

# 形成的評価の位置と方法

－ 総括的評価と形成的評価の差異に着目して－

## The Roles and Methods of Formative Evaluation － Differences Between Summative Evaluation and Formative Evaluation －

北 川 剛 司

Takeshi Kitagawa

( 要 約 )

本研究の目的は、形成的評価の位置を明らかにすることをとおして、その方法について考察することである。本稿では、まず、教育評価における形成的評価の提唱者であるブルーム (Bloom, B. S.) の言論に主に依拠しつつ、同じくブルームの提唱した総括的評価との対比によって形成的評価の位置を明らかにした。次に、形成的評価の位置をふまえてその方法について検討し、形成的評価の方法として対話を用いるという一つの示唆を得た。

(キーワード)

形成的評価、ブルーム (Bloom, B. S.)、対話型評価

### I. はじめに

ブルーム (Bloom, B. S.) は、教育評価における3つの機能を提唱したことで、わが国においてもよく知られている。すなわち、診断的評価、形成的評価、総括的評価と称される3つの機能である。

ブルームは教育評価の機能をこのように3つで捉えたが、それらを互いに異なる別個のものとは捉えていなかった。例えば、「診断的評価は、形成的評価や総括的評価と区別されると同時に、それらと密接な関連を持っている」<sup>1</sup>、「生徒に対して彼の進み具合に関する情報をできるだけ迅速に与えてやることは、まさに形成的評価の本質であるが、総括的評価もまたこういうふうで使用できる」<sup>2</sup>と述べられるように、教育評価の機能は、相互に関連性があるとされている。

だが一方で、「実際には、教師は、しばしば1つの試験から両方の評価（総括的評価と形成的評価—引用者註）を引き出そうとする。しかしそのような場合には、形成的テストだけの場合と異なった複合的な刺激を学習者に与えるおそれがあり、また、非常に長い複雑なテストを必要とすることになるであろう」<sup>3</sup>と述べて、総括的評価と形成的評価が混同されることの否定的側面についても言及している。

教育評価におけるこれら3つの機能それぞれが関連性をもつということと、それぞれの機能を混同するというとは異なるということに、我々は自覚的であればならないのではないだろうか。つまり、関連性をもつということは、同じ評価方法をとるという意味ではないということに自覚的であるべきだろう。教育評価の方法について議論する際、それを一枚岩でとらえて議論するのではなくて、いま一度、教育評価の機能ごとの違いに着目しつつ、各々について丁寧に議論する必要があるのではないかと思われるのである。

形成的評価とは、学習の途中において実施される評価で、教授活動および学習活動へのフィードバックのための評価であると理解されているが、本研究においては、この形成的評価を対象として、形成的評価の方法について考察する。本研究の目的は、形成的評価の位置を明らかにすることをとおして、そ

の方法について考察することである。

そこで、まず、教育評価における形成的評価の位置を、目的、方法、アカウントビリティというそれぞれの側面から、ブルームの言論を中心に依拠しつつ総括的評価と対比させることによって明らかにする。そして、形成的評価の位置をふまえたうえで、主観的な評価の理論に関する動向を中心に、特にその方法について検討する。

## Ⅱ. 教育評価における形成的評価の位置 — 総括的評価との対比をとおして —

### 1. 目的の側面

すでに述べたように、総括的評価の結果もまた形成的評価のように、教授・学習へのフィードバックとして用いられるのであり、総括的評価と形成的評価の目的は、それぞれに共通する部分をもっている。しかしながら、その主目的に焦点化すると、それぞれの差異を明確に述べることができる。

ブルームは、「形成的な観察の主要な目的は、・・・（中略）・・・完全に習得するためにはどんな課題が必要かを、学習者と教師の双方に明らかにすることが目的なのである」<sup>4</sup>と述べる。また、「形成的評価のテストは学習過程の一部とみなされるべきであって、けっして生徒の能力判定と混同されたり、また成績づけの過程に含まれたりすることがあってはならない」<sup>5</sup>と形成的評価が成績づけの目的で使用されることを否定している。すなわち、形成的評価の主目的は、成績づけといった価値判定を行うことではなく、完全習得にむけた教授・学習の手がかりを得ることである。

これに対して、総括的評価は、「全課程あるいはその1部分について達成された学習成果の程度を把握するというはるかに一般的な評価である」<sup>6</sup>とされ、そうした総括的評価の目的については、「成績をつけることが目的となろう」<sup>7</sup>と述べられている。すなわち、総括的評価の場合は、形成的評価のように、教師や学習者へのフィードバックということにその主たる目的があるのではなく、あくまで学習成果の程度の判定、すなわち成績判定にこそその主たる目的がある。<sup>8</sup>

上で述べた、形成的評価および総括的評価の目的の議論は、現在なされているそれらの目的の議論と基本的に同じである。

### 2. 方法の側面

ブルームは、それ以前の教育評価において主流であった正規分布曲線にもとづく相対評価法ではなく、形成的評価においても、総括的評価においても、目標に照らして達成状況を判定するという方法を提唱した。

ブルームが、目標の達成状況をはかるのに用いる手段として想定していたのはテスト法である。形成的評価の手段は「特別に作られた形成的テスト」<sup>9</sup>、総括的評価の手段は「期末試験、あるいは総括的テスト」<sup>10</sup>と述べているように、形成的評価も総括的評価もともにテスト法を評価方法としていることがわかる。

さらに、そのテスト方法についてくわしくみていくと、ブルームは、「客観テスト」の問題、および、問いに対してある程度の長文で答えさせる記述式のテスト問題を紹介している。しかしながら、形成的テストの問題と総括的テストの問題をそれぞれ区別しては述べていない。

ブルームの述べるテストの方法をみると、そのほとんどが、評価者の解釈の余地のない「客観テスト」であることが分かる。こうした評価方法には次のようなことを指摘することができる。すなわち、「客観テスト」によっては、把握できる能力に限界があるということである。

ブルームは、自身のタキソノミー研究において、認知領域における教育目標を分類して「知識」・「理解」・「応用」・「分析」・「評価」という5つの能力を挙げていることから分かるように、子どもの学習を、単なる知識量の蓄積とはとらえていなかったが、「知識」以外の応用的な能力さえも、「客観テスト」を用いて評価しようとしている。このように「客観テスト」をそうした応用的な能力の把握にまで用いることは現在批判的にとらえられている<sup>11</sup>。

一言で「テスト」と言ってもさまざまなテストがある。「客観テスト」といわれるような、評価者の解釈の余地のない記述式および選択式で解答するテストもあれば、評価者の解釈を必要とする、ある程度長い文章で記述解答させるテストや実技のテスト等のパフォーマンスのテストもある。子どもの学力が基礎から応用までを含む広いものであるならば、その評価においては、これらのテスト法を複合的に用いることが求められるだろう<sup>12</sup>。

ブルーム以降、応用的な能力を評価するための新たな動向が生まれている。すなわち、評価において評価者の解釈の余地を積極的に認め、その解釈の方法を研究しようとするものである。その一つが、ルーブリックづくりである。ルーブリックによって、評価者間の解釈の基準を統一し、そのルーブリックにもとづいて判定を行おうとするものである。現在、わが国の教育評価の議論において、パフォーマンス評価を提唱する論者の多くが、パフォーマンスを解釈する手段としてルーブリックを提唱している<sup>13</sup>。ただし、これは、総括的評価の方法か形成的評価の方法かということを確認にした上で議論したものではない。

### 3. アカウンタビリティの側面から

教育評価におけるアカウンタビリティとは、かならずしも一義的ではないが、近年においては、直接的な関係者以外にも、外部の第三者、つまり保護者や地域住民も含むすべての人・組織などに対して説明する責任という意味をもつようになってきている。教育評価におけるアカウンタビリティを果たすためには、評価の目的や対象、方法も含めて、評価の手続き全体が明確であることが求められる。評価は一般に、ハイ・ステイクス（high stakes）であればあるほど、このアカウンタビリティが求められる<sup>14</sup>。つまり、個人の将来の教育機会や職業機会に大きく影響したり、学校の成果を示すものとして社会の注目を集めたりする評価であるほど、その手続き全体が明確であることが求められる。

この評価のステイク（利害に関わる程度）という観点に着目して形成的評価と総括的評価を比較すると、形成的評価はその性質から、ロー・ステイクス（low stakes）な評価であるといえる。というのも、「形成的評価は成績付けには使われない」と明言されるように、形成的評価の結果はその後の教授活動や学習活動へのフィードバックとして用いることを第一の目的として利用されるからである。したがって、形成的評価においては、評価の手続きを当事者と第三者に対して明確に説明することはさほど問題にならない。

一方、総括的評価はその性質から、ハイ・ステイクスな評価であるといえる。というのも、総括的評

価は、学期末等に行われることから、成績評価と直接的に結び付けられやすく、子どもの進路決定等に強い影響を及ぼしやすいからである。この意味で、総括的評価はその手続きを明確にし、それを説明することが求められる。

#### 4. 形成的評価と総括的評価の比較

まず、目的という側面における形成的評価の位置についてであるが、形成的評価の主目的が子どもの成績づけということにはなく、その主目的は教授・学習の改善にあるという、教育評価における形成的評価の独自の位置が明らかとなった。

次に、方法という側面における形成的評価の位置についてであるが、これについては、総括的評価との明確な方法上の違いを見いだせなかった。教育評価の先行研究において、形成的評価と総括的評価を区別してその方法を論じているものはない。

三つ目に、アカウンタビリティという側面における形成的評価の位置であるが、「成績づけに使う」総括的評価とは違って、「成績づけに使うべきでない」とされる形成的評価の特徴から、そうしたロー・ステイクスな評価においては、アカウンタビリティもそれほど求められないということが明らかとなった。

以上をもとに、形成的評価と総括的評価を比較すると表1のようになる。

表1：形成的評価と総括的評価の比較<sup>15</sup>

	形成的評価	総括的評価
目的	子どもの学習の進展に関する教師と生徒へのフィードバック (→成績をつけることではない)	単元、学期、課程の終りに、単位を認定したり成績をつけたりすること。
方法	特別に作られた形成的テスト (→あくまでテスト法による)	期末試験、あるいは総括的テスト
アカウンタビリティ	ロー・ステイクスであるがゆえに、アカウンタビリティはそれほど求められない。	成績づけ等に用いられることから、ハイ・ステイクスとなり、アカウンタビリティが要求される。

### Ⅲ. 主観的な形成的評価の方法

「国際的な研究者たちは、現在のところいかなる『形成的アセスメントの理論』もないことに気づいている。したがって、効果的な形成的プロセスの要素を理解することは、まだ非常に多くのデータを積み上げ推論する努力の段階である」とされるように、形成的評価の研究はまだ発展途上である。というのも、総括的評価とは違って形成的評価は、アカウンタビリティから解放されている分その方法の自由度が高いがゆえに、たとえその方法が教師の目で見えた判断といった必ずしも手続きが明確とは言えないものであったとしても、教授・学習の改善に生かされるならば、有用だと言えるからである。ゆえに、こうした教師の目で見えた判断にもとづく評価も、形成的評価の理論研究において取り扱っていかねばならないトピックであるといえる。

そこで、本章では、形成的テストの方法というよりも、主観的な判断にもとづく形成的評価（以下、主観的な形成的評価）に関する研究の動向を中心に、特にその方法について検討する。

形成的評価には、形成的テストだけでなく、教師が目で見えた子どもの様子に関する判断、具体的に言えば、子どもの理解に関する授業中のメモ書き等も含まれると考えると、そうしたある意味主観的と言える教師の判断を、その教師固有の名人芸に終わらせるのではなく、共有財産にしていく作業が必要である。つまり、主観的な形式的評価の方法について議論される必要がある。その際、Ⅱ. で述べた形成的評価の位置に立ち返って、総括的评价における主観的な方法の議論とは区別しつつ論じる必要があるだろう。

主観的な総括的评价の方法については、これまで筆者の拙稿の中で一部明らかにしてきた。主観的な総括的评价の方法については、総括的评价はアカウントビリティーを重視されるということもあって、その手続きに関する議論ができた。すなわち、その手続きを明文化することは一定程度可能であった。しかしながら、主観的な形成的評価の方法については、手続きを明らかにすることは難しい。難しいというより不可能である。あるいは、手続きを決めるべきでないというべきかもしれない。というのも、教師が目を見て子どもの様子を判断するという行為は、その場その時とその対象に応じた極めて状況依存的・対象依存的なものだからである。

「形成的評価の特質は、・・・(中略)・・・まさしくすぐれた教師たちに内在していた教育技術のひとつに合理的な自覚化を促そうとしたものである」といわれるように、すぐれた教師が普段やっている授業中の机間指導や赤ペン指導といったものから「褒める」といった指導的评价に至るまで、実に多様なものが主観的な形成的評価に含まれると考えられる。

このような形成的評価の方法のあり方について、一つの興味深い提案がなされている。それは OECD 教育研究革新センターの形成的アセスメントに関する次のような定義の中に表現されている。「形成的アセスメントとは、生徒の学習ニーズを確認し、それに合わせて適切な授業を進めるための、生徒の理解と学力進歩に関する頻繁かつ対話型（インタラクティブ）なアセスメントを指す」。すなわち、OECD 教育研究革新センターによると、形成的な評価は対話型をもとに進められるべきだと述べられているのである。形成的評価においては教師が子どもの学習の様子について一方的に判断を下すのではなく、対話という評価方法を通じて、子どもの考えにも向き合いつつ評価を行うという態度が目指されるべきことが提案されている。

#### Ⅳ. おわりに

本研究を通じて、形成的評価の方法について次のことが明らかとなった。まず、形成的評価の目的を鑑みて、ロー・ステイクスな形成的評価においては手続きが明確であるというよりも、いかに多面的に対象をとらえ、フィードバックとなる情報を多く引き出せるかという部分に主要な関心があつてよいということである。このように考えるならば、形成的評価における方法の議論に対して、アカウントビリティー、すなわちその方法・手続きが明確で結果が客観的であることを強く要求するのはあまり意味をなさないということになるだろう。

次に、形成的評価においては、対話型の方法がとられるべきということである。評価といえば、教師が子どもの学習に対して行う価値判断であるというように一面的に捉えるのではなく、形成的評価は子

どもと対話することを前提として、教師の考えと子どもの考えとの突き合わせの中で、子どもに必要な指導・支援を行う行為であると捉える必要があるだろう。

#### 主要参考文献（註にあるもの以外）

- ・ Bloom, B. S. (1980). 'The New Direction in Educational Research: Alterable Variables.' In *Phi delta kappan*, vol.2, no.1, pp. 382-385.
- ・ 梶田叡一（2002）『教育評価[第2版補訂版]』有斐閣双書。

#### 註

- 1 B.S.ブルーム、J.T.ヘスティングス、G.F.マドゥス著、梶田叡一ほか訳（1973）『教育評価法ハンドブック—教科学習の形成的評価と総括的評価—』第一法規、126頁。
- 2 同上書、103頁。
- 3 同上書、89頁。
- 4 同上。
- 5 同上書、81頁。
- 6 同上書、89頁。
- 7 同上。
- 8 総括的評価の結果の利用法として「成績の評定、技能及び能力の認定、後続課程での成功の予測、後続課程での教授の出発点の決定、生徒へのフィードバック、グループ間の結果の比較」（同上書、96頁。）といった様々な目的が挙げられている。たしかに、これらの中には形成的評価の目的になっているものもある。だが、総括的評価において、第一番目の目的として、「成績の評定」が挙げられていることから、ブルームの考える総括的評価の主目的は「成績の評定」であると推察できる。
- 9 B.S.ブルーム、J.T.ヘスティングス、G.F.マドゥス著、梶田叡一ほか訳（1973）前掲書、130頁参照。
- 10 同上。
- 11 例えば、松下佳代は、評価の対象とその評価方法の関係について「重要なのは、〈評価したいと思っている学力ができるだけ直接的に表れる課題〉（直接性）であることです」（松下佳代（2007）『パフォーマンス評価—子どもの思考と表現を評価する—』日本標準ブックレット、16頁。）と述べ、対象に応じた評価方法の必要性を論じている。
- 12 有本昌弘は「形成的評価から形成的アセスメントへ」（有本昌弘（2008）「監訳者あとがき」OECD教育研究革新センター編著、有本昌弘監訳『形成的アセスメントと学力—人格形成のための対話型学習をめざして—』明石書店、272頁。）と述べて、「形成的アセスメントとは、形成的テストのみに止まらずさまざまなデータ収集手段を取り入れ」（同上。）ると述べている。
- 13 例えば、松下佳代（2007）前掲書や西岡加名恵（2003）『教科と総合に活かすポートフォリオ評価法—新たな評価基準の創出に向けて—』図書文化など。
- 14 田中耕治（2008）『教育評価』岩波書店、139頁参照。

## 形成的評価の位置と方法

- 15 B.S.ブルーム、J.T.ヘスティングス、G.F.マドゥス著、梶田叡一ほか訳（1973）前掲書、130-131 頁参照。
- 16 OECD 教育研究革新センター（2008）「形成的アセスメントの範囲と方法論」OECD 教育研究革新センター編著、有本昌弘監訳『形成的アセスメントと学力—人格形成のための対話型学習をめざして—』明石書店、272 頁。
- 17 田中耕治（2005）『よくわかる教育評価』ミネルヴァ書房、9 頁参照。
- 18 北川剛司（2006）「コンテンツ・リテラシーとしての書く力の評価に関する研究—自然主義的アプローチを中心として—」日本教育方法学会編『教育方法学研究第 32 巻』49-60 頁など。
- 19 スクリヴァン（Scriven, M. S.）のチェックリストによるゴール・フリー評価や、アイズナー（Eisner, E. W.）の芸術批評アプローチ、グーバとリンカーン（Guba, E. G. & Lincoln, Y. S.）の第四世代評価などがそれにあたる。
- 20 田中耕治（2008）前掲書、123 頁。

