

誌上シンポジウム

テストの常識と非常識

企画者：村 上 隆

企画のねらい

昨年度から開始された本紀要の誌上シンポジウムの2回目は、当教室の通称「統計系」で企画することになった。教育心理学における統計的な話題にこと欠くわけでは決してないのであるが、近年この領域における数理的な面の高度化は著しく、教育心理学研究者の中に多くの読者を見出すことが期待されるこの誌上シンポジウムにふさわしいテーマは、そう容易には見出せない。結局、内容的には、比較的広く関心を集めそうなトピックスであるテスト、あるいはテスト理論をとりあげることにした。その上で、異分野あるいはそれに近い研究者を提案者とし、伝統的な心理測定研究者である本学の野口裕之助教授を討論者に配して、その間の対話の実現を目指すこととした。

さて、異分野の研究者としては、山梨医科大学の平野光昭教授と、関西医科大学の有田清三郎教授にご寄稿をお願いした。奇しくも、お二人はともに単科の医大に所属される数学者であり、本来の研究テーマの追求のかたわら、テストの研究を進められてきたという共通点をおもちである。しかしながら、とりあげられる論点は、お二人の間で全く異なっている。

まず、平野教授は、国立大学入学者選抜研究連絡協議会を舞台に、入学者の多面的な追跡調査にもとづく入試の妥当性の研究を精力的に発表し続けられる一方、入試研究に対して、常に冷静で透徹した批判を展開してこられた方である。今回は、入試データの解析では常に問題になる、選抜効果と回帰効果を中心に論じていただく。この二つは、初等的な記述統計学の課程において、まず必ず扱われる素材であり、いわば関連データを扱う際の常識である。しかしそれらが無視される結果、前者は入試の妥当性について過度の失望を、後者は特にいわゆるユニーク入試の結果について極度に楽天的な評価をもたらしてきた。ここでは、事柄の性質上、入試データをそのままご提示いただくわけにはいかないが、事例に近い人工データにもとづいて、問題点が明らかにされる。

ところで、こうした「誤り」の頻発は、単に、入試研究がしばしば統計的知識に乏しい担当者によってなされるためだけではない。この二つの効果は、特にその両方が共存している場合には、専門家によってもしばしば見逃される。つまり、常識は決して常識となっていないのである。当然のことながら、選抜効果も回帰効果も入試研究にのみ存在するのではない。一般の教育心理学研究の中にも、気付かぬうちに忍び込んでくる。検定によって高度に有意な結果を得たことを喜ぶ前に、データ分析の手続きの中にこの種の誤りがないかどうか確かめること、あるいは確かめる「眼力」をもつことは、教育心理学者にとって必須の素養と言えよう。また、こうした誤りを犯しえないところまで受講者を鍛え上げることが、大学における教育統計学の担当者の責任でもあろう。討論者には、特にこのあたりを中心に議論していただければと考えている。

次に、有田教授には、多肢選択形式のテストにおける、受験者の部分的知識の効果について述べていただく。統計理論に詳しくない人達は、客観テストは「まぐれ」でも正答できてしまうから不合理だ、と単純に考えることがある。それに対して、計量心理学者は、選択肢の数は四つか五つであっても、問題項目の数を十分多くすれば、「まぐれ」による得点の影響はテストの信頼性と妥当性を著しく危うく

するものではないことを示してきた。また、彼らは論文体テストにも問題の偏りや採点者間の不一致等の問題が多いことを指摘して、客観テストをより強く支持する議論を展開することも多い。しかしながら、この議論には、(1) 受験者は、問いに対して正答するための知識・技能を完全にもっているか全くもたないかのどちらかである、(2) 知識・技能をもつ受験者はかならず正答を選ぶ、(3) 知識・技能をもたない受験者は、ランダムに一つの選択肢を選ぶ、という三つの前提がある。しかしながら、実際には大部分の受験者は、問いに対して全くの無知の状態にあるわけではなく、また、仮にそうであったとしても、受験技術その他により、幾つかの選択肢を排除できることが多い。計量心理学者が、これらの点を全く無視していたわけではない。しかし、それがテストの得点の上にもたらす影響の正確な査定、さらには、それが出題傾向に間接的に及ぼす影響といった面については、必ずしも十分な議論がなされてこなかったのではなかろうか。有田教授は、この部分的知識の効果について精緻な理論を展開されているが、ここではそのエッセンスをわかりやすい形で紹介していただく。

なお、これは単なる理論のための理論ではなく、現在の医師国家試験の在り方に対する強い懸念の表明としてなされた研究が端緒となっている。また、有田教授のグループが、この内容について数回にわたって日本教育心理学会で発表されたことは、この問題への教育心理学者の何らかの対応が期待されることと思われる。筆者の知る限り、教育心理学者からの明確な反応はまだほとんどなされておらず、それが、有田教授に今回寄稿をお願いした一つの理由である。

第三の論文は筆者が執筆する。筆者は、学生時代を実験心理学の世界の中で過ごしたから、テスト理論は、本学に着任して以後主として独学で学んだ。その上で、理論の講義と、実習や共同研究を通じて応用的な仕事にも携わってきた。その際、理論の中に大変教えにくい部分があることに気づいたし、応用研究において決断に迷う事態にしばしば出会った。前者の代表は、古典的テスト理論の無作為誤差の原因の説明および、構成概念妥当性の考え方である。後者に関しては、因子分析における因子数の決定の問題がある。テスト理論とその応用について「まっとうな」トレーニングを受けてこなかった筆者は、それらを自分の知識がたりないせいであると思ってきたのであるが、今ではそれらが理論の、あるいは理論の枠組みの難点である、とほぼ自信をもって言うことができる。要するに、心理学的特性は、物理量や筆者がかつて携わっていた精神物理的測定値とは違った性質をもっている。それは、ガリレオの言う第二性質、すなわち、認識主体によって異なり得るものである。また、個人差の次元は事実上無数にあり、それを少数の特性にまとめあげているのは、意識的、無意識的になされる項目の選択と、尺度構成の手続きである。それは、本質的に連続な様相をもつ現実の世界を、離散的な概念へとまとめあげている言語の働きと似ており、それと同様に、根底には言語と同様に（ソシユールの意味での）恣意性が存在している、というのがおおよその論旨である。

結論的には、従来、必ずしも論理的に正当化されないままに広く用いられてきた、項目水準の「因子分析」にもとづく尺度構成手続きを肯定しており、その意味で特に斬新な方向を示唆するものではない。しかし、背景にこの種の理念をもっていることは、理論的にも経験的にもより柔軟な行き方が可能になる点で意味があるものと考えている。

有田教授の議論と拙論との共通点は、ともに「テストで本当に意味のある個人差が測れるのか」という「素人」の常識から出発して、ある意味でその疑問を正当化する方向の結論を導いていることである。この点は、平野教授によってとりあげられた、選抜効果と回帰効果の場合とは、常識と非常識の所在が逆である。選抜効果と回帰効果の場合には、専門家の常識が「素人」の非常識を打破できないことが問題であった。

もちろん、以上の議論は過度に単純化されたものであるが、討論者の側からは、歴史的必然性や、実用的な観点等から従来理論を擁護する議論を期待したい。そのことは一方で、なぜ通常の古典的テスト理論（あるいは、項目反応理論のような新しいテスト理論）が現在のような形をとっているかを、少し違った面から明らかにすることにもつながるであろう。

誌 上 シ ン ポ ジ ウ ム

こういった点で、今回のシンポジウムは多分に教育的側面を含むものとなる。そうした内容を研究紀要において扱うことを、筆者はある意味で自然なことと考えている。本来、この領域はそれ自体として自立して存在し得るものでなく、内容領域との密接なかわりを必要としている。そうであるとすれば、いわゆる啓蒙活動が当然に研究活動の一環とならざるを得ないからである。

(1992年6月記)

この「企画のねらい」の執筆後、筆者はメニエル氏病で倒れ、原稿を執筆することができなかつたため、他の執筆者、および討論者に多大なご迷惑をおかけすることになりました。今年度、新たに改稿された論文をお寄せ下さったお二人の執筆者にはお礼のことばもありません。また、結果的に、誌上シンポジウムを1年休載する結果になったことについては、教室スタッフに、この場を借りてお詫びしたいと思います。

(1993年9月記)