

# 教育評価の最近の動向

藤岡 秀樹

(京都教育大学 教育学科)

## Recent Trends in Educational Evaluation

Hideki FUJIOKA

2007年11月30日受理

**抄録**：最近の日本内外の教育評価の動向について、①指導と評価の一体化、②教育評価の技法、③指導方法の改善—の3領域にわたって紹介し、併せて課題についても言及した。①では「P-D-C-A」「評価規準と評価基準」を、②では「ルーブリック」「モデレーション」「パフォーマンス評価」「オーセンティック評価」を、③では「逆向き設計」「ポートフォリオ検討会」を取り上げた。

**キーワード**：教育評価の改革、指導と評価の一体化、教育評価の技法、指導方法の改善

## I. はじめに

2001年に指導要録が改訂され、「評定」欄が「絶対評価を加味した相対評価」から「目標に準拠した評価（いわゆる絶対評価）」に転換され、従前から絶対評価で評価されてきた「観点別学習状況」との整合性が取れるようになった。併せて、個人内評価の重視も行われるようになった。目標に準拠した評価を実施するに際しては、妥当性のある評価規準と評価基準の作成が必要であり、形成的評価を通した「指導と評価の一体化」が強く求められる。このような教育評価の改善が各地で進められるようになった。

他方、OECDのPISA調査やIEAの学力調査においては、日本の子どもの学力は概ね良好であるものの、「関心・意欲」が学力上位国と比べて低く、経年的変化では読解力や数学的リテラシーの相対的順位の低下が見られていた。これらの結果を受けて「学力向上」の方針が文部科学省から提起された。さらに、2007年4月には文部科学省が小学校6年生と中学校3年生を対象に「算数・数学」と「国語」の全国学力調査を悉皆調査の手法を用いて行われた。子どもの学習意識や生活調査も併せて実施されたが、応用力や活用力に課題があることが示された。

このような教育評価の改革をめぐる流れは加速度的であり、検討すべき論点も多い。山森（2006）は2004～2005年を中心に、教育評価についての教育心理学的研究や国際的学力調査をレビューしている。また、筆者は2000年代前半の日本の教育評価の動向と課題を教育政策との関連で論じている（藤岡，2004）。本論では、日本国内外で近年話題となっている教育評価の諸概念について取り上げ、今後の研究の方向性について論じてみたい。

## II. 指導と評価の一体化

### 1. P-D-C-A

P (Plan) - D (Do) - C (Check) - A (Action)は、もともとは経営学者のドラッカー (Drucker, P.) によって提起されたマネジメントサイクルに由来する。当初は、P-D-S (See)であったが、評価を次の計画に活かすことができていることが指摘され、Actionを追加してP-D-C-Aとなった。P-D-C-Aの学校現場への導入は、当初は学校経営や学校評価が中心であったが、近年、指導過程にも導入されるようになった。つまり、1単位時間の形成的評価から児童・生徒の学習到達度を把握し、それに伴って次時の指導計画を組み立てるとい

うサイクルである。結果の評価だけではなく、過程（プロセス）の評価の重視もこのP-D-C-Aモデルと関係している。形成的評価が重視され、その結果を指導に活かすことが強く求められる。「C（未到達）」の児童・生徒への学力回復指導も念頭に置くことが大切である。

授業終了時に行う形成的評価は、時間も短いことから評価観点を多面的に捉えることは困難である。1～2の評価観点到って実施し、中単元を終えた段階で全観点を網羅できればよいだろう。評価技法も小テストだけではなく、発表・ワークシート・観察など多面的な技法を用いることが肝要である。

## 2. 評価規準と評価基準

目標に準拠した評価の導入により、妥当性のある評価規準と評価基準づくりが強く求められている（藤岡，2003）。評価規準は「何を評価するか」といった質的なものであり、評価基準は「どの程度到達しているか」といった量的なものである（藤岡，2006）。国立教育政策研究所や都道府県教育センターでは評価規準表を作成し、各学校で活用されている。しかしながら、各校独自の創意工夫ある評価規準とはなっていない。例えば、「総合的な学習の時間」や中学校の選択教科に課題を残している。また、高等学校での評価規準表の作成は十分なものとは言えない。

評価基準については、「A（十分に到達）」と「B（概ね到達）」の差異が判りにくいとの現場の声をよく耳にする。カッティングポイントをどう設定するのかについては、経験則に委ねるのではなく、客観的・科学的なものにすることが必要である。

中学校の実技教科では、一人で全学年の評価規準と評価基準を作成しなければならないことが多いが、労力もかかり、適切なものになっていない場合が散見する。市や郡単位の公的研究会に各校の教師が評価規準表を持ち寄り、相互に検討する機会を持つことが望ましいであろう。

# Ⅲ. 教育評価の技法

## 1. ルーブリック

ルーブリック（rubric）は評価指標とも呼ばれ、児童・生徒の学習達成状況を評価するための評価基準であり、数段階に分けられた尺度と学習内容の質的な特徴を記載した記述語から成るものである。ルーブリックは質的な判断もすることができるため、パフォーマンス課題のような評価にも適用することができる。ルーブリックは2000年代になって日本に紹介され、実践が緒についたばかりである。

ルーブリックは事前に教師が評価基準を設定するのではなく、児童・生徒の作品などを収集し、複数の採点者が独立して採点し、特徴を抽出していく必要がある。各段階にふさわしい児童・生徒の作品を添付することで、評価者による差異が生じることを防ぐことができる。採点を通して段階分けや記述語の作成などを行う必要があるが、教師間の合意形成を大切にすることが何よりも必要である。児童・生徒の表現力や思考力、課題解決力を評価するためにパフォーマンス課題が今後多く用いられることになる想定できるが、ルーブリックは最適の評価技法である。有効性は高いが、時間と労力がかかる点が課題であろう。

## 2. モデレーション

モデレーション（moderation）とは、複数の評価者間で評価基準を共通理解し、同一基準で採点し、評価の一貫性・信頼性を確保する手法のことである。モデレーションの方法としては、①参照テストやスケーリングの方法を用いた統計的なモデレーション、②査察に基づくモデレーション、③再調査委員会、④コンセンサス・モデレーション、⑤グループ・モデレーション、⑥機関レベルの承認、⑦本質的モデレーションなどに分類できる（Gipps, 1994）。

また、モデレーションは評価過程を統一する方法と評価結果を統一する方法に大別することができるが、評価過程を統一する方法の方がよいとされている（西岡，2005a）。各評価段階ごとに典型的な児童・生徒の作品例（anchor）を添付しておくことが望ましい。

グループ・モデレーションは、教師が児童・生徒の課題の事例を使って討議し、評価基準についての共通理解を図るものであり、システムレベルでの判断の一貫性を高めるためのものである。イギリスでは複数の学校や機関からメンバーが参加することもあるが、このグループ・モデレーションを通して教師の評価の技能を高めることができるという効果があり、日本でも導入を期待したい。

### 3. パフォーマンス評価

客観テストで測定されるものには限界があるため、近年、パフォーマンス評価が行われるようになった。一口にパフォーマンス評価と言ってもその内容は多岐にわたり、実験・観察、演技、作品作成などが挙げられる。研究者によって、その捉え方は微妙に異なっているが、実際の活動を通して評価が行われることが共通した特徴である。

パフォーマンス評価は、形成的評価のような短時間の学習活動における評価ではなく、長期にわたる学習活動の評価である。複合的・総合的な技能や能力を用いる課題を通して行う評価でもある。Gipps(1994)によれば、イギリスでパフォーマンス評価に当たるのは、標準評価課題(Standard Assessment Tasks)とGCSE試験のコースワークであるという。

松下(2007)は、パフォーマンス評価の難点として、①多くの時間と労力を要する、②課題数が制限される、③客観テストに比べると「信頼性」が低い、④ルーブリック作りが難しい、⑤言語による場面設定にはいくつかの問題がある—の5点を挙げている<sup>1)</sup>。このような難点や制約があるにも関わらず、パフォーマンス評価を実施することによって、教師の教材研究が進み、専門的力が培われるという点で意義がある。

パフォーマンス評価においてモデレーションは柱であり、採点者間の評価の一貫性を向上させるだけでなく、評価過程の統一の面でも必要とされている(Gipps, 1994)。客観テストでいくら信頼性と妥当性を高めても、パフォーマンス課題に置き換わるものではない。「総合的な学習の時間」の評価のみならず、様々な教科学習においてもパフォーマンス評価の可能性を示唆している。

### 4. オーセンティック評価

オーセンティック評価(authentic assessment)とは、「真正の評価」とも呼ばれ、1980年代後半にアメリカで登場した概念である(Wiggins, 1998)<sup>2)</sup>。標準テスト、とりわけ択一式テストは生活場面を離れた作成的な問題を含み、真の生きた学力を測定していないという批判から導き出された考え方である。オーセンティック評価は、現実の子どもの生活に近い場面での課題が出題され、授業内容と評価が密接に結びついている。

オーセンティック評価の評価方法は、パフォーマンス評価、ポートフォリオ評価、ウェビング、概念地図法などがあり、学習内容との関連で適切な評価方法を選択することになる。一部の研究者では、オーセンティック評価とパフォーマンス評価を同一の評価技法として論じているが、これは誤りであることも指摘しておこう。オーセンティック評価の技法の一つにパフォーマンス評価が挙げられるからである。オーセンティック評価は、長期にわたるものであり、労力と時間がかかるものである。教師の主観にとらわれず、客観性のある評価にするためにはモデレーションを含めた教師間の相互理解が求められる。学力の広範な領域を評価するには適していないことも難点である。

オーセンティック評価を進めるためには、リアリティのある生活場面に関わりの深い授業を計画することも大切であろう。言い換えれば、現実社会で必要とされる技能や能力を活かす学習活動を実施することによって、児童・生徒を評価することである。

## IV. 指導方法の改善

### 1. 逆向き設計

逆向き設計(backward design)とは、Wiggins & McTighe(1998)によれば、①望ましい結果を明確にし、②そのことを容認できる証拠を決定してから、③学習経験と指導の計画を立てるというものである。結果から始

めるという点、あるいは指導計画の前に評価の構想を行うという点が、従来のカリキュラム設計とは逆になっているためこのように呼ばれている（遠藤，2005）。

逆向き設計は、以下のような各段階から構成されている（遠藤，2005）。

- ①望ましい結果を特定する
- ②容認できる証拠を決定する
- ③学習経験と指導の計画を立てる

①の段階は目標設定の段階であり、カリキュラムの優先事項が決定される。②の段階は、特定の単元や授業を設計する前に「評価者」のように考えてみる段階である。③の段階は指導計画を立てる段階であるが、望ましい結果と評価を特定した後で行うことが留意点である。

各段階に対応して次のような設計スタンダードが設定されている。

- ①その設計は、目標にしている内容の重大観念にどのくらい焦点化しているか？
- ②その評価は、求めている結果についての公平で妥当で信頼できる十分な測定尺度をどのくらい提供しているか？
- ③学習計画はどのくらい効果的で魅力のあるものか？
- ④どの程度その単元全体は上述の3段階の諸要素を伴いながら一貫しているか？（全体）

この中で、②がオーセンティック評価と関連が深いことが見いだせる。

逆向き設計を行う際には、子どもたちが現実の文脈の中で使いこなせるような位に深く扱うべき知識を、目標設定において絞り込むことが重要であると西岡（2005b）は指摘している。さらに、カリキュラム設計に当たっては、教えるべき知識のうちどれが「事実に基づく知識」「個別的スキル」に当たり、何が「転移可能な概念」「複雑なプロセス」「原理と一般化」なのかを明らかにする必要があると指摘している（西岡，2005b）。

逆向き設計は、発想の大きな転換が必要であり、学校現場ではなかなか受け入れられにくいものである。しかしながら、授業改善に資するところが大きい指導法である。日本では、この考え方が導入されたばかりであり、まだ実践はほとんど進められてはいない<sup>3)</sup>。今後、指導法の研究開発を進めていく必要があるだろう。

## 2. ポートフォリオ検討会

ポートフォリオ評価は普及は著しく、「総合的な学習の時間」を始め、様々な教科学習でも評価技法として用いられるようになってきた。ポートフォリオ検討会は、ポートフォリオに収められている子どもの様々な作品を素材として、子どもが学習を振り返り、教師と共に評価や確認を行い、新たな課題を設定する場である。

ポートフォリオ評価法の原則として、西岡（2002）は次の6点を挙げている。

- ①ポートフォリオ作りは子どもと教師の共同作業である。
  - ②ポートフォリオ作りの過程で、子どもと教師は具体的な作品を蓄積し、取捨選択したり整理したりする。
  - ③ここでいう作品には、子どもの作った完成品だけでなく、完成品を作り出す過程で生み出される下書きやメモなども含まれる。
  - ④子どもと教師はポートフォリオを使いつつ話し合う。
  - ⑤ポートフォリオ検討会は、ポートフォリオ作りの初めでも途中でも締めくくりでも行われる。
  - ⑥ポートフォリオ作りは長期にわたるものであり、継続性をもつ。
- ④と⑤がポートフォリオ検討会に相当する。

アメリカで構想・実践されているポートフォリオは、①子ども自身が自分なりの評価基準を設定し、自己アピールをするために作る「最良作品集ポートフォリオ」、②予め決められた評価基準を教師が提示する「基準準拠型ポートフォリオ」の2つの展開が見受けられる（西岡，2002）。いずれの形態であれ、子どもの参加が保障されなければならず、フィードバック機能を有効に働かせる必要がある。また、評価に際しては、ルーブリックの作成<sup>4)</sup>が必須となる。

教師がポートフォリオを見て評価する段階に留まっている学校が多いが、是非とも時間を捻出してポートフォリオ検討会を設定することが大切である。さらに、クラスメートも発表会に参加し、「学びの共有化」を図ることや相互評価を実施することも有意義である。

## V. 結びに代えて

小論では、近年話題となっている教育評価の諸概念を取り上げ、紹介と課題について言及した。「学力低下」問題が様々なところで論じられるようになった今日、指導に結びついた教育評価のあり方が問われている。客観的で妥当性のある教育評価の技法を開発することは大切であるが、指導のあり方と結びつけなければ「心理測定 (psychometric) 主義」に陥る危険性がある。「教育測定 (edumetric)」の視点を忘れないことが大切であろう。

教育評価は子どもを評価するだけでなく、指導の改善に役立てることが大きな目的である。カリキュラム評価や指導法の評価についても、研究の推進が求められる。このような論点については別の機会に譲りたい。

### 註

- 1) パフォーマンス評価の概説と採点の具体例については、松下 (2007) が詳しい。松下は、お茶の水女子大学の 21 世紀 COE プログラム「誕生から死までの人間発達科学」のプロジェクトⅢ「子どもから成人への移行に及ぼす社会・文化的要因の影響」のサブプロジェクト (JELS : Japan Education Longitudinal Study) の「算数・数学」学力調査グループに関わり、パフォーマンス課題の作成や評価を担当した。
- 2) Wiggins のオーセンティック評価における「真正性」の概念については、遠藤 (2003) が詳しい。
- 3) 逆向き設計を実践している学校としては、京都市立衣笠中学校や鳴門教育大学附属中学校が挙げられる。
- 4) ポートフォリオ評価におけるルーブリックの位置づけについては、西岡 (2001) を参照されたい。

### 文 献

- 遠藤貴広 2003 G. ウィギンズの教育評価論における「真正性」概念—「真正の評価」論に対する批判を踏まえて— 教育目標・評価学会紀要, 13, 34-43.
- 遠藤貴広 2005 G. ウィギンズのカリキュラム論における「真正の評価」論と「逆向き設計」論の関連—「スタンダード」概念に注目して— 京都大学大学院教育学研究科紀要, 51, 262-273.
- 藤岡秀樹 (編) 2003 小学校新指導要録記入文例 1000 日本標準
- 藤岡秀樹 2004 教育評価の動向と課題 教育目標・評価学会紀要, 14, 3-10.
- 藤岡秀樹 2006 評価規準と評価基準 辰野千寿・石田恒好・北尾倫彦 (編) 教育評価事典 p.80. 図書文化
- Gipps, C. V. 1994 *Beyond Testing : Towards a theory of educational assessment.* The Falmer Press. 鈴木秀幸 (訳) 2001 新しい評価を求めて—テスト教育の終焉— 論創社
- 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 No.7
- 西岡加名恵 2001 ポートフォリオ評価法におけるルーブリックの位置づけ 教育目標・評価学会紀要, 11, 2-12.
- 西岡加名恵 2002 教育評価への子どもの参加—ポートフォリオ評価法— 教育方法, 31, 110-125.
- 西岡加名恵 2005a 比較可能性とモデレーション 田中耕治 (編) よくわかる教育評価 Pp.72-73. ミネルヴァ書房
- 西岡加名恵 2005b カリキュラム適合性 田中耕治 (編) よくわかる教育評価 Pp.70-71. ミネルヴァ書房
- Wiggins, G. 1998 *Educative Assessment : Designing Assessments to Inform and Improve Performance.* Jossey-Bass.
- Wiggins, G. & McTighe, J. 1998 *Understanding by Design.* Association for Supervision and Curriculum

**Development.**

山森光陽 2006 学力低下論争, 目標準拠評価の定着, 学力テストブームの狭間で 教育心理学年報, 45, 92-103.