điểm, trả lời đúng mỗi ý được 2 điểm (ý trả lời gần đúng nhất 1 điểm)
- Câu 15: 8 điểm, trả lời đúng mỗi ý được 2 điểm (ý trả lời gần đúng nhất 1 điểm)
- Câu 16: 8 điểm, trả lời đúng mỗi ý được 2 điểm (ý trả lời gần đúng nhất 1 điểm)
- Tổng điểm của các trắc nghiệm là 82 điểm. Điểm

này sau đó được quy đổi sang thang điểm 10 (dựa theo điểm trung bình và độ lệch chuẩn) cụ thể như sau:

- Từ 79 - 82 tương đương 10 điểm.
- Từ 74 - 78 tương đương 9 điểm
- Từ 69 - 73 tương đương 8 điểm
- Từ 61 - 68 tương đương 7 điểm
- Từ 53 - 60 tương đương 6 điểm
- Từ 41 - 52 tương đương 5 điểm
- Từ 29 - 40 tương đương 4 điểm
- Từ 16 - 28 tương đương 3 điểm
- Dưới 16 tương đương 2 điểm

Thời gian làm trắc nghiệm: 65 phút.

Tuỳ theo mục tiêu đánh giá, trắc nghiệm này có thể được chia thành hai tiểu trắc nghiệm, mỗi tiểu trắc nghiệm làm hết khoảng 30 - 35 phút. Đây là trắc nghiệm đánh giá năng lực, kiểu tiêu chí dùng đánh giá kiến thức môn học tại lớp, khoá học, do vậy các item (câu hỏi) có độ khó khác nhau (từ dễ đến khó), trong đó có những item rất khó đòi hỏi khả năng vận dụng sáng tạo. Trắc nghiệm này đã được thử nghiệm trên ba khoá sinh viên. Kết quả điểm trắc nghiệm có tương quan đáng kể với các hình thức đánh giá kiến thức khác (thi viết, thi vấn đáp, xêmina). Điểm trắc nghiệm có khác biệt đáng kể giữa hai lần đánh giá: trước khoá học và sau khoá học, nhưng không có sự khác biệt đáng kể giữa nam/nữ, không khác biệt giữa hai lần đo liên tiếp trên cùng một nhóm sinh viên. Những thông tin vừa nêu rất quan trọng vì chúng cung cấp thông tin về các đặc tính đo lường của trắc nghiệm (độ khó, độ phân biệt, độ tin cậy, độ hiệu lực).

Các nhóm đối tượng:

- Học sinh lớp 5 - 9 (4 test)
- Học sinh lớp 10 - 12 (4 test)
- Sinh viên (4 test)
- Lao động trẻ (3 test)

Các công đoạn xử lý: (sử dụng phần mềm thống kê SPSS, phiên bản 11.0)

- 1 - Lập bộ mã để nhập số liệu
- 2 - Chấm điểm thô các trắc nghiệm
- 3 - Nhập phiếu
- 4 - Làm sạch số liệu
- 5 - Xử lý tình:

a/ Đánh giá độ tin cậy, độ hiệu lực của từng item, từng trắc nghiệm.

b/ Xác định các mối quan hệ giữa các chỉ số trí tuệ: IQ, EQ, CQ.

c/ Xác định những điểm mạnh/thiếu hụt về trí tuệ ở từng nhóm đối tượng trên số liệu điều tra.

d/ Phát hiện "quy luật" phát triển trí tuệ, so sánh, tìm mối quan hệ giữa sự phát triển trí tuệ và nhân cách (ở một số nhóm khách thể chung).

e/ Kiểm tra ảnh hưởng của các yếu tố:

- Giới - Vùng/ miền - Gia đình (nghề bố/mẹ)
- Tuổi - Khu vực - Thành tích học tập

g/ Tìm những ca điển hình để phỏng vấn sâu.

Phụ lục 2
CÁC CÔNG ĐOẠN XỬ LÝ SỐ LIỆU
(Đề tài KX-05-06)

1. Xử lý thô: chấm điểm thô các trắc nghiệm (xem bản hướng dẫn)

2. Lập mã

- Lập mã cho từng đối tượng điều tra (theo số thứ tự)
- Lập mã cho từng biến theo từng item của mỗi trắc nghiệm theo 4 nhóm mẫu điều tra: học sinh lớp 5 - 9, học sinh lớp 10 - 12, sinh viên và lao động trẻ.

3. Nhập số liệu vào máy

4. Làm sạch số liệu

Làm sạch số liệu trước khi xử lý tính 7 file số liệu trắc nghiệm

Gồm các công việc cụ thể sau:

- Làm sạch các phiếu trả lời ầu, phát hiện các phiếu trả lời mâu thuẫn của 7 trắc nghiệm để loại bỏ.
- Kiểm tra số liệu vào máy theo xác suất (10%) để phát hiện lỗi khi nhập số liệu.
- Điều chỉnh các số vào sai để chính xác hoá các file số liệu của 7 loại trắc nghiệm sau khi nhập thô.
- Sử dụng các phép kiểm định thống kê mô tả để

xử lý điểm thô nhằm phát hiện các điểm số không thích hợp (cực trị/ngoại lai) để loại bỏ số liệu sẽ phân tích. Chuyển số liệu về dạng phân bố điểm chuẩn để có thể áp dụng các phép kiểm định thống kê đặc trưng khi xử lý tình các file số liệu theo từng nhóm mẫu: học sinh lớp 5 - 9, học sinh lớp 10 - 12, sinh viên và lao động trẻ để có thể tính được các chỉ số trí tuệ: IQ, EQ, CQ khi xử lý tình.

- Hoà các file số liệu của từng trắc nghiệm thành file số liệu tổng hợp để có thể xử lý liên biến nhằm phát hiện mức độ quan hệ/ảnh hưởng của từng biến đến sự phát triển trí tuệ của các nhóm đối tượng điều tra.

5. Xử lý tình

- ❖ IQ Munzert - (TE-NL)
- ❖ IQ Sternberg - TE
- ❖ IQ Sternberg - NL
- ❖ EQ - i BarOn - TE
- ❖ EQ MSCEIT - NL
- ❖ CQ TSD-Z
- ❖ CQ Bremner

Gồm các công việc cụ thể:

- Phân tích độ tin cậy của từng item trong mỗi trắc nghiệm, độ tin cậy của từng tiểu thang đo trong mỗi trắc nghiệm, độ tin cậy chung của từng bộ trắc nghiệm trên 4 nhóm mẫu: học sinh lớp 5 - 9,

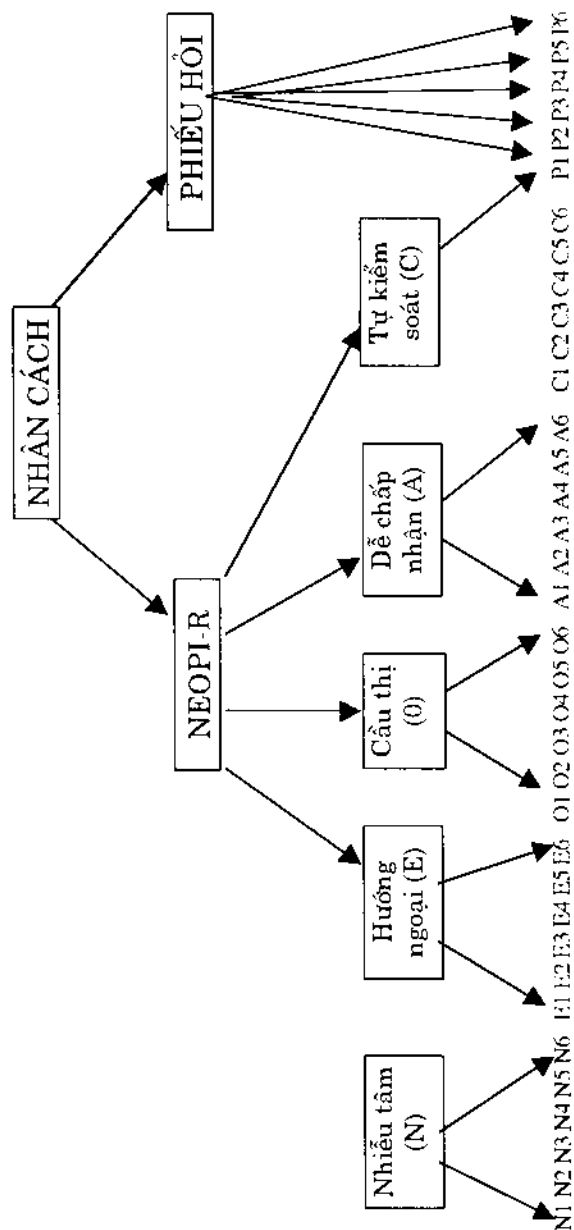
học sinh lớp 10 - 12, sinh viên và lao động trẻ (công nhân, nông dân, giáo viên, tiểu thủ công, trí thức)

- Đánh giá độ khó, độ phân biệt của từng item trong từng bộ trắc nghiệm để xác định các thông số đo lường phù hợp ở mức độ nào của từng trắc nghiệm.
- Đánh giá độ hiệu lực: Sử dụng phương pháp phân tích yếu tố để đánh giá độ hiệu lực cấu trúc, độ hiệu lực nội dung, độ hiệu lực dự báo của 7 bộ trắc nghiệm.
- Đánh giá tính chuẩn của phân bố điểm của 7 trắc nghiệm (xác định mức độ thoả mãn giả thiết phân bố điểm của 7 trắc nghiệm phải là phân bố chuẩn - điều kiện cần và đủ khi muốn áp dụng các phép toán kiểm định thống kê suy luận) để xác định mức độ phù hợp khi tính tương quan điểm số ở từng cá nhân trên từng tiểu thang đo, trên từng trắc nghiệm.
- Tính điểm IQ, EQ, CQ cho từng cá nhân, tính điểm trung bình, độ lệch chuẩn, sai số ngẫu nhiên ở 7 bộ trắc nghiệm theo từng nhóm đối tượng điều tra.
- Thành lập các bảng điểm chuẩn theo lứa tuổi, theo vùng, theo khu vực đối chiếu với bảng điểm chuẩn quốc tế để phát hiện những khác biệt đặc trưng.
- Phát hiện các điểm mạnh, điểm yếu/thiếu hụt ở

từng chỉ số trí tuệ: IQ, EQ, CQ trên từng nhóm đối tượng điều tra.

- So sánh sự khác biệt về các chỉ số trí tuệ trên 7 trắc nghiệm theo giới, tuổi, vùng, khu vực.
- Lập ma trận tương quan hợp biến và đánh giá mức độ tác động của từng yếu tố ảnh hưởng (gia đình, nhà trường...).

MÔ HÌNH XỬ LÝ DỮ LIỆU
(Đề tài KX-05-07)



Các nhóm đối tượng:

- Học sinh
- Sinh viên
- Lao động trẻ (công nhân, nông dân, giáo viên)

Các bước xử lý:

1 - Lập bộ mã để nhập số liệu

2 - Xử lý các câu hỏi mở

3 - Nhập phiếu

4 - Làm sạch số liệu

5 - Xử lý tình:

a/ Đánh giá độ tin cậy, độ hiệu lực của trắc nghiệm

b/ Xác định các mối quan hệ giữa các mặt của nhân cách

c/ Xác định những điểm mạnh/thiếu hụt về nhân cách ở từng nhóm đối tượng trên số liệu điều tra

d/ So sánh, tìm mối quan hệ giữa sự phát triển nhân cách và phát triển trí tuệ ở một số nhóm khách thể chung

e/ Kiểm tra ảnh hưởng của các yếu tố:

- Giới - Vùng/miền - Gia đình (nghệ bố/mẹ)

- Tuổi - Khu vực - Thành tích học tập

g/ Tìm những ca điển hình để phỏng vấn sâu

Phụ lục 3

PHIẾU HỎI

Lời hướng dẫn:

Dưới đây là 127 câu hỏi. Chúng sẽ giúp bạn xác định một số những đặc điểm của cá nhân. Các câu hỏi này không nhằm đánh giá bạn trả lời "đúng" "sai" mà chỉ tìm hiểu ý kiến của bạn thôi. Vì vậy mong bạn hãy trả lời *thật đúng như mình nghĩ*.

Trước hết, bạn hãy cung cấp các thông tin cần thiết về bản thân vào phần trên của tờ phiếu trả lời. Sau đó, bạn hãy đọc kỹ và trả lời từng câu hỏi như sau:

- Từ câu 1 - 120, mỗi câu có 3 hoặc 4 đáp án (a, b, c hoặc a, b, c, d) để bạn chọn lựa. Bạn hãy chọn lấy *một* đáp án nào phù hợp nhất với mình và đánh dấu "X" vào cột tương ứng trên trang 1 của tờ phiếu trả lời.
- Từ câu 121 - 124 là những câu hỏi mở hoặc câu hỏi được lựa chọn nhiều đáp án. Vì vậy bạn hãy điền hoặc chọn cho mình *những* câu trả lời nào phù hợp nhất vào *trang 2* của tờ phiếu trả lời.
- Từ câu 125 - 127, mỗi câu có 3 hoặc 4 mức độ trả lời, bạn hãy chọn lấy *một* mức độ nào phù hợp nhất với mình và đánh dấu "X" vào cột tương ứng trên *trang 2 hoặc 3* của tờ phiếu trả lời.

Xin bạn lưu ý những điều sau đây:

1 - Chỉ chọn *một đáp án cho một câu hỏi* (hoặc một mức độ cho từng mục). Riêng với câu số 122 và 123 có thể chọn nhiều đáp án (khuyến tròn vào *các chữ a, b, c...* nếu phù hợp). Trả lời xong câu hỏi này mới chuyển sang câu hỏi khác. Không bỏ sót bất cứ câu hỏi nào.

2 - Không cần phải để thời gian suy nghĩ về các câu hỏi. Hãy trả lời bằng ý nghĩ trong óc mình đầu tiên. Tốc độ trả lời khoảng 2 - 3 câu trong 1 phút.

3 - Gặp những câu đề cập đến những vấn đề không quen thuộc thì hãy hình dung và trả lời theo cách suy nghĩ của bạn.

Xin chân thành cảm ơn sự hợp tác chân tình của bạn!

Phần nội dung câu hỏi

1. Những triết lý đạo đức: "Ăn quả nhớ kẻ trồng cây", "uống nước nhớ nguồn" vẫn giữ nguyên giá trị trong điều kiện xã hội hiện nay, có đúng không?

a - Đúng b - Phân vân c - Không đúng

2. Bạn rất ghét những kẻ vô ơn, bạc nghĩa phải không?

a - Đúng b - Phân vân c - Không đúng

3. Mọi người cần phải sống và làm việc theo pháp luật:

a - Đúng b - Phân vân c - Không đúng

4. Bạn cảm thấy ân hận mỗi khi vi phạm một điều quy định chung nào đó, đúng vậy không?

a - Đúng b - Đôi khi đúng c - Không đúng

5. Bạn cảm thấy cần phải sống trung thực, thẳng thắn dù có lúc bị thua thiệt phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

6. Bạn cảm thấy khổ tâm khi buộc phải nói dối ai một điều gì đó, đúng vậy không?

a - Đúng b - Đôi khi đúng c - Không đúng

7. Khi gặp các thầy, cô giáo cũ bạn:

a - Luôn chủ động chào hỏi niềm nở, dù họ không nhận ra bạn

b - Chào hỏi khi họ nhận ra bạn

c - Tìm cách tránh mặt nếu có thể.

8. Mỗi khi được giao giải quyết những công việc đòi hỏi cá nhân phải chịu trách nhiệm về hậu quả của nó, bạn thường:

a - tìm lý do để từ chối b - Miễn cưỡng c - Vui vẻ và sẵn sàng

9. Triết lý: "Thương người như thể thương thân", "lá lành đùm lá rách" không còn phù hợp trong điều kiện xã hội hiện nay, đúng vậy không?

a - Đúng b - Phân vân c - Không đúng

10. Bạn chưa bao giờ làm hại ai, ngay cả những người đã từng có những hành động xấu đối với bạn, đúng thế không?

a - Đúng b - Không rõ c - Không đúng

11. Sống với nhau cốt ở chữ tín, theo bạn điều đó có đúng không?

a - Đúng b - Phân vân c - Không đúng

12. Khi thất hứa với người khác, bạn hay bị dằn vặt hoặc có cảm giác áy náy khó chịu phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

13. Bạn cảm thấy làm việc, lao động, học tập là một nhu cầu tự nhiên, không cần phải thúc giục hay bắt buộc, có đúng vậy không?

a - Đúng b - Phân vân c - Không đúng

14. Có phải trong học kỳ vừa qua, bạn chưa tự giác và cố gắng trong học tập không?

a - Đúng b - Không rõ c - Không đúng

15. Thực tế đã giúp bạn nhận ra rằng: "Một cây làm chẳng nên non, ba cây chụm lại nên hòn núi cao" là

a - Hoàn toàn đúng b - Có thể đúng c - Không đúng

16. Bạn thích những công việc làm một mình hơn là làm chung với những người khác phải không?

a - Đúng b - Phân vân c - Không đúng

17. Có đúng là bạn đến trường theo ý nguyện của gia đình hơn là của bản thân không?

a - Đúng b - Phân vân c - Không đúng

18. Khi nhìn thấy người khác vứt rác bừa bãi ra đường, bạn:

a - Cho là bình thường b - Làm ngơ, mình không làm, còn ai làm kệ họ c - Khó chịu, bất bình hoặc nhắc nhở

19. Bạn tự nhận thấy mình thiếu nhiệt tình và niềm tin vào đường lối phát triển kinh tế thị trường theo định hướng xã hội chủ nghĩa phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

20. Tăng cường cải cách kinh tế, ổn định chính trị

theo định hướng xã hội chủ nghĩa là sự lựa chọn đúng đắn của Đảng và nhân dân ta, theo bạn điều đó có đúng không?

a - Đúng b - Phân vân c - Không đúng

21. Theo bạn có đúng là học sinh/ sinh viên cần phải tích cực, tự giác tham gia vào các hoạt động xã hội để học hỏi và rèn luyện mình không?

a - Đúng b - Phân vân c - Không đúng

22. Mỗi khi được yêu cầu tham gia vào các hoạt động xã hội ở nơi bạn ở, bạn thường:

*a - Sẵn sàng và nhiệt tình b - Miễn cưỡng
c - Tìm lý do để từ chối*

23. Bạn cho rằng học sinh/sinh viên nên quan tâm đến những vấn đề khác hơn là thời sự và chính trị phải không?

a - Đúng b - Phân vân c - Không đúng

24. Bạn có hay quan tâm và bàn luận về những vấn đề liên quan đến chủ trương chính sách của nhà nước không?

a - Không b - Thỉnh thoảng c - Thường xuyên

25. Tinh thần yêu nước và tự hào dân tộc là những giá trị đạo đức quan trọng, theo bạn điều này có đúng không?

a - Đúng b - Phân vân c - Không đúng

26. Theo bạn nền văn hoá Việt Nam thực sự có một bản sắc riêng, rất đáng tự hào phải không?

a - Đúng b - Phân vân c - Không đúng

27. Bạn tin rằng mỗi chúng ta cần phải tích cực, chủ

động, linh hoạt theo phương châm: "Hãy tự cứu lấy mình trước khi trời cứu" phải không?

a - Đúng b - Phân vân c - Không đúng

28. Có phải trong hoàn cảnh xã hội có nhiều tiêu cực, tệ nạn như hiện nay mỗi học sinh/sinh viên cần giữ mình sao cho "gần bùn mà chẳng hôi tanh mùi bùn" không?

a - Đúng b - Phân vân c - Không đúng

29. Bạn thích cộng tác với những người tích cực, chủ động, tháo vát dù họ là người khó tính, đúng vậy không?

a - Đúng b - Phân vân c - Không đúng

30. Bạn chưa bao giờ mua sắm một thứ gì vượt quá khả năng kinh tế của bản thân, dù rất thích nói phải không?

a - Đúng b - Phân vân c - Không đúng

31. Mục đích học của bạn là để có một trình độ khoa học, có tay nghề, sau khi ra trường có thể lao động bằng nghề của mình, đúng vậy không?

a - Đúng b - Phân vân c - Không đúng

32. Bạn cho rằng giữ vệ sinh môi trường là công việc riêng của từng cá nhân: Ai giữ được vệ sinh môi trường thì người đó sẽ khoẻ mạnh sống lâu, ai không làm như vậy sẽ phải gánh chịu hậu quả xấu phải không?

a - Đúng b - Phân vân c - Không đúng

33. Có phải những người gần gũi bạn đều cho rằng bạn là người dễ đồng cảm với người khác, biết chia sẻ với họ những niềm vui và nỗi buồn hàng ngày không?

a - Đúng b - Phân vân c - Không đúng

34. Khi người thân hoặc bạn bè gặp chuyện không vui, bạn luôn thăm hỏi, động viên hoặc giúp đỡ theo khả năng của mình phải không?

a - Đúng b - Phân vân c - Không đúng

35. Bạn cảm thấy bối rối mỗi khi phải giải quyết những vấn đề mới lạ phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

36. Theo bạn, có đúng là khả năng thích ứng, tự điều chỉnh là một phẩm chất quan trọng trong thời kỳ công nghiệp hoá, hiện đại hoá không?

a - Đúng b - Phân vân c - Không đúng

37. Có phải bạn thích một lối sống tự do, ít ràng buộc hơn là một lối sống có nề nếp quy củ không?

a - Đúng b - Phân vân c - Không đúng

38. Bạn có thấy khó chịu nếu phải sống hoặc sinh hoạt với những người không có tính ngăn nắp, không có giờ giấc quy củ không?

a - Có b - Lúc có, lúc không c - Không

39. Bạn tôn trọng sự tự do tín ngưỡng của người khác, nhưng không thích những người quá sùng bái, quá tin vào thần, phật, chúa trời, có đúng vậy không?

a - Đúng b - Phân vân c - Không đúng

40. Khi đi tàu xe, bạn phải chọn ngày giờ tốt, tránh ngày giờ xấu, đúng vậy không?

a - Đúng b - Đôi khi đúng c - Không đúng

41. Theo bạn, hôn nhân có cần dựa trên tình yêu không?

a - Có b - Phân vân c - Không

42. Sự hiểu biết của bạn về thể chế và các đạo luật cụ thể (như luật giáo dục, luật hôn nhân và gia đình) có thể được đánh giá ở mức:

a - Ít, lơ mơ

b - Tam được, tuy chưa cập nhật và chưa sâu sắc

c - Khá tốt, cập nhật và sâu sắc

43. Bạn thích làm những công việc đòi hỏi phải tận tâm, tận lực, có tinh thần trách nhiệm cao phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

44. Bạn đã làm ngơ khi thấy người khác làm sai luật pháp phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

45. Phải luôn kính trọng bố mẹ, ông bà, dù giữa ta và họ có những khác biệt nào đó về học vấn, đạo đức hoặc lối sống, có đúng vậy không?

a - Đúng b - Phân vân c - Không đúng

46. Khi đi tàu xe, nếu không bị kiểm tra thì ta không dại gì mà mua vé phải không?

a - Đúng b - Phân vân c - Không đúng

47. Khi cần tìm một người "đã nói là làm" thì tập thể thường chọn bạn):

a - Đúng b - Đôi khi đúng c - Không đúng

48. Bạn sẽ thể hiện thái độ như thế nào khi thấy những học sinh/sinh viên có hành vi gian lận trong thi cử:

a - Đồng tình b - Làm ngơ c - Khó chịu bất bình

49. Bạn cảm thấy lo lắng và cố gắng nhiều hơn khi làm một việc gì đó mà kết quả của nó liên quan hoặc

ảnh hưởng đến nhiều người khác phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

50. Đã bao giờ bạn cảm thấy day dứt và sẵn sàng nhận hậu quả về mình khi có những quyết định sai lầm ảnh hưởng đến mọi người chưa?

a - Thường xuyên b - Đôi khi c - Chưa bao giờ

51. Khi làm một điều gì đó tổn hại đến người khác, bạn có thấy lương tâm mình dằn vặt không?

a - Có b - Lúc có, lúc không c - Không

52. Bạn giúp đỡ người khác ngay cả khi điều đó gây cho bạn những khó khăn, thiệt hại, có đúng như vậy không?

a - Đúng b - Đôi khi đúng c - Không đúng

53. Khi hoàn cảnh không thuận lợi, để tránh hậu quả, bạn đã thay đổi lời hứa và những cam kết của mình phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

54. Không có nghề gì là thấp hèn, chỉ có những kẻ lười biếng "vô công rồi nghề" mới đáng xấu hổ, theo bạn điều đó có đúng không?

a - Đúng b - Phân vân c - Không đúng

55. Bạn thích làm một công việc an nhàn, ít phải cố gắng mà có thu nhập cao phải không?

a - Đúng b - Phân vân c - Không đúng

56. Khi làm việc, bạn có kết hợp hài hoà nhu cầu và lợi ích của mình và của người khác không?

a - Có b - Lúc có, lúc không c - Không

57. Khi giải quyết công việc chung, bạn thường:

a - Làm theo ý của mình, ít để ý đến ý kiến của người khác

b - Chiều theo ý kiến của người khác, để tránh sự phạt ý, dù biết là họ sai

c - Cùng bàn bạc với mọi người để tìm ra giải pháp tốt nhất

58. Bạn chỉ mong muốn làm một người tốt, còn có phần đấu trở thành đảng viên hay không thì không quan trọng:

a - Đúng b - Phân vân c - Không đúng

59. Tích cực học tập là mục tiêu phần đấu quan trọng nhất lúc này của bạn phải không?

a - Đúng b - Phân vân c - Không đúng

60. Những người xung quanh bạn thường có chung nhận xét rằng bạn là người luôn sẵn sàng và nhiệt tình trong các công việc của lớp, của trường:

a - Đúng b - Không rõ c - Không đúng

61. Bạn có hay theo dõi các chương trình truyền hình về thời sự trong nước và quốc tế không?

a - Đúng b - Thỉnh thoảng c - Thường xuyên

62. Bạn có hay đọc báo và nghe đài không?

a - Đúng b - Thỉnh thoảng c - Thường xuyên

63. Mỗi khi được tiếp xúc với những giá trị văn hoá của thế giới, bạn lại có mặc cảm, tự ti về những gì gọi là giá trị văn hoá của Việt Nam.

a - Đúng b - Đôi khi đúng c - Không đúng

64. Có lúc bạn mơ ước sau này mình được mang quốc tịch nước ngoài như Anh, Mỹ, Pháp hoặc Ôxtrâyliia hơn

là quốc tịch Việt Nam phải không?

a - Đúng b - Phân vân c - Không đúng

65. Trong cuộc sống, đã bao giờ bạn trì hoãn hoặc né tránh những vấn đề rắc rối mà lẽ ra bạn phải chủ động giải quyết ngay chưa?

a - Thường xuyên b - Ít khi c - Chưa bao giờ

66. Những người biết bạn cho rằng bạn là người tháo vát và hay có sáng kiến phải không?

a - Đúng b - Phân vân c - Không đúng

67. Bạn cảm thấy khó chịu với những lời nói, hành vi thiếu văn hoá, kể cả trong nhóm bạn thân quen phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

68. Bạn đã tìm cách tự kiểm soát mình để không bị lôi kéo vào con đường cờ bạc, nghiện hút, có phải vậy không?

a - Đúng b - Đôi khi đúng c - Không đúng

69. Mục tiêu học tập của bạn chỉ là cần có tấm bằng, còn sau này làm nghề gì sẽ tính sau

a - Đúng b - Phân vân c - Không đúng

70. Trong cuộc sống, bạn hầu như luôn luôn đạt được các mục đích đã đề ra

a - Đúng b - Không rõ c - Không đúng

71. Trong số những người thân và bạn bè, có rất ít người hiểu được bạn, đúng vậy không?

a - Đúng b - Đôi khi đúng c - Không đúng

72. Mỗi khi có chuyện buồn vui cần tâm sự, bạn cảm thấy không có ai để cùng chia sẻ, cảm thông phải

không?

a - Đúng b - Đôi khi đúng c - Không đúng

73. Khi bước vào một môi trường chưa quen biết bạn thường cảm thấy:

a - Khó thích nghi b - Lúc dễ, lúc khó c - Dễ thích nghi

74. Bạn có thấy lung túng, bị động mỗi khi công việc diễn ra trái với dự định không?

a - Có b - Lúc có, lúc không c - Không

75. Bạn tự thấy và được bạn bè công nhận là chưa bao giờ sai phạm nội quy nhà trường phải không?

a - Đúng b - Không rõ c - Không đúng

76. Trong học tập, sinh hoạt tập thể, bạn đã chú ý giữ đúng nề nếp tác phong theo quy định để tránh bị nhắc nhở, đúng không?

a - Đúng b - Đôi khi đúng c - Không đúng

77. suy cho cùng, không ai có thể tránh được số mệnh

a - Đúng b - Phân vân c - Không đúng

78. Bạn có tin rằng "nhân định có thể thắng được thiên định" và bằng đức độ, khả năng của mình, con người có thể khắc phục được hoàn cảnh không?

a - Có tin b - Nửa tin, nửa không c - Không tin

79. Bạn không bỏ lỡ một cơ hội nào để giúp đỡ hoặc tham gia vào các hoạt động "đền ơn đáp nghĩa" những người có công với nước, đúng không?

a - Đúng b - Phân vân c - Không đúng

80. nếu làm sai pháp luật hoặc một quy định chung

nào đó mà không ai biết, là bạn yên tâm đúng vậy không?

a - Đúng b - Phân vân c - Không đúng

81. Nếu như bạn vi phạm một cách không chủ định các quy định ở những nơi công cộng thì bạn cho rằng điều đó không sao cả phải không?

a - Đúng b - Phân vân c - Không đúng

82. Bạn có hay ân hận với những lời nói và việc làm mà đáng lẽ không nên nói hay làm như vậy không?

a - Thường xuyên b - Ít khi c - Chưa bao giờ

83. Những người gần gũi với bạn, đều cho rằng bạn là người sống có tình nghĩa

a - Đúng b - Phân vân c - Không đúng

84. Bạn hay giận dữ và muốn trừng phạt hoặc trả thù những ai cố ý làm hại mình phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

85. Đã làm việc, dù là công việc chân tay hay trí óc, bạn đều tận tâm, tận lực, đúng vậy không?

a - Đúng b - Phân vân c - Không đúng

86. Khi cần hoàn thành một việc quan trọng đòi hỏi ý thức trách nhiệm cao, người ta hay đề nghị bạn đảm nhiệm phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

87. Đã có những hoàn cảnh nào đó, bạn đánh mất tính nghiêm túc và trung thực của mình trong học tập phải không?

a - Đúng b - Không rõ c - Không đúng

88. Trước khi bắt tay giải quyết một công việc, bạn

luôn cân nhắc lợi hại để thấy rõ hậu quả và trách nhiệm cá nhân đúng vậy không?

a - Đúng b - Đôi khi đúng c - Không đúng

89. Khi được tiếp xúc với trình độ nhận thức, khoa học công nghệ của thế giới, bạn cảm thấy:

a - Muốn học hỏi để sau này giúp ích cho đất nước

b - Bi quan, không muốn làm gì cả.

c - Muốn được ra nước ngoài sinh sống và làm việc

90. Bạn mong muốn, sau khi ra trường có một cuộc sống tự lập, được lao động bằng tay nghề của mình phải không?

a - Đúng b - Phân vân c - Không đúng

91. Trong công việc, bạn giữ kín những bí quyết riêng của mình, tìm cơ hội để đạt được kết quả hơn người khác, đúng vậy không?

a - Đúng b - Đôi khi đúng c - Không đúng

92. Biết đưa ra yêu cầu với người khác và chấp nhận những yêu cầu của người khác là cơ sở để tạo ra sự nhất trí trong hợp tác, theo bạn điều đó có đúng không?

a - Đúng b - Phân vân c - Không đúng

93. Đối với các phong trào "giữ vệ sinh chung", "trồng cây xanh", "giữ sạch đẹp cho đường phố, bản làng", bạn:

a - Không tham gia

b - Tham gia miễn cưỡng

c - Tích cực tham gia

94. Bạn có thói quen chuẩn bị, cân nhắc và suy tính trước khi hành động phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

95. Những người xung quanh cho rằng bạn là người có lối sống giản dị, tiết kiệm

a - Đúng b - Phân vân c - Không đúng

96. Có đúng là bạn rất quan tâm và mong muốn đài, báo, các phương tiện thông tin đại chúng khác của xã hội đề cập nhiều hơn đến vấn đề ô nhiễm và bảo vệ môi sinh không?

a - Đúng b - Phân vân c - Không đúng

97. Bạn cho rằng "dân giàu, nước mạnh, xã hội công bằng, văn minh" là lý tưởng phấn đấu của mình, đúng vậy không?

a - Đúng b - Phân vân c - Không đúng

98. Nhiều bạn trẻ hiện nay đang khao khát, trăn trở, muốn đóng góp sức mình xây dựng đất nước, bạn cũng có những ý nghĩ nguyện vọng như họ phải không?

a - Đúng b - Phân vân c - Không đúng

99. Khi xử lý những vấn đề khó, bạn cân nhắc những giải pháp khác nhau và kết hợp chúng lại để có được một giải pháp tốt hơn, có đúng vậy không?

a - Đúng b - Đôi khi đúng c - Không đúng

100. Khi phải quyết định một việc gì đó, bạn thường:

a - Tự quyết định

b - Chần chừ, không quyết định ngay

c - Hoàn toàn dựa vào ý kiến của người khác

101. Bạn cảm thấy khó chịu với những người làm việc không có kế hoạch phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

102. Bạn cảm thấy khó chịu với những người sống

"vô tâm" theo kiểu "dèn nhà ai nhà nấy rạng" đúng vậy không?

a - Đúng b - Đôi khi đúng c - Không đúng

103. Mỗi khi gặp thất bại, bạn tự hỏi mình: "Liệu ta đã nhận thức và hành động phù hợp chưa" để sau đó, cố gắng tự điều chỉnh cho phù hợp với hoàn cảnh thay đổi phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

104. Khi chuyển từ trung học phổ thông lên đại học, bạn phải mất bao lâu mới có được sự thích nghi (học sinh trung học phổ thông không phải trả lời câu hỏi này)?

a - Dưới một học kỳ

b - Khoảng hai học kỳ

c - Trên hai học kỳ

105. Bạn có bao giờ nghỉ học không có lý do không?

a - Thường xuyên b - Ít khi c - Chưa bao giờ

106. Những người biết bạn đều cho rằng bạn là người thẳng thắn, trọng lẽ phải, đúng vậy không?

a - Đúng b - Phân vân c - Không đúng

107. Đối với bạn, không thực hiện được lời hứa vì những lý do "khách quan" là điều:

a - Thường xảy ra

b - Ít khi xảy ra

c - Chưa bao giờ xảy ra

108. Bạn căm ghét sự lợi dụng công sức, tài năng của người khác để sống sung sướng và tạo nên sự thành đạt cá nhân phải không?

a - Đúng b - Phân vân c - Không đúng

109. Thực tiễn của bản thân đã giúp bạn nhận ra rằng triết lý "lao động sáng tạo nên con người" là:

a - Hoàn toàn đúng

b - Có thể đúng

c - Không đúng

110. Chân thành trong hợp tác và biết lắng nghe, cân nhắc các ý kiến khác nhau là những đặc điểm nổi bật trong cách làm việc của bạn, đúng thế không?

a - Đúng b - Phân vân c - Không đúng

111. Bạn chưa thực sự có ý thức đấu tranh với những hành vi phá hoại môi sinh phải không?

a - Đúng b - Đôi khi đúng c - Không đúng

112. Bạn có cho rằng yêu nước tức là yêu chủ nghĩa xã hội không?

a - Đúng b - Phân vân c - Không đúng

113. Gần đây trong học sinh/sinh viên, quyền lợi được đề cao hơn trách nhiệm và nghĩa vụ cá nhân, theo bạn điều đó có đúng không?

a - Đúng b - Phân vân c - Không đúng

114. Đa số học sinh/sinh viên hiện nay học tập:

a - Tích cực, chủ động

b - Tích cực, nhưng nhiều khi chưa chủ động

c - Không tích cực, thiếu chủ động

115. Có thể đánh giá sự chấp hành kỷ luật của học sinh/sinh viên trường bạn là:

a - Tốt b - Tương đối tốt c - Không tốt

116. Số học sinh/sinh viên có những biểu hiện yêu

nước và tự hào về truyền thống dân tộc là:

a - Nhiều b - Tương đối nhiều c - Ít

117. Trong lớp hoặc trường của bạn, số học sinh, sinh viên có lối sống lành mạnh, giản dị và tiết kiệm là:

a - Nhiều b - Tương đối nhiều c - Ít

118. Số sinh viên có "lời nói không đi đôi với việc làm", "hứa nhiều làm ít" ở lớp của bạn là:

a - Nhiều b - Tương đối nhiều c - Ít d - Không có

119. Số sinh viên tích cực phấn đấu rèn luyện để có được kỹ năng hợp tác ở lớp của bạn là:

a - Nhiều b - Tương đối nhiều c - Ít d - Không có

120. Nhìn chung, số học sinh/sinh viên có nguyện vọng vào Đảng là:

a - Nhiều b - Tương đối nhiều c - Ít d - Không có

121. Hằng năm nhiều người hành hương về đâu để giỗ tổ Hùng Vương

122. Trước khi có công việc hệ trọng phải giải quyết bạn thường:

a - Đi xem bói

b - Xin quẻ

c - Cầu khẩn thần thánh, trời phật phù hộ

d - Cầu chúa

e - Hỏi ý kiến của người khác

f - Ý kiến khác (ghi cụ thể).....

123. Khi đọc báo hoặc nghe đài, bạn quan tâm nhất đến lĩnh vực nào sau đây:

a - Thời sự, chính trị

b - Kinh tế, sản xuất

c - Văn hoá - thể thao

d - Giáo dục - đào tạo

e - Khoa học kỹ thuật

g - Y tế

124. Xin bạn hãy kể tên 6 sự kiện thời sự và chính trị đang được quan tâm nhiều hơn cả, trong năm

a - Trong nước

b - Ngoài nước

1.

1.

2.

2.

3.

3.

125. Xin bạn hãy cho biết ý kiến của mình về những quan niệm sau đây:

Tán thành Phản văn Phản đối

a. Yêu nhiều để lấy một

☐☐☐

b. Đã yêu là cho nhau tất cả

☐☐☐

c. Tình yêu phải dẫn đến hôn nhân

☐☐☐

126. Bản thân bạn tham gia như thế nào vào các hoạt động sau đây:

	Không tham gia	ít tham gia	Thường xuyên tham gia
a - Công ích, từ thiện			
b - Bảo vệ môi trường			
c - Chống tệ nạn xã hội			
d - Chính trị, xã hội của các cơ quan đoàn thể			
e- Văn hoá - thể thao			

127. Những hiện tượng sau đây biểu hiện như thế nào trong học sinh/sinh viên ở lớp của bạn, trường của bạn?

	Nhiều	Tương đối nhiều	Ít	Không có
1 - Lơ là trong học tập				
2 - Không có mục tiêu và kế hoạch học tập rõ ràng				
3 - Lười biếng, ít tự giác cố gắng trong học tập				
4 - Gian lận trong thi cử				
5 - Vi phạm các quy định của nhà trường đến mức phải kỷ luật				
6 - Mê tin dị đoan				
7 - Lợi dụng tình yêu để thỏa mãn nhu cầu cá nhân (có tiền tiêu xài, có việc làm...)				
8 - Sống cho riêng mình, ít quan tâm giúp đỡ người khác				
9 - Không đúng giờ, đúng hẹn				

	Nhiều	Tương đối nhiều	ít	Không có
10 - Bỏ học không có lý do				
11 - Nói chuyện, làm việc riêng trong giờ học				
12 - Không học bài, làm bài				
13 - Mua điểm hoặc xin điểm				
14 - Quay cóp				
15 - Vứt rác bừa bãi				
16 - Không giữ vệ sinh chung khi đi đại, tiểu tiện				
17 - Khạc nhổ, xì mũi bừa bãi				
18 - Đánh nhau				
19 - Xích mích cái cọ				
20 - Không quan tâm giúp đỡ nhau				
21 - Mắng chửi nhau				
22 - Nghiện thuốc lá				
23 - Nghiện bia, rượu				
24 - Thích ăn diện				
25 - Tiêu xài lãng phí				
26 - Nghiện ma túy				
27 - Đánh bạc, chơi số đề				
28 - Mua, bán dâm				
29 - Buồn lậu				

TÀI LIỆU THAM KHẢO

Tiếng Anh

- 1- Aiken, L.R, *Psychological Testing and Assessment*, 10th, Ed, 1999. Allyn and Bacon, London.
- 2- Aiken, L.R, *Questionnaires & Inventories: Surveying opinion & assessing Personalities*, 1997. John Wiley & Sons, Inc.
- 3- Babbie, E, *The Practice of Social Research*. 6 th Ed, 1992, Wadsworth Publising Company, Inc.
- 4- Grimm, L.G. *Statistical Applications for the Behavioral Sciences*, 1993, John Wiley & Sons, Inc.
- 5- Murphy, K. R & Davidshofer, C. O, *Psychological Testing: Principles & Applications*. 3th Ed, 1994, Prentice-Hall, Inc.
- 6- SPSS Base 10.0 User's Guide. 1999, SPSS Inc.

Tiếng Việt

- 1- Baker.L.T, *Thực hành nghiên cứu xã hội*, Nxb Chính trị quốc gia, Hà Nội, 1998
- 2- Nguyễn Công Khanh, *Các phương pháp xây dựng trắc nghiệm: quy trình và kỹ thuật thiết kế trắc nghiệm trong khoa học hành vi*. Tài liệu chuyên khảo dùng cho giáo viên Trường Cao đẳng Sư phạm Nhà trẻ - Mẫu giáo

Trung ương Hà Nội. 2001.

3- Nguyễn Công Khanh và Nguyễn Văn Liệu, *Ứng dụng SPSS cho Windows: Xử lý và phân tích dữ liệu trong khoa học hành vi*. Tài liệu chuyên khảo, Viện Khoa học giáo dục, 2001.

4- Trần Trọng Thuỷ và Nguyễn Công Khanh, *Các phương pháp nghiên cứu: thiết kế, thích nghi và chuẩn hoá công cụ đo lường trong khoa học xã hội*. Tài liệu tập huấn của Trung tâm Nghiên cứu tâm lý học - sinh lý học lứa tuổi, Viện Khoa học giáo dục, 2000.

Chịu trách nhiệm xuất bản

TRỊNH THỨC HUỠNH

Chịu trách nhiệm nội dung

HOÀNG PHONG HÀ

Biên tập nội dung: NGUYỄN KHÁNH HOÀ

ĐINH THỊ MỸ VÂN

Biên tập kỹ, mỹ thuật: NGUYỄN PHƯƠNG MAI

Trình bày bìa: NGUYỄN PHƯƠNG MAI

Chế bản vi tính: PHẠM PHƯƠNG NHUNG

Sửa bản in: NGỌC BÍCH - PHƯƠNG NHƯ

Đọc sách mẫu: ĐINH THỊ MỸ VÂN

In 1.040 cuốn, khổ 14.5 x 20.5 cm, tại Nxb. Chính trị quốc gia.
Giấy phép xuất bản số: 38-1570/CXB-QLXB, cấp ngày 6-11-2003.
In xong và nộp lưu chiểu tháng 5 năm 2004.

NHÀ XUẤT BẢN CHÍNH TRỊ QUỐC GIA

24 Quang Trung - Hà Nội

ĐT :9422008

Fax :84-4-9421881

E-mail :nxbctqg@hn.vnn.vn

Website:www.nxbctqg.org.vn

TÌM ĐỌC

Vũ Cao Đàm

- NGHIÊN CỨU KHOA HỌC - PHƯƠNG PHÁP LUẬN VÀ THỰC TIỄN

Trần Xuân Sầm

**- TÌM HIỂU PHƯƠNG PHÁP LUẬN VÀ PHƯƠNG PHÁP
NGHIÊN CỨU KHOA HỌC QUA MỘT SỐ TÁC PHẨM KINH ĐIỂN**

Mai Quỳnh Nam (Chủ biên)

- TRẺ EM - GIA ĐÌNH - XÃ HỘI



Giá: 21.000đ