

くらもと なお き
倉 元 直 樹

学位の種類	博士（教育学）
学位記番号	教 第 107 号
学位授与年月日	平成16年11月17日
学位授与の要件	学位規則第4条第2項該当
最終学歴	平成2年11月 東京大学大学院教育学研究科博士課程後期3年の課程退学
学位論文題目	ペーパーテストによる学力評価の可能性と限界 —大学入試の方法論的研究—
論文審査委員	(主査) 教授 荒井克弘 教授 宇野 忍 教授 小野寺 淑行 教授 村木英治 (教育情報学教育部)

論文内容の要旨

本論文は、序章および第1章から第7章までの本論と、結章の全9章から構成される。序章では、研究の出発点として本論文の主題となった問題の所在を述べている。わが国の近代学校制度の成立以来、大学入学者選抜は常に社会的な関心の高い問題であった。学力中心の入学者選抜制度は実質的にわが国の教育水準を支えるという機能を果たしながら、他方では教育制度そのものへの不満を生じさせる原因でもあった。ペーパーテストは選抜の主要な手段であり、批判の矢面に立たされることが多いが、ペーパーテストは教育測定・評価の手段としてその技術的發展の中で生まれてきたものである。ところが、日本では入学試験制度に対する関心の高さと対象的にペーパーテストへの技術的側面にはあまり関心が払われることがなかった。入試改革が受験者、関係者の批判への対応を専らとしてきたことからその事情は推察できる。入試制度に対する批判は、主に試験の「負の遡及効果」という教育上の認識から発しているものが少なくない。しかし、大学入試には、じつは高校以下の教育に対して正負両面の遡及効果が存在する。その事実が本論文を組み立てるうえで基本的な視座となった。

第1章では、筆者自身が調査した諸外国の大学入学者選抜制度の事情を解説し日本への示唆を述べている。わが国でも諸外国でも、近年の大学入試は高等教育の大衆化と教育の質の維持という両立し難い問題への取り組みが共通課題となっている。ただし、具体的な方略は国や地域によって異なり、韓国はわが国と類似した教育環境にあるが、受験競争の苛烈さはわが国の比ではなく、その対策もドラスティックである。北米では、一部の選抜型大学では学力以外の要素を重視した丁寧な入学者選抜が行われていて、結果的に卒業率も高い。一般の州立大学では簡単な学力指標で選抜を行い、入学後に不適格者をふるい落とすシステムとなっている。オセアニアでは、ニュージーランドで新保守主義による急進的な改革が行われたが、その結果、高等教育の水準を著しく低下させてしまった。一方、オーストラリア・クィーンズランド州で実施されている入学者選抜の判定資料作成のプロセスは、わが国の入試にとって有効なヒントが数多く含まれている。ヨーロッパもまた、伝統的な大学の理念と急速な高等教育の拡大の矛盾に苦しんでいる。ボローニャ宣言の下、EU圏内での知的人的流動性を高める政策が進められており、各国の制度に互換性を持たせる方向の改革が取られている。オランダでは、複線型の教育システムの単線化が進んでいるが、CITOが管轄する各種のテストが進路選択の場面で大きな役割を果たしている。各国の教育制度にはそれぞれ特有の歴史と環境があり、そのままではわが国の範とはなり難いが、そこに多くのヒントが隠されていることは確かである。その観点で筆者が掴み得た示唆がまとめられている。

第2章以降は数量的なデータに基づく実証的な研究が中心となっている。第2章では、教育測定の理論と多肢選択式のペーパーテキストに焦点を絞って検討を行っている。最初に、本論文のいわば縦糸とも言うべき教育測定の理論について紹介がなされた。大学入試センター試験などで取り入れられている客観式テスト、多肢選択式テストは大量の答案処理に向くと同時に、教育測定の理論との親和性が高い。前者のメリットは周知されているが、後者はあまり関心が払われてこなかったというのが著者の認識である。教育測定の理論では、アメリカ社会において、その社会的必要性の下に発達してきたと考えられるが、現状は「テスト漬け社会」というべき状況にあり、逆に、わが国では教育測定の理論が積極的に受容されることもなく、また発展することもなかった。

次に、教育測定論の1つである項目反応理論を応用した日本語基礎能力テストの開発経験を紹介し、項目反応理論に基づくテストの作成と実践例を示している。また、教育測定の理論をほとんど用いないで実施の行われた例として、大学入試における多肢選択式の共通テストである共通1次試験、大学入試センター試験をとりあげ、そのデータを実際に用いて大学入試1年後の学力水準についての分析を行った。最後に、それらの研究の結果をもとに現行のセンター試験の成績算出システムとその問題点を論じている。筆者の理解によれば、戦後日本の教育言説を支配してき

たのは「能力主義的差別教育」論にもとづく序列化批判であり、その種の批判は社会全般に流布されたある種の部分的な公平性概念を代弁しているにすぎない。筆者の依って立つ教育測定の理論は能力に基づく分配的公平性を基本概念とするために、従来の教育言説と原理的に相容れないのだという解釈が上記の根底にある。例えば、センター試験の得点調整問題を例にとれば、教育測定の理論から見れば到底合理的とはいえない方法が用いられるが、社会的にはそれがあたかも公平な方法であるかのように受容され、批判もない。これも1つの典型であるといえる。

第3章では論述式のペーパーテストについて検討を行っている。教育測定の理論から見て、採点を主観的に行う論述式テストには客観式テストにはない固有の問題が多く存在する。そこで、まず過去の実証的研究のレビューを行い、その中から高い信頼性の確保が難しいと思われる事例を見つけ、評価観点の構成を工夫すればある程度高い信頼性を確保できることを紹介した。その方法を下敷にして大学個別試験の小論文試験の採点を行ってみると、大学入試の実際の採点においても同様の結果を導くことができた。この研究により、内容領域の専門性に関わらない作文能力といった測定を目的とするならば、評定者の専門性の影響もさほど大きくないと結論した。

遡及効果の観点からは多肢選択テストは論外とされ、論述式テストが望ましいとされる。逆に、教育測定論の観点からは、論述式テストは時代遅れのものであり、多肢選択形式のテストが望ましいとされる。このジレンマを前提としながらも、信頼性を確保しつつ、遡及効果の観点を重要視した問題作成の工夫が必要であり、また可能であることも示唆している。

第4章では、テストの妥当性に関する教育測定論の研究の中で、将来に応用可能性が期待される2つの検討を試みている。最初に取り上げたのは、複数の採点者が小論文を採点するケースであり、単純に合計点ないしは平均点を用いると評価観点の多次元性が失われる。そこで多次元アルファ係数の考え方を用いて、それを補う方法を試みた。次に、ルールスペース法を用いて、項目反応理論に準じて作成されたテスト結果から学習者の学力診断情報を抽出する方法を試みた。データには第2章で用いた日本語基礎能力テストを用いた。これらの2つの方法は論述式、多肢選択式のペーパーテストの分析にそれぞれ有効であることが示されたが、そのままでは、現実の大学入試に利用するにはまだ多くの障害があることも同時に述べている。

第5章では、現実の大学入試における臨床的な問題を検討している。内容は東北大学のAO入試を扱った諸研究である。最初に、東北大学AO入試の性格として研究大学の立場から学力を大切にした選抜の重要性を論じた。次に、歯学部のAO入試については、歯学部固有の問題とそれに対する取り組みを導入初年度に関して分析した。さらに、工学部のAO入試を題材にした箇所では、選抜資料としての調査書の問題点、志願倍率の変化と志願者の質の関係を検討した。これらの研究を具体的な材料として、個々の大学入試現場の力学的なダイナミクスとペーパーテスト以外の選抜資料に関する技術論的な問題点から、現実の選抜の場面において生じる大学入試のさ

さまざまな限界と制約について考察した。

第6章では、本論文で述べたさまざまな分析をもとに、わが国の文化的土壌と現実的制約条件の中で、ペーパーテストを中心とした大学入学者選抜方法の技術的改良の可能性を論じている。まず、共通試験に供することを目的として開発された教科科目複合型総合試験問題の開発研究を紹介し、最初に問題作成上での具体的な障害、実施結果から得られた統計的な性質を整理した。次に、オーストラリア・クィーンズランド州のQCSテストをヒントに問題作成を行い、総合問題の試行調査を行った際の分析結果を示した。調査では、数学的分野、科学的分野の総合問題を用いて、物理の選択者と生物の選択者のスキルの違いを検出することに成功した。これにより試験的にはあるが、総合問題の実施が具体的に可能であることを示した。総合問題の研究は、「ペーパーテストは教科知識を測定するだけ」という通説を覆し、より教育的に意義のある試験を行える可能性を示唆した。

第7章では、本論文で展開した諸研究の成果を踏まえて、わが国の大学入試制度についてその改善の方向性を述べている。

わが国の教育現場は小学校から大学まで、教える側と教わる側とを含めて約2,000万人に達する巨大な規模である。その中でも「児童・生徒・学生/教員」比を見る限り、大学は相当に劣悪な教育環境に置かれていることがわかる。高等教育の大衆化の中で、大学教育の質を維持するためには入学者の学力向上が必須要件であることを指摘し、また、高校以下の教育改革の論議では高校教員に対する配慮を欠いた乱暴な提案がなされることがあり、現場の教員が納得でき、現実的に対応可能な範囲で大学入試の改革を進めることが必須条件であることを述べた。

次に、学力選抜が入試に占める割合が小さくなり、個別大学の入試の影響力が低下していることを示し、このため、学力水準の維持向上のためにはセンター試験の担う役割が大きいことを述べた。さらに、これまで大学入試に教育測定の理論が用いられることが少なく、現在のわが国には将来の教育政策を策定するに足る学力データを収集する技術がないことを論じた。その例証として、学力低下に関連した2つの研究の結果を示している。新しい大学入試制度は、まず「能力主義的差別教育」論と素得点主義から脱却しなければならないこと、実施面では現行のセンター試験のノウハウを生かしながら、全ての大学進学志望者にとって努力目標となり教材ともなりうることを、そして教育測定論の観点から公平性が担保されることを前提としなければならないことを述べている。

最後に、結章では、それぞれの章の内容の総括、要約を行った。

論文審査の結果の要旨

わが国の「入試試験」ほど、世間の注目を浴び、またさまざまな批判にさらされてきた教育問題も少ない。入学試験にこれだけ多くの社会的関心が寄せられ、国民の多くがその関係者であるという社会も少ない。その意味で日本はまさに試験大国ということになる。しかしながら、実施されている試験の内容を科学的見地からみれば、それは必ずしも高い水準にあるとはいえない。入試の実践と教育評価・測定論研究とが乖離している現状は深刻である。入試はいわば聖域であり、評価・測定の専門家といえども入学試験問題（項目）を事前にチェックすることもできず、事後にも入試データを分析し公表することは容易ではなかったのである。

倉元氏はこれまでタブー視されてきたこの入試の領域に果敢に攻め込み、各種の試行（モニター）調査を実施するだけでなく、現実の入試データに直接アクセスし、それに教育評価・測定論の最先端の手法を適用することで、さまざまな新しい学術的成果をあげた。教育測定論の有効性を実践に活かした功績はきわめて大きいと言わねばならない。

博士論文の主題に揚げられた「ペーパー試験」への取り組みでは、日本語基礎能力テストのデータに項目反応理論などさまざまな手法を縦横に適用してみせたほか、実際に一定の項目プールを作成することによって大学入試センター試験を大学入学後の学力の保持と変化を捉えるためのテストとして有効活用することの可能性を示した。また小論文試験など論述式のペーパー試験については、従来から評定者の誤差要因による変動が大きく、信頼性を欠くことが指摘されてきたが、これに対しても、事前に具体的な評価の観点を明確にすることにより、必ずしも専門家の評定によらなくとも信頼性の高いデータの得られることを示した。このほかルールスペース法など、世界でも最先端の評価法を積極的に利用することで、教育現場における評価研究の可能性を広げた功績も大きい。現実に多用されているペーパーテストの問題点を教育測定論にもとづいて明らかにし、また同時に実践性に富む改善方策を具体的に示した点は、測定論研究にも、現行の入試制度の改善にも資するところ大であり、その成果は高く評価できる。

本論文は倉元氏のこれまでの研究の集大成を示したものであり、多数の研究分野にわたる多彩な研究業績は氏の知的好奇心の旺盛さとその資質の高さを示すものである。しかし、氏の力量をもってしても豊富すぎる題材を博士論文にすべて盛り込むには構成上の難を生じさせる。各章を構成する個別研究の質の高さに比べて論文全体の印象が冗長に感じられたのはそのためである。とはいえ、その点を差し引いても、本論文に納められた研究業績の数々は教育評価・測定論研究の最先端を示す水準を備えたものであり、また現在の入試研究をリードするに足る優れた業績であると判断できる。

よって本論文は博士（教育学）の学位論文として合格と認める。