

教育評価におけるルーブリック作成のための いくつかのヒントの提案

—— パフォーマンス評価とポートフォリオ評価に着目して ——

山口 陽 弘

群馬大学教育学研究科専門職学位課程教職リーダー専攻

(2012年9月26日受理)

Suggestions for Creating Rubrics of Educational Evaluation : A Focus on Performance Assessment and Portfolio Assessment

Akihiro YAMAGUCHI

Program for Leadership in Education, Professional Degree Course,

Graduate School of Education, Gunma University

(Accepted on September 26th, 2012)

1. 本稿の目的

本稿の目的は、近年、小中学校においても教育評価の領域で重要視されている「ルーブリック (rubric)」および、その作成法について解説することである。また、この「ルーブリック」が必要となるような実際の教育評価の場面となる「パフォーマンス評価」や「ポートフォリオ評価」についても併せて解説する。

本稿では、これらの用語の理論的な背景について焦点を当てて、作成のためのヒントをいくつか提示したい。ルーブリックという用語は耳慣れないものであるだろうし、はじめて聞く読者もいると思われる。最初にこのルーブリックという概念に焦点化して論じていくことにする。

ルーブリックのそもそもの語源は、赤いインクを意味するラテン語の *rubrika* に由来するとされる。赤いインクは祈禱書などに書き込まれる重要な指示を記すために使用され、そこから転じて日々の努力の積み重ねを示すための目標を指すようになったらしい。現行の英和辞典にも宗教用語として「典礼法

規」(=儀式の指図書き)という語義が残っている。同時に、英和辞典の語義で最初に出てくるのは、「(試験用紙に書かれている)受験心得」とされている (e.g. 「新英和中辞典」第6版 研究社)。

しかし、教育評価における語義としては、これらとは少し異なる特別な意味を持つ。なるだけ簡潔に述べれば、ルーブリックとは、「学習者の達成度を示すための基準」を意味する。

教師に求められているのは、教師自身でこのルーブリックを、自分が教える単元において作成することであり、そう簡単なことではない。ルーブリックを深く理解していれば、ある教育内容を理解した上で、児童生徒の実態に合わせて、学習達成状況を把握することができることを意味するが、これが難しいのである。

このように、ルーブリックという概念を深く理解することが難しい理由は、それが教育内容(算数や国語など)を踏まえた上で作成される必要がある点であろう。本来ルーブリックは、まず教えるべき内容や目の前の児童生徒が存在して、はじめてその内容が決定され、作成されるものであろう。ルーブリッ

クというものが抽象的、一般的なものとして存在するわけではない。この点を最初に確認しておきたい。

もう一点、本稿を執筆するにあたっての筆者の方針であるが、このループリック作成については主として教育学の観点からの提案がほとんどであり、心理学や教育測定学的な観点からの提案は相対的に少ないように思われる。特に教育学の観点からの論者は、やや理想論を述べる傾向があり、それが実現可能であるならば大変望ましいことであるのだが、現場の教員が現実的にそのレベルに達するのは、かなり困難な印象を筆者は持っている。実際に小中学校の教員が教育現場に適用可能なレベルでのループリックとは、どの程度のものであるのかを、本稿では明らかにしたい。このため、心理教育測定学の立場からの観点を踏まえ、さらに筆者自身が、実際に現職教員と一緒にループリック作成を考えた経験を踏まえた上で、作成のためのヒントを述べていくことにしたい。

2. 「基準」と「規準」の違いについて

ループリックを論じる前に、しばしば教育現場で混同され、時に混乱することがあるが、しかし非常に重要な教育評価における概念として、「キジュン」の問題がある。この「キジュン」には「基準」と「規準」という異なる漢字が当てられるのだが、両者の違いについて論じてこう。

現在の日本の義務教育においては、戦後の「相対評価」から、「目標に準拠した評価」へと、教育評価のあり方が大きく変化してきたという背景が、大前提としてあることを確認しておきたい。日本の戦後における、これまでの教育評価の変遷を、小学校の指導要録の変遷に絞ってまとめる形で、以下に表1をあげておく。

上記の表1の中の重要な点は、H13(2001)年改訂の小学校の指導要録では、「目標に準拠した評価」が求められるようになってきている点であり、それはH22(2010)年、すなわち現行の指導要録に至るまで引き続いており、変化はないところである。

この「目標に準拠した変化」こそが、戦後長期間

表1 小学校指導要録における評価法の変遷

(樋口, 2005)

見直しの年	教科等の指導の評価法
昭和23年	評価の客観性が重視され、相対評価法が行われた。学習の記録は「+2, +1, 0, -1, -2」とし、「学習指導上とくに必要と思われる事項」欄、「全体についての指導の経過」欄が設けられた。これは、いわゆる個人内評価にあたる。「行動の記録」欄の評価でも同様な相対評価法がとられた。
昭和30年	「評定」欄の設定。総合評定「5, 4, 3, 2, 1」の相対評価法。 「所見」欄の設定。観点その他について、その児童としての特徴があれば○×を記入する。個人内評価。 「行動の記録」欄は、ABCの絶対評価。
昭和36年	「各教科の評定」欄は5段階の相対評価法。 「各教科の学習についての所見」欄、個人内評価。
昭和46年	「評定」欄は絶対評価を加味した相対評価とした。5段階に機械的に割り振ることがないようにともある。 「所見」欄、個人内評価。
昭和55年	「評定」欄は絶対評価を加味した相対評価。小学校低学年は3段階に変更する。「観点別学習状況」欄は絶対評価とする。 +「十分達成」、空欄「達成」、-「達成不十分」の3段階。 「所見」欄、個人内評価。
平成3年	「観点別学習状況」欄を基本とし、絶対評価を行う。A「十分満足できる」、B「おおむね満足できる」、C「努力を要する」。 観点は「関心・意欲・態度」、「思考・判断」、「技能・表現」、「知識・理解」の順である。 「評定」欄は、低学年は廃止。中・高学年は3段階相対評価。「所見」欄は個人内評価。
平成13年	「観点別学習状況」欄を基本とし、学習指導要領に示す目標に照らしてその実現状況を評価する。 ABCの3段階並びに観点の内容は先に同じ。「評定」欄は中・高学年において、目標に準拠した3段階の評価。 「総合所見及び指導の参考となる諸事項」欄、児童の成長の状況を総合的にとらえる。
平成22年(現行)	「観点別学習状況」欄、引き続き目標に準拠した評価を行う。ABCの3段階、学習評価の観点は、「関心・意欲・態度」、「思考・判断・表現」、「技能」、「知識・理解」の順である。「評定」欄は中・高学年において、目標に準拠した3段階の評価。「総合所見及び指導の参考となる諸事項」欄。

出典：樋口太郎 2005 「指導要録」よくわかる教育評価」田中耕治(編) ミネルヴァ書房

採用されてきた「相対評価」の問題点を改善するために、それと代わって採用されたものであり、まず両者を対比させて理解しておく必要がある。

問題があるとされる「相対評価」も、現実的には、たとえば高校、大学入試などの選抜目的の試験など

では、今も使用されているものである。集団の中での個人の位置づけを明らかにするというのが相対評価であるのだが、では、その方法のどこが問題であるというのだろうか。

この相対評価には、次の四つの問題点があるとされている。その第一は、ある集団の中で「できない児童生徒」が、確実に存在することを前提とした評価法であり、非教育的な評価ではないかという点である。つまり五段階評価なら、「1」をつけられる者が確実に何パーセント存在するという点を前提としているという問題点である。

第二は、児童生徒の間での排他的な競争をあおるのではないかという点である。「勉強とは勝ち負け」とする学習観を生み出すこと、(中略)「他人の不幸はおのれの幸福」とする心情を形成することになる」(田中, 2006) という指摘もある。

第三が、学力の実態を示すものではなく、児童生徒の集団の中での相対的な位置づけを、単に示しているに過ぎないのではないかという点である。つまり上位層になったとしても、当該単元の教育目標を、児童生徒が理解したことに必ずしも繋がらないのではないかという問題点である。

第四が、評価の結果、特に低位層とされた児童生徒への対処法が不明確であり、ともすると、その学業不振の原因が本人に帰せられてしまうという問題点である。

以上の四つの問題点をまとめると、いずれも「非教育的」とあるという点である。本来あるべきである「エバリュエーション」の持つ優れた側面が乏しく、旧来型の「上からの」視点で断罪するニュアンスが強く、教育的な機能を喪失してしまっているという批判である。こうした批判が、相対評価に対して、日本では1970年代以降、繰り返し投げかけられてきた。

その結果、相対評価から「到達度評価」という評価へと転換されていく。この到達度評価は、教育目標に応じて、目標を設定し、その目標が達成されたかどうかを評価するというものである。この到達度評価は、「目標に準拠した評価」とほぼ同義である。歴史的に言えば、目標に準拠した評価の原形にあた

るものである。いわば到達度評価の概念を整理し、そこに教育的・診断的な機能を付加して創造的に発展してきたのが「目標に準拠した評価」と言えるだろう。

このように、相対評価の持つ上記の四つの問題点を克服するべく、現行の日本の義務教育ではその評価方針として「目標に準拠した評価」が提案され、採用されている。

これは、到達されるべき目標を明確にした上で、目標を「規準」として設定して、そのための教育をなすべきであるということである。つまり、教育活動に生かすための評価をするべきということであり、指導と評価を一体化するための手がかりとしての「目標に準拠した評価」なのである。

この「目標」にあたる部分が「規準」であり、英語ではクリテリオン (criterion) と言われるものである。たとえばある算数の単元で、教えるべき到達点が「分数の足し算が、通分した上でできる」というものであったとすると、これが「規準」であり、クリテリオンである。いわば当該単元での「ねらい」にあたるようなものであると理解すればよかろう。

この「規準」すなわち単元のねらいを明確にすることは、その単元での教育内容をより明確にして充実させるためのものであり、これこそ、形成的な評価にもつながるものである。

しかし、この「ねらい」が明確になっただけでは、まだ教師は、児童生徒への評価を十分にはできないだろう。実際に児童に与える分数の足し算の問題をどう設定するのか、さらにその問題がどの程度解ければ合格にするのか、さらにはA, B, Cの三段階なら、どう判定するのかという次の問題が生じる。この細目化された具体的な問題、およびその評定のレベルを、「基準」と呼んでいる。この「基準」は、英語ではスタンダード (standard) と呼ばれるものである。たまたま日本語での訳語が、両者ともに「キジュン」という音になっており、混同、混乱しやすいので注意してほしい。

このように、「規準」から「基準」へと細目化されていく流れで、評価方法は具体化されていくことが、現行の小中学校では、求められているのである。

しかし、次なる問題が生じる。「基準」が具体化され、明確であったとしても、その際の実例が、客観テスト（この場合では、たとえば業者テストなどの〇×式の分数の問題）の採点結果のみというのは、問題ではないだろうか。こうした客観テストの結果から、教師が次なる教育活動を準備するための、十分な情報が得られるのだろうか。

確かに客観テストの結果は明確であり、そこから何段階かに児童生徒を評価することは手続き的には問題なくできるだろう。否、優れた教師であれば、児童生徒が、簡単な計算問題は解けているが、文章題になると解けなくなるなどの誤答分析などの「みとり」によって、こうした〇×式の客観テストからも、次になすべき教育活動を考案していくことは十分可能な場合もあるだろう。

ここで最初に述べておきたいのは、客観テストは、決して一律に否定されるべきものではないということである。客観テスト（業者テスト）であったとしても、単にその採点結果の点数にのみ着目するのではなく、どのような問題にどう解答（誤答）しているのかということにまで踏み込めば、次の教育活動のヒントは得られるということは、ここで確認しておきたい。

しかし、未熟な教師や、当該単位に関する研究が不十分な教師であると、単にこうした客観テストの点数や正答率などの「量的」な側面にのみ着目して「できる児童」「できない児童」という単純な見方になってしまう可能性がある。つまり、相対評価の問題を脱するために目標に準拠した評価をうたっているが、相対評価の持つ問題点を相変わらず脱していないということになるのである。

相対評価の問題を克服するために、目標を設定し、その目標を「規準」から「基準」にまで具体化したのに、再び相対的に「できる」「できない」という集団内での比較、評価の割り振りしかできないというのは本末転倒している。また、非常に高次の思考力や判断力は、こうした客観テストでそもそも測定できるのかという根本的な問題もある。

以上の議論を踏まえて、「できる」「できない」、あるいは「90点」「50点」というレベルではなく、

より深いレベルで児童生徒をみとることを目的として提案された評価指標が、「ルーブリック」なのである。つまり、「量的」な見方を脱して、「質的」に、思考力や判断力などの高次の目標の達成状況を「みとる」ために提案されたのが「ルーブリック（評価指標）」という考え方である。次章ではルーブリックに関して再度論じることにする。

3. ルーブリックの作成にあたっての注意点

ルーブリックという用語について再度丁寧に定義しておこう。

ルーブリック (rubric) とは、いくつかの段階に分けて教育上の達成度の目安を記述して、学習者の達成度を判断する基準を示したものである。

つまり、児童生徒が学習した結果、当該単元を代表するようリアルな課題を解決する際に、実際に行える振る舞い（＝パフォーマンス、次章で再度詳述）のレベルの目安を、いくつかの段階に分けて記述して、学習の達成度を判断する基準を示すために作成されたものである。

このルーブリックと対比されるのが、前章でも触れた客観テスト（いわゆる〇×式のテスト）の量的な結果（70点とか60%解答したとか）である。もちろん、既に述べたように、この評価法でも、知識・理解は確かにある程度は評価できるだろう。

しかし、よりリアルな課題であるような高次の思考・判断、スキルなど（＝パフォーマンス）の評価は、量的な形では難しい。したがって、そのために考え出された質的な、主として行動的な面で、第三者にも観察可能な行動指標上の特徴をもとにして作成されるのである。

以下に一例を示そう。これは小学校5年生の理科で、コオロギの世話をする単元での児童生徒の「実験」（＝パフォーマンス）を評価するために作成されたルーブリックの例である。縦軸にとっている1から4までの四段階で4が最高レベル、1が最低のレベルとなる。横軸にとっている観点が二つあるが、児童自身にも自己をしてもらうための観点と、教師の側の観点とが二種類ある。

表2 子どものルーブリックの例
—コオロギの世話について (小川・片平, 2005)

	子どものルーブリック (よい実験をしましたか?)	評価者のルーブリック (実験計画)
4	私は、コオロギが生育環境を自由に移動できるような実験を計画し実施できた。その生育環境は、条件ごとにはっきりと区別されていた。実験の条件を自分で変えることができ、教師の助けは必要なかった。	実験計画には、児童が問題を分析し、よく考えられた実験を独自に計画し、実施したことが示されている。
3	私の実験は、いつもコオロギに生育環境を自由に移動させることができなかった。	実験計画には、児童が的確に条件制御した実験を実施し、科学的プロセスについての基本的な考えを把握していることが示されている。
2	私の実験は、コオロギに生育環境を自由に移動させることができなかった。また、私は実験の条件を変えるとき、先生の手助けを必要とした。	実験計画には、児童が科学的プロセスについての基本的な考え方を把握したことが示されている。しかし、的確な条件制御を必要とするときには支援を必要とした。
1	私の実験は、コオロギに生育環境を自由に移動させることができなかった。また、実験計画を立てるとき、先生に多くの助けを借りた。	実験計画には、児童が教師によって積極的な支援があったとき、実験できたことが示されている。

出典：小川博士・片平克弘 2005 「オーセンティック・アセスメントにおけるオーセンティック・タスクの事例分析」『理科の教育』

このルーブリックは、基本的には横軸を「観点」、縦軸を「段階」として、観点ごとに段階が一目でわかるように作成された評価基準表となっているのが標準的な形態である。これを作成する目的は、観点別での具体的な教育のねらいを、「規準」のレベルではなく、さらに「基準」のレベルで明らかにすることである。

こうしてみるとルーブリックは、児童生徒の学習達成度をみとる際に有効であり、毎時間ごとの児童生徒や教師の授業の反省を促し、さらには次の授業につながる補償教育を考える有効な手がかりとなることがわかるだろう。さらには、複数の教師間で評

価基準を共有できるというメリットもある。あるいは、未熟な教師が授業を進行していく際に、児童生徒をみとるポイントをつかむためにも役立つだろう。

ルーブリックの作成方針は次のような手順に基づくものである。まず、第一に、事前に予想される児童生徒の様々な振る舞い方（問題、課題への解法）を、可能な限りリストアップしておくことである。いわば子どもの間違い方まで含めた思考を、事前にできる限り可視化しておくことである。

第二に、可能であれば複数の教師（できれば3人くらい）で、そうしてリストアップされた解答パターンや振る舞い方を段階に分けて採点する。その際に、複数の教師の間で、その段階にした理由をつき合わせて、合議の上で練り上げていく。

最後に、これらの作業をしながら、ルーブリックの各段階のすべての段階を埋めていく。ある程度、リストアップされた解答パターンを採点していくとルーブリックがほぼでき上がる。新しい解答パターンが生まれた場合、ルーブリックに追加する作業を継続していく (cf. 松下, 2007)。

このように、ルーブリックは、「それぞれの尺度に見られるパフォーマンスの特徴を示した記述語 (descriptor)」と「達成の度合いを示す数値的な尺度 (scale)」から、評価指標を設定しようという考えに基づいている。「記述語」というのは、評価の視点、あるいは観点にあたるもので、「尺度」というのは段階にあたるものとされる (cf. 高浦, 2004)。

こうした作業は、心理学の測定技法の中では、歴史的にもきわめて古典的な行動チェックリストの作成法と言えよう。

教育学の立場から田中 (2006) は、次のように述べている。「このようにして作成されるルーブリックは「尺度」、評価基準を示す記述語、具体的なサンプルによって構成されている。(中略) この場合の「尺度」は合議によって確定された記述語とサンプルを伴うことから、たんなる「名義尺度」ではなく、「順序尺度」さらには「間隔尺度」をめざすものと言えよう。」

上記の田中の理想論には筆者も同意するのだが、

現実的には「間隔尺度」にまでルーブリックを高めることを現場の教師に求めるのは、非常に酷であるように思われる。

たとえば表2の例を参照して頂ければわかるように、この1～4までの四段階は、いかにしても間隔尺度に洗練させることは困難である。むしろ筆者は四段階という段階数にすること自体かなり疑わしいと考えている。

無理に間隔尺度に高めていくことを目標とせず、せいぜい二三段階の順序尺度にとどめておくことが賢明であり、かつ有用性が高いのではないかと。これは、筆者自身も、現職の教員の方と一緒にルーブリック作りに関わった経験から提案するものである。

測定論的な立場からして、間隔尺度にまで尺度を高めるためには、十分なサンプル数（少なくとも百以上）やその分布（正規分布性の確保など）に関する様々な考察を必要とする。特に一つのクラス単位（せいぜい三十名程度）で考える小学校の教員が、間隔尺度にまでルーブリックを練り上げるのは、ほとんど不可能に近いのではないだろうか。そもそも間隔尺度を作成するためには、統計的な諸知識を勉強した上で、作成にもきわめて労力を要するものであり、各单元ごとに、そうしたルーブリックを作成することは、評価のためのエネルギーを割きすぎているのではないかと考えるのである。

また、ルーブリック作成にあたって、できれば三名以上の教員による合議によって「記述語」や「尺度」を練り上げていくことは、それが可能ならばルーブリックの信頼性と妥当性の向上のために、大変望ましいことではあるが、現実的にはかなり難しいだろう。

現実的に可能な範囲で考えると、一人ではなく、複数（二人）で見直してみるだけでも、その評価の信頼性が格段に向上することが、山口・清水（2009）の一般化信頼性の研究から提案されている。可能ならば教務主任や教頭などの第三者によるスーパーヴァイズ（高い見地からのアドバイス）を一人で行い、受けるのが現実的ではあろう。それですら実施は困難で、行えば大変立派なことである。

したがって、第三者が同一の児童生徒を評価しても、あまりぶれないような、すなわち評価者間の信頼性が高いことを目標とするレベルで、また行動指標を可視化するレベルで記述できれば、現場の教員が利用するためのルーブリックとして十分ではないだろうか。

そして、これは古くからある、行動チェックリストの作成方針とも繋がるものである。通常の行動チェックリストを、間隔尺度のレベルにまで高めることは、きわめて困難なのである。

ただし、そこで注意すべきなのは、その尺度に、確実な順序性や段階があることは確認しておきたいということである。たとえば三段階でのルーブリックを作成した際に、もっとも低いレベルの行動指標が、学習の結果、高いレベルに移行していなくては行動指標と尺度を作る意味がない。1→2→3という場合に、その段階を追って行動レベル（パフォーマンス）が発達していくことが前提として作成されるべきである。

ピアジェが唱えた発達段階説などは、こうした行動レベルでの発達の段階性（順序性ということである）を、実に緻密で丁寧な観察結果から、ピアジェ自身が言語化、可視化したものである。その点では、このような発達心理学の知見も踏まえることこそが、ルーブリック作成には役立つであろう。つまり、ある单元における発達・学習の段階的な観点での道筋（順序性）を、教師が自分なりに見いだすということである。これだけでもかなり大変なことであることがわかるだろう。間隔尺度にまで順序尺度を持ち上げるのは、こうした順序性をしっかり確認した上でなければできないのである。

もしある行動ができる、できないというレベルしか教師が見いだせないのなら、そのルーブリックにおける「尺度」の段階数は、せいぜい二段階で十分であろう。尺度の段階数を多くすることがルーブリック作成の目的ではない。むしろ教師が確実にみとれる行動指標を見いだすことの方が重要である。段階数を増やしたことで、段階のどこに児童生徒が位置づけられるのかを、教師が判断できないというのは本末転倒である。

たとえ二段階であっても、それが教師の確実なみとりと、次の指導に結びつく情報たり得ることが重要なのである。次のアクションを教師が起こすための情報となれば、ルーブリックを作成した意味はあるのである。

心理統計学の観点からしても、段階数や「尺度」のレベルを練ることが重要なのではなく、それはせいぜい「順序尺度」のレベルで十分であり、それでも大変なことである。むしろ「行動指標」=「記述語」=「みとりのための児童生徒の振る舞いの指標」を、明確にすることの方を重要視するべきであると、ルーブリック作成に当たっての注意点としてまとめておきたい。

4. パフォーマンス評価について

これまで述べてきた「ルーブリック」を、評価のために作成することが必要となる「パフォーマンス評価」について、他の評価法との対比をすることによって、その長所と短所を論じてみよう。

例によって、パフォーマンス評価をはじめに定義しておこう。

パフォーマンス評価とは、「知識を応用・活用・統合することを要求する「真正の課題」に挑戦させ、実際の完成作品を生み出させたり、実演を行わせることによって子どもたちの理解の様相を把握しようとする方法である」(田中, 2006)。

ここでわかりにくいのが「真正の課題」という用語である。これは「リアルな課題」ということであり、これがパフォーマンス評価のために望ましい「パフォーマンス課題」なのである。

つまり「真正の課題」=「リアルな課題」=望ましい「パフォーマンス課題」ということになり、これらを実践する方法が「パフォーマンス評価」なのである。

「真正」や「リアル」という言葉の含意するところはなかなか難しい。どこまで言ってもどうどう巡りになってしまう部分が残るほど、定義が難しい概念であるが、山口・石川(2012)が「真正の評価」や「リアル」であることについてかなり丁寧に述べ

ているので、参照されたい。

本章では、「パフォーマンス」という言葉で言いたいことが何なのかに力点を置いて述べておこう。パフォーマンス評価やパフォーマンス課題で言うところの「パフォーマンス」とは、現実に近いような場面を想定した「リアルな」課題を解くときの「振る舞い」すべてを総称するものである。

最終的な解答のみ、すなわち算数ならその解答結果だけを、「パフォーマンス」とは言わない。身体や五感を使って解答していくプロセスまでも踏まえたものであり、「その方法としては、筆記による自由記述問題から完成作品や実技・実演による評価、日常的な対話や観察による評価までも含むものであって、とりわけ高次の学力の様相としての「思考力、判断力、表現力」をとらえようとするものである」(田中, 2006)。

このように、そもそもパフォーマンス評価は、客観テストとの対比から、その問題点、限界点を克服するために提案された評価方法である。客観テストとは、いわゆる紙と鉛筆で測定され、しかも模範的な解答が一つに決まっており、その解答を採点するようなテストである。

客観テストとは、大学受験の際にしばしば用いられるマークシート型の大学入試センター試験などを想像してもらえばよいだろう。小中学校に即して言えばいわゆる「業者テスト」などがそれに当たる。

この客観テストは「真正の評価」という目的からすると、しばしば否定的なニュアンスで扱われるが、長所もちろん、存在することはここでも再確認しておきたい。まず、第一に、客観テストは信頼性が極めて高いという長所がある。信頼性とは評価結果の安定性を意味する。このこと自体は推奨されるべき立派な長所である。注意すべきなのは、妥当性も高い、すなわち、当該単元の学習者の達成度がすべてここに反映されていると即断することである。

しかし、知識・理解面で言えば、十分に練られた業者テストであるならば、妥当性も高いことが往々にしてありうる。また、妥当性はともかくとして、同年齢の児童生徒におけるだいたいの正答率などの目安なども、テストの実施以前に分かっていること

も重要な利点である。これらが第二の長所ともなるだろう。

第三の長所は、教師の問題作成、採点などの手間を省くことができる点である。この点が客観テストのもっとも重要な長所といえるだろう。この点は現実的にはきわめて大事な点である。

そして、これらの客観テストの長所は、同時にパフォーマンス評価の短所にも繋がってくるのである。つまり、パフォーマンス評価の最大の問題であり、欠点であるのが、その課題の作成および採点に非常に手間がかかる点である。また、「ループリック」が然るべき手続きに基づいていない場合に、その信頼性（評価の安定性）が低い場合もしばしば生じる。

パフォーマンス課題の作成過程においては、より「リアル」であるような場面を想定して、しかも当該單元の中で、中心的な思考や判断や表現力などを評価するのに適した課題を作成する必要がある。これは教師にとってきわめて創造的で、挑戦的なことだが、同時に難しいことである。

また、その課題を評価するために「ループリック」を作成する必要がある、その上で採点していく必要がある。これらの労力は、教師にとっては現実的に負担が重いことである。

目の前の児童生徒を具体的にイメージして、パフォーマンス課題を作成した結果、教科書や学習指導要領に準拠することを、失念したり、逸脱してしまう可能性もある。これらもパフォーマンス評価をする際には十分注意すべき点である。

こうした問題を克服するためには、何よりパフォーマンス評価をする際の原則として、目の前の教育活動を充実させるためであることを忘れぬことだろう。評価のための評価になったり、突飛な課題作りや段階数の多いループリック作りのために、教師のエネルギーの主要部分が割かれることがあってはならないということである。教育することが目的であり、評価することが目的ではないことに注意する必要がある。

こうした原則を踏まえた上で、パフォーマンス評価をするためのポイントは何かを再度まとめておこう。パフォーマンス評価は、「真正の課題」に挑むこ

とで、五感で表現される学習の豊かな様相を把握するためになされる必要がある、そのような評価方法を創意工夫することである（田中、2006）。

ここで最優先事項は、当該單元における「真正の課題」を工夫することである。上述したように、それが学習指導要領などを外れたものであってはならないが、「真正の課題」作りが、パフォーマンス評価をするための必要条件となるだろう。

この「真正の課題」とは、日常的なものとも繋がるようなリアルな課題ということである。内包（いわばその概念を獲得するための性質）が貧弱であったり、外延（概念の具体例である事例）がわずかしかな存在しないような課題は、リアルな課題とは言えない。その概念の内包（性質、ルール）は豊かであり、多くの外延（事例）が含まれるような課題、つまり、日常と結びついているような概念を獲得するための課題が、「真正の課題」となっていくのである。

とはいうものの、やはり「真正の課題」というのはわかりにくい。パフォーマンス評価に適するようなその真正の課題はどのような特徴があるのだろうか。以下の表でその特徴をまとめておこう。

表3 パフォーマンス課題の特徴

- | |
|---|
| 1. 生徒はその課題を選ぶ際には、いくつかの選択肢を持っている。
(一つの解答に限定されないし、解答の表現法も複数あるということを意味する) |
| 2. その課題は、單元の中で中心となる知識内容の精選と特定のプロセスの使用を要求する。
(一定の学習内容、教育目標が設定されているということの意味する) |
| 3. その課題は、明確な採点システムを持っている。
(明確にループリックが作成されているということの意味する) |
| 4. その課題は、教師に限定されず、より広い聴衆、教室外で、実際に仕事に関わるような人にも繋がるように作成されている。
(その課題がリアルで、真正(オーセンティック)な課題であるということの意味する) |
| 5. その課題は、それが本来的に測定されるべきことを測定していること。
(ループリックに妥当性が備わっているということの意味する) |

出典：Burke, K (2006) を田中 (2006) が訳したものを参照し、筆者が改訳、() 内で補筆。

もう一つの重要な真正の課題の重要な特徴は、それが、クローズドエンドな課題、つまり正答が明瞭なものではなく、オープンエンドな課題であることが多い点である。もちろん、オープンエンドであれば、真正の課題になると即断してはならないが、課題作成のヒントにはなるだろう。このオープンエンドな課題とは、解答は一つに決まらず、様々なものがあるような課題を意味する。

米国の教科書にも掲載されている一例を出してみよう。

表4 オープンエンドな課題の例

クラスでお金儲けをするためにグリーティング・カードを作って、売ろうとしている。その材料費として50ドルある。どのようにお金を使えばよいのか。ちなみに封筒つきの無地のカード10枚組で4ドル、絵の具は1瓶2ドル、9本1組の絵筆5ドルである。完成したグリーティング・カードは1枚3ドルで売ることになっている。

(田中, 2006より引用)

上記の問題にはある程度文化的な要因も関わっているので、そのまま日本で導入することには注意を要する。しかし、たとえばこれを作り替える形で、中学校数学などで、実際に研究授業などで実施される例に変換して考えてみよう。

複数の携帯電話の料金体系のプランがあるとすると、いずれも利用時間とその料金体系は、一次関数で表現することができる。その一次関数の切片と傾きとがプランによって異なっているということである。あるプランは利用時間が少ないときには安上がりであるが、利用時間が長くなると損をする。別のあるプランはそれとは逆であったりする。このような複数のプランがあったときに、どのプランが、どの程度の利用時間の人にはもっともお得であるのかということ、自由な手法で生徒にプレゼンテーションしてもらおうというパフォーマンス課題を想定して欲しい。

このプレゼンテーション(説明)の方法に特に制約はない。表で説明してもよいし、図やグラフで説明してもよいし、場合によっては数式で説明してもよい。どのような説明法でも構わない。いわば自分

が携帯電話の会社の人間になったつもりで、それを顧客に対してプレゼンテーションすることをイメージして、説明させるという課題は、まさにリアルな課題と言えないだろうか。このプレゼンテーションの中でどれがもっともわかりやすかったかをクラスの中で相互評価させるという場面を想定したらどうだろうか。これは優れたパフォーマンス課題であり、かなり適切な真正の課題となっているのではないかと考える。

こうしたパフォーマンス評価の要件として、そこにそのパフォーマンスを演じることになるための「聴衆」や「状況」が伴う必要があるとされている。上記の携帯電話のプラン説明も、なるだけうまくこの「聴衆」や「状況」を設定しようとして考えられたものであり、その部分に注意して欲しい。この二点を考慮に入れることが、パフォーマンス評価を行うために重要な点なのである。

ここでいう「聴衆」の意味するところは、教育活動を行っている当該の教師のみが評価するのではなく、評価者がより広い聴衆へと広がっていくことを目標としている。つまり教室内に限定されての評価ではないということで、「聴衆」が教師、「状況」が教室場面に限定されて評価されるのは、パフォーマンス評価としては望ましくないということを意味する。

理想を述べれば、その課題が、教室の外に向けて、保護者や、その課題を仕事としている人とも結びつく、つまり現実の問題を解決するような手がかりを与えることまでを、最終的な目標としているということである。もちろん、実際の教育場面では非常にそれは難しいことであろう。

しかしたとえば、小学校社会科で事故の防止や安全といった単元を学習する際に、その仕事に実際に関わっているお巡りさんや消防士さんに話を聞くことはよくあることである。この単元の学習成果、すなわち自分自身がどうやって事故を防ぐのか、安全を確保するのかを、クラスの仲間に対して発表し、さらにはお話をうかがった人たちにも報告することは、パフォーマンス課題の一例であろう。少なくとも、お話をうかがったお巡りさんや消防士さんに、

報告することを目指すという形で、子どもたちに学習成果をまとめさせ、発表させることは、一つの立派なパフォーマンス課題である。これこそ、まさにリアルな「聴衆」と「状況」を兼ね備えたパフォーマンス評価となり得るのではないだろうか。

ある単元の中で、なるだけ必然性がある形でパフォーマンス課題は実施されるべきであるということである。それが不自然な形で実施されては「真正の課題」ではなくなってしまう。したがって、それは「真正の評価」には結びつかない。パフォーマンス課題は、学習指導要領を踏まえるとともに、児童生徒の実態（それは学力のレベルであるだけではなく、生活者としてのどのような環境で生きているかということ）を踏まえる必要がある。両者の高いレベルでの統合が必要になってくるのである。

以上のように「真正の評価」の典型例となるべく提案されたのが「パフォーマンス評価」であるのだが、「真正の評価」＝「パフォーマンス評価」と即断することには慎重になってほしい。それは確かにおおむね当たっていることであるし、期待される新しい評価法として、この「パフォーマンス評価」が教育学者たちから提案されて、検討されていることは間違いないことである。

しかし、これまでも繰り返し述べてきたように、客観テストにも長所が存在する点を忘れてはならない。その信頼性の高さや、教師の労力を省けるという利点は十分活用してほしいし、熟練した教師であれば客観テストからでも、様々な深い情報を得られるはずである。

客観テストとパフォーマンス評価とは、評価のあり方が両極として極端な様相として現れたものと考えべきものであり、どちらかが正しいと考えるのは誤りであろう。すべての評価がパフォーマンス評価に収斂することが望ましいとは筆者は考えない。もし評価がすべてパフォーマンス評価になったら、それもまたリアルな課題ではなくなってしまうだろう。両者の長所と短所を踏まえた上で、併用しながら指導に活用して貰うことが学校教育としては「リアル」なのではないだろうか。

また客観テストではないが、いわゆる紙と鉛筆に

よるペーパーテストでは、パフォーマンス評価ができないわけではないことにも注意しておこう。ペーパーテストでもそれがよく練られた算数などの問題であり、自由記述式で、プロセスまでも踏まえて評価するという場合であれば、それが「パフォーマンス課題」となりうるということが十分生じうる。

実際に算数の問題などで「パフォーマンス評価」の例として紹介されているのは、先に示した表4の課題のように、自由記述式のものであったりする(松下, 2007)。いわゆる「パフォーマンス」で言うところの中心的なイメージでは、五感や身体を使った実技・実演型の振る舞いが、確かに中心的なものとなる。しかし、課題(この場合の算数のように)によっては、それがかえって不自然な振る舞い＝パフォーマンスを児童生徒に求めることになりかねない場合もある。これは好ましいことではない。紙と鉛筆を使つてのテストから、望ましい「パフォーマンス」を引き出すことも可能であることは強調しておきたい。

5. ポートフォリオ評価について

やはりルーブリック作成が必要であり、また、いわゆる新しい評価として、パフォーマンス評価と並んで位置づけられる、もう一つの評価が「ポートフォリオ評価」である。このポートフォリオ評価について簡単にまとめておこう。

「ポートフォリオ」とは、そもそも「紙挟み」「書類カバン」という意味であり、子どもたちの日々の創造の記録やそれへの評価記録を、その「紙挟み」にまとめておくことで評価活動にしていこうとするものである。もちろん、「紙挟み」というのはたとえであって、それを入れるのがボックスや棚であっても何の問題もない。また、日々の創造物は「日記」や「ワークシート」、場合によっては子どもたちを撮った記録となるビデオであったりすることもありうる。

以下にこの「ポートフォリオ評価」と通常の標準テストとを対比させた表を示そう。

表5 ポートフォリオ評価と標準テストとの対比
(De Fina, 1992)

ポートフォリオ評価	標準テスト
子どもが生活している自然な環境で行われる。	不自然な出来事として子どもに与えられる。
子どもが自分の弱点だけでなく、長所を表現できる機会を提供する。	特定の課題について、子どもが、どこで失敗したかを概観させる。
現実的で意味のある日常的なりテラシーにかかわる課題を評価する。	子どもにとっては、意味はない可能性がある、人工的な課題について評価する。
子どもが自分の作品や知識について反省する（メタ認知を獲得する）ように導くためのものである。	期待された単一の解答を出すように、子どもに求めるものである。
親が子どもの作品や知識について熟考するように促すものである。	本質的には意味がない。しばしば恐れを抱かせるような数値を親に提供する。

出典：De Fina (1992) を田中 (2006) が訳したものを参照し、筆者が部分的に改変。

ポートフォリオ評価で必要とされる要件は、上記の表5の五点ほどに代表されるが、さらにそれらに付け加えられるべき、重要な点を以下にまとめておこう。

第一に、ポートフォリオは、日常的な教育活動のプロセスを記録することを目的としている点である。日常の学習過程で生み出されてくるものを、収集・蓄積することを重要視しており、結果主義ではなく、過程主義である点が重要な点である。

また、この過程で生み出されるものは、児童生徒から提出されたものに限定されず、教師や親からのものも入ることを想定している。文字情報にも限定されず、図や描画や場合によっては映像、音声情報なども含まれることを想定しており、この点では「パフォーマンス評価」で想定される幅広い「パフォーマンス」にあたるものである。したがって、その評価のためにはルーブリック作成が必要になってくるのである。

第二に、その評価の半ば、あるいはまとめとして、「検討会 (conference)」を実施することが求められる点である。この検討会では、子どもたち自身の学習状況に関して振り返り活動がなされ、メタ認知が促されることが求められる。その際の振り返り活動

も、子どもに一方的に「強いる」ものではなく、教師自身の教育活動を振り返るためにも、なされる必要がある。こうした振り返りの行為は、教師と子どもとの共同作業であることが望ましいとされる。

つまり、教師の想定している教育目標と、児童生徒がつかんだ学習のめあてが擦り合わされて、新しい評価基準、規準が生み出されていくことまでを理想としている。

また、その検討会では、可能ならば、教師と子どもたちだけではなく、保護者や地域住民も招かれ、まさに教室内に限定されない「リアルな課題」であることを確認していくという遠大な理想があるのである。

こうしてポートフォリオ評価の特徴や目標、それがなされる目的を述べていくと、それはパフォーマンス評価の特徴ともほぼ同じものであることが分かるであろう。両者が新しい評価の二本柱として位置づけられるゆえんである。

しかし、やはりこのポートフォリオ評価も、パフォーマンス評価の持つ欠点、すなわち非常に手間がかかるという問題点がある。さらに、その理想、特に地域住民や保護者も踏まえて検討会を実施すべきであるというのは、現実的にはあまりに高邁な理想である印象を筆者も感じる。

とはいえ、たとえば、保護者の授業参観の際に、過去の学習を振り返る形でこのポートフォリオ評価を使ってみるということは可能なレベルではないかと考える。

また、科目によってはポートフォリオが有効なものとなりうるものも確かに存在する。たとえば総合学習などの、ある程度長期的に、幅広いテーマで学習していくような科目がそうである。さらには、小学校社会科などで「パフォーマンス評価」で例に出したように、「事故や事件から暮らしを守る」というような単元で、地域との結びつきを児童が調べたりするようなものの場合には、授業の一環として、お話をうかがった学校外の方たちへの報告という形で、日々の学習成果を、ポートフォリオの形でまとめていくことはありうるだろう。

また、表5では、「標準テスト」のもつ悪い面を強

調しすぎている印象を筆者は抱いている。たとえ標準的なテストであっても、そのテストがより自然な形で児童生徒に与えられ、彼らの長所も見いだされるように設計され、日常的なりアルな課題になっていて、メタ認知を促すように使用され、親にも子どもの知識を熟考させるように作成・返却されるのであれば、標準テストであっても、「真正の評価」へと近づいているのではないかと思われるのである。逆に、ポートフォリオ評価であったとしても、表5の要件を満たしていないときには、それは非常に不自然なものとなり、標準テストに劣るものとなり、「真正の評価」とは言えないものになってしまうだろう。

なお、実際のポートフォリオ評価に、筆者も関わった例がある。これは小学校社会科の時間で、一枚のポートフォリオに複数の時間で児童自身が学んだことをまとめて、単元全体（十数時間）の中で振り返る教育活動（OPP（＝One Page Portfolio））の実践例である。この実践例が、本学の教職大学院のストレートマスターであった日部によってなされているので、興味のある方は参照していただきたい（日部・山口・石川，2012）。

引用文献

- Burke, K. 2006 *From Standards to Rubrics in 6 Steps*, Corwin Press.
- De Fina, A.A. 1992 *Portfolio Assessment*. Scholastic Professional Books.
- 樋口太郎 2005 「指導要録」 p.139
よくわかる教育評価 田中耕治編 ミネルヴァ書房
- 日部貴博・山口陽弘・石川克博 2012 「わかる授業により児童の学習意欲を高める社会科学習指導—授業間のつながりに着目した振り返り活動の工夫を通して—」 群馬大学教育実践研究, 29, Pp.201-210.
- 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
- 高浦勝義 2004 絶対評価とルーブリックの理論と実際 黎明書房
- 田中耕治 2006 教育評価 岩波書店
- 小川博士・片平克弘 2005 「オーセンティック・アセスメントにおけるオーセンティック・タスクの事例分析」 理科の教育 2005年4月号 東洋館出版社
- 山口陽弘・清水真紀 2009 「英語学習者のための音読テストの信頼性の検討—一般化可能性理論の適用と第二言語習得への示唆—」 群馬大学教育学部紀要 人文・社会科学編, 58, Pp.155-168
- 山口陽弘・石川克博 2012 「教育評価の理論と実践—真正の評価をめざして—」 群馬大学教育実践研究, 29, Pp.187-200.