

## 講座

# 教育評価 測定と評価

吉岡昌紀

教職課程のカリキュラムでは、教育評価を扱うことになっており、筆者が担当する「教育方法論」で5回ないし6回分の授業をこれに充てている。本稿は、教育評価について、授業で使用している資料をもとに、その導入部分をまとめたものである。当初は、授業の資料に簡単な説明を添えて、授業の紹介をしつつ内容の説明を行うという、講座と実践報告を合わせたものを考えたが、文章としてまとめたものにしようとするうちに、教科書的な内容に近づいた。しかし、講座あるいは実践報告の要素も残すために、授業で使用している例題やそれに対する略解、質疑応答のような授業でのやりとりを含めた形でまとめた。

## 教育と評価

教師にとって、児童・生徒を評価することは重要な職務である。職務といっても、授業を担当したり生活指導をおこなったりすることと別に評価という職務がある、というわけではない。評価は教育という営みに本来的に組み込まれているもので、教育に携わる以上、必然的に評価に携わっている、という意味で、評価は教師にとって重要な職務である。

いま、評価は教育に本来的に組み込まれていると述べた。しかし、一般には、教育場面での評価に対して、否定的な意見や、否定とまではいかなくても、ある程度は必要だが強調しすぎるのはよくない、点数をつけることが教育の本質ではないといった、評価に対する懸念を示す意見も多くあるように感じることもある。実際の教育評価にさまざまな問題があり、それが教育全体に好ましくない影響を及ぼすこともある。しかし、評価に対する懸念や否定的意見には、誤解も含まれているように思われる。

教育という営みは様々な場面で行われる。親が子どもを育てる、幼い兄・姉がつたない手つきや言葉で弟・妹に何かを教え伝える、学校で教師が児童・生徒に教科書を手に黒板の前で授業をする、クラブで先輩が後輩を指導する……これらの教育の営みは、内容も場所も教え方も大きく異なっているが、そこには共通点がある。いずれも、現在よりも望ましいと思われるところに近づこうという、教える側、学ぶ側双方の願い、思い、目標（意識的な場合も無意識的な場合もあるだろう）に根差している、という共通点である。新しいことを身につけたり理解したりする必要を感じず、現在のままでいることだけしか念頭にないならば、このような営みは生じない。このように、教育は、現在よりもより「よい」状態に近づきたい、近づけたいという願いを含んだ営みであり、より「よい」状態が（無意識的にであっても）目標として置かれている。

そうであるならば、その願いがどのくらい叶ったのか、目標にどのくらい近づくことができたのかを振り返って考えることもまた、教育という営みには必然的に含まれているはずである。何かを願い、その実現に向けて何らかの行動をしているにもかかわらず、その願いの実現について確認しないとしたら、それは自己矛盾であろう。このような、願いを叶え、目標に近づこうとして行う教育という営みが、実際にどの程度実現できているのかを確認することが、教育評価である。これが、教育という営みは本質的に評価を組み込んでいるという意味である。以下、教育は評価を本来的、必然的に含む、ということを前提に、評価を考えていくこととしたい。

評価が教育に必要なとすれば、当然、評価に対して消極的な姿勢をとるわけにはいかない。評価が難しい場面はたくさんあるだろうし、評価をしたくないと感じる場面もあるかもしれないが、そのような場合であっても、適切な評価のあり方を求めながら評価に取り組むべきである。このような姿勢は、特に教師に求められる。学校という制度は、あらゆる教育的営みの中で、もっとも意図的、組織的に、多大な労力を注いで行われる営みである。それに伴って、目標や願いもまた、意図的、体系的に、長期的な視点の中で設定されている。教育評価もまたしかりである。学校での教育に携わる教師は、教育評価についても、広い見地に立って深く理解する必要がある。

質問： 学校の現状をみると、テストをはじめ、評価が多すぎるように思う。これらの評価が進路指導や生徒の序列化のために利用されて、学校が生徒を管理するような窮屈な場所になっている気がする。

回答： 評価に関する問題には、評価の問題というよりも、評価の前提として存在する教育の目標の問題として捉えるべきものが多い。評価は目標に従って変わる。目標が異なれば評価の基準や方法が変わり、その結果、一人ひとりに対する評価も変わる。評価に問題があると思ったら、教育の目的にさかのぼって、問題の原因や解決法を考える必要がある。たとえば、いわゆる有名大学への進学者を増やすことだけを教育の目標としたら、入試に関係する科目の成績が重視され、その成績によって生徒を評価することになるだろう。そして、進学と関係のないことがらは評価の対象外となるかもしれない。このような評価はこの教育目標にとっては合理的で適切であろう。問題があるとすれば、評価法よりもむしろ、成績だけを評価の対象とする教育の目標のほうにあるというべきだろう。

## 測定、評価、統計

評価について考える際には、測定と評価を明確に分けることが重要である。

どのような評価であれ、評価を行う際には、評価の根拠となる事実を知る必要がある。教師としてある生徒を評価しようとしても、その生徒のことを何も知らなければ、評価のしようもない。生徒をよく知ることによって始めて、その生徒を適切に評価できるようになる。

生徒をよく知るためにはさまざまな方法がある。たとえば、観察は情報を得るための重要な方法である。生徒のあいさつの仕方、授業中の様子、休み時間の過ごし方、友人関係などをよく観察す

るだけでも、心理的状态や健康状態、勉学の様子について大事な情報を得ることができる。

対象を知るための重要な方法の一つに測定がある。生徒を知る場合も、日常的な行動の観察によることはとても重要だが、それだけでは十分に生徒の様子を知りえないこともある。たとえば、生徒がだるそうで熱がありそうだと感じたら、体温計を使って正確な体温を知ろうとするだろう。このように、対象のある特性について正確に記述することを「測定」という。今あげた例では、体温という特性を、体温計を使って正確に記述したわけである。

学校にはさまざまな測定がある。学校でもっともよく行われる測定は、勉強に関するさまざまなテストであろう。授業内の小テスト、学期の中間や期末の定期テスト、入試、模擬試験、各種の検定試験など、さまざまなテストが学校では実施される。これらにはいくつかの目的があるが、主要な目的は、生徒の学力を測定することである。

ところで、本論の初めから「評価」という言葉を用いてきたが、その意味を少し明確にしておきたい。評価とは、対象のある特性について価値判断を行うことを言う。価値判断とは、その特性について「よい」「悪い」といった判断をすることである。教育は何らかの願いの実現、目標の実現に向けての活動であると、本論の初めで述べた。当然、このような願いが実現した状態、目標が達成された状態は「よい」状態である、という価値判断がある。教育評価は、その「よい」状態にどのくらい近づいたかを判断している。

評価を行う際には、その前提として、測定などによって対象の特性を正確に把握する必要がある。測定と評価の関係について、体重という特性を例に取り上げることにする。漠然と「A 君は重い」と言わずに、体重計を用いて「A 君の体重は 100.0 キロである」と正確に述べる作業が測定である。一方、この測定結果をもとに「よい」「悪い」という判断を加えることが評価である。ここでの「よい」「悪い」はいろいろで用いられる。100 キロという体重を、健康という観点からよい、悪いと判断することもあるだろうし、何らかの美的基準から美しい、美しくないという意味でよい、悪いと判断することもあるだろう。小さな飛行機にあと一人乗客を乗せるか否かを判断するときに、100 キロの人を乗せるのは危険だからよくない、避けるべきである、という判断をするかもしれない。これらはいずれも価値判断を含む評価の例である。

事象の正確な記述を目的とする測定（ここには価値判断は含まれない）と、正確な記述をもとに価値判断を行う評価とを区別することは重要である。

また、「統計」という言葉の意味についても、ごく簡単に触れておくことにする。ここでは簡略に、統計とは、たくさんの測定値を簡単な数値等に集約し、測定値全体の特徴などを分かりやすく示す方法である、としておく。

質問： 測定は正確に行うことができるだろうか。たとえば、学力はどの程度正確に測ることができるのだろうか。高校の定期試験で、苦手な分野なのに、たまたまヤマがあたっていい点をとったことがある。これは、テストが学力の測定を正しくできてなかった例だと思う。

回答： 対象の特性によって、測定の可能性・容易性は大きく変わる。身長や体重を測定するのは難しくない。適切な器具があれば、手軽に正確に測定できる。しかし、学力を測

定するのはそれほど簡単ではない。質問にあるように、たまたま正解をすることも、たまたま不正解になってしまうこともある。こういう測定の誤差は、身長や体重の測定に比べて、学力の測定では大きい。

より根本的な問題もある。たとえば、そもそも何を学力というのだろうか。また、何を測定したら学力を測定したことになるのだろうか。学力とは何かを明確にしておかないと、どのように測定したらいいのかもわからない。しかし、学力とは何かを明確に述べるのは簡単ではない。学力に限らず、教育に関わる人間の能力や資質を測定するのは、難しいことが多い。

この問題は、信頼性、妥当性という二つの概念を使いながら改めて考察する。

## 測定と評価 練習問題

ここにあげる練習問題は、実際に授業で使用している資料を修正したものである。

問 次の作業は、測定と評価のいずれだと思うか。その理由は？

1. 身長をはかった。
2. 体重をはかった。
3. 血液型を調べたら A 型だった。
4. その結果を友だちに話したら、「血液型、A 型なの？ いいなー。」と言われた。
5. 身長と体重から BMI (Body Mass Index) を計算した。

参考：BMI = 体重[kg] ÷ 身長[m] ÷ 身長[m]

6. BMI から、自分は標準体重だと考えた。

参考：肥満度の判定基準（日本肥満学会 2000）

低体重（やせ）	BMI が 18.5 未満
普通体重	18.5 以上 25 未満
肥満（1 度）	5 以上 30 未満
肥満（2 度）	0 以上 35 未満
肥満（3 度）	35 以上 40 未満
肥満（4 度）	0 以上

7. ある高校で、1 年生 1 学期中間テストの平均点を計算したところ、数学 I は 53 点だった。
8. この数学 I のテストで 76 点をとった生徒の偏差値は 67 だった。
9. この生徒は、このテストに関して、5 段階の 5 だとされた。

以下、各問の解答と簡単な解説である。

1. 身長をはかった。
2. 体重をはかった。

この 2 問は迷いなく判断できるだろう。健康診断で身長測定、体重測定というように、いずれも測定である。

もっとも、これらのように測定であることに疑問の余地がないようなことがらであっても、我々の実際の受け止め方はそれほど単純ではない。頭の中で何らかの評価を添えていることが

多いからである。子どもの時に健康診断で身長を測った時に、「背が伸びてよかった」「背が伸びすぎていやだ」などと思った記憶がある人もいるだろう。体重を測って「よかった。増えてなかった。」「やばっ。少し甘いものを控えよう。」など考える人は多いに違いない。このような思いには、測定結果に対する何らかの価値判断が含まれているのであり、したがって、身長や体重に対する評価を行っていることになる。日常生活では、測定と評価を峻別せず、両者が混ざり合って捉えられていることがよくある。

また、評価という意図をはっきり持つて行う測定も多い。ダイエット中に体重を測定するのは、体重〇〇キロという目標にどれだけ近づいたかを評価するためである。勉強して練習問題を解いてみるのは、その内容を修得するという目的を達成したかどうかを評価するためである。これらの場合は、測定と評価はほとんど一体のものと捉えられている。

このように、測定と評価が混ざり合っていることや、一体となっていることがあるが、しかしそういった場合であっても、両者を切り分けて捉えることができる。測定と評価は別のものであり分けて考えることができる、ということを確認することが重要である。

### 3. 血液型を調べたら A 型だった。

血液型の検査結果は数字では表されず、A、B、O、AB という文字で表されるが、この検査は測定である。

### 4. その結果を友だちに話したら、「血液型、A 型なの？ いいなー。」と言われた。

A 型が一般にうらやましがられる対象であるかどうかはともかく、この友だちの「いいなー」という言葉は、明らかに、その友だちが A 型に「よい」という評価を与えていることを示している。このように友だちは価値判断を示しているので、この言葉は評価である。

### 5. 身長と体重から BMI (Body Mass Index) を計算した。

### 6. BMI から、自分は標準体重だと考えた。

先に 6.について。BMI を使ったこの分類は、肥満度という観点から、健康状態が望ましいものであるのか否かを判断している。この判断は健康状態に関する価値判断を含んでいるから、評価にあたる。

では、5.の BMI の算出はどうだろうか。

身長も体重も測定の結果得られた数値である。それを組み合わせて計算した結果得られた数値も測定の結果と考えられる。したがって BMI の算出は測定のうちに含まれる。

これに対して疑問を感じる人もいるだろう。この計算は、健康状態を評価する BMI を得る、という意図をもって行ったのであり、そもそも BMI の評価を行う意図がなければ、このような計算をすることもないだろう。そう考えると、この計算は評価に含まれるのではないか。このような反論があるかもしれない。

たしかに BMI の計算は BMI の評価と分かちがたい。しかし、問題 1.、2.の説明で述べたように、測定と評価が密接に関わっている場合であっても、両者を分離して捉えることが可能である。BMI に関しても、数値自体は測定の結果なのであり、それを BMI の判定基準に照らす作業が評価なのである。

### 7. ある高校で、1 年生 1 学期中間テストの平均点を計算したところ、数学 I は 53 点だった。

実は、この質問はやや不適切である。平均は統計値の一種であり、測定の結果とはやや異なるし、評価とも異なるからである。もっとも、すぐ後で述べるように、統計は測定と深く関係している。もし、この問いに対して測定、評価の二分法で答えるよう求められたとすれば、測定と答えるほうが適切な選択肢を選んだことになる。

統計とはデータの集約作業であると、前節の終わりのところで述べた。平均は、もっともよく使われる統計の一つである。ある試験の受験者が 10 人だけならば、10 人の得点を眺め渡せば、集約作業を行わなくても全体の特徴はわかるだろう。しかし、受験者が 100 人、1,000 人となった場合、数値を眺め渡しても全体の特徴をつかむのは難しい。このようなときにデータ全部を集めて平均を計算すれば、一つの数値によってデータ全体の中心という特徴をわかりやすく示すことができる。

平均は、測定結果全部を合計し、データ数で割る、という集約作業を行って得られた数値である。つまり、平均は複数の測定値から得られる値であり、一つ一つの測定とは異なる。もっとも、測定結果がなければ平均は計算できないから、両者は深く関係している。

平均を求める計算には価値判断は含まれていない。したがって、平均と評価とは異質のものである。

8. この数学 I のテストで 76 点をとった生徒の偏差値は 67 だった。

この問いも、前問と同じくやや不適切な問いである。偏差値は、平均と同じように測定結果をもとにした統計値だからであるが、もし二分法で答えるよう求められたら、測定と答えるほうが適切である。

この質問は、学生がもっとも間違える質問である。偏差値を評価と考える学生が多いのである。

偏差値は、個々の得点を、平均と標準偏差という二種類の統計値を用いて換算した数値である（標準偏差や偏差値についても教育評価の授業で扱っているが、ここでは省略する）。平均や標準偏差は、データ全体を一つに集約した値として示されるが、それに対して偏差値は、平均と標準偏差を使って個々のデータに対して計算される。生徒や受験生にとっては、偏差値はつねに自分の得点に対して与えられるある数値であり、この数値は、自分の得点が評価されるときに用いられる数値である。また、高校や大学をランキングするための評価指標としても偏差値は用いられる。そのため、学生にとって、偏差値は、データ全体を参照した、自分の得点についての評価と感じられるのであろう。実際に、偏差値をもとに評価された経験はだれにもあるだろう。

しかし、5. の質問にあった BMI の数値と同様、偏差値自体に価値判断は含まれていない。もっとも、BMI の数値から自動的に健康状態の評価が得られるように、日本社会の中では、偏差値から自動的に何らかの評価が得られてしまっている、ということかもしれない。偏差値は、統計の基本的な概念であり、多くの統計的な概念や手法が、偏差値をもとにしている。このような偏差値が評価を示す数値だと考えられているのは大きな問題である。

9. この生徒は、このテストに関して、5 段階の 5 だとされた。

これは評価である。5 をもっとも上の評価とする 5 段階の成績評価はおなじみであろう。こ

こでの「5」は、数字ではあるが、数や量を示しているのではなく、「よい」「わるい」という価値判断を表す語として用いられている。