

高大接続で問われるべき 能力と適性とは？

村上 隆

(中京大学・現代社会学部)

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)話をさせていただく。

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

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？

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?

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。

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・・・。

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

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・・・。

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。

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, 今日指されているテストとは?

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。

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。

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 ではなく、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。

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。

問題は評価にかける時間

- 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？

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？

日本的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，選抜に用いられてはならない（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）。

日本的公正観の「穴」

- 日本的公正観は、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。

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。
- 大問によるいわゆる 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。

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・技能であると言えよう。
- 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。

抽象的な内容への理解力

- 動機づけ(興味の方角)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(ますます)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)が大きなたまら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。

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、社会に出て不適應になるかと言えは、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。

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りませんと、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。

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。

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るだろう。