

## 調査研究における質問紙の作成過程と適用上の諸問題

山崎 勝之\*, 内田 香奈子\*\*

(キーワード: 質問紙, 調査研究, 標準化, 信頼性, 妥当性)

心理学の研究や応用の分野では、質問紙検査ほどよく使われる検査法はない。しかし、使われる検査のほとんどには問題があり、まるで、適当につけられた目盛りでものの長さを測るような現状が続いている。本論文においては、質問紙検査に備わるべき要件について、質問紙の作成過程を追いながら説明し、実施上の諸注意についてもふれる。

### 1. 質問紙法を使うとき

質問紙検査は何を測定するのか。質問紙の良し悪しの見きわめはどのようにするのか。本節では、質問紙法を使ったり、作成したりする前に、質問紙の基本的な役割と特徴について詳述する。

#### (1) 質問紙法の役割

本論文で説明する質問紙とは、性格などの構成概念(construct)を測定するための質問紙であり、単なる調査紙のことではない。「今朝、朝食を食べましたか」などという、実体のある事実をたずねることは難しい問題は何かもない。質問紙法の難しいところは、目に見えず、客観的にとらえることの困難な、操作的に作られた概念を扱うことにあり、この点から検査自体に確認しておくべき事柄が多々生まれるのである。たとえば、性格の一つである外向性(extraversion)は心理学的に構成された概念で、ある人物について誰もが簡単かつ客観的にその外向性の大きさを決定することはできず、質問紙の登場ということになる。

質問紙についての解説書は多いが、本論文はその中級的な解説を試み得る。検査作成のための計算や分析方法、検査結果の解釈など基本的な手法は初級にあたるが、ここではその内容にはほとんどふれられない。しかし、以下に展開される中級的な解説こそが最も重要で、しかも、この内容を的確に紹介した論文や書物はこれまでには見あたらない。

#### (2) 質問紙法のチェックポイント

質問紙法は手軽なためか、頻繁にしかも安易に使われている。しかし、目に見えない構成概念を測るだけに、質問紙法ほど難しい検査はないことを忘れてはいけない。本論文を通して、質問紙法を利用することの難しさと注意点を説明するのであるが、質問紙法を使用する際の問題は、使用する検査を決定するまで、あるいは検査を作成するところまでに大半が集中し、この時点でのチェックがどれほど厳しく行われるかが重要である。たとえば、自分で何れかの質問紙を利用しようとする場合のチェックポイントを紹介すると次のようになる(表1)。

表1 質問紙検査選択時の主要チェックポイント

- |                                |
|--------------------------------|
| ・信頼性と妥当性(入念にチェック)              |
| ・検査がターゲットとする対象者(年齢、男女、定職の有無など) |
| ・検査時間や項目内容などの調査対象者にとっての適切性     |
| ・製品化の有無                        |
| ・(未製品化なら)使用許可など、作成者との直接的接触     |

まず、以下に詳しく説明する検査の妥当性(validity)と信頼性(reliability)、すなわち標準化(standardization)の出来映えをチェックする。妥当性とは、測ろうとする概念をその検査で本当に測れるかどうかということ、そして信頼性とは、その概念が安定して測れるかどうかということである。あたりまえのようなチェックポイントであるが、このチェックを怠ることが日常茶飯事のように起こっている。また、この標準化の実施年齢幅も問題で、たとえば、検査によっては幼児から中学生あたりまでのデータを一緒にして標準化しているものがあるが、幼児と中学生がその検査について同じ特徴をもつはずがない。さらに、どの年齢層を対象にしているのかというポイントも大切で、大学生用に標準化された検査を高校生に使用しても問題はないという主張は根拠がない。他にも、検査によっては項目数が多すぎたり、男性や定職

\*鳴門教育大学人間形成講座

\*\*鳴門教育大学大学院学校教育研究科(研究生)

をもつ者だけを対象にしていたり、過激で回答しづらい項目内容であったりして、目的とする調査対象には使用できないことがあるので、この点も見逃せず、強引な使用は避けなければならない。

製品化の問題も大切で、製品化されているのなら、その販売元から購入せざるを得ない。検査は、できれば製品化せず、フリーにするべきだという意見をよく聞くが、まったく同感である。製品化されていなくとも、作成者へコンタクトをとって使用することが推奨される。思わぬ改訂が進められていたり、作成者からしか聞き得ない情報を得ることがある。

この場合、米国における著作権意識はきわめて高いことには注意を要する。質問紙を無断で印刷して使用することが禁じられていることは当然で、著作権がある印刷物を無断で雑誌などに掲載することも許されない。その厳しさを目の当たりにした例をあげると、わが国のタイプA性格・行動 (Type A personality or behavior) に関する学術雑誌「タイプA」に米国のタイプA質問紙である Jenkins Activity Survey (JAS; Jenkins, Zyzanski, & Rosenman, 1979) の翻訳版の項目の掲載が拒否されたという事実がある。日本のタイプA行動パターン (Type A Behavior Pattern; TABP) カンファレンスのメンバーの一人が、雑誌に JAS 翻訳版の掲載への許可を得るために、使用権をもつ Psychological Corporation に連絡をとったところ、その返事は、質問項目の1~2問を翻訳し掲載することは認めるが、全問の掲載は許可できないというものであった。そこで今度は、JASの作成者である Jenkins, C.D. 博士から直接お願いして頂くという、これ以上ない手段をとったが、最終的には教科書や雑誌等に JAS の全文の掲載は認めることはできないという結論をつきつけられた。

日本であれば、項目の紹介なら宣伝にもなるし、ましてや作成者自身の口添えがあつて掲載を断ることなど考えられないが、米国では一度著作権が確立されると作成者の意のままにならないほど著作権意識は高いらしい。著作権が確立されれば、このようなことは起こっても何ら不思議はなく、Psychological Corporation が批判されるべきことでもない。ただ、著作権のある質問紙の扱いには注意をしなければならないということである。ところが、著作権の確立していない質問紙の扱いは驚くほど問題がない。作成者に許可をとらなければならないことは当然であるが、その使用や日本版への標準化が断られることはまずない。筆者らはこれまでに米国の10人近くの質問紙作成者に日本版作成への許可を求めたが、一度も断られたことはない。例外なく、日本版を作ってくれることは光栄であるとの返事をいただき、同時にその検査について最新の情報を受け取った場合も少なくない。著作権の確立を批判することはできないが、使う側からすれば、作成者

には自由に無料で使える状態を保ってほしいと願うばかりである。

### (3) 質問紙法はおすすめではない

どんなによくできた質問紙でもエラーはつきものである。たとえば、質問紙検査は集団で実施する機会が多いが、そのときの受検者の体調や構えまで察知し、それを統制することなどできない。体調をくずしている者もいれば、やる気のない者もいる。つまり、エラーどっさりの状態で検査は回収され、一人ひとりの結果など信頼できるはずがない。質問紙は数百人単位で実施して、そこからの平均値などの全体像でものを言うたぐいの検査である。このことを忘れて、学校などで実施した子ども個人の質問紙結果を担任などにわたすことがあるが、百害あって一利なしである。

そもそも短時間に大量のデータが得られる方法論を安易にとる姿勢はいただけない。質問紙法は、他に選択する方法がないために、しかたなく実施するぐらいの心構えが欲しい。特に若手研究者は、体力があるうちに、人間を直接みて、実験など方法論的に複雑な領域でもまれるべきであろう。質問紙は質問紙なりの役割と長所をそなえていることは間違いないが、どちらが良いか決定できそうにはないものの、実験や観察などで人間を間近でみる手法にはかなわないという印象を受ける。

## 2. 標準化の出発点

質問紙検査をどうしても使いたい、使わざるを得ない場合がある。そのとき、わが国で使用できる検査がなければ、自分で検査を作成する必要が生まれる。本節では、外国に適当な検査があり、それをもとに日本用への標準化を行う際の諸問題について詳述し、同時に標準化の中心となる信頼性と妥当性の概念を説明する。

### (1) 外国の質問紙を利用する

さて、質問紙を利用することになると、関連する検査を国内で探し出せることは少なくない。しかし、その検査には標準化に大きな不備があり、使用に耐えるものではないことが多い。こうなると次の手段は、自分で質問紙を作成することであるが、途方もない労力と時間がかかることを覚悟で作成に入る。

このとき、自分で独自の検査を作成することは考えず、まず諸外国の質問紙を調べ、測定したい概念が測れる検査を探してみる。米国などではおびただしい数の質問紙が用意されているから、適合する検査を探し当てることは容易である。そして都合のよいことに、米国では標準化への考え方が比較的厳しいから、公表されている検査の多くが高い信頼性と妥当性を備えている。自分で独自

の検査を作成するより、たとえ英語版であっても完成された検査をもとにして日本版を作る方が労力と時間が節約されることになる。ただ、国が違えば、風土、社会、文化が違い、概念によっては、最初からわが国独自の検査を作った方がよいときもある。

## (2) 日本版作成時の注意点

そこで、米国などで検査が見つければ、それを利用することになるが、まず製品化されているかどうか、著作権が確立されているかどうかを調べる必要がある。製品化されていると手続きは煩雑になるので、適当な出版社を国内で探して交渉をまかせるのがよい。先述したように、米国での著作権意識はわが国とは比べものにならないほど高く、注意が必要である。しかし、製品化されている検査の数に比べると、未製品化の検査の数の方が圧倒的に多い。そこで直接作成者にあたってみて、日本版の標準化への許可をもとめてみるとよい。製品化されていなければ、了承の返事が来る。作成者は検査が広く使われることを望んでいるから、断る理由がないのである。

許可が得られれば、標準化を始めることになるが、その場合、元の検査の作成のプロセスを熟知し、完成された検査から出発するのではなく、元の検査が標準化の過程で削除した項目を含めることや、元の検査にはない独自の項目を挿入することも考える必要がある。そして最初の邦訳にあたっては、元の項目の内容を損なわないことも大切だが、文化・社会差などを考慮し、わが国にあった項目にする姿勢も必要である。この意味では、日本版の作成は単なる和訳ではなく、日本語にした項目を元の言語に直してもその通りにならないことがあっても当然である。標準化の過程は元の検査で採用された方法を基本的には踏襲すべきであるが、その方法をベストと鵜呑みにすることは危険である。検査によっては米国版などにも不備が多いので、改善するぐらいの姿勢をもって作成にあたるべきである。

ここで邦訳の難しさを例示すると、攻撃性 (aggressiveness) などはその際たる例となり、米国などでは攻撃の直接的な表現が多いのに対して、わが国では直接的な表現を避け、受身-攻撃型 (passive-aggressive type) (加賀, 1979) の攻撃性を示すことが多い。受身-攻撃型とは、直接的に攻撃を表出しないで、他人のすることを妨害したり、不平を言ったりして間接的に攻撃を表出する人格タイプをさす。かつて筆者らは、タイプA性格・行動の質問紙の一つであるKG式日常生活質問紙 (山崎・田中・宮田, 1992) の作成にあたって、最初米国にある何れかの検査の日本版作成を考えたが、タイプAにある攻撃特徴の表現がむずかしく邦訳は断念し、わが国独自の検査を作成することとし、KG式の質問紙が出来上がった経緯がある。今1つ例をあげると、数年前、敵意

性と健康との関連をみる研究会 (HP2000研究会) が攻撃性を測定する質問紙を作成した。攻撃性といえば、人間に生来的に備わり、われわれの生活を特徴づける重要な心理特性である。それにもかかわらず、わが国では成人用に標準化が完成された質問紙は皆無であったし、質問紙大国米国でも攻撃性に関する数十という検査の中で標準化、特に妥当性が備わった検査は1, 2というありさまであった。この日本版の攻撃性質問紙 (日本版 Buss-Perry 攻撃性質問紙; 安藤・曾我・山崎・島井・嶋田・宇津木・大芦・坂井, 1999) は、米国で標準化が完成された質問紙である Buss & Perry (1999) の検査をもとに作成されたが、29項目あった米国版も日本版への標準化の過程の中で24項目に減り、訳出や回答分布の検討により内容ががらりと変わった項目も少なくない。表2には、米国版と日本版のいくつかの項目を比較し、その変化を紹介している。副詞や語尾の表現に注目し、他国の検査を日本版にすることは検査を一から作り直すことに等しい苦勞であることの一端でも感じて欲しい。

表2 Buss-Perry 攻撃性質問紙における米国版項目と日本版項目の比較

- |  |
|--|
| <ul style="list-style-type: none"> <li>• If I have to resort to violence to protect my rights, I will.<br/>権利を守るためには暴力もやむを得ないと思う。</li> <li>• I know that "friends" talk about me behind my back.<br/>友人の中には、私のことを陰であれこれ言っている人がいるかもしれない。</li> <li>• When frustrated, I let my irritation show.<br/>いらいらしていると、すぐ顔に出る。</li> </ul> |
|--|

さらにここで、逆訳 (back translation) の問題の困難さも指摘する必要がある。既存の米国などでの検査を日本語版にするとき、バイリンガルの方の協力を得て、日本語化した項目からふたたび英語に訳し直し、その整合性をみることがよく行われる。この逆訳は、国際比較研究では常套的な手段である。しかし、筆者らの経験からすると、この逆訳がスムーズに進むことはまずない。完成された論文をみると、「逆訳を行い整合性を確認した」という記述が簡単になされているが、この逆訳が問題なく進むことはない。項目内容にも依存するが、英語にあるが日本語にはない表現が多数存在するのである。また、その逆の場合も少なくない。たとえば、Scollon, Diener, Oishi, & Biswas-Diener (2004) は、「親しみ」、「ふれあい」、「負い目」を英語にはなく、日本語独自の表現だと指摘している。そこで、バイリンガルの方たちとの長時間にわたる論議がなされ、なんとか日本語を探し出したり、場合によっては、その項目の日本語版への採用は断念せざるを得なくなる。逆訳の真の状況は、このような紆余曲折を経る難関であることを承知して論文を読み進めることが推奨される。

### (3) 信頼性と妥当性

さて、項目がそろい、邦訳が整えば、標準化の手続きに入るが、その前に標準化の中心的概念である信頼性と妥当性についてここで簡単に紹介したい。信頼性と妥当性については、あまり実用的でない細部概念まで紹介することが多いが、ここでは実用的な概念のみを中心に紹介する。

信頼性とは、質問紙によって得られる得点が安定していることで、たとえば攻撃性格の質問紙であれば、攻撃性の高い者がいつも高い得点をとると信頼性が高い検査ということになる。信頼性の高さは信頼性係数として算出されるが、係数値の出し方はいろいろある。まず、同じ概念を測定する複数の項目が等質であることが信頼性を高めるとする観点から、 $\alpha$ 係数や、折半法 (split-half method) からスピアマン・ブラウン (Spearman-Brown) の公式を適用した信頼性係数があげられるが、最近では $\alpha$ 係数がよく用いられている。他にG-P分析 (good-poor analysis; 上位-下位分析) や尺度総得点とその構成項目との相関なども、精度は低い信頼性の高さを示すことになる。信頼性の次の観点は再現性で、ある期間をはさんで同一人物に検査を2回行い、2回の検査の相関の高さをみる再検査法 (test-retest method) がある。この場合、設定する期間がむずかしく、短いと信頼性の検討にならないし、長すぎると個人のもつその特性自体が変化する。一般に数か月の期間が用いられ、確固とした根拠はないが3か月前後がよく採用されているようである。

この信頼性について $\alpha$ 係数や再検査法での相関係数が得られた場合、数値がどれほど高いと十分な信頼性と言えるのかは明確ではない。しかし、厳しいことにこしたことはなく、.8以上の値が得られることが望ましい。

次に妥当性であるが、信頼性と比較して妥当性の検討はきわめてむずかしい。それは、妥当性としてあげられている概念の多様さと重複性をみても想像される。予測的妥当性 (predictive validity)、併存的妥当性 (concurrent validity)、内容的妥当性 (content validity)、構成概念的妥当性 (construct validity)、基準関連妥当性 (criterion-related validity)、収束的妥当性 (convergent validity)、弁別的妥当性 (discriminant validity) など数多くの概念があり、しかも、これらは内容的に大きく重複した概念になっている。この中で、最も重要で包括的な妥当性が構成概念的妥当性であろう。構成概念的妥当性は、大きく言えば、ある質問紙の測定しようとする概念がその質問紙によってどの程度正確に反映できるか、その程度を示す概念である (Aiken, 1996)。他の妥当性の概念は、ある意味ではこの構成概念的妥当性に含まれると言え、この概念の広範性からその検討は困難を極める。この妥当性を検討するにはものさしが必要になるが、そのものさしとしては何らかの基準が採用されることが多い。その基準は、

実験場面における行動であったり、仲間からの評定が採用されることが多い。つまり、ある特徴 (質問紙の測定内容) を高くもつ者は特定の行動を示し、仲間からの特定の評価が高くなるという基準を採用するのである。仲間評定法 (peer rating method) では、当該の概念そのものを提示し、多数の他者の判断結果により、妥当性を確認するという方法をとることも多い。この内容からして、上述の予測的妥当性や基準関連妥当性がここに適用できることがわかる。集団においてある特徴について際立つ小数のメンバーを集団の代表者が指名する仲間指名法 (peer nomination method) も、広く言えば、仲間評定にあたる。また、既存の同種の検査との相関を見るという併存的妥当性による方法も採用できるが、基準となる検査の妥当性が曖昧な場合が多いので注意が必要である。なお、ある検査の妥当性が高いということは、信頼性も同時に高いことを意味することを付記しておく。

## 3. わが国におけるオリジナル検査の作成

### — 項目の作成から信頼性の検討の前段階まで —

わが国独自の質問紙を作成する場合を考え、その手続きを詳細に紹介する。この節では、信頼性や妥当性の検討に入る前までの、質問紙の作成では最も重要な段階について詳述する。ここで紹介する質問紙の作成方法を学ぶことによって、既存の質問紙の良し悪しをみきわめる目も養うことができる。

#### (1) 概念の熟知

前節に外国の検査を利用した邦訳版の作成までを紹介したが、その後の方法は、わが国で独自の質問紙をつくる場合と同じであるので、その内容は独自の検査の作成と合わせて次節で紹介する。ここでは、外国の検査が利用できず、わが国独自の検査を作成する場合をとりあげ、信頼性と妥当性検討の前段階までを詳細に紹介する。実は、この前段階こそが検査の良し悪しのほとんどを決定することになるが、論文などではこの部分がほとんど取りあげられていないので、注意が必要である。また、独自の検査の作成では多くの労力が必要とされ、多人数の目で誤りなどをチェックする必要からも、複数名のチームを組んであたることが望ましい。

まず質問紙の作成では、測定したいと考える概念や特性を熟知することである。この部分の重要性はいくら強調しても強調しすぎることはない。心理学における性格関係の研究などは、研究が開始されてほどなく袋小路に入ることが多いが、その原因のほとんどは研究対象の概念が曖昧で、誰もが認める合致した内容になっていないことにある。そこで、当該領域の文献を精読し、概念の把握に努める。ここで明確な概念像がつかめず、あるい

はもともと明確な像がない場合は、検査の作成は断念した方がよい。

(2) 検査仕様の決定

(a) 自記式か、他記式か

項目の作成以外にも、検査の仕様で決定しなければならないことは多数ある。成人の場合は自記式に落ち着くが、子どもの場合は小学校3年生以下であると他者が評価する方法をとらざるを得ない。他者の場合も、学校や幼稚園の担当教員か、親かの選択がある。どちらの場合も問題があり、またどちらも同じ結果が出ることはない。測定内容や状況により、そのつど精度の高い方を選ぶしかない。

表3は、タイプA性格についてその下位特徴である競争性や焦燥・攻撃性を含めて、父母と教員が子どもについて幼児用のタイプA検査であるMYTH (Mathews Youth Test for Health; Mathews & Angulo, 1980)に回答した結果の相関を示している。父親は母親や教員との相関は低く、子どもへの評価がおぼつかないことが推測され、教員と母親の相関は全般的に高いが、焦燥・攻撃性では低い相関が認められている。

表3 親と教員の評定一致度 (Vega-Lahr & Field, 1986)

		父	母
タイプA	母	.41	
	教員	.23	.61
競争性	母	.61	
	教員	.26	.66
焦燥・攻撃性	母	.34	
	教員	.16	.11

(b) フェイスシートと教示文

質問紙においては実際の質問項目が大切であるが、フェイスシートや教示文も大切な役割を担っている。フェイスシートとは質問紙の表紙にあたり、質問紙の名前や年齢等の個人的情報の記述欄そして教示文も含んでさすことが多い。

まず質問紙の名前に注意を払いたい。測定しようとする内容を直接表すような名前をつけると被検査者の構えができて正確な回答が得られない場合がある。たとえば、攻撃性を測定する場合に攻撃性質問紙では問題があり、関連する範囲であたりさわりのない名前をあげておくのがよい。このことは教示文でも同様で、攻撃性について測定することをあらわに説明するよりは、感情表現を問うぐらいの説明に抑えるべきであろう。個人的情報については最低限にすべきで、性別や年齢は必須であるが、調査目的によって職業や既婚かどうかを問うことになる。

利用するかもわからない程度の意図で数多くの事柄をたずねるのは、被検査者の負担になったり、たずねる事柄によっては防衛心を高めてしまう。

教示文では、質問紙の場合、記述しておく必要のある事柄がいくつか指摘される。まず、回答内容のプライバシーは完全に守られることを約束する。これが約束されなければ、正直な回答など得られるはずがない。次に、回答の主観性や一般性の構えを促すことである。構成概念を扱う質問紙では、状況によって回答が異なることは当然であるが、各項目でその状況を細かく記述することはないので、普段の一般的な回答を主観的に行う姿勢をつくる必要がある。考えすぎないようにという教示内容も入れるが、これもこの姿勢を促すためである。他には、順番にすべての項目へ回答することなど記入上の細かい指示などが盛り込まれる。

フェイスシートではないが、すべての項目のあとには記入もれの項目がないか、また内容や状況に応じて協力への謝辞なども簡単に付与することも考えたい。とにかく、質問紙の場合は、集団の中で各自が一人で行うことがほとんどであるから、すこしでもエラーを減らし、被検査者が負担少なく、迷うことなく回答できるようにすることが必要である。

(c) 記名か、無記名か

一般に質問紙では回答者の名前を記入するが、調査によっては無記名で実施される。問題となるのは、標準化が記名で行われ、それをある調査では無記名で行う、あるいはその逆の場合である。記名と無記名では、被検査者の構えは大きくかわる。攻撃性など社会的に望ましくない特性では記名時に自己防衛が起り、無記名であると回答への真剣さが低下する。しかし、自己防衛のマイナス点を考慮しても記名の方が精度が上がる印象をもつが、これも個々の調査状況によって異なる。標準化の過程で、記名時と無記名時で別々に標準化するのが望ましいが、そこまで考慮した試みはなく、このあたりが質問紙の標準化の限界かと思われる。

(d) 回答方法 一選択肢数と言葉一

回答方法では、段階点を定めたりカートタイプ (Likert type) の回答方法をとることが多い。その場合、選択肢の数はパラメトリックな統計分析の使用に耐えるよう4件以上が望ましいが、選択段階が多すぎても段階間の差違が曖昧で意味がなくなる。また、段階すべてに選択用語を付与する場合、段階間の心理的距離は心理学的にも統計学的にも等間隔であることが望ましい。たとえば、1 まったくあてはまらない 2 あまりあてはまらない 3 どちらともいえない 4 だいたいあてはまる 5 非常によくあてはまるの5段階の場合、1と2の心理的な距離は4と5の距離と等しいことが望まれる。

これらの条件を満たす選択段階が決まったとしても、

人によっては高い目につけたり、低い目につけたり、あるいは平均につけたりしがちな、恒常的エラー (constant error) がこの種の検査ではつきものである。また、個々人の違いだけでなく、国によっても選択方法に特徴があるようで、たとえば、Chen, Lee, & Stevenson (1995) は、日本、台湾、米国、カナダで、7 選択肢の場合に、中央選択肢と両端選択肢を選ぶ傾向を指摘しているが、わが国では自己主張の低い国民性を反映してか中央の選択頻度が高く、反対に米国では両端の選択頻度が高くなっている。これは質問紙の拭いようのないエラーのひとつであると考えられるが、この領域での国際比較の障壁がここに明確に認められる。この種のエラーは再検査の場合にも認められ、短期間で繰り返される 2 度目の検査では種々の再検査効果が指摘され、2 回目の検査で適応の良い方向へ回答の変化が生じる傾向などはその一例である (たとえば、Windle, 1954)。多段階の評定は微妙な差違をとらえる利点があるが、その反面このようなエラーが数多くともない、一長一短といったところである。

その他、被検査者の受検の構えがかたよらないように、調査特性とは無関係な項目や逆方向から (いいえと答えるほど得点大) 答える逆転項目を入れることも大切である。被検査者の虚偽反応傾向をチェックするための項目なども可能なら入れたいが、この項目自体の内容が難しいので最近の質問紙ではあまり採用されていない。

### (3) 標準化のための質問紙原型の作成

概念を把握した後は、実際の項目の作成や収集に入る。項目の作成では、測定する概念を把握し、その概念にかかわる研究領域に精通した研究者が、独自にあるいは過去の関連質問紙を参照しながら項目を作成したり、多数名の一般回答者から項目を収集して精選するなど、多様な方法がある。いずれの方法においても、細心の注意を要する事柄が数多く指摘されるが、紙幅の関係から本論文ではその内容は割愛したい。作成した項目については、重複項目の整理や文表現の適切性などを修正し、イメージする最終版の項目よりも多めの項目を準備する。いずれにしても、ここで用意する項目が検査の良し悪しを決定する最大の要因となることには留意されたい。

教示文その他の仕様を含め、これで検査の原型が出来上がったことになるが、信頼性の検討に入る前に予備的な調査をする必要があり、この予備調査からは外国の検査をもとに日本版をつくる場合も同じ手続きとなる。この予備調査は、項目文を中心にその理解の容易さを確認するとともに、回答のかたよりをなくすためのもので、できれば集団ではなく個別に検査を実施し、教示や項目のわかりにくさなどの情報をそのつど詳細に収集したい。また調査対象者も、質問紙が対象とする年齢層や男女、職種などに広く対応するよう、少なくとも 50 名ほどの

データの収集をめざしたい。

収集されたデータから、まず項目表現などの問題を探り、同時に選択肢ごとの回答人数を集計し、回答のかたよりをみる。回答は、たとえば 5 件法であれば、真ん中を最多度数としてそこから両端に度数が順次減少することを理想とするが、この段階では厳密に考える必要はなく、かたよりの大きな項目の内容や表現方法を語尾や副詞を中心に修正する。

## 4. わが国におけるオリジナル検査の作成

### —信頼性と妥当性の検討—

次に、標準化の中心となる信頼性と妥当性の検討に入る。この段階までに質問紙の良し悪しの大半が決定しているはずであり、後は作成されたものがどれほどの出来映えかを中心にチェックする方法を適用することになる。しかし、その方法は煩雑であり、ここでその詳細にふれる。

#### (1) 調査の実施

質問紙では調査は集団で実施されることが多いが、この実施方法によって回答結果が大きく左右される。実施時の注意は次節で説明するが、調査人数は十分にとり、必要な分析区分で分析を行うことを可能にする必要がある。多くの質問紙では、男女ごとの分析は必須であろうし、子どもの検査では年齢段階の区別も必要となる。この区別を行わずにまとめて分析をしている質問紙ほど精度が落ちるが、調査対象の人数が少なければ分析対象の細分化を行うことはできない。

#### (2) 因子的妥当性と信頼性の検討

信頼性の検討に入る前に、質問項目群がどのようなまとまりになっているのかを調べる必要がある。質問群に異なる複数のまとまりがあれば、そのまとまりごとに信頼性を調べることになる。質問紙によっては、最初から複数の概念上のまとまりを想定しているときもある。そこで、回答結果に因子分析を実施する。最近では簡単に扱える統計ソフトが出回り、因子分析なども簡単にできるが、その方法も、主成分解か主因子解か、因子数の決定方法、回転のかけ方など詳細な方法を選択する必要がある。このあたりの判断のためには因子分析に習熟する必要がある。特に因子数の決定は難しく、固有値の変化や解釈の可能性などを考慮しながらいろいろ試みて決定すべきである。もちろん、単因子構造という結果も多いから、あえて複数の因子に分ける必要はない。また、解釈可能性を前面に出す研究がしばしば見かけられるが、因子を構成する負荷量の低さやまとまりの悪さから次に検討する信頼性や妥当性が低くなるので、解釈可能性の扱

いには注意を要する。

とにかく因子分析は項目のまとまりに目安をつけるほどに考えるべきで、出されたまとまりで良いかどうかは、信頼性と妥当性の高さによって決定される。そこで、因子が確認されると、因子の命名とともに、因子負荷量の大きさなどを考慮して各因子を構成する項目を決定する。この後ようやく、抽出された因子（尺度）の信頼性の検討に入ることができる。この信頼性は抽出された因子ごとに行い、先に紹介したように、信頼性係数を出すために $\alpha$ 係数や再検査法による相関係数の算出が行われる。信頼性係数はいずれの場合も高いことが求められるが、信頼性の検討方法ごとにその結果は微妙に異なる（表4参照）。

表4 日本版幼児用タイプA検査（MYTH）の信頼性（山崎・菊野，1990から改変）

	$\alpha$ 係数	折半法	再検査法
タイプA			.91
競争性	.87	.77	.87
焦燥・攻撃性	.91	.92	.89

また、信頼性の他に得点の分布形態をこの段階で調べておく必要がある。正規分布していることが望ましいが、正規性が得られなくとも分布が単峰性で尖度や歪度が正規性からそれほどずれていなければ、パラメトリックな統計検定で処理するのにさほど問題はない。

### (3) 構成概念的妥当性の検討

妥当性の中では構成概念的妥当性の検討が最も重要となることは先述した。しかもこの検討が最も難しく、妥当性の基準となるものの設定が難題である。ここでは、主として実験的方法や仲間評定法を利用することが推奨

される。これらの方法では、検査で測定しようとする特性の高い者が示すとされる行動特徴について、この検査の得点の高い者ほどその特徴をよく示すか、また仲間からその特徴が高く評定されるかどうか調べられる。この場合、採用する行動特徴の選択には注意することが必要で、過去の研究で周知となっている事実を反映させなければならない。

図1は、日本版幼児用タイプA検査（MYTH）の構成概念的妥当性の検討の一部で、幼児を対象にタイプA得点の中央値を境にタイプA群とB群を設定し、3つの課題に対する遂行時間を比較している。これらは、名前等の記入された用紙を設置された郵便箱に入れてくるまでの（歩行）時間(a)、2つの絵の異同判断における反応時間(b)、そして、木片入れを2人で競争する場合の課題遂行時間(c)の平均値を群ごとに示している。いずれの課題でもタイプA群がB群より時間が短くなり、この検査の妥当性の一部の証明となっているが、これらの課題は、タイプA児はB児よりも遂行時間が短くなること周知の事実であることを前提とした検討である。

構成概念的妥当性は検査の中核ともいえる部分で、この妥当性が確認されていなければ、何を測定しているのかわからなくなる。

## 5. 質問紙の実施ならびに分析時の諸注意

質問紙の標準化が完成し、ようやく質問紙を使用することが可能となった。ここでは実際に質問紙を使う場合を想定し、その際注意しなければならない諸点を整理して紹介する。

### (1) 集団実施時の諸注意

本論文の最初にも説明したが、質問紙にはエラーがつきものである。エラーをなくすことはできないが、少し

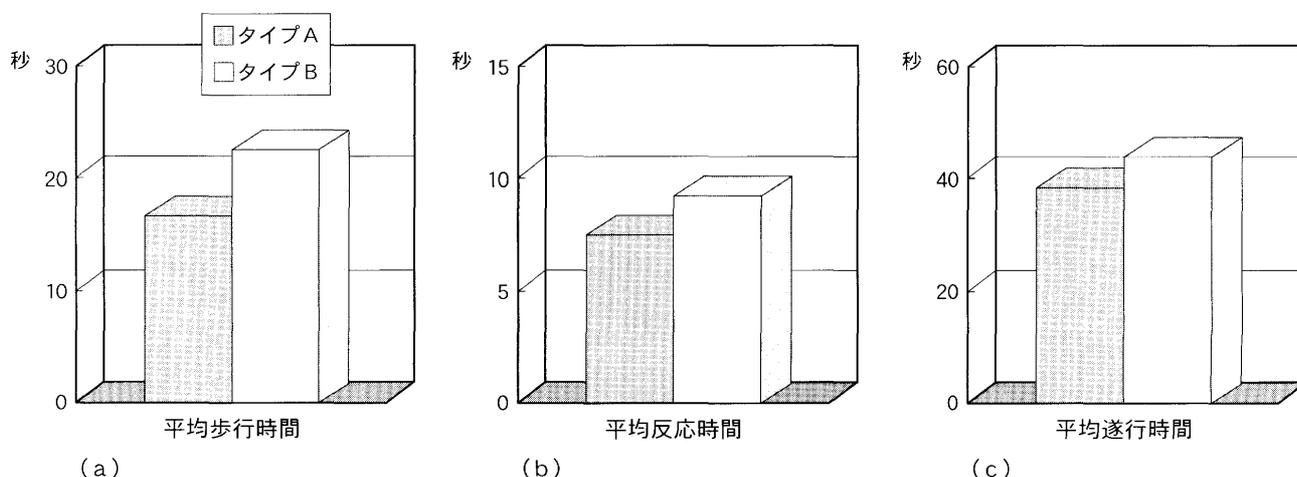


図1 タイプA児とB児の平均課題遂行時間：(a)郵便課題における歩行時間、(b)異同判断課題における反応時間、(c)木片入れ課題における遂行時間（山崎・菊野，1990）

でも減らすために、実施上気をつけなければならないことが数多くある。

質問紙はほとんどの場合集団で実施することになるが、大学などでは授業の時間を使って実施することが多い。この場合、その授業の担当の教員が検査を実施するのと、卒業論文などで学生が実施するのとでは、受検者の態度がかわり、教員が実施する方が学生は真剣に回答する。他に、単位の一環としてするのか、授業の最初にするのか、最後にするのかによっても受検者の態度は大きく変わる。また、複数の質問紙を一度に行うときは、実施順序は大きな影響因となるから、順序効果を相殺するなどの工夫が必要である。さらに、他者評価で幼稚園や小学校の教員が受け持ちの子どもについて評価をする場合などでも、せめて夏を迎えるころでないといふと教員は子どもの特徴を十分につかんでいないことなど、細心の注意を必要とする事項が山積みである。

数百人の大集団にざっと質問紙を配って、喧噪の中で自由に回答させ、なだれるように回収する調査をよくみかけるが、やるだけ無駄な調査になる可能性が大きい。

## (2) 複数の質問紙を実施するときの諸問題

質問紙を実施する場合、複数の質問紙を同時に実施することが多い。これは、2つ以上の特性や事柄を比較したり、関連を調べたりすることが多いことからである。しかし、実施する質問紙の数が多ければ多いほど問題が生まれる。

まず質問紙の数がふえるとそれだけ項目数が多くなるので、回答者への負担がふえる。多数の項目を一度にしようとする、検査の精度が落ちるのは言うまでもない。このようなときは、別の日を設けて質問紙を実施すべきであろう。また、質問紙の実施順序には注意する必要がある。実施順序が先であるとか、後であるとか、推奨される順序がある質問紙でない場合は、実施順序を無作為にしたり、前後同数にしたりして、順序効果を相殺しなければならない。特に後で行う検査は前の検査からの影響を受けることは避けられない。

さらに指摘すると、複数の質問紙を実施する場合、類似した特性間で比較することがよく行われるが、この場合の質問項目の構成には注意しなければならない。たとえば、攻撃性と抑うつ傾向の関連を調べる場合、攻撃性を構成する項目と抑うつ傾向を構成する項目に一部類似したものが入っていたりすると、当然のことながら、2つの特徴の関連が高まる。これは、はじめから関連があるように作っておいて、関連があるという結論をわざわざ出すようなものである。標準化された質問紙を使う場合、その内容を変えることはないので、このような問題を見過ごしてしまうことが多い。

以上は実施時の諸問題であるが、結果の分析に際して

も複数の質問紙の場合にはいろいろと工夫が必要になる。関連をみる場合も、単なる相関や群差の分析ではなく、調べる特性が多い場合や、その特性間に何らかのモデルが想定されれば、多変量解析、とくに共分散構造分析を駆使した因果関係を積極的に検討する必要がある。しかし、同時に実施した質問紙であるから、因果関係の決定はむずかしく、関連があるという結果も、ある特性から別の特性への影響を明示するものではないことには注意を要する。

## (3) 調査研究と実験研究

質問紙では回答内容そのものを最終データとする調査研究が多い。実施が終われば分析であるが、質問紙は実施が簡単なのでにせめて分析には手間暇をかけ、多面的に行いたい。単なる平均値や項目への回答人数、あるいは相関分析などに終始するのではなく、上述のように、多変量解析の手法も存分に駆使して結果を処理したい。

質問紙調査ではその調査を利用して実験研究などに入ることよくある。たとえば、ある性格特性について、その高い人と低い人で何らかの行動特徴を調べたいとする。この場合、高い人と低い人を弁別するために質問紙を利用することが多い。弁別は質問紙の代表値を用いて行われるが、中央値基準でその上下2群を設定したり、基準を厳しくして中央値や平均値から上下に1標準偏差分離して、その離れた上下で2群を設定したりして実験を行う。厳しい基準が望ましいが、1標準偏差分離すると調査したデータの70%弱が利用できなくなるほど厳しい研究となる。この場合の結果の分析では、分散分析を中心とした群差の比較を行うことが多く、相関分析などのように関連の大きさを測定することはほとんどない。分散分析では交互作用が確認しやすい長所があったが、最近では重回帰分析でも交互作用を確認する手法が考案され(Aiken & West, 1991)、実験の労力の大きさもわざわざいか、最近はこの種の実験的研究は敬遠される傾向にある。

## 6. 標準化に終わりはなし

これまでに説明してきたように、質問紙検査の利用はそれほど簡単なものではない。その作成や標準化にいたってはなおさら煩雑な仕事となる。完成された検査はその出来映えにいろいろと批判が浴びせられるが、一度でも質問紙を標準化したことがある者は、その苦労の方が先に脳裏に浮かび批判の手をゆるめてしまうほどの大仕事である。しかも、検査の標準化というものはこれで終わりという時点がない。つまり、完璧という検査はあり得ず、何らかの欠点を備えているということである。できれば、精度を高めるために時をおいて検査の見直し

を実施して欲しいのだが、そこまで手をかける質問紙にはめったに出会わない。

質問紙は調査や研究の道具である。その作成は困難をきわめるが、ひとたび出来上がれば、多くの人に利用され、貢献度の高い道具となる。今後も、精度の高い質問紙が、利用しやすい環境で提供されることが期待される。

## 引用文献

- Aiken, L.R. (1996). *Rating scales and checklist: Evaluating behavior, personality, and attitudes*. New York: John Wiley & Sons.
- Aiken, L.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage Publications.
- 安藤明人・曾我祥子・山崎勝之・島井哲志・嶋田洋徳・宇津木成介・大芦 治・坂井明子 (1999). 日本版 Buss-Perry 攻撃性質問紙 (BAQ) の作成と妥当性, 信頼性の検討 心理学研究, 70, 384 - 392.
- Buss, A.H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, 63, 452 - 459.
- Chen, C., Lee, S., & Stevenson, H.W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Measurement*, 48, 29 - 41.
- Jenkins, C.D., Zyzanski, S.J., & Rosenman, R.H. (1979). *Manual for the Jenkins Activity Survey*. New York: Psychological Corporation.
- 加賀多一 (1979). 受身-攻撃型人格とその臨床 原俊夫・鹿野達男 (編), 攻撃性 (pp.51 - 76) 岩崎学術出版社.
- Matthews, K.A., & Angulo, J. (1980). Measurement of the Type A behavior pattern in children: Assessment of children's competitiveness, impatience-anger, and aggression. *Child Development*, 51, 466 - 475.
- Scollon, C.N., Diener, E., Oishi, S., & Biswas-Diener, R. (2004). Emotions across cultures and methods. *Journal of Cross-Cultural Psychology*, 35, 304 - 326.
- Vega-Lahr, N., & Field, T.M. (1986). Type A behavior in preschool children. *Child Development*, 57, 1333 - 1348.
- 山崎勝之・菊野春雄 (1990). 日本語版幼児用 Type A 検査 (MYTH) の作成 心理学研究, 61, 155 - 161.
- 山崎勝之・田中雄二・宮田 洋 (1992). 日本版成人用タイプA質問紙 (KG 式日常生活質問紙) - 標準化の過程と実施・採点方法 - タイプA, 3, 33 - 45.
- Windle, C. (1954). Test-retest effect on personality questionnaire. *Educational and Psychological Measurement*, 14, 617 - 633.

# **Problems of the development and administration of questionnaires in psychological survey studies**

Katsuyuki YAMASAKI\* and Kanako UCHIDA\*\*

(Key words: questionnaire, survey study, standardization, reliability, validity)

Questionnaires are often administered in survey studies in psychology. The reason why they are used more frequently than the other methods, such as observations and interviews, is that they have greater merits in time and costs. However, the questionnaires have low accuracy to measure psychological characteristics, specifically psychological constructs, such as personality, affection, and cognition. So, in order to improve their measurement accuracy, it becomes essential to be careful to develop and administer them. In this paper, after indicating the low measurement accuracy of the questionnaires, we presented various methodological considerations when developing the Japanese versions of the questionnaires based on their original versions in foreign countries, and furthermore the original Japanese versions. Above all, we explained in great detail the process to develop the questionnaires that are original to Japan. The process contained the precise understanding of concepts to measure, how to answer the questionnaires (e.g., self-report or report by others), collection of items, and general standardization methods concerning reliability and validity. In addition, we called attention to a number of points to notice when administering the questionnaires to groups. Lastly, we closed the paper, emphasizing that the attempt to enhance the level of standardization is never ending.

---

\* Department of Human Development, Naruto University of Education

\*\* Graduate School of School Education, Naruto University of Education (Research student)