

## 評価とアカウンタビリティにおける教育現場の重要性

中野 真志 (愛知教育大学生活科講座)

磯谷 みどり (愛知教育大学大学院)

(2007年10月31日受理)

## An Importance of the Classroom in Assessment and Accountability

Shinji NAKANO (Department of Life Environment Studies, Aichi University of Education)

Midori ISOGAI (Graduate Student, Aichi University of Education)

**要約** 日本の学校教育における評価が目標標準評価によって行われるようになってから5年以上経過し、その評価観は広く知られるようになった。しかし、国の施策と教育現場とのずれも浮き彫りになっている。本小論では、マーク・ウィルソン (Mark Wilson) の判断のコミュニティに関する研究成果を学びながら、教育現場の重要性について論じる。また、日本において教育評価、教育のアカウンタビリティに対する意識が高まってきているが、教育現場とのずれも生じているため、このようなテーマについて論じることは価値があるだろう。

**Keywords** : 評価, アカウンタビリティ, 判断のコミュニティ

### I はじめに

2000年12月の教育課程審議会答申<sup>1)</sup>において、公の教育機関である学校の評価の重要性が述べられ、従来の集団標準評価と異なる、目標に準拠した観点別評価を積極的に取り入れることが求められた。また、評価の妥当性・信頼性を高めるために、各学校での評価規準を作成することが求められるなど、日本の学校教育における評価活動に大きな変化をもたらした。このような評価に対するとらえ方やアプローチは、過程を大切に生活科の授業実践において特に有効であると考えている。また、この答申で述べられた評価観は、現在では広く知られるようになってきている。

一方、国際学力調査での学力や意欲の低下が指摘され、学校教育の現状について把握することが求められるようになった。また、公教育の質を保証することも求められている。そこで、2005年10月の中央教育審議会答申<sup>2)</sup>において、子どもたちの学習到達度・理解度についての全国的な学力調査を実施することが適当であることが述べられ、2007年4月に小学校6年生と中学校3年生で全国一斉に調査が行われた。この調査では、子どもの学力や学習状況をテストと質問紙によって把握し、国、教育委員会、学校レベルで教育の改善を図ることが目指されている。

このように、日本において教育評価、教育のアカウンタビリティに対する意識が高まってきている。しかし、教育現場とのずれも生じているのではないか。アメリカでも日本と同様の状況がみられ、それを教師集団の取り組みによって解決していこうとする研究がある。そこで、本小論では、マーク・ウィルソン (Mark Wilson) の研究成果を学びながら、日本での評価とアカウンタビリティに対する示唆を得たい。

教育的なシステムは、子どもの教育に対してアカウンタビリティを負わなければならない。つまり、1つに子どもが教育を受けて発達したということを示すこと、2つにある教育プログラムが子どもの発達を促進させるものであるかどうかをチェックし、より発達を促進させる教育プログラムを選択していくことが必要である。

さて、教育評価とアカウンタビリティの関係について考える前に、教育現場からシステムへ、そしてシステムから教育現場へという2つの情報の流れがあることを確認したい。教育現場からシステムへの情報の流れとは、教育現場での具体的な授業実践や、その中での子どもの様子などから、システムを見直し、改善していくという情報の流れである。一方、システムから教育現場への情報の流れとは、システムによって決められたことが教育現場に伝えられ、システムに沿った授業実践に取り組むなどの情報の流れである。現在合衆国では、スタンダードに基づく枠組みを使用することが、最も一般的なアカウンタビリティのアプローチとして取り組まれている。スタンダードに基づく枠組みを使用したアカウンタビリティは、まず、いくつかの標準テストの成績を州のなかから抽出し、その結果を学校などで作られているアカウンタビリティ・システムに基づいて分類する。それを教育現場へ戻し、さらにそれを親や学校管理者のような第三者へ提供するという仕組みになっている。

このように、スタンダードに基づく枠組みを使用した場合の情報の流れは、主にシステムから教室へというものである。標準テストから引き出された評価情報は、教育現場で集められた評価情報の一部でしかない。しかし、評価は教育現場で教師がデザインし、評価し、実践するという毎日の積み重ねから構成されて

いる。それゆえ、学校において評価を有用なものにするためには、第一に毎日の実践を助け、実践の中で活用されるものにならなくてはならない。つまり、教育現場の状況にあわせてシステムを改善していく必要があるのだ。それにもかかわらず、システムから教育現場へ、という情報の一方通行が存在し、教育実践現場にあまり目が向けられていない現状がある。

## Ⅱ 評価をデザインするための観点と課題

本題に入る前に、評価とアカウンタビリティの問題を考えるための評価に関する観点を提示したい。ペレグリーノ (Pellegrino, J.) らは、評価は3つの側面から構成されていると述べている<sup>3)</sup>。

- ①認識の側面 (教師がもつ子どもの認識のモデル)
- ②観察の側面 (教師が子どもの認識を評価するのに使う方法)
- ③解釈の側面 (教師が「観察」と「認識」を関連付け、解釈する方法)

具体的には、①はカリキュラムと教授計画、②は評価課題 (テスト、宿題など) ③はある課題における子どもの得点などの結果を解釈することである。これは、「アセスメント・トライアングル」(図1) で示される。

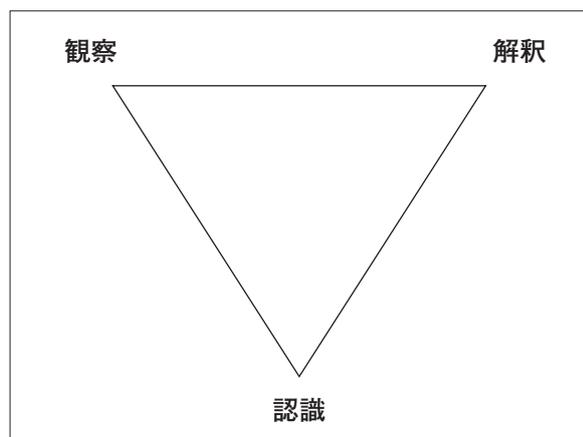


図1 アセスメント・トライアングル<sup>4)</sup>

「アセスメント・トライアングル」の観点で考えると、教育現場での評価をデザインする上で、3つの課題が明らかになる。

1つ目の課題は、「システムレベルの評価」の方法を見つけることである。つまり、システムレベルで教育現場において直接使用できる結果を生み出すことのできるテストをデザインすることが必要である。教育現場でテストの結果を直接活用できるなら、そのテストの内容と構成は妥当であるといえる。しかし、そのようなテストによって、教師自身が評価をデザインしたり、解釈したりするときに考え方が固定されてしまう恐れがある。

2つ目の課題は、「教室で使うための評価」をデザ

インする際に気をつけなければならないことである。つまりそれは、評価がカリキュラムと矛盾しないこと、評価とカリキュラムが相補的であること、子どもの学びを個人でもクラス全体でも筋道立て、学習を解釈するのに能率的で、学習計画と課題に便利な評価方法を見つけることだ。

3つ目の課題は、アカウンタビリティ・システム全体に関する問題である。それは、教育現場レベルでの評価とシステムレベルの評価の両方に効果があり、便利な方法を見つけることと、有効な情報を州、市、学校、学級といったさまざまなレベル間で伝えていくことだ。具体的には、以下の3点が目指される。

- アカウンタビリティ・システムをより妥当にするために、子どもの学びが反映されているパフォーマンスに対する教師の判断を向上させること
- アカウンタビリティ・システムにおいて、教師が専門性や立場をいかせるようにすること
- システムレベルの評価で筋が通っており、同時に教育現場でなされている実践と矛盾がない評価の枠組みとスタイルを使うことを教師に促すこと

これら3つの課題全てを扱うことで、教育現場からシステムレベルまでの全ての評価システムにわたる評価の枠組みと方法の統合がなされるだろう。

## Ⅲ 課題への取り組み ～判断のコミュニティ～

教育的なシステムは、先述の通り、子どもの教育に対してアカウンタビリティを負わなければならない。つまり、①子どもが教育を受けて発達したということを示すこと、②ある教育プログラムが子どもの発達を促進させるものであるかどうかをチェックし、より発達を促進させる教育プログラムを選択していくことである。合衆国では、①を教科のある範囲での標準テストの結果を集めること、②を①の結果を子ども集団単位で比較し、検討することのみで完了している。その結果、標準テストは、毎日の教育環境の中にとっても深く浸透してきている。

このように、1つの方法で評価を行うことは、教育プログラムの有効性を比較しやすいという利点がある。そのため、選択式の標準テストでの評価が用いられるようになり、広がっている。しかしながら、子どもが特定の方法以外のやり方で表現することもまた重要であるにもかかわらず、1つの方法で評価を行うことで、その評価では測定できないものを排除してしまうことになる。そのため、伝統的な選択式の標準テストの長所を大切にしながらも、最高の教育実践を映し出すタイプの評価を促進するシステムが求められる。

この問題に対するアプローチの1つに、「判断のコミュニティ」とは、よい教育実践を共有したり、模範的なカ

リキュラムを考案したりするなど、子どもの学習に関する諸問題を扱う教師集団である。この判断のコミュニティのアプローチは、子どもの教育を一般に任せられている専門家、すなわち教師が、アカウンタビリティ・システムにとって必要な情報に最も多く直接アクセスしているという考えに基づいている。教師がもっている情報や専門的な知識を直接使うことで、妥当性、信頼性、そして能率性を最もよく達成することができると思う。さらに教師は、教育的に確かな方法で、子どもの学びに関する情報を解釈するのに最も適している。そのため、判断のコミュニティのアプローチは、アカウンタビリティ・システムの中心に教師と教師による判断を据えているのである。

このように、判断のコミュニティは伝統的な選択式の標準テストの長所を大切にしながらも、最高の教育実践を映し出すために、子どもの取り組みと、取り組みに対する判断をつなぐ解釈のシステムの中心に教師をおいている。そのアプローチは以下のようなものである。

1. 達成変数を設定すること。達成変数とは、ある教科における子どもの到達度から、子どものパフォーマンスのレベルを述べるための実質的な枠組みのことである。
2. 子どもの返答を評価するための評点ガイドラインに付随した達成変数のために、評価形態のセットを構成すること。具体的には、達成変数を実現したものであり、私たちの最高の教授活動を具体化した広範囲の課題、プロジェクト、そしてパフォーマンスである。
3. 上述の達成変数や評価形態のガイドラインを使う上で、判断の「適切さ」が求められる。つまり、評価する教師の力量の向上と、教師間で判断をすり合わせる調節機能の2つのことが必要となる。
4. 評価の証拠を評定する際に、妥当で信頼できることを保障する手段と、評価の証拠をしっかり見取ることができていることを裏付ける手段の2つの手段を用いて、教育の質を監督し、保証していくことが必要である。この方法は、2の評価形態からの証拠を尊重する。

この判断のコミュニティが示した4つのアプローチは、評価のどの部分に対するアプローチだろうか。先述の評価トライアングルの3つの頂点にあてはめると、1の枠組みは子どもの「認識」のモデル、2の評価形態は「観察」のモデル、3の適切化と4の質のコントロールは「解釈」の側面であり、3は教師レベルの「解釈」、4はシステムレベルの「解釈」に対するアプローチである。

このシステムのための評点の構成要素は、個々の子どもや学校、または他の子どものグループの達成変数における位置をそのまま扱うだろう。そして、その変数に沿ったそれらの進捗を分析するだろう。この情報

は、授業が生じる諸条件および教育的な介入の領域を考慮することによって、評価されるだろう。

このようにして判断のコミュニティが集めた適切な評価は、目的がどれくらい達成できたかを報告する基準として役立つだろう。また、それらの評価は特定の教育プログラムを評価する目的にも適しているといえる。

このような判断のコミュニティによるシステムは、過去のシステムと比べて大きく優れている点が2つある。1つ目は先述の通り、広範囲の評価モデルを受け入れることができるようにデザインされている点である。つまり、判断のコミュニティによるシステムはあらかじめ仕組まれ、価値づけられた教育活動の範囲外の評価タイプを受け入れうるということだ。それゆえ子どもの発達をより真正に評価しうるのだ。これは、教育において価値のあるものを標準テストによって決められることがない、ということの意味する。判断のコミュニティによるシステムを活用することで、今まで標準テストが実施される学年で標準テストのために使われていた時間や労力を通常の教育活動に使うことができると同時に、その通常の教育活動が、そのままテストのための実践になる。

2つ目は、システムの中心に教師を据えることで、子どもを直接教える立場の人をシステムに引き込んでいる点である。そうすることで教師が以下の点を得ることができると考えられる。以下で言う「枠組み」については次節で述べる。

- 枠組みの知識
- 枠組みと関連した新しい教育実践と、新しい教育実践と関連した新しい評価モデル
- 教師が協力して取り組むための新しい方法
- 教師が自分の実践を評価する手段

つまり、システムを教師が開発することで、教師の専門性を高めることにもなるということだ。また、上述の4点は判断のコミュニティをつくっていくうえで乗り越えなければならない課題でもある。

しかし、このシステムに対する不安の声も聞かれる。1つ目は、コストに関することである。システムをつくるためには、システムにかかわる人を育てることやシステムのために必要な技術や物品を開発する必要がある。そのためには、大規模な投資が必要である。

ただ、このシステムの開発は、アカウンタビリティのためのものであると同時に、子どもへの教育を向上させるための教師の専門性の開発につながると考えられる。そのため、多少コストがかかっても得られるものは大きい。また、一度軌道に乗れば、判断のコミュニティのために情報を集める実践は、通常の授業の一部になっていくだろう。そうすれば、現在テストとテストの準備に費やされている時間を生産的に使うことができる。そして、今まで標準テストにかかって

きた費用は、教師の専門性の向上のために使うことができる。このように、システムの開発に費用がかかっても、全体としてはマイナスにならないといえる。

2つ目は、教師の判断を信頼できるかどうかである。これは、このアプローチに関する最も大きな問題である。確かに、教師に子どもを教えることを任せ、そのことに関して信頼しているのに、子どもを評価することに関して信じていることができないのはおかしな話かもしれない。しかし、子どもの学びに対する教師の評価が信頼できるかどうかは大きな問題である。そのような不安の声があっても、判断のコミュニティは教師を中心としたシステムを構築している。そして、教師の判断が教育的であると思われ、教育的であることが確認でき、それが保障されるためにシステムはデザインされている。そのデザインは具体的に、以下の4点のようである。

1. 子どもの学習について測定したものの全てをアカウンタビリティ・システムに用いるための枠組みが用意されている。全てを用いることで、子どもの学習を担任教師だけでなく、親、管理職、政策決定者がよりよく理解することができる。
2. 評価のタイプは、子どもの反応を観察することができ、比較できることが確認されるものを選択し、使用する必要がある。
3. 評価を適切にする過程では、教師が判断をするうえで、他の教師といっしょに取り組むようになっている。つまり、教師の判断を適切化する方法は、綿密な調査とは異なる方法である。
4. 教育の質を保証するために、システムを超えて比較と確認ができることや、公平さを保障することが必要である。

つまり、判断のコミュニティは、コミュニティの中でなされる判断の有用性と妥当性に寄与するようにデザインされている。

#### IV 「枠組み」へのアプローチ

では、判断のコミュニティの4つのアプローチについて詳しく述べていく。

判断のコミュニティが活動する際にはまず、コミュニティのメンバーのなかで考え方や用いる言葉について共通理解を図る必要がある。たとえば教育は、「カリキュラム上のさまざまな過程において子どもが発達すること」が目指されるという共通理解が必要である。それを前提として、「発達」には何が含まれるかなど、子どもの「発達」についての共通理解、達成変数のセットに関する共通理解、そして達成変数に沿ったパフォーマンスを識別するためのレベルのセットへの共通理解が必要である。

「枠組み」とは、パフォーマンスの限定されたレベ

ルが位置づけられた、そして子どもが発達すると予期される、カリキュラムの限定された範囲の中でのさまざまな達成のセットである。また、「達成変数」は、より基本的な知識から、より進んだレベルの理解、そして技術へ広げたものである。そして「レベル」は、達成変数に沿って質的に区別できるパフォーマンスによる理解を段階的に示している。

枠組みの概念は以前から用いられており、現代の「基準」は枠組みの概念を単純化したものである。基準が教育の個々の目標のリストからなっているのに対し、「枠組み」の概念は基準と同様のリストをもち、さらにそれらを、意味をもった、そして解釈できる子どもの教育的な発達の進行に組織化している。枠組みで示される子どもの発達は、基本的に質的なものである。

#### V 「評価形態の範囲」へのアプローチ

子どもの発達を科学的な観点から見ようになったのは、19世紀の中ごろからである。子どもの発達過程についての科学的な研究が進むにつれ、量的に発達を示す必要性が生じ、子どもの発達を年齢順に記録にとどめる方法の開発や、個人差についての理解もなされるようになった。それが、テストの始まりである。一方、学校では子どもの発達に関して、学校の管理運営者、親、議員、そして社会全体に対してアカウンタビリティを負っている。子どもの発達を説明するために、学校では、さまざまな評価方法が用いられるようになった<sup>5)</sup>。

1970年代以降、学力低下が指摘されるなかで、学力保障を行っていることについてのアカウンタビリティが学校に求められるようになった。そのため、国や州が中心となって標準テストに基づくアカウンタビリティが促進された。しかし、1980年代頃から、アカウンタビリティの手段として標準テストが多く用いられていることに対する批判が高まってきた。標準テストは、たとえば「算数で80%の得点だ」と言うように、量的な測定はできるが、質的な測定には向かない。そのため、「学力とは何か」と、学力への問い直していく運動があり、子どもの学力を総合的に捉える真正評価の重要性が受け入れられるようになった<sup>6)</sup>。

真正評価の考え方を発展させたアサッチベイカー（Aschbacher, P.R.）は、新しい評価の鍵となる特徴を以下のように表現した<sup>7)</sup>。

- 子どもが単に1つの答えだけでなく、高いレベルの思考や問題解決技能を要求する何かをしたり、創造したり、生産する
- 評価課題自体が子どもの学びにとって意義深く、魅力的な教育活動になっている
- 子どもが生活する実際の社会の中にある問題や、完

全にそのものでなくても、実社会に近い問題を評価課題にする

- 結果だけでなく、結果に至るまでの過程も評価される
- パフォーマンスの規準（criterion）と基準は、事前に子どもに知らせ、教師と子どもで共通理解をする  
このような評価観を前提に、真正評価を支持する人たちは、さまざまな方法で評価情報を集めることを述べている。

では、実際にどのような評価方法があるだろうか。評価情報を収集するためのさまざまなアプローチについて考えるために、図2を示した。この図によって全ての評価形態を表現することはできないが、ここで述べる論点に関する評価のさまざまな側面を描写できると考える。

この図の縦軸は、評価課題にかかわる統制の大きさを示している。統制が大きい評価課題は、子どもが決まった答えでしか解答できない「外部から決められた課題」である。なかでも、多肢選択式のテストは、決まった言葉以外で答える可能性のある短い答えのテストと比べてもより統制が大きく、最も統制の大きい課題であるといえる。逆に統制の小さい評価課題は、「課題がない」のが特徴である。つまり、教師の主観

によって子どもの全体的な印象を判断しているのだ。この2つの間に、「状況に合わせて教師が作成した課題」と「外部のガイドラインの範囲内で教師が作成した課題」が当てはまる。

横軸は、子どもの学習をどのように判断するかについての統制の大きさを示している。統制の大きい判断は、機械的に評点化できるものを評点化することである。逆に統制の小さい判断は、「基準のはっきりしない教師の判断」である。この2つの間の判断は、判断する人がどれくらい熟練しているかと、判断するための計画案がどれくらい規定されているかによって変わってくる。たとえば、評価材料について研修を受け、熟練した判断者だと認められた教師が判断を行う場合、そうでない教師が評価材料を自分で解釈して判断を行う場合よりも判断にかかわる統制は大きくなる。

この図の右上隅の評価は、全ての子どもに対して同様の客観的な基準を使って評価することができる。また、心理測定のモデルの歴史的に顕著なタイプと一致する課題で構成される。そのため、評価の信頼性が大きいといえる。それに対して左下隅の評価は、実際の教育内容に近く、教師の累積された経験に基づき、またその地域の状況に柔軟に対応するものである。その

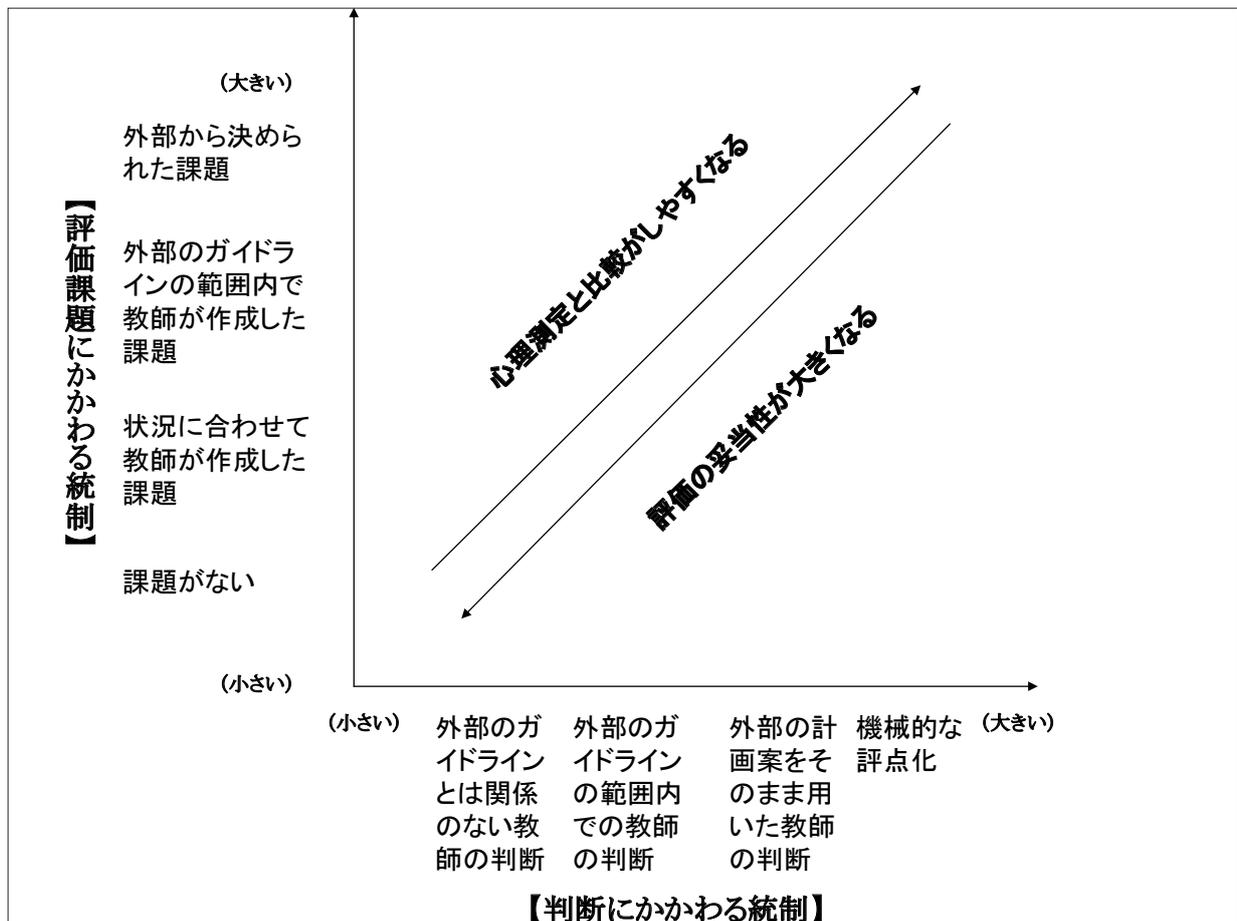


図2 評価形態のための統制に関する図<sup>9)</sup>

ため、個々の教育や子どもに対する評価の妥当性が大きいといえる。このように、図の中の異なる位置にある評価は、異なる理由のために異なる利点をもっている。この2つの利点を両方とも含むことができる評価が理想的だが、図が示すように単一の評価では両方の利点を含むことはできない。

では、よく知られている評価形式である、外部の選択式テストと教師の評定はどうだろうか。外部のテストは、評価課題とパフォーマンスの結果をどのように評価するかが一貫しているのが特徴だ。そのため、学級の異なる子ども同士を比較することができ、比較的评价の公平さについて批判されることがないという利点をもっている。対照的に、教師による評定の利点は、教師が授業をし、その子どもをずっと見ているため、その子どもにとってふさわしい評価課題を選ぶことができるということである。また、教師はその子どもの学習を近くで見ているため、子どもを判断するのにふさわしい人物であるという利点もある。つまり、図2の中では外部のテストが右上隅に、教師の評定は左下隅に位置づけられ、この2つの評価形式は対角線の両端になっている。

さまざまな異なる形式の評価（図2のさまざまな位置にある形式）からの情報を利用する試みの1つとして、信頼性を増大させるための機会と同様に、評価が妥当性を増大させるための機会を利用しうることがあげられる。このように、評価形態の範囲を使用する上で、教師にとって最も重要なモチベーションになることは、1つの評価方法で評価の妥当性（既存の評価用語で言えば“真正さ”）と信頼性の両方の必要性を満たすことはできない、ということだ。

## Ⅵ 「適切な評価システム」へのアプローチ

教育関係者に「教育と評価を結びつけることが重要だ」と言いながら、評価形態の範囲を規定することは、矛盾している。つまり、評価形態の範囲を決めることで、その範囲外の教育は評価されなかったり、範囲外の教育がなされなくなる可能性があるということだ。また、さまざまな評価形態への子どもの返答に対する教師の判断を、他の教師の反応と、熟練教師の反応と、評価の形態とを照らし合わせ、その評価形態に組み込んでいく方法が必要とされている。このように、評価形態に関して、さまざまな課題がある。以下に述べる「適切なシステム」は、アカウンタビリティシステムに子どもの返答や反応を組み込んでいくためのものとしてのみでなく、教室に新しい評価の本質を伝える方法としてもまたみられるべきだ。そのためには、このシステムが単なる評価の見本であってはならないし、そうならないためにも、情報ネットワークと教師のための支援を設立しなければならない。

### 1 適切な評価システムの概要

適切な評価システムは、教師がより首尾一貫した判断をするのを助けること、教師の判断を実証することの2つの過程がある。この適切なシステムは、統制の大きい評価方法のみに基づいた伝統的なシステム以上の利点をもっている。たとえば、教師がその子どもたちにとってよりふさわしい評価形態を決定できることがあげられる。そして、評価を適切にする過程に教師が直接かかわることで、教師が同僚と一緒に自分の評価実践を見直し、改善してだけでなく、評価を超えた教育実践全体の専門性を向上していくことになる。

### 2 教師の判断を適切にすること

適切なシステムの1つ目は、教師が子どもの学習を判断する“過程”である。

教師がシステムの、ある評価モデルや評価のガイドラインを使う前に、まず、その枠組みに含意されていることについて学び、評価についての専門性を高める必要がある。そのために、イギリスでは「モデレーション・グループ」という地方の教師のリーダーのグループが結成されている<sup>9)</sup>。このグループは、先述の「判断のコミュニティ」のスタッフの育成、評価活動の実践、その地域に合った評価課題と授業実践などを行っている。また、そのなかで教師同士が助け合いながら、子どもたちの学習の支援を考えていく。このような話し合いを通じて、教えることの核となる問題について話すことになり、教師の専門性が高められていく。

### 3 教師の行った判断が適切であるかの確認

適切なシステムの2つ目は、教師の判断が適切であるかの確認である。この確認には2つの方法がある。その方法の1つは、教師が子どもの学習のデータを集め、評価し、そのデータをもとに熟練した教師が再び評価する、というものである。この方法により、教師は自分の評価と熟練教師の評価を比較することができ、より真正な評価にすることができるようになる。

もう1つの方法は、子どもが統制の大きさの異なるタイプの評価法で同時に評価され、その2つの情報を解析するというものである。たとえば、「教師の評定」と「外部のテスト」という図2で示した最も極端な2つの評価の使用があげられる。教師の評定は、学級内の評価としては適しているし、ある子どもの中での変容についてはよく示すことができるが、他の学級との関連についてはほとんど示すことができない。また、外部のテストは他の学級の子どものと比較することに適しているが、学級内の評価を行うには適しているとは言いがたい。

## Ⅶ 「教育の質の保証」へのアプローチ

教育の質の保証のために、以下の4項目の手続きが必要とされている。

- 異なる評価形態を使って集めた情報の首尾一貫性を検査すること
- 子どものパフォーマンスを達成変数上に示すこと
- 達成変数に置き換えて、アカウントビリティ・システムの構造上の要素（課題と評価者）を述べること
- 妥当性と信頼性についての研究と偏りと公平さの研究することで、システムの役割の基準の特徴を確立すること

この項目のように、評価形態を選択し、適切なシステムをうまくかみ合わせることで、アカウントビリティの要求とスタッフ育成の要求を満たすことができる。それだけでなく、教育的なプログラムを評価するための基礎と、教育システム全体を供給しうる評価システムを構築していくことができる。ただし、このシステムは全ての州で同じものである必要はない。ある州は統制の小さい評価としてポートフォリオを使えばよいし、別の州では教師の評定を使えばよい。また、統制の大きい評価も、それぞれの州で異なる方法を使うことができる。このように、州の特色にあった評価形式を選べばよいのである。一方、同じ州であっても、時代によって評価形態を変えていく必要がある。評価形態を変えるためには、必ず移行のための負担がある。そのため、徐々に評価形態を変えていくことで、子どもや保護者も新しい形の評価を受け入れやすくなるだろう。評価形態のそのような変化の、長期間にわたる比較可能性は、本質的に同じものが異なる評価形態によって測定されているということの証拠を要求するだろう。そしてそのデータは、2つの評価形態の評価のもとで行われた子どもの見本サンプルから集められているということも示さなければならない。形態が変わっていなかった場合、長期的な比較可能性は、返答モデルの基準項目によって維持されえた。

判断のコミュニティは、子どもの学びを達成変数に沿ってその変化や成長を測定できるようにしてきた。ただし、子どもの成長には学校での学び以外にその子どものもつ背景や、その学校の子どもの平均的な社会的な地位や地域の環境といった学校自体の特色も大いに影響するだろう。このような背景は、学校教育として扱うことができる範囲を超えているにもかかわらず、子どものパフォーマンスには影響を与える。そのため、最終的な発達のみを取り上げて、学校間で比較することは不合理である。学校間で評価システムを比較する際は、先ほど述べたような影響を加味する必要がある。

## Ⅷ おわりに

本小論では、判断のコミュニティの評価計画の適切化に関するシステムへのアプローチを中心に論じてきた。評価の適切化はアカウントビリティのためのシステムを堅実で公平に実施していくために欠かせない。また、教師の判断を中心としてシステムを構築していくためには、その移行を促進させるために「判断のコミュニティ」が必要である。教師をこのアプローチの中心に据えることで、

- 全ての学級でカリキュラムの革新に対して具体的に適応していくこと
- 授業実践と評価実践において教師間で専門性の向上を促進すること
- 教師と子どもという教育の主要な参加者が、その成功において、与えられた利益をもつアカウントビリティシステムを創造すること
- 教師集団のなかで、専門的な基準を自分たちで維持していくという責任を含んだ、より専門的な役割を開発すること

を成し遂げることができる。そのようなシステムを実施するためには、コストがかかるだろう。しかし、そのコストは、教師により証拠に基づいた評価を促し、自分の教育を振り返る時間を与え、教師が共に専門性の向上を目指していくための必要経費である。長期間にわたる実施にむけて、教師の働く状況とキャリア体系を変えていくことに加えて、教師がこのような活動について理解することも必要だろう。このような教師の判断を中心としたアカウントビリティ・システムの状況と、標準テストのためにわざわざ学級での授業時数を割かれ、学校の基金が使われているという現在の状況とを比較してみると、どうだろうか。現在の標準テストを中心としたアカウントビリティ・システムにかかった費用は、テストを作っている出版社の利益になる。対照的に、判断のコミュニティは、今日的で公平な評価を保障し、新たな教育を広め、教師集団の専門性を高めることによって、学校での学び自体を向上させ、直接子どもの利益として戻すことをねらいとしている。

このような判断のコミュニティを実施するための第一歩は、先述のようにすでに整えられてきている。では、今後判断のコミュニティは、その思想と活動について、どのような形をとっていく必要があるだろうか。

1. 先述の判断のコミュニティのそれぞれの構成要素について、引き続き改善していく必要がある。評価の枠組みとさまざまなモデルへの取り組みの継続、そしてより適切にするための取り組みと、適切にすることで教育の質を保証するシステムのための技術的な側面の拡大が求められる。
2. 構成要素をばらばらに扱うのではなく、一体的に

取り組む必要がある。構成要素をばらばらに実施しても成果は上がらない。また、構成要素の相互作用は、全ての要素を実施するまでは分からない。そのため、小規模であっても、判断のコミュニティでもたらされた考えを実践し、その中で発展させていくことが求められる。

3. このような判断のコミュニティの考えを多くの人に広めることもまた、大切な要素である。

こうした取り組みには、それを実行するための展望、能力そして責任が必要となるだろう。この100年、真正評価に関する数多くの研究がなされてきたが、それがばらばらであったため、実践レベルでの成果が上がらなかった。それゆえ、これらの新しい考えをどのようにアカウンタビリティの作業システムに統合していくかという展望をもたなければならない。

さて、日本では、文部科学省によって観点別評価が示されてから5年以上たった。しかし、筆者のインタビュー調査によると、全ての学校で評価規準を用いているわけではない。また、評価規準を利用して評価活動を行っているという回答は4割程度にとどまり、評価規準を使用したほうがよいという意識はあっても、比較的实际の評価に活用されていない現状がある。つまり、国の施策レベルと教育実践レベルでのずれがあることが分かる。このようなずれを減らすためにも、より教育現場を重要視した評価システム作りが求められる。

判断のコミュニティのアプローチでは、「枠組み」「達成変数」「レベル」を教師が中心になって作成している。日本でもこれほど細かくはないが、学校ごとに評価規準の作成が求められている。しかし、一度作成したもののその後手を加えられず、使われていない状況もみられる。評価規準は授業を実践する中で改善され、評価規準を見直すことで授業実践をよりよくしていくことが目指される。判断のコミュニティを形成することは難しいかもしれないが、学年集団で評価規準を作成・改善していく必要があるのではないか。

また、日本において、子どもの活動を教師が見取って評価する生活科は、評価が難しいという声がある。しかし、評価に「子どもの活動の様子を見取った記録」、「子どものレポートやワークシート」、「子どもの絵・作品」など、評価に3つ程度の資料を用いており、多面的に子どもの学習を捉えようとしている。判断のコミュニティの評価形態へのアプローチで述べたような、より妥当で信頼性のある評価が日本において行われているのではないか。

現在日本では、教育の質の保証を全国的な学力調査によって行おうとしているが、小学校6年生と中学校3年生で調査を実施するだけで50億円の費用がかかっている。判断のコミュニティの教育の質の保証で述べたように、学力調査の費用を、教師を中心に据えた評

価のための費用にすることで、教師の専門性を高めることにもなる。日本での教育の質を保証していく方法は、再考の余地があるのではないか。

(注)

- 1) 教育課程審議会「児童生徒の学習と教育課程の実施状況の評価の在り方について（答申）」2000年。
- 2) 中央教育審議会「新しい時代の義務教育を創造する（答申）」2005年。
- 3) Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.) *Knowing what students know: The science and design of educational assessment*. National Research Council Committee on the Foundations of Assessment. 2001, Washington, DC: National Academy Press, pp.44-45.
- 4) *Ibid.*, p.44
- 5) 田中耕治『ポートフォリオをデザインする』ミネルヴァ書房, 2001年, pp.4-9.
- 6) 西岡加名恵『教科と総合に生かすポートフォリオ評価法』図書文化, 2003年, p.43.
- 7) Aschbacher, P.R., "Performance assessment: State activity, interest, and concerns." *Applied Measurement in Education*, 4(4), 1991, pp.275-288.
- 8) Wilson, M., "Assessment Accountability and the Classroom: A Community of Judgment", *Towards Coherence between Classroom Assessment and Accountability*, 2004, Chicago, Illinois: The University of Chicago Press, p.10.
- 9) Bell, A., Burkhardt, H., & Swan, M. *Balanced assessment and the mathematics curriculum*. Nottingham, UK: Shall Centre for Mathematical Education, in Wilson, M. *Ibid.*, p.12.

付記：本小論は、中野の監督・指導のもとで磯谷が作成したものである。