

客観テストの特徴

研究部客員教授 池田 央
(評価研究部門)

大学入試センターで実施している共通第1次学力試験では、いわゆる客観テスト形式がとられ、各国公立大学で行う第2次試験と合わせて受験生の選抜に用いられている。

客観テストは、しばしば論述テストと対比されて議論されるが、両者は互いに補完的な関係にあり、それぞれの長所・短所を正しく認識し使いわけることは、受験する側の人にとってもまた実施する側の人にとっても重要なことである。ここでは客観テストの特徴を理解し、とくにその利点を引出すように心がけるという意味で、なるだけ公平な立場に立ちつつ客観テストの特徴を考えてみたい。

1 客観テストの設問構成

客観テストの作り方には、基本的には次の二つの方向が考えられる。一つは階層型構造をもったもので、もう一つは並置型構造とでもいうべきものである。

a) 階層型構造のテスト

階層型構造は、図1のように、一つの課題条件を設定した少数の大問があり、それを中心にいくつかの小問一枝問と展開していく形式をとるものである。この方法は一つのテーマを中心に深く掘り下げていくのに適しており、よい課題材料さえ選べば、客観テストでも奥深いところまで、質問を切込んで行くことができる。解答者が問題の内容を本当に理解しているかどうかを確かめていくのに適している。

しかし反面、それぞれの小問は前後の小問と無関係ではなく、とくに相互に関連性の強い質問の場合は、前問の出来・不出来が、後問の出来・不出来に影響する場合もあり得る。したがって、そういうことが起らないように、作題者は細心の注意を払う必要があるのである。また与える課題テーマが大きければ、一定時間に与えられる課題個数には限りがあるため、この形式のテストでは、すべての側面について、万遍なく受験生の能力を確かめることが難しい。評価される受験生の能力が、

そこでたまたま設定された課題テーマに依存した形で測定され、それに左右される度合いが大きいともいわれている。どうしても尋ねるべきと思われる質問が、課題上の制約から質問し得なかつ

たということが起り得る。したがって、能力評価のためのサンプルとしてのテーマの選定が難しいところである。偏りを少なくするためには、どうしても課題数をふやすことが必要である。

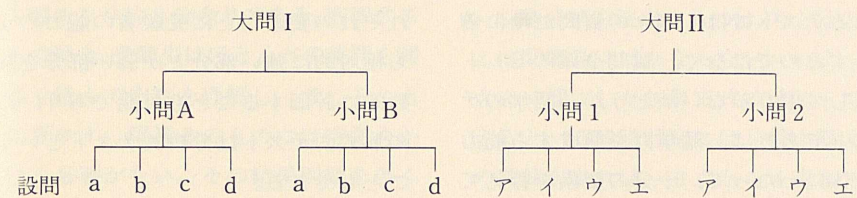


図1 階層型構造の設問形式

b) 並置型構造のテスト

これに対して、並置型構造のテストは、図2のように幅広い領域にわたって、数多くの独立した小問を並置した形をとるもので、アメリカ式の試験にはよくみかけるものである。並置型テストでは、それぞれの質問項目は、他の質問と関係なくそれぞれ個別に考えることができる。そういう意味では他の質問の出来・不出来の影響を受けないうように設けることができる（もちろん作題にあたっては、うっかりして他の質問が解答のヒントになるようなことが起らないよう注意する必要があるが）。

並置型テストの質問はそれぞれの質問がただ一つの事項をチェックする目的で用意される。たとえば外国語の問題で、ある単語の発音なら発音、アクセントならアクセントと標的をしぼ

って尋ねる。そのとき、語の意味とか活用とかの問題とは一応切りはなして考える。出題者としては、同時にそれも尋ねたくなるが、それをしないのは、その単語そのものを知らなければ、単語も発音も意味も活用もすべてできない可能性があり、たまたまその語を知らなかった受験生にとっては（他の語ならいくら知っていても）、損失があまりに大きいからである。このことは、のちに触れる尺度の目盛りを小さくみにする上で具合が悪い。一つの質問では一つの要素に目標をしばって尋ねるという意味で、この形式は要素型のテスト形式と考えることができる。

もちろん、たとえば語学の力がそうした単なる要素の集合によってだけ決まるとはいえないように（必要条件であることにはちがいないが）、こうした

並置型テストでは、全体的にどういった総合力をみているのかわかりづらいところがある。個々の問題をみると、一見何だか小間切れな部分的知識ばかり問うているように見えるかも知れない。したがって、こうした質問形式をとるテストでは、個々の質問だけに着目するのではなく、質問全体のセットとして偏りがなくどうか、語学のテストであれば、語学力を構成する能力全体にわたって、十分の要素を押えて

問1	問2	問3	問4	問5		問n
						

図2 並置型構造の設問形式

階層型テストでは、質問が与えた課題テーマに依存することが大きいので、一度使用した質問は繰返して使用することが難しい。よくどこかの試験でかつて出題されたと同じ教材が出題されたということで、問題になることがあるのはそのためである。一問一要素形式の並置型テストの小問は、それ自身独立しているので、別のテスト枠組の中での再使用も可能である。たとえばある単語の発音は、ほかに変えようがないし、その単語が重要な単語であれば、本来何度テストしても問題はな

いかどうかに注目する必要がある。テストの質を決めるのは、個々の質問（もちろんそれも大事であるが）というよりは、それらを集めた総合的な組合せであるからである。また慎重に問題の組合せを配慮したテストでは、それぞれの要素ごとに受験者の能力のうち相対的に強い部分とか弱い部分を診断的に評価することも可能であり、この形式のテストは分析的テストということもできる。

テスト項目は、質問項目のプール制を考えたときには、これになじみやすい形式ということができる。

いずれにしても、実際の試験では、両方の要素を加味した試験問題が課せられている。階層型と並置型をどのような配分で、どう組合せていくかも客観テストの作成上で、難しいところである。

2 客観テストと問題量

客観テストにあつては、問題量が多いということがよくいわれる。そのかわり1問あたりの解答時間は比較的短い。たとえば昭和59年度の日本史・世

界史受験生は、120分の試験時間に合わせて113問のマーク解答をしなければならぬ。平均するとマーク1個あたりに換算して、約1分余りである。また問題数が少いと思われる物理・化学受験生の場合でも59個のマーク解答があり、1個あたり約2分余りである。客観テストの場合、受験生はほとんどの時間を問題を読んで内容を理解し、考えることに費やす。解答そのものには時間をかける必要がない。その結果問題を多くすることができる。一方論述テストの場合は、答案を書くことに多くの時間が費やされるため、課題をそれほど多くすることができない。

受験生が何千人何万人という大規模テストになると、問題の量を多くしないと、テストの信頼性という面からいって具合の悪いことがおこる。たとえば、受験者数が数十人程度の少数の場合には、A-B-Cの3段階評定とか、5-4-3-2-1の5段階法による成績評価でもそれほど大きな支障はきかさない。それで十分なことが多い。しかし、それが数百人となり、同点者が100人以上ともなると、ランク境界付近の人は、微妙な1点差によって、1ランク上か下かに入れられてしまう。その結果、順位に100番以上のちがいが生ずる。一方で同一ランク内にいる人は、かなり力に差があっても、それは同じ能力として扱われてしまう。しかも、

この1ランクの差は、1題できるかできないかといったわずかな差で容易に起り得る。それでは不公平であるので、受験者の数が多くなるにつれて、それだけ細かい区分けが要求されることになるわけである。

段階分けを決定するのは、もとはといえば問題数にはかならない。いま、配点を一律にした加算方式のテストで考えると、5段階評定には、最小限4個の質問があれば、0点から4点までの5区分が理論上可能である。しかし、先ほどのべたように何万人もの受験生では、1点のちがいが何千番もの順位差を生んでしまうので、それが1個の問題の出来・不出来で決定されてしまうのは好ましくない。大ざっぱに言って、N人の受験者があって、標準偏差が σ である試験では、同点者の一番多い分布の山の付近では、ほぼ、 $0.4 \times N / \sigma$ 人程度の同点者がいると覚悟しなければならない。35万人の受験者で、標準偏差が140前後の試験で一番多いところで約1,000人ほどの同点者がいることになる。いいかえれば1点のちがいが1,000番の順位差を生むのである。しかも1題の配点は小さいものでも2点、大きいもので10点位はあるから、1題の出来・不出来が2点から10点の差を生むことを意味している。順位にして、数千人の順位差を生むことがあることを意味している。

こうした影響を少なくし、得点の安定性を高めるには、どうしても問題数を多くし、1題の配点を小さくして、その影響を軽くする必要がある。このように、問題数を多くすることは、ある1題の出来・不出来が、決定的な役割をする危険を避ける上での安全弁の働きをしているのである。

選択式解答をとる客観テストの避け得ない欠点の一つとして、当て推量による得点の問題がとりあげられる。当て推量の可能性を避ける意味で、選択肢の数を多くすることが考えられるが、それらの中から、受験生が全く純粋に当て推量を行うことは少なく、実際問題として、受験生の迷うあやふやな選択肢は、2・3個に限られることが多い。したがって、選択肢の数をむやみにふやしてもあまり効果はなく、実際は4・5肢で十分なことが多い。当て推量の影響を小さくするには、1題の選択肢の数をふやすよりは（そこから新たに獲得できる情報量は少ない）、問題数を多くすることの方が効果的である。少数の問題では、当て推量で、偶然よい成績を得る人があったとしても、すべての問題について当て推量が成功する可能性は少ないからである。その意味でも客観テストにおいて、問題数を多くすることは重要である。

3 テストの構造化と解答・採点の自由

客観テストにおいて、問題数を多くすることからくる最大のメリットは、上述のようなテスト技術の面だけでなく、そのことによって出題範囲を広くとり、計画的・組織的に問題を構成することが可能な点である。このことは多くの教科課程を一定水準までマスターしているかどうかを確かめる上で重要なことである。そのかわり、この形式は、解答者・出題者ともにある種の制約を課することになる。

客観テストの解答形式はいくつかの許容された選択の中から判断し、最良の解答を選ぶ形である。そのため受験生に許される解答の自由度は制限されたものになり、受験生の思うままに自由に記すことは許されていない。同様に採点者にとっても、正解や配点はあらかじめ事前に設定されており、受験生の答案を比較しながら、事後に採点基準を設定するという自由さはない。このことは、論述式テストとちがって、受験者・採点者ともに自由を制限されていて、不満の種になりやすいが、反面、明確な評価目的・システム設計が事前になされているときには、その長所を生かすことのできる極めて構造化されたテストであるということができ

る。このようにテストの構造が定められているため、テストの結果、たとえば問題の難易度とか、得点分布の形などは、問題作成の時点ですでに決定されているといつてよく、テスト後あれこれ調整することはむずかしい。したがって、事前の検討をおろそかにすると、問題が難しすぎたり、また易しすぎたりして、最初の意図とちがった不本意なテスト結果になる危険も同時にかかえている。そのために問題作成者に高度な作成技術が要求されることになるのである。

予備テストを事前に実施することが可能であれば、問題の適否を経験的に確かめることが可能であるが、実際はそれもできない。しかし、客観テストの長所の一つとして、たとえ事後であっても、解答結果を統計的に分析して設問の良否や受験者の思考様式を判断する資料が得られる。このフォーラムにもみられる「日本史」「国語」の解答分析などはその例である。こうした客観的資料はのちのちのテスト作成に有力な参考資料を提供するものである。マークの形でしか解答者の応答をみることができない出題者・採点者にとって、受験生の語る応答の様子はこうした解答分析表をたんねんに見ていくことによって、はじめて得られるものだからである。

4 テストの間接的効果

a) 学習目標のたてやすさ

入学試験の直接目的は、明らかに「大学教育を受けるにふさわしい資質を備えた人材を選抜する」ことである。したがって、ここで用いられる試験は選抜に用いられるテストとしての条件を備えたものでなければならないことはいうまでもない。このことは入学試験である以上避けて通れない条件であり、たとえテストの内容が教育的に優れたものであったとしても、試験の問題として不適当であれば、やはりそのテストはテストとしての第一の資格を欠いたものといわなければならない。

しかし、そうはいっても、テストはやはりそれに付随する様々な間接的効果をもっているものであり、それを全く無視してテストの作成を考えるわけにはいかない。テストには先ず教育的な見地から考えて、いくつかの条件にこたえるものでなければならない。

その第一の条件は努力目標がたてられやすいということである。テストが選抜テストである以上競争は避けられない。競争というのはある目標に向かっ

受験生の場合の努力目標はそれぞれの教科内容の習得である。過度の競争を避けるため、受験準備を必要としない問題を課せばよいという考え方もあろう。適性検査の中には、それを意図した問題を用意することがあるし、論述テストの課題の中には、特定の知識や教科内容に限るのでなく、総合的な力をみるということを目指したものがある。

しかし、入学試験が本来競争試験である以上、それを目指して受験生が勉学にいそむことはむしろ望ましいことであり、準備が意味をなさない試験は必ずしも競争試験に最適というわけにはいかない。また受験準備をすると、答案の書き方を準備した論述テストはかえって個性に乏しく、本来の目的から次第に逸脱していく傾向があることはよく経験することである。

勉学の目標を立てさせることは客観テストでなければ難しいというわけではないが、教科内容に則した出題を用意する上で、学習しておいて貰いたいことを、組織的・系統的に万遍なく出題するには、客観テストは計画がたてやすいということはいえるであろう。

b) 結果の公開性

またテスト結果については、自己診断が可能なものでなくてはならない。いいかえれば受験生本人に手ごたえのわかるものでなくてはならない。通常

学力試験は受験生本人が受験した後、出来たか出来なかったかについては、ある程度本人自身が判断できるものである。共通第1次学力試験のように正解が発表されればなおさらのことである。逆に自分がどのくらい出来たかについて、自分自身判断のしようがないという試験では困るのである。本人自身答案を書くには書いたが、それがどの位の成績であるのか予想できないとすれば、それは勉学の目標が定められないのに等しい(成績の順位については、また別である)。受験生の勉学に対する意欲をそぐだけでなく、不明朗なものを残すものにもなる。試験としてはそういうことを避けねばならない。

入試に対する信頼感は、本人自身の手ごたえだけでなく、周囲のものの期待感とも一致することによって生れる。高校の教師はふだんの授業から一応期待される生徒とそうでない生徒についてのある程度の正確な認識と予想を持っている。そして通常の公正な学力テストではその予想はほぼ当たっているといえよう。まして、現代は受験産業等が提供するより精微な予測情報によって、その期待は一層動かしがたいものとなっている。そのような中であって、ごく少数の番狂わせや意外な結果が起り得ることは避けられないにしても、そのようなケースが頻繁におこるとすれば、それはやがて選抜する側に対し

て、その選抜方針についての疑問をいだかせる結果に至るであろう。つまり周囲の人が納得する形で受験生の合否が決められるという仕組みが、本人の手ごたえと平行して存在しないと選抜テストは成り立たないのである。いいかえれば、明文化されていないとしても、選抜のルールが暗黙のうちに確立しているということが必要なのである。

c) 学習習慣の形成

よいテストの条件として、最後に良い学習習慣を形成できるかという問題がある。残念ながらこの点に関しては客観テストはその力が弱いかも知れない。前にも述べたように、客観テストは問題を読むこと、考えることには時間を費やすが解答を書くことそのものには時間をあまり費やさない。受験生の知識や理解力について、できるだけ多くの情報を短時間に得るということが、客観テストの一つの目的であるから、そこでは書くことそのものに多くの時間を費やし、その結果、必要な他の情報を引出す時間が少なくなることを避けている。限られた時間内に多くの情報を系統的、組織的に得るという目的には、構造化された客観テスト形式は向いているが、それは学習そのものをそういう形にすることをすすめているわけではない。本来学力は、骨の

ある長い一冊の本を読破したり、それを要約整理したり、論旨のはっきりした文章を展開するといった地道な訓練を経て次第に形成されていくものである。客観テストはそうした努力の結果をできるだけ効果的に偏りなく推測するために用意された道具に過ぎない。そして、そうした地道な基礎訓練を重ねたものが、高得点を得ることを期待して作成されるものである。それがいつしか、テスト練習そのものが学習目標になり、解答に時間を要しないテスト・サンプルの練習イコール教育ということになれば、それはテストの本来の目的からはずれることになる。教育目標は、テスト問題の背後にある教科内容の理解であり、また良き読書習慣の形成や発表能力の育成であるはずである。ところが、短時間にできるだけ多くの内容を吸収しようとするあまり、もとの基礎訓練に時間を費やさないまま、短絡的にいきなり、問題練習ばかりに力を入れるとなると、それは教育本来の姿からはずれることになる。もし現在客観テストがかかえる最大の悩みをあげるとすれば、こうした良い学習習慣の形成に役立っているかどうかという反省である。それは実際の教育の現場に立つ方々との協力の上ではじめて改善されることでもある。