

小論文、面接の評価方法

—米国イリノイ大学 在外研究報告—

研究開発部進学適性研究部門 豊 田 秀 樹

イリノイ大学は、米国の中部にあるシカゴから南へ500キロほど下ったアバナ、シャンペーンという街にある。というよりも、イリノイ大学があるためにアバナ、シャンペーンという街が成り立っていると言った方がいいかもしれない。街のかなりの部分を研究棟が占め、その周りに大学関係者が住み、街を一步出ると地平線の彼方まで畑が続いている。小さな街だが、空港もあり、日本からの友人を空港で見送って、それから20分後の授業に間に合ってしまうという絶好の環境にイリノイ大学はある。

住んでいる人たちは、礼儀正しく気さくで、自分たちがカントリーピープルであることに誇りを持っている。日本にいてアメリカの大学の噂を聞くと、大抵は、日本よりアメリカの学生の方が熱心に勉強するというものだが、これは半分しか当たっていない。確かに大学院生にもなると将来のことを考えて必死に勉強していたようだが、学部生、特に1、2年生はよく遊んでいた。

私が部屋を与えられていた研究棟の前は酒場だったが、夜になると日本と同じようににぎわっていた。また、大学対抗のフットボールやバスケットボールの試合のある日は、寄宿舍の周りは大騒ぎである。親元を離れている学生がほとんどだったから、そのせいもあるだろうが、イリノイの学生たちは勉強も遊びも本当に青春を楽しんでいるようにみえた。

私は文部省長期在外研究員として平成4年7月から平成5年5月まで、この愛すべきイリノイ大学で客員研究員として心理教育測定の研究をしてきた。私が頻繁に出入りしていた教育心理学科は教育評価、教育統計に関しては米国でも有数のレベルにあり、古くはクロンバック教授たちがテスト理論を構築した学部としての伝統を誇っている。本稿では、イリノイ大学を中心に開発され、その後全米に広がっていった「一般化可能性理論」という小論文、面接の評価方法について紹介する。

小論文や面接の試験を実施した際に

は、様々な形式のデータが得られる。ここでは最も一般的な「受験者」×「採点委員」×「課題（あるいは読み返し）」という3相データを考える。3つの相の組み合わせでは1つの得点しかないから、このデータはちょうど分散分析における変量モデルの「繰り返しのない3因子実験」と見なすことができる。

繰り返しのない3因子実験は

$$y_{ikl} = \mu + a_i + b_k + c_l + (ab)_{jk} + (ac)_{jl} + (bc)_{kl} + e_{jkl}$$

のように書かれる。右辺には8つの項が出てきたが、最初の μ はデータの平均値だから特に気にする必要はない。残りの7つの項は実施されたテストの性質を調べる際に、とても役に立つので、以下に一つ一つ説明する。

式の中の a_i は i 番目の受験者の実力である。テスト理論の分野ではテスト得点の分散に占める実力の分散の割合を、測定の信頼性の指標として利用する。つまり、 a_i の分散が他の項の分散よりも大きいほど測定の信頼性が高いと解釈する。

b_k は k 番目の採点委員の採点の甘さである。つまり、 b_k の分散が大きい場合は「採点委員によって、厳しい人と、甘い人がいる」ということを意味する。全員の採点委員がすべての受験者の答案を見る場合は、この分散が大きいてもあまり問題はない。ただし、現実には限られた時間内に採点処理を行うの

で、一部の採点委員が一部の受験者の答案を見る場合も多いだろう。しかし、この項の分散が大きいときには、甘い採点委員に当たった受験者が有利になるので注意が必要である。

c_l は l 番目の課題のやさしさである。この値が小さい課題ほど難しいと解釈する。全員が同じ課題を解答する場合は、この分散が大きいても構わない。しかし、一部の課題を選択して解答する場合は難しい課題を選んだ受験生が不利になる。

また、3番目の因子を読み返しとした場合には、 c_l は l 番目の読み返しにおけるやさしさを意味する。例えば、この値が段々大きくなるようなら、読み返すうちに採点委員の基準が甘くなっている証拠である。

このように、小論文や面接のデータを分散分析の変量モデルと見なすと、いろいろな解釈ができるが、興味深い解釈ができるのは、ここまで説明した3つの主効果ばかりではなく、むしろ交互作用にある。繰り返しのない2因子実験には交互作用の項はなかったが、3因子実験では繰り返しがなくとも2次の交互作用が3つも登場する。

例えば交互作用 $(ab)_{jk}$ の分散は、受験生ごとの採点委員の採点の傾向が異なる度合いと解釈される。受験生の立場からこれを解釈すると採点委員の個人的好みによって、悪く言えば、えこ

ひいきがあるということである。

ただし、大学の立場からは採点委員の好みになうことは、必ずしも誤差とはいえない面もある。そういう意味からはこの分散を実力と解釈することもできる。

交互作用 $(ac)_{jl}$ の分散は、課題ごとの受験者の点数が異なる度合いを意味している。つまり、受験者にとっては課題ごとの（例えば理科系の課題と文科系の課題ごとの）得意と不得意の差の程度と解釈する。

この分散があまり大きいようなら、何のために一つのテストにまとめているのか分からなくなるので注意が必要である。しかし、それらの課題が出題者によってきちんと吟味され、別々の課題の得意不得意を克服した受験生にこそ入学して欲しいという確固たる共通認識がある場合は、この分散が多少大きくても問題はない。

そして、交互作用 $(bc)_{kl}$ の分散が大きい場合は、課題ごとの審査委員の重視度の傾向が異なることを意味している。もし、この分散が大きいようであれば、前もって課題間の採点の基準を統一するように採点委員の間での共通認識作りが必要であろう。

最後に、 e_{jkl} の分散は、このデータを3因子の変量モデルと見なした場合の純粋な測定誤差の大きさと解釈される。もちろん、この分散が小さい方が

望ましい状態と言える。

このように、分散分析の変量モデルを使うと、面接、小論文のテスト得点を変動させる要因の大きさを分解して推定することができる。この方法を一般化可能性理論におけるG研究という。

実際には、各分散の大きさを推定した後に、来年の入試に向けてどのような改善を行ったらいいかという意思決定を行う必要がある。意思決定とは、例えば、採点委員の人数は増やす必要があるか、採点委員のグループごとの平均点を調整する必要があるか、採点の基準を更に明確にする必要があるか、一人の受験生の答案を何回読み直す必要があるか、作題の際に個々の課題の平均点をそろえるように配慮する必要があるか、採点委員が答案を読む際にどのように休憩を入れればよいか、などである。その際に用いられる方法は、一般可能性理論におけるD研究と呼ばれるが、紙面も尽きてきたのでD研究の紹介は、別の機会に譲ることにしよう。

一般化可能性理論について、応用的な側面から体系的に論じた教科書としては、

R.L.Brennan, Elements of Generalizability Theory, American College Testing, 1992.

を薦める。この本は、私がイリノイ大学に滞在していたときに修士課程の授

業で教科書として使われていた。基礎的なレベルから実際に使えるようになるまでが丁寧に解説されており、値段

も18ドルと安く、学生に評判のよい教科書だった。