

妥当性を重視した尺度構成の試み ——以下の3論文の序論に代えて——

村 上 隆

以下に続く3つの論文は、1999年度前期に著者が担当した大学院の授業「教育統計学研究I」において行われた研究の成果をまとめたものである。

この授業では、3つのグループに分かれた受講者が、それぞれのグループで決めた測定対象となる特性(trait)について、過去の研究を調べた上で理論的フレームワークを定め、それにもとづいて項目を書き、それらを含む質問紙を実施して、得られたデータに基づいて尺度を構成することを求められた。その際、特に次の点を強調した。

- (1) 尺度の信頼性だけでなく妥当性、特に、構成概念妥当性を重視すること。
- (2) そのために、調査票には今回構成する尺度に含める項目だけでなく、妥当性検討(validation)のために既成の尺度の項目も含めること。
- (3) 妥当性検討のための尺度を適切に選択するために、測定しようとする特性に関する過去の研究の蓄積に、十分な注意を払うこと。(このことは、トピックとして選ばれる特性の選択を大いに制約する結果にはなったが。)

こうした要請をしたこと的背景には、以下のような思いがあった。

今日、社会・教育系心理学において、質問紙を用いた測定尺度の使用は隆盛を極めているように見える。そこでは、従来のような項目分析や探索的因子分析のような方法を用いた、いわゆる内的整合性(internal consistency)にもとづく尺度(それには、 α 係数が添えられるのが普通である)が構成される。最近では、構成された尺度が測定していると想定される特性の間の因果関係を想定したモデルが立てられ、パス解析等の手法でそれが「検証」されることもある。さらに、それらを一体化した方法としての潜在方程式モデル(共分散構造分析)の使用も目立つようになってきた。すなわち、数理統計学的手法の適用に関しては、大きな進歩が認められると言ってよい。

しかしながら、こうした最近の動向を、心理尺度が本

来依拠すべき信頼性と妥当性という評価基準に照らして考えてみたとき、はたして無条件に歓迎すべきものと言えるであろうか。

まず、信頼性の概念は、 α 係数等を根拠にして評価される内的整合性と、再検査信頼性にもとづく安定性(stability)とに分離して考えられる必要があり、1回限りの調査では後者を評価することができない。つまり、尺度の信頼性を評価するには、内的整合性と安定性の双方を評価する必要があり、もし、ある尺度が後者の点で十分でなければ、その尺度が反映しているものは、特性でなく状態(state)である、といったことが判断される。

ただ、この点については、近年、わが国でも認識が深まってきており、再検査信頼性係数が併記された論文が増えてきている。

しかしながら、妥当性の方はどうであろうか。測定の対象を、直接観測できない構成概念(construct)と考え、それについての理論から幾つかの命題(仮説)を導き、それらの検証を通じて尺度の妥当性を徐々に確立していくという構成概念妥当性の考え方は、わが国ではまだ十分に理解されるに至っていないように思われる。実際、多くの測定に関するテキストには、いまだに妥当性といえば、基本的に単一の外在基準(external criterion)との相関に基づく基準関連妥当性と、項目内容を専門家が適切性(relevance)と代表性(representativeness)にもとづいて評価するという内容的妥当性だけが論じられている。せいぜい、構成概念妥当性の形式的な評価方法としての多特性-多方法行列が論じられていれば良いほうである。

著者自身、こういった面については決して十分なトレーニングを受ける機会をもつことはできずに現在に至ってしまった。十数年前、Cronbach and Meehl (1955)の論文を、数日かけてやっとの思いで通読したときの驚きは大変なものであり、それ以来、折りに触れて大学院の授業では、構成概念妥当性をトピックとして取り上げるようにしてきた。その過程でも、Messick (1975)や

Messick (1981) から受けた衝撃は大きく、これらの論文には、二読、三読しても汲み尽くせない含蓄があると思う。

もちろん、理論に関する知識がありさえすれば、乏しいデータからでも妥当性検討が可能になるというものではない。「理論から導かれた仮説」を経験的に検証するためには、そのための証拠が収集されていなければならない。そうした証拠として代表的なものは、理論的に意味があると考えられる外部の変数との相関の検討である。(それだけではないが。)理論的に高い相関が得られるべき変数との相関が実際に高ければ、それを収束的証拠 (convergent evidence) といい、理論的に低い相関が予想されるところで実際に低い相関が得られれば、それは弁別的証拠 (discriminant evidence) と呼ばれる。(ここで、「外部の変数」と呼んだものは、もはや外的基準と呼ぶわけにはいかない。相関が高ければ良いというものではないからである。)

そうした仮説と証拠の数には限りがないから、構成概念妥当性検討は、事実上、エンドレスな過程となる。もちろん、実際的には、どこかの段階で検討を打ち切って、研究論文の出版に向かわざるを得ないわけであるが、それにしても、論文にかなりの数の検証結果が載せられていなければ、それは尺度の提案のために、十分なものとは言えないであろう。

Murakami, Suzuki, and Nogami (2000) は、限られた分野に関してではあるが、わが国の研究論文において、こうした証拠の数が欧米の論文より決定的に少ないことを見出した。こうした「証拠不十分な」段階での出版の伝統は、どこかで断ち切らなければならない。

また、数年前、ある経営学者から、「日本の心理測定尺度には、信頼性の検討はなされているけれども、妥当性の検討がほとんどなされていないがそれはなぜか？」とたずねられたことも、こうした方向に踏み出さなければならないと考えた大きな理由である。そのときは、「日本の研究は、実質的に大学院生によって支えられており、彼らには、時間と資源を必要とする妥当性検討を行う余裕も社会的能力もないのが原因ではないか？」などと、はなはだ責任回避的な答えをしてしまったのであるが。

ただ、こうした問題の解決には、当面の研究の担い手である大学院生に対してキャンペーンするのも一つの方法であろう。著者が我が身の不始末も省みず、受講者に最初に述べたような要請をしたのには、こうした理由があったのである。

実際に授業を始めてみると、受講者にはいつにない積極的な参加意識が見られ、この試みには一定の成果があっ

たように思われる。この点は、かつて学部で行っていた「教育心理学実験演習一測定一」の実績と比べても、やはり大学院生の実力には目を見張るものがあった。著者は、ときどき全受講者の前で簡単な方向づけのための講義を行ったり、どこかのグループのディスカッションに参加したり、授業後に各グループのリーダーやサブリーダーから進行状況を聴いたりしたが、基本的に各グループの自主的進行に任せていた。リーダー、サブリーダーの性格、メンバーの構成によって、進行の状況が異なっていくことも興味深く眺めていた。(ご本人たちにとっては、興味深いどころではなかつただろうが。)完成度にはまだ問題はあろうが(構成概念妥当性検討の性質から考えれば、そもそも「完成」は不可能であろう)、3つともこのような形で刊行されたことを喜びたい。また、そのうちの少なくとも2つについては、各グループの有志の手で更なる展開が見られるのは心強い。

ところで、ここで用いられている方法、たとえば伝統的に行われている全項目の総点を用いた項目分析を省略していきなり因子分析している点、あるいは、もっと新しい数理モデルを使っていない点等、我が同業者からはいろいろと不審の目で見られるところもあろうかと思う。そうした点については、著者自身、こうした行き方を支持する一定の論拠をもっているつもりであるが、詳細については、機会を改めることにしたいと考える。

最後になったが、著者の一方的な授業計画に、辛抱強くつきあってくれた受講生諸君に心からお礼申し上げたい。

文 献

- Cronbach, L. J., & Meehl, P. E. (1955) Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Messick, S. (1975) The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30, 955-966.
- Messick, S. (1981) Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin*, 89, 575-588.
- Murakami, T., Suzuki, Y., and Nogami, Y. (2000) Issues in attitudes and personality measurement in educational psychology: Validity of measurement scales and causal modeling. *教育心理学年報*, 39, 200-216.

(2000年9月16日 受稿)