

3. 「高大接続で問われるべき能力と適性とは？」

中京大学教授・現代社会学部長、名古屋大学名誉教授
村上 隆

はじめに

とりあえず自己紹介

- 名古屋大学(教育学部・大学院教育発達科学研究科)で31年半勤務
- 中京大学現代社会学部に11年半勤務。今年(2018年)3月に定年(70歳)退職する。
- 専門領域は、名古屋大学時代には、計量心理学(心理・教育測定学)、中京大学に異動後は、社会統計学も加わる。前者は広い意味でのテスト、後者は社会調査を対象とする分野。
- 本日は、研究成果というよりも、国立大学と私立大学、心理学と社会学を渡り歩いた経験にもとづいた(あまり学術的でない)話をさせていただきます。

私は名古屋大学教育学部・大学院教育発達科学研究科で31年半勤めました。その後、中京大学の現代社会学部に移り、11年半になろうとしています。この3月に定年で退職します。私の専門領域は、名古屋大学時代には「計量心理学」というように称していました。教育学部なので「心理教育測定学」と称していたこともあります。次に中京大学の現代社会学部に異動しました。この学部は社会学を中心とするので、「社会統計学」が私の専門に加わりました。私は、これらは同じことだと思っていたのですが、実は全くそうではありませんでした。これについては後でお話しします。

「計量心理学」いうのは、広い意味でのテストの研究です。これは能力や学力のテストだけでなく、性格検査といわれるもの、あるいは態度測定といわれるものまで含まれると思いますが、要するに個人差を査定(assess)しようとするものです。「社会統計学」の対象は社会調査です。社会そのものを対象にするような調査です。つまり「計量心理学」は“個人”、「社会統計学」は“社会全体”を相手にするという違いがあります。この講演は私自身の研究成果というよりも、私の経歴、つまり国立大学と私立大学、そして心理学と社会学という分野を渡り歩いた経験に基づいた話ということで聞いていただければと思います。

1. 能力と適性とは？

能力と適性とは？

- 能力: 一般的, 適性: 状況依存的？
- 生得的？(教育・訓練・学習によって)可変的？
- 大学教育に「適応」できる準備状態ができているかどうか、と考えられる。
- 高大接続の現在の流れ
学力の3要素を(全部)評価
- この流れの必然性と問題点

まず能力と適性についてですが、文部科学省のWEBページを見ても「入学者選抜は本人の能力と適性に基づいて行う」と当たり前のように書かれています。わざわざ能力と適性と分けているのですが、これらはどこが違うのかというと、おそらく能力というのは一般的知能というように、特に対象があるわけでもない一般的な能力、ある意味ではどこにでも通用するような能力ということでしょう。適性とはむしろ“文系の適性、理系の適性”あるいは“教師としての適性”というように、いわば状況依存的であるというようなことではないかと思います。そもそもそれは生得的なものであるのか、あるいは教育、訓練、学習によって変わりうるものなのかということですが、今のところほぼ半々くらいだと考えていいのではないのでしょうか。特に遺伝的なものではないというわけではないですし、それを全く変えられないということでもありません。

では能力と適性に基づいて選抜をするということはどういうことかということですが、これは結局、大学教育に「適応」できる、いわゆる準備状態ができているかということを確認する、そこを判断して入学者を選抜するというのではないかと思います。

一方で、最近の高大接続の流れを見ると、学力の3要素という話が出ています。これを全て評価しないとイケないというようなことであるようです。こうした流れにはそれなりの必然性がありますが、その一方で問題点もあります。このあたりがこのお話のテーマになると思います。

2. 本講演の柱

本講演の柱

- 現行の知識・技能中心の評価方法の限界はどこか？
- 思考力・判断力を問う共通試験は可能か？
- 高大接続における本当の問題点は何か？

まず学力の3要素の中で一番基盤になるものが「知識・技能」であるといわれていますが、現行の知識・技能中心の評価方法には限界があります。その限界がどこなのかを少し探ってみようというのが、1つ目の柱です。2つ目の柱は、大学入試は共通試験と各大学が行う個別試験の2段階構成になっており、特に国公立大学の場合は、完全にこの2つが積み重ねられる形になっている場合が多いですが、では「知識・技能」を基礎とした「思考力・判断力」を共通試験で問うことができるのかということです。そして、先ほどの2つの柱も問題があると思いますが、「高大接続」における本当の問題点は、それと少し外れたところにあるのではないかと私は思っていますので、その話を3つ目の柱とさせていただきたいと思います。

3. 学力の3要素で入学者選抜を行う危険性

学力の3要素で入学者選抜を行う危険性

- 学力を3つに整理する
 - ① 知識・技能の確実な習得
 - ② ((①を基にした) 思考力・表現力・判断力
 - ③ 主体性を持って多様な人々と協働して学ぶ態度
- 学力の整理の仕方としては理解できる。
ただし③は??? 妥協の産物?
- それにしても、どうやって評価するのか?

学力の3要素とは何かというと、まず「①知識・技能の確実な習得」といわれています。そして「②(①を基にした) 思考力・表現力・判断力」、それから「③主体性を持って多様な人々と協働して学ぶ態度」といわれています。学力という概念を整理することは必要ですし、その整理の仕方としては、これは、私は理解できないことはないと思っていますが、ただそれにしても③とは一体何なのでしょう。これは妥協の産物ではないのでしょうか。

学部長をしていると色んなところからアンケートの依頼があります。先日、文部科学省からアンケートの依頼がありました。どのような内容かという「学生が主体性を持っていると判断したのはどういう状況であるかを、学部から選んだ2名の教員にそのことについて幾つか例示させてくれ」というものでした。おそらく喧々諤々と議論したうえで、③のような表現になったのでしょう。どうやらこの「主体性」というところが一番大事なようで、「主体性」とは何であるのかを、密室で議論ばかりせずに現場に聞いてみようと思われたことは良いことかもしれませんが、こうして「主体性」について色々と聞き、それを基に「主体性」を評価する尺度を作るといっているようなのです。一番の基盤である「知識・技能」の評価さえある意味怪しいですし、「思考力・表現力・判断力」の評価はもっと難しいのに、それに加えて「主体性」の評価の尺度を作るというのです。一体どうやって評価するのでしょうか。

4. 心理測定と社会調査の共通点と相違点－2つの世界にかかわってわかったこと

心理測定と社会調査の共通点と相違点 2つの世界にかかわってわかったこと

- どちらも、外的に設定された設問への、個人の解答(回答)にもとづく。
- どちらも、解答(回答)から個人に(複数の)数量が付与されるルールがある。
- 心理(教育)測定では、個人の特性に関するassessmentが行われる。
- 社会調査は、社会の実態、仕組みのdescriptionを目的とする。

ここからは私の個人的な経験から来ることなのですが、先ほど言いましたように、私は計量心理学の世界、つまり“個人をassessmentする世界”と、“社会調査”の二つの世界に関わってきました。どちらにしても個人に対して設問をし、それに答えてもらうことから始めます。“問題”ならば“正しい答え”ということになりますし、“態度・意見”を聞いているのなら“各人の回答”となります。さらにそれを数値するようなルールがあるということなので、計量心理学と社会調査はそれほど変わらないだろう、同じことじゃないかと思っていたのですが、一見形態は似ていても双方が目指しているところが違いました。それはどういうことかという、繰り返しになりますが、心理(教育)測定とはあくまでも個人の特性、個人というものに関してそれを査定する、assessmentが目指されます。個人を正しく測れるかどうかということが問題になるのに対して、社会調査は社会の実態がどうであるか、あるいは社会の仕組みがどうなっているのかをdescriptionつまり記述します。assessment「査定」に対してdescription「記述」というように目的が違います。つまり社会調査は集団全体のことが分かればよいのに対して、大学入試もそうですが、計量心理学、つまり教育測定では個人を正しくassessできなければいけません。この違いがかなり大きいということに気が付いてきました。

5. Inventoryが備えるべき性質

Inventoryが備えるべき性質

- **Inventory:** 心理学では質問紙, 社会学では調査票, ここでは, 試験問題冊子も含める。
- **心理測定**では, 高い信頼性, つまり個人の得点が時間間的に安定していることが求められる。
- 似たような質問(問題)項目が多数集められる。
- **社会調査**では, 広い「帯域幅」, 内容の多様性が求められる。
- 多様な質問項目が多数集められる。
- 一定の多様性は心理測定でも必要。その理由は・・・。

「Inventory」いう言葉は心理学では“質問紙”と訳されています。社会学では“調査票”といいますが、原語は同じです。ここでは試験問題冊子も含めて考えたいと思います。

どんな性質を備えていなければならないかという、心理測定ではまず“高い信頼性”が求められます。10年も経てば個人の能力はもちろん変わりますが、せいぜい数日・数週間・数ヶ月間繰り返し測っても、それほど変わらずに安定しているというのが信頼性です。つまり測定誤差が小さいということです。出された問題にたまたま正解したとか、不注意で選択肢を間違えたということもありますから、一問一問でいうとあまり高い信頼性は期待できません。そこでどうするかというと、たくさん問題を出す、それもだいたい似たような領域から出すことにより、信頼性を高めようとします。態度調査や性格検査をすると「なんでこんな同じ事ばかり何回も聞くのだ」と対象者から抗議をされることがあるのですが、これはなぜかという、同じような問題や例や質問をたくさんすることによって、測定の信頼性が上がるとされているために、内容的に非常に似たようなことが繰り返し問われるのです。それに対して社会調査は信頼性よりは帯域幅 (bandwidth) の広さ、つまり“内容の多様性”が求められます。社会とは非常に多元・多様なものですから、それを記述するためにはいろんな角度からの質問が必要です。したがって心理測定のように同じ事を何度も聞くということはなく非常に質問の種類や数が多く、内容的にはバラバラの質問をします。

6. 信頼性と妥当性

信頼性と妥当性

- テストによる assessment は、この2つの側面から評価されてきた。
- 信頼性とは、得点が時間的に安定している程度のこと。1つの問題項目への解答は不安定だが、複数の項目への反応を加算すれば、安定したものになる。
→ 小問主義への傾斜
- 妥当性とは、得点が測定を目指したものを実際に反映している程度のこと。出題範囲は中等教育で学習される範囲を超えてはならず(領域適切性)、出題範囲から満遍なく出題する必要がある(領域代表性)。
→ 小問主義への傾斜

学力を測るテストの場合には信頼性と妥当性の両方が必要です。信頼性とは、繰り返しになりますが、“得点が時間的に安定している程度”のことです。ひとつの問題項目への解答は不安定でも、複数の項目への反応を加算すれば安定したものになります。これは結局、小問主義であるといえます。試験問題というと大学のテストでは「～について述べよ」という問題がよく出てきますが、そういったものではなく、小さい問題がたくさん並んでいるセンター試験あたりをイメージすればよいと思います。高い信頼性を得るために、とにかく繰り返しが必要だということで、小問主義に傾斜していく傾向があります。

また妥当性とは、極めて普通の言い方をすると、“テストの点数が実際に測定を目指しているものを反映している程度”ということですが、これは結構微妙な問題ではあります。例えば“性格

の内向性”と“情緒不安定性”は非常に重要な2つの次元ですが、意外とこれらは区別が難しいです。内向的な人はストレスに弱い傾向を持っている場合が多いので、この2つは相関しています。内向性を問う質問が、実は情緒不安定性を問うているということが結構あります。このように、測定を目指す個人差を正確に反映するものさしを作ることはなかなか難しいのですが、入試で要求される学力の場合には、ある程度確立された方法があります。試験問題の検討によってある程度見当がつけられる内容的妥当性という考え方です。

内容的妥当性には、2つの側面があります。まず、出題範囲は中等教育で学習される範囲を超えてはいけません。この基準は「領域適切性」といいます。つまり教えていないことは試験に出してはいけないということです。実際問題としては「範囲」の判定は結構難しく、例えば、世界史ではどこかの教科書の写真に写っていればそれは出題範囲になるという話を聞いたことがあります。いずれにしても日本では中等教育で教えらるる範囲にない問題を出してはいけないと一応定義することができるでしょう。

もうひとつ特定の単元、授業からではなく、出題範囲とされている中からまんべんなく出題する必要があります。これは「領域代表性」といいます。例えば全15回の授業を行ったのに、11回目の授業の内容だけから出題しているということになると、これは領域代表性を著しく欠いていることになります。15回授業を行ったのなら、その全15回分の内容から、そこで学習した知識・技能を問う問題をまんべんなく出さなければなりません。これと領域適切性、この2つの面から内容的妥当性というものが評価されます。こうするとやはり色々な分野から出題する必要があります。領域代表性ということを考えると、大学の試験によくあるような「11回目の講義で教えたこと（もちろん、具体的に「援助行動を促進する複数の要因」とか「同一労働同一賃金が実現しない理由」とか）について、知るところを述べよ」というのは、その1問だけでは領域代表性を欠いていますので、基本的には小問主義に傾斜するということになります。

7. 試験は常に不完全である

試験は常に不完全である

- 完全な信頼性と妥当性をもつテストを作ることは不可能である。
- 小問主義は、それでも信頼性と妥当性の度合いを高めようとする努力の方向としてやむをえない（マークシートは実用上の問題）。
- それでも、次の「常識」は打破できない。
秀才に／「試験は水物だからね、気をつけたまえ」
鈍才に／「試験は水物だからね、頑張りたまえ」
— 50年以上前の「蛍雪時代」受験ユーモアより
- だが、それ以上に小問主義には問題が…。

そうはいつても、完全な信頼性と妥当性を持つテストは作ることはできないでしょう。今回のお話ではこれが重要だと思いますが、完全なテストが作れない理由のひとつは試験時間です。限られた試験時間の中では出題される問題の数は限られます。無限にたくさん出題するというわけ

にはいかないのです、そういうことでは信頼性が限られてきます。妥当性に関しては領域代表性、つまり教えてきたことをすべて網羅して問題を作るということはできません。つまり信頼性、妥当性に関して完璧なものをつくることはできないということです。

私は“常に試験は不完全”ということを何回かいろんなところで言ったのですが、そのたびに非常に強い反発・抵抗を受けました。「公的機関が実施する試験が不完全とはどういうことなのだ」「我が組織が実施する試験というものは完全なものでなくてはならない」というような非常に強い反発です。それはそうなのですが、私が言っている“不完全”の意味をもう少し説明する必要があります。「使い物にならない」とか「出題ミスである」とかいった意味ではありません。つまり“完全な信頼性”とは“何度やっても全く同じ点数になる”ということですし、“完全な妥当性”とは“その分野に関しては完全に全てを網羅した試験”ということですから、それは1時間で実施できるような試験では不可能であろうということなのです。

しかし、そういう意味でより完全なものに近づきたい、信頼性と妥当性を高めたいと考えたときに、やはりに小問主義という方向になるのは、努力の方向としてはやむを得ないと思います。世間では共通試験がマークシートであるということの問題にする方がいらっしゃいますが、これは実用上の問題であって、基本的には小問主義であることがマークシートを使いやすくしているということは事実だとしても、マークシート自体が問題なのではありません。

そうはいつでも“不完全とはどういうことなのか”ということをもう少し説明してみましよう。私の受験時代の雑誌「蛍雪時代」の読者投稿のコーナーの「受験ユーモア」というコラムに50年以上前の投稿されたもので、私が未だに覚えているものがあります。「先生は秀才には『試験は水物だからね。気をつけたまえ』と言ひ、あまり普段から成績が良くない生徒には『試験は水物だからね。頑張りたまえ』と言う』というものでした。つまりこれが要するに試験、テストの不完全性ということではないでしょうか。これは常識ですね。それにしても、小問主義という方向性は、打破するのはなかなか難しいです。しかし信頼性、妥当性を高めようということと言うならば、努力の方向として間違っていない、やむを得ないと思います。しかし小問主義にはそれ以上の問題があると思います。

8. やはり現行の共通試験に問題はある

やはり、 現行の共通試験に問題はある

- 小問は、相互に独立な知識技能を問うことになる(そうでないと「共倒れ」など、大問と同じことが問題になる)。
- したがって、複数の知識や技能を組み合わせないと解けない問題(思考力、判断力)は作りにくい。
- 特に、特定のパターンで解答できる知識・技能の小問だけの試験が定着してしまえば、そうしたテスト専用のさまざまな知識・技能が発達する。
- そんな試験に特化した能力は、一般性を欠き、大学教育への準備状態の assessment にはならない。
- 大学教員による、こうした経験の物語には事欠かない。筆者は、数学の「確率」の分野について大きな問題を感じているが、それはまた別の機会に。

まず小問は、お互い独立した知識・技能を問うこととなります。そうでないと1つ目の問いの選択肢を誤ったために2つ目、3つ目、4つ目と、すべて間違いになるという、いわゆる「共倒れ」が起きてしまいます。それは信頼性の低下を招くこととなりますので、お互いに独立の知識・技能を問わないといけません。そうするとある一連の問題の中で、複数の知識や技能を組み合わせ、それを推論していくというような内容にして、その思考力と判断力を問うということになると思うのですが、そういう問題は非常に作りにくいです。結局は、断片的な知識・技能の評価に終わりがちです。

典型的なパターンですが、昔から「の」がけ」というのがあります。「算数の文章題で“の”と書いてあったら掛け算しろ」ということです。特定のパターンで回答できる知識・技能の小問だけの試験が定着すると、実はその全体が分かっているわけではない、もしくは基本の原理が分かっているわけではないのですが、“こういう問題が出たらこうする”というように、入試技術や入試指導の技術が進み、そしてテスト専用のさまざまな知識・技能が発達します。しかし、そういった試験に特化した能力というのは一般性を欠いていますから、大学教育を受ける準備状態、つまり個人の能力と適性を正確にassessmentしているということにはならないのです。

9. 思考力を問うには大問が必要

思考力を問うには大問が必要

- 思考力、表現力、判断力を問うためには、大問が必要。しかし、そうした大問からなる試験は、問題数をそれほど多くできない。
- したがって、高い信頼性は望めず、領域代表性も低くなる。
- 換言すれば、このような試験を短時間で行うことには無理がある。
- では、時間を延ばせるか？ 大学の協力といった現実の問題以上に、多くの障害が予想される。
- それにしても、今目指されているテストとは？

では思考力を問うにはどうしたらよいかということですが、やはり大問が必要です。複数の事実を組み合わせ、論理的な繋がりを推論するということがないと、思考力や判断力を測るということは難しいと思います。そうすると、たまたま出た問題がその受験者にとって得意だったとか、他方でたまたま不得意な分野だったとか、得意分野でも、最初の問題で間違ってしまったためにその後を全部間違えるといった共倒れをしてしまう場合は、その人の得点が一気に下がるので、高い信頼性は望めません。それから、大問の場合は広い範囲のあらゆることを問うことは出来にくいので、領域代表性も低くなります。つまり信頼性も妥当性も低い、あまり高くないだろうということです。それならばたくさん問題を出せばいいじゃないかという、一方では試験時間を短くしないといけないという圧力があります。現行のセンター試験でも大学教員や大学の事務職員にとっては大変な重労働です。これ以上試験時間が延びたら大学が協力するのかという問題以上に、多くの障害が予想されます。

10. PISA型の試験でassessmentは可能か？

大問を使うということになると、そのお手本の1つとしてPISA型の試験があります。PISA型の問題は大問中心です。今回の新試験にもこれの影響は現れているようです。

まずPISA型の問題では、例えば“受験者はあらゆる中等教育の修了者”ということと考えれば、正答できる受験者とできない受験者ということを区別することは可能です。能力が広い範囲に及んでいればPISA型のような問題でも、ある程度の信頼性は得られるだろうと思います。しかし現実には、大学間の格差が非常に大きいので、ある特定大学の入学者選抜では非常に幅の狭い集団で相対的な差を競うことになりまますから、信頼性が低くなると言わざるを得ません。

PISA 型の試験で assessment は可能か？

- PISA (Programme for international student assessment)の問題は、大問が中心であり、共通試験の1つの規範のように言われている。新試験にもその影響はあらわれているようだ。
- PISA 型の問題でも、かなり高い確率で正答できる受験者と、ほぼ確実に正答できない受験者はいる。複数の設問があれば、この2つのグループは得点によって区別できるであろう。
- 能力が広い範囲に及ぶ受験者群の間では、そうした問題集合でも、一定の信頼性は得られる。
- しかし、大学間の格差は大きく、現実には非常に能力の幅の狭い(分散の小さい)集団での相対的な差を競うことになる。信頼性は低い。
- つまり、上の問いへの一般的な答えはない。

実はアメリカのあるところでも、こういう方向で共通試験をつくるというような試みが出されています。そういう話を聞いたところで、私がディスカッションをしていたある討論で、会場からの「PISA型の問題でassessmentはできるのでしょうか？」という質問がありました。私は「アメリカでのデータで一定の信頼性が認められたのであればできるでしょう」とつい言ってしまったのですが、現実問題としてそれを入学者選抜に使うということになった時、これでいいのかということになれば、それは能力が広い範囲に及ぶ受験者の能力はある程度区別することができて、「assessmentが可能だ」ということは簡単には言えないのです。そのテスト、どういう使い方をするのかということなのです。

11. 信頼性はテストが使われる集団次第

信頼性はテストが使われる集団次第

- 信頼性係数の定義

$$\text{信頼性係数} = 1 - \frac{\text{誤差の分散}}{\text{得点の分散}}$$

- どんな集団でも、誤差の分散はそれほど変わらない。他方、得点の分散は、競争の激しい集団ほど小さくなる。
- 第2項の分母が小さくなるため、テストは同じでも、競争の激しい集団では、信頼性は低下する。

信頼性は得点全体の分散、つまり全部の受験者の点数の散らばり（得点の分散）と誤差の散らばり（誤差の分散）からなる式で表わされます。専門分野ではここをどう評価するかということが問題になるわけですが、実際にはどこの集団で計算しても誤差の分散（分子）はそれほど大きく違いません。同じ人の点数がどのくらい変動するかということですが、できる人もできない人もこの誤差はそれほど違わないと考えられます。ところが得点の分散（分母）はどういう集団で計算しているかに影響されます。つまりその集団があらゆる中等教育修了者だとすると、得点分散（分母）がとても大きくなるので、誤差の分散（分子）が同じだとすると信頼性係数が高くなります。ところが競争が激しい大学、競争的な大学では、みんなの能力が非常に狭い範囲に集中しているので、得点の分散（分母）は小さくなります。そうすると“得点の分散”分の“誤差の分散”の比は“1”に近づいていきますので、1からこれをマイナスする信頼性係数の式では、信頼性がどんどん下がってくるということになります。つまり競争の激しい集団では信頼性は低下することです。PISA型のテストの信頼性はどのくらいかということは、どういう集団で行ったのかということを示さなければ、その信頼性が高いとか低いとかはいえません。つまり集団次第ということなのです。

12. なぜPISA型のテストがあるのか

なぜ、PISA 型のテストがあるのか

- PISA が目指しているのは、個人の assessment でなく、集団の状況の記述である。要するにこれは社会調査なのである。
- 一部の領域に偏った（領域代表性を欠く）問題は、端的に言って不公平であり、個人の評価には適さない。しかし、（たまたま）そうした問題に正答できる個人が多いか少ないかは、教育課程・方法の評価、集団間の比較には使える。
- 領域代表性をもつ assessment に使うためには、問題数を増やす必要があるが、それには、長大な試験時間が必要になる。

では、なぜPISA型のテストがあるのかについてですが、PISA型が目指しているものは実は個人のassessmentではなく、集団の状況の記述です。適切なサンプリングがなされているとすれば、国際比較や地域間比較であれば、集団の状況の記述としては可能です。要するに心理測定か社会調査かという2分類でいえば、これは社会調査なのです。ですので、一部の領域に偏った問題、つまり領域代表性の低い問題であっても、その集団の中にその問題が解ける人がどれだけいるかという観点からは有効な作問と言えるでしょう。しかし、個人のassessmentであるとするれば、こういう問題は不公平と言わざるを得ないでしょう。

つまり、PISA型のテストは個人の評価には適さないですが、そういう問題に正答できる人が多いとかそういう人が増えているということが分かれば、教育課程や方法の評価、あるいはその集団が優れている、またはその教育プログラムが優れているかどうかという評価には使えると思います。誰がそれに正解したか、あるいは誰が高得点を取ったかはそれほど重要ではありません。

ん。PISA型の問題は、それを見るべきものではないのであって、高得点を取る人が増えてきた、平均点が上がったというように、その教育プログラムや方法が有効であったというような言い方はできるだろうと思います。ではPISA型で領域代表性を満たす問題、つまり広い範囲の問題を作ればいいのかということになりますが、それには問題数を増やす必要があり、個人のassessmentに使うには、1人の受験者に広い範囲から出題されたたくさんの問題を解いてもらわないといけないので、非常に長い試験時間が必要になります。

13. よく誤解されること

よく誤解されること

- 現行のセンター試験の「数学」は、必ずしもここでいう小問主義の試験ではない。
- いわば大問を、適当に誘導しながら小問に分解している形。平行した記述型の試験に書き換えることができる。
- マークシート＝小問主義ではない。
- 困難度は違うが基本的に同じ能力(思考力を含む)を測っていることが実証されている。

ここで、よく誤解されていることについてお話しします。現行のセンター試験はマークシートだからダメだと言う方がよくいらっしゃいますが、センター試験の数学の作問のやり方はここでいう小問主義ではありません。言わば大問を適当に誘導しながら小問に分解している形なのです。それなりに技術のある人は平行した記述型の試験に書き換えることができます。つまりマークシート＝小問主義ではないのです。センター試験の問題を記述に書き換えたとき、どうなるかというと、困難度は違いますが、基本的に同じ能力を測っていることが実証されています。全く違う能力を測っているということではありません。非常に相関が高いです。つまり現行のセンター試験の「数学」には思考力が含まれているのです。その分、共倒れの危険性というものがありますから時々番狂わせが起こり得ます。

14. 問題は評価にかかる時間

問題は評価にかかる時間

- 以上をまとめると、一般に、思考力、表現力、判断力を問うためには、大問が必要。しかし、そうした大問からなる試験は、問題数をそれほど多くできない。
- したがって、高い信頼性は望めず、領域代表性も低くなる。対応するには、問題数を増やすしかない。
- 換言すれば、このような試験を短時間で行うことには無理がある。
- では、時間を延ばせるか？ 大学がそうした負担に耐えられるか、といった現実の問題以上に、多くの障害が予想される。
- が、その前に……。これは学力評価だけの問題か？

もう一度戻りますが、結局、問題は「評価にかかる時間」です。今までの話をまとめると、思考力、表現力、判断力を問うためには大問が必要です。小問主義の試験では問えません。しかし大問ではそれほどたくさんの問題を使うことができません。なぜなら時間がそれほどに多く取れないからです。したがって、それほど高い信頼性は望めませんし、領域代表性も低くなります。つまり短時間でを行うということは無理があるということになります。それなら試験の時間を延ばせるかという、先ほど言ったようにこれは大学側に大変な負担がかかるので難しいですし、そういうこと以上に多くの障害が予想されます。しかしその前にこれは学力の評価だけの問題か？と私は思っています。

15. 医療分野のassessment evidence

医療分野の assessment evidence

- 血圧の測定の誤差は、心理測定より大きい？
- 実際、ランダム誤差も恒常誤差も大きい。病院での1回の測定で assessment はできない。
- 血圧は自宅で毎日複数回測るものになりつつある。
- しかし、血圧計は血圧以外のものを測っているわけではない(妥当性の問題はない)。
- 病気の原因を知るための医学的 evidence としては、健康診断での1回の測定値の平均値や相関係数でも十分とされる(らしい)。こちらは assessment ではなく、社会への対応が問題なのである。
- PISA 型テストによる assessment は、毎日実施？

心理測定の説明をするために少し他の分野のことを知ろうと思い、比較的単純なのが血圧の測定ということで、最近、血圧について勉強してみました。実は血圧を問題にしている心理測定家がいまして、違う検査のほうが血圧の測定よりもずっと信頼性が高いと言っています。本当かなと思いましたが、血圧の測定値には確かに“ランダム誤差”があります。つまり繰り返して測っても、必ずしも安定しているわけではないということです。それから“恒常誤差”というのはある一定の

偏りを持つということなのですが、その例として“白衣効果”というものが知られています。白衣を着ている人が血圧を測る、あるいはその人がそばにいるというだけで、血圧が15ほど上がるという現象です。また電話をかけていると血圧が30上がるといいますが、確かに電話中に脳出血を起こした方がいらっしゃいます。では病院で測ったとき15だけ高くなるからその数値から15引いておけばよいかというと、その一方で“仮面高血圧”という方がいらっしゃいます。これは先ほどの逆で、病院では血圧の数値が低いのですが、自宅に戻ってリラックスして測るとかえって高くなるという現象です。やはり恒常誤差は避けられません。

今はどうやっているかという、血圧に問題がある人は自宅で毎日、朝昼晩と血圧を測定することになっているようです。さらに測ったときはどういう状況だったか、例えば、特別体調が悪いということではないが、「だんなが競馬に行っても戻ってこない」というような状況だと、やはり数値は高くなります。つまりその時どういう状況だったかということを書き留める“血圧測定日誌”をつけることになっているようです。

そうはいつでも血圧計というものが血圧以外のものを測っているとは考えにくいですが、ここで誤差とっているのは実際のその人の体調の変化等であり、測定値の誤差もあるかもしれませんが、その時の状況によって変化しているということなのです。この点は学力の測定よりはるかにばんましな状況ですね。

それにしても、このように実際にランダム誤差も恒常誤差も大きいので、病院での一回の測定でassessmentはできませんが、一方で病気の原因を知るためのevidence、つまり健康診断での1回の測定値を地域内で平均をする、あるいは他の何かとの関連をみる、たとえば、その地域の食生活との関連をみるということについてなら、evidenceとして十分機能します。つまり血圧測定も1回だけでなら個人のassessmentではなく、いわば社会調査だということです。assessmentとしてPISA型のテストを使用するなら、血圧測定日誌のように毎日実施するののかという話になりかねません。

16. 日本の公正観が究極の障害か？

日本の公正観が究極の障害か？

- 受験者のだれにとっても、教授されたことがある(はずの)知識・技能だけで(原理的には)正答できる問題でなければならない(思考力、表現力、判断力はそこから外れることにならないか)。
- 試験場に入った後は、全員が同じ条件で受験できなければならない。
- 試験時間は、能力評価が可能である範囲において、できるだけ短くなければならない(受験生の負担は試験時間で測られる?)。
- 本人の能力と適性以外のものは、選抜に用いられるはならない(昔は、面接資料に、父親や親族の職業を書く欄まであったが)。

そういうことを考えていくと、「日本的な公正観」というものが究極の障害ではないかと思えてきます。まず日本では「誰にとっても教えられたことがあるはずの知識・技能で正答できる問

題でなければならない」とされています。そうすると、新しい試験が作られていくと、思考力・表現力・判断力というものには教授される、あるいは学習されるようなものなのかどうかということについて批判が出てくる可能性があるのではないかと思います。それから「試験場に入った後は、みんな同じ条件で受験しなければならない」ということは、ほとんど何も持ち込めないということですが、また「試験時間はなるべく短くなければいけない」ということですが、受験生の負担というのは、試験時間で測られるということになっているようです。それから「本人の能力と適性以外のものを選抜に用いてはならない」ということです。昔、私がある大学で面接を受けたときに、父親の職業は何だとか親族の職業は何だとか聞かれましたが、今はそういうことは絶対聞いてはいけないということになっています。そうすると例えば、アメリカのやり方、つまりアドミッションズ・オフィサーが家庭訪問までして入学者の選抜をするということになれば、これは本人の能力、適性以外のものを判断しているわけですから、日本的な感覚からは受け入れにくいと思います。

17. 日本の公正観の「穴」

ただこの「日本的公正観」にはいろいろな「穴」があります。それは実は目に見える部分（表面的な部分）ばかりに集中している、限定されるからです。

まずセンター試験の得点調整は毎年問題になりますが、ではなぜ科目間の平均点差だけが問題にされるのでしょうか。例えば、ある種の証拠がありますが、理科で物理を受けている受験者と生物を受けている受験者ではもともと学力の差があり、物理の受験者の方が平均的に他の教科の成績が良いことが分かっています。その場合に、同じ平均値にそろえるということは生物を受けた受験者にとって有利になるのではないかとということがあります。

日本的公正観の「穴」

- 日本の公正観は、実は、目に見える(表面的な)部分に限定される。
- (センター試験の得点調整で、)なぜ科目間の平均点差だけが問題にされるのか。得点を低下させる調整はなぜ認められないのか。
- 平均値以外にも不公平を生み出す得点分布の側面はある。得点調整で不利になる受験者もいる。
- 1種類の試験でも、予想外の難易差の変化(分布の変化)は、実は不公平を生み出す。
- 試験時間・時期の設定も、評価しようとしている能力とは無関係な変動を付け加えている。
- さらに目に見えないのは、経済格差による受験前の(大きな)格差。

それから、なぜ得点を低下させる方向での調整は認められないのでしょうか。これは平均値以外にも不公平を生み出す得点分布の側面があります。まず、例えばある科目の試験が非常に易しかったとすると、その科目が得意な人にとっては非常に不利になります。つまり、今の入試は基本的には相対評価で行われているので、他の受験者と差をつけるということが合格の絶対条件になるわけですが、皆にとって易しい、皆がいい点を取るような問題では、本当にできる人はかえって不利になります。そういう易しすぎる科目が1つあると、それに平均値がそろえられる他の科

目でも、上位層の間での差が縮まってしまう。だから得点調整で不利になる受験者も実質的にはいるわけです。それなのに、見かけ上の点数が下がることは認めない。これは不合理です。

それから同一教科内に複数の科目があるから公平・不公平があるようにいわれますが、今の話からすると1種類の試験であっても予想外な難易差の変化があれば、実は不公平を生み出しているのです。それから試験の時間や時期の設定も、余計な変動を付け加えているということもあります。そして何よりも目に見えない受験前の大きな格差として経済格差があります。そういうことに関してはあまり問題にされていません。ともかく「試験場に入ったらみんな公平」という感じ。つまり非常に視野が狭い公正観と言わざるを得ないと思います。

18. ここまでの話をまとめると

ここまでの話をまとめると

- 受験者が同一条件で、比較的短時間で受験できる客観テストは、小項目を多数並べることによって、信頼性と妥当性を高められる。
- しかし、それによって測れる学力は、ほぼ知識と技能に限られる。
- 大間によるいわゆる PISA 型の問題を短時間で選抜目的に用いると、著しく信頼性が損なわれる恐れがある。
- 思考力・判断力・表現力にもとづく assessment には十分な時間をかける必要があるが、日本的公正観の下では、その実現は困難である。

受験者が同じ条件で比較的短時間で受験できるというのが、試験の望ましい性質であり、誰が採点しても同じ点数になるということのも公平性の条件だとすると、それは結局、小項目をたくさん並べて信頼性と妥当性を高めるということになります。しかしそれで測れる学力というのは、ほぼ知識と技能に限られてしまいます。大間を使いたいいわゆる PISA 型の問題を短時間で入学者選抜の目的に用いると、少なくとも著しく信頼性が、さらに領域代表性という意味での妥当性も損なわれる恐れがあります。思考力、判断力、表現力にもとづく assessment、個人を査定するには十分時間をかける必要がありますが、日本的な公正観の下ではその実現は難しいと思います。

19. 本当の問題点は？

本当の問題点は？

- 初等・中等教育で習得が目指されている「学力」の多くの部分は、3要素の中の知識・技能であると言える。
- そうした知識・技能自体は、すべての学習者を動機づけるような性格をもっているとは言えない。
- 知識・技能の獲得そのものへの動機づけが、内発的動機づけ (intrinsic motivation) と呼ばれるものであるが、これには個人差がある。
- このタイプの動機づけの強さは、いわゆる受験勉強には有利であり、いわゆる難関大学には、内発的動機づけの強いタイプが多く入学する。

初等・中等教育の目的は、基本的には知識・技能を教授することです。これはもう少し広い意味で言えば、我々の社会の持っている常識や文化を伝授しているというあたりに目的があるということなのではないかと思います。ただ、その初等・中等教育での知識・技能は、全ての学習者を動機づけるような性質を持っているのでしょうか。つまり全ての学習者にとって“面白いのか”ということです。

「内発的動機づけ」とは“知識・技能を獲得することそのものが個人にとってモチベーションになる”ということなのですが、これにはかなり大きな個人差があります。つまり、知識・技能の取得の能力あるいはその獲得された結果以前に、そういうものを学ぶ気になるか、学ぶ方向へ動機づけられるかどうかということについて、非常に個人差があるのです。この内発的動機づけが強いタイプは、受験勉強的には有利です。そのこと自体が何に使えるのか、そういうことを知って何になるのかという疑問を持つ子どもは結構多いですが、とにかく学ぶこと自体が面白い、知識が増えるあるいは技能を身につけることが面白いと考えられれば、受験には有利なのです。いわゆる難関大学の学生は内発的動機づけが強いタイプが多いように思います。

20. 抽象的な内容への理解力

抽象的な内容への理解力

- 動機づけ(興味の方向)と、かなりの相関をもっているのが、一般的知能であり、さらに近年ではその内容は、抽象的思考能力であるというのが有力な見解である。
- 講演者が私立大学に異動して、教えるのに(まずまず)苦勞するようになったのが、推測統計学である。特に、帰無仮説 (null hypothesis) が大きなつまずきの石となるようだ。
- 端的に言えば、仮定の話を理解できない、あるいは受け入れない一群の人たちがいるのであって、これが大学教育においては問題になる。

内発的動機づけについての問題に加えてもう1つお話しします。このあたりはかなり私の印象

に近いものなのですが、いわゆるIQで測られる一般的知能といわれるものと、その動機付けの方向はかなり相関しているように思われます。これはどちらが原因でどちらが結果かということは分かりませんし、そこを問う必要もないのですが、現実には内発的動機付けを持っている人間は一般的知能が高いようです。では一般的知能とは何かということですが、その中心になるのが「抽象的思考能力」であるということ、多くの心理学者が合意しています。

私立大学に異動して、なかなか教えるのが難しいと私自身が感じているものに（名古屋大学で教えていた時も難しかったです）、「推測統計学」があります。特に「帰無仮説（null hypothesis）」は非常に難しいです。例えば2つのグループを比較するとします。今までの教え方と新しい教え方で教える2つのグループを比較するために、1つのクラスを等質的に2つに分けます。ただしその時は、本人がどちらのグループで学びたいかは一切無視してランダムに分けます。これをrandom assignmentといいます。そうして授業が終わった後で試験をし、平均点はどちらのグループのほうが高いかを比較します。新しい教え方のほうの平均点が高ければ、こちらのほうが効果はあったと、とりあえずは考えることができます。

その時に推測統計学の理論では普通どうするかというと、まず2つのグループの間に差がないと仮定します。これを帰無仮説といいます。その上である計算をして、差がないとしたときにこれほどの平均差がおこるような確率は非常に低いということで、この「差がない」とした帰無仮説を否定して、「確かに新しい方法に効果がある」「新しい方法のほうの平均点が高い」という結論を出すのですが、ともかくいったん2つのグループの間には違いがないのだと仮定します。

ここで「仮説」という言葉を使うところに問題があるのかもしれませんが。仮説というと、今までの知見であるとか、あるいは理論的な根拠に基づいて立てるものですが、それはいわゆる研究仮説です。ところが帰無仮説はそうではありません。「この新しい教え方を開発したのは、理論的にあるいは今までのデータに基づいて考察した時、この教えの方が、習得度が高くなると考えたから」というのは研究仮説です。帰無仮説でこの研究仮説を否定します。せっかくそう考えていたのに、帰無仮説では全く何にも考えていないことと同じだと考えます。ここで学生はこれを理解しようと努めますが「それはおかしい」と理解できない、ここからなかなか先に進めないということがよくあります。「これは仮定の話であって、その後の計算を見れば分かるように、その検定統計量、今のような例では“t統計量”、これを使うのが“t検定”ですが、そういったもの前提の式の中に未知数が入ってきてしまうが、帰無仮説を立てることによってそれが無くなるので、やっとな計算することができる。一見回りくどいようだけど、こういうふうに計算しないとイケないのだ」ということを、かなり一生懸命に説明するのですが、なかなかそのあたりをうまく理解できないということがあります。

今の推測統計学の話は非常に端的な例ですが、それ以外の場合のケースについても、仮定をともかくいったん受け入れて、そうしたらどうなるだろうかと考える思考方法というのは大学では非常によく行われることだと思いますが、この仮定の話なかなか理解できにくい人たちがいるのです。

21. 社会が大学に求めているもの

社会が大学に求めているもの

- それでは、そうした学生が、就活で苦勞するか、あるいは、社会に出て不適応になるかと言えば、必ずしもそうとは言えない(決して、有利でもないとは思わが)。
- つまり、大学がもっている研究と教育という機能からみればかなり異質な能力の持ち主が、実は社会の求めている人材であることは、決して例外的な事態ではないように思われるのである(こちらの方が社会的に有能であるとも言えないとは思わが)。
- 要するに、従来高等教育機関が果たすと考えられていた役割と相反するとまでは言わないものの、ほぼ「直交する」ような役割が、実は求められるようになったように思われる。

ではそういう人たちは就職活動で非常に苦勞するのか、あるいは社会に出て不適応になるのかという、必ずしもそうではありません。一流企業や優良企業に就職しているケースも結構あります。その後そこで困るのではないかという、そうでもなく、順調に社会で活躍しています。今、大学がもっている機能は研究と教育であり、基本的に大学教員は研究者ですから、抽象的な思考をしながら仕事をしていますが、このような現状からすると、大学の機能からみればかなり異質な能力を持っている人たちが、実は社会が求めている人材であるということは決して例外的なことでは無いように思われるのです。つまり抽象的思考能力というものは、社会で活動するために必要でも十分でもない条件ではないかということです。今まで大学教員が“高等教育機関が果たすべき役割”と常識的に考えていたことと相反するとまでは言いませんが、ほぼ直交するような役割が実は求められるようになったと思われます。抽象的な能力が無いほうがいいということではありません。それがあということが社会貢献に繋がるのではないのかもしれないということです。

22. 3要素は順に習得されるとは限らない

3要素は順に習得されるとは限らない

- 3要素(知識・技能／思考力・表現力・判断力／主体性)を持って多様な人々と協働して学ぶ態度)という階層性は理解できる。たとえば、知識なしでの思考力はある得ないと、一応は考えられる。
- しかし、習得の順序はどうであろうか？ たとえば、フィールドワークや社会調査の実習を通じて、初めて、データ分析の理論やスキルの獲得の意味を理解し、まじめに学習に取り組むようになるとか、マニュアルにしたがって統計ソフトウェアを操作していた学生が、座学としての社会統計学で、「やってきたことの意味がわかった」と言うといった経験を講演者はたびたびしてきた。

最近の議論を聞いて少し違和感を持っていることについてお話します。学力の3要素ですが、

まず「知識・技能」があって、それに基づいて「思考力・表現力・判断力」、さらにそれが身についた上での「主体性」という階層性を持っているというのは、これは整理の仕方としてはそうなのだろうと思います。例えば、何かしらの知識があるから、それに基づいてものを考えることができるのであり、知識無しに思考ということはないでしょう。しかし、習得の順序としてはどうなのでしょう。なぜこの話をしたのかというと、必ずしも知識・技能の習得に向けて努力するような内発的動機付けを持っているとは限らない人たちが大学に入学してきた場合は、これに当てはまらないということを私自身が経験しているからです。

私が所属している中京大学現代社会学部では、フィールドワークや社会調査の実習等を行いますが、それらを通して初めてデータ分析の理論とかスキルを獲得することに意味があることに学生が気づき、その後真面目に学習に取り組むようになるということがあります。また、先生の指示でマニュアルどおりに統計ソフトを操作していた学生が、座学として学んだ社会統計学で「今まで何をやっているのか全く訳が分からず統計ソフトの操作をしていたが、それには意味があった」と気付く、これは比較的稀ですが、そういう学生は毎年いるのです。

つまり協働して学ぶ経験をすることにより、それを実現していくために基本となる知識・技能が必要であることに初めて気が付くということは十分あり得ますし、現にあります。いわゆる上位大学ではない大学で、知識・技能の習得に向けて努力するような内発的動機付けを持っていない学生だから、それらが身につかないのかということでは必ずしもないと思います。3要素というものの使い方ですが、まずはとにかく知識・技能を無理やりにも詰め込むべきだということを言う人もいますが、そういうことではないと思います。

23. 高大接続の難しさ（まとめ）

高大接続の難しさ(まとめ)

- 現行の共通試験の限界は明らかだとしても、学力の3要素を網羅するような(一発勝負型の)共通試験を作ることは、ほぼ不可能である。信頼性・妥当性が低くなりすぎるからである。
- では、(何らかの改善をしながら)主として知識・技能を測る試験を前提として、個別試験は思考力や表現力を中心とするものにしてはどうか？
- これについては、現在の大学がそうした負担に耐えられるか、いや、そもそもその能力があるかと問われることになるであろう。

これまでの話をまとめます。とにかく現行の共通試験の限界はあります。それは知識・技能のかなり特殊な学び方に限定されてしまうところです。新試験では3要素を網羅するような一発勝負型の共通試験というものを作るのではないかと、何となく期待されているようですが、これは信頼性・妥当性が低すぎて不可能です。では知識・技能を測る試験を前提として、個別試験で思考力や表現力というものを中心に測るものにしたらどうかということですが、これは少し申し上げにくいことですが、今の大学がそうした負担に耐えられるのでしょうか。つまり研究者でもある

教員が、研究者として競争し、その競争力を問われるという現状で、入学試験を作る仕事に時間をかけて果たして取り組めるのかということです。能力がないとは言いません。これについて今、あちこちで弊害が起こってきているということからも、なかなか難しいのではないのでしょうか。

高大接続の難しさ(まとめ)

- さらに、そうしたやり方がもし可能だとしても、それは受験生の多くが、そこで問われる知識技能に対して内発的動機づけをもつと想定される難関大学で有効であるにすぎない。
- それ以外の大学で、どのような入学者選抜が有効か(そもそも、選抜が可能かという問題もある)は、まさに現場が悩んでいる問題である。
- たとえば、表立って議論されることは少ないが、体育会系課外活動の人材育成能力は、大学が今日まで暗黙裡に依拠してきた大きな財産であるように思われる。その高大接続における位置づけも、改めて真剣に問われる必要があろう。

さらに言えば、そういうやり方ができたとしても、それは受験者の多くが知識・技能に対して内発的動機づけを持つと想定される難関大学での話になるでしょう。そうではない大学ではどんな入学者選抜が有効なのでしょう。実を言うと、難関大学ではない大学、いわゆる“Fランク”といわれているような大学が私立大学かなりの部分を占めてきていますが、そのような大学はそもそも選抜が可能なのかどうか、この先の入試制度そのものをどうしていくのかということも、非常に悩んでいると思います。

例えば、表立って議論されることはありませんが、体育会系の課外活動の人材育成についてです。常に目立つアスリートが“大学の看板”として活躍をすると、大学側としてはどうしてもそちらのほうに重点をおいてしまうということがあります。しかしその一方で、ずっと下積みにいる選手もたくさん存在しており、そうした人たちは卒業時にはそれなりに成長しています。これは必ずしも大学教育が果たした役割というのではなくて、やはり課外活動が効いたのだと言わざるを得ません。これは大学が暗黙のうちに依拠してきた“大きな財産”になるのではないのでしょうか。ではこれを「高大接続」でどう位置付けるのか。もちろんスポーツ推薦ということになると思います。一方で卒業時の学力をきちんと担保する“出口管理”というような、今の大学に対するいろんな意味での厳しい流れがあります。体育会系課外活動と卒業時の学力担保は、必ずしも両立しないわけであって、そのあたりも含めた「高大接続」の議論というものが、これから必要なのではないかと私は思っています。