

教育評価に関する認知発達心理学的考察

— 新学習指導要領に伴う再整理と展望 —

佐々木 典彰, 森 和彦

An Educational Evaluation in Cognitive Developmental Psychology

— Review for the New Course of Study —

I. はじめに

2002年4月より新学習指導要領が完全実施されている(高等学校においては2003年4月より完全実施)。その大きなねらいは「生きる力」の育成であり, その実現へ向けて, 例えば, 「総合的な学習の時間」が新設されている。「総合的な学習の時間」の大きなねらいは児童生徒の主体性であり, 具体的な活動内容としては, 例えば, 児童生徒が自分のテーマについて自ら追究していくという問題解決的学習や体験的・実践的学習(児島, 1998; 他)が期待されている。したがって, 児童生徒の学習意欲をいかに喚起させ, 児童生徒自らの学習へと結びつけていくかが教師の役割となる。そうすると, 評価の場面においても, 児童生徒の主体性を適切に評価しながら, 学習や指導に還元していくことが求められる。しかし, 主体性の評価, すなわち「関心・意欲・態度」の評価はどのようにして行うのかという大きな問題点が発生しており, 多くの議論が交わされている。例えば, 田中(2002)は, 「関心・意欲・態度」を「知識・理解」や「思考・判断」と関連づけることを主張しており, そのためには学力に関するたしかな構造認識が求められることを示している。さらに, 相対評価から絶対評価の重視に伴い, 多くの困惑が見られている。例えば, 授業中の発言や発表の回数, 宿題の提出状況などを点数化して評価している一方で, 教師への忠誠心を競わせることになるとのではないかと懸念されている(東奥日報, 2002)。これらのことは, 教育評価について, 改めて検討し直すことを求めているのではないだろうか。

そこで本論では, 新学習指導要領に伴う教育評価について, ① Measurement (測定), ② Assessment (査定), ③ Evaluation (評価: 重要性の判断を伝達して共有する) の3つの側面に注目し, 認知発達心理学的な視点から再整理し, 考察していきたい。

II. 評価とは

II-1. Measurement (測定)

Measurement (測定) とは, あるものさしで対象を測ることであり, 例えば, 定規を用いて長さを測ったり, 体重計を用いて体重を測ることなどである。これらの物理的な測定に関しては, 信頼性が保証されていなければならない。信頼性(reliability)とは, 誰が何度測っても同じ結果が得られることであり, 1 cmは誰が何度測っても1 cmでなければならない。ところで, 教育評価における測定には, どの程度, 信頼性が保証されているのだろうか。教育評価において, 測定の信頼性を考えるときには, 物理的・客観的なものさしが存在しない。そのため信頼性があるということは, 一貫性のある測定結果が得られるということに置き換わる。すなわち, 測定の場所や時間, あるいは測定者の違いにもかかわらず, 測定結果の一貫性が保証されているということである。例えば, 仮想的に, 児童生徒全員が等質の学力を持っていたとする。そして, この児童生徒を2組にランダムに分けて, 同一の学力テストを実施したときに, 両組の点が等しい場合には, このテストは一貫性があり信頼性があると言えよう。しかし, 大きな差があった場合, このテストは一貫性がな

く信頼性がないということになる。

次に、あるテストの信頼性はあるが、そのテストが測りたいものを本当に測っているのかという問題も浮上してくるであろう。これが妥当性 (validity) の問題である。例えば、1桁の整数の掛け算ができるかどうかを測ろうとしたときに、その掛け算の計算を意図した問題が作られていても、実際には掛け算ではなく、足し算でも十分解答できる場合は、その問題の妥当性は必ずしも高いとは言えない。そもそも学力は、物理尺度ではなく、心理尺度 (Psychological Measurement) で測られているということは留意すべきことである。学力テストに関して言えば、それは、学力特性と数値とを対応 (間接測定) させたものあり、その得点を学力の指標として用いているにすぎない。むしろ、学力テストの得点だけで、その領域の全ての学力を代表させようとするには問題がある。

ここで、測定から得られる数値の尺度水準についても触れておきたい。定期考査などで用いる教師作成テストの点数には、当然、等間隔性が保証されているべきである。例えば、100点と90点の差 (100点-90点=10点) と、10点と0点の差 (10点-0点=10点) は、本来、同程度の学力差を示している。しかし、その根拠は十分証明されているとは限らず、単に見なしている場合が多い。必ずしも等間隔性が十分に保証されていない教師作成テストは、むしろ順序尺度として捉えることができる。ところが、そのように扱えば、中間考査や期末考査の得点を合計したり、平均値を算出することなどは妥当とは言えなくなってしまう。Stevens, S.S. (1951) によれば、ここで用いられるはずの間隔尺度とは、数値間に等間隔性が保証されている尺度のことであり、加算が可能になる。一方、順序尺度とは、数値間に等間隔性が保証されておらず、大小関係だけを示す尺度のことであり、したがって、足し算はできない。

以上のことから、Measurement という視点から考えてみると、児童生徒から産出される数量的結果には、常に信頼性や妥当性の観点から再検証が行われるべきであり、同時に、数値の持つ意味 (尺度水準) についても、幾度となく再吟味される必要があろう。この点は、特に、後述する内容基準による評価において重要なポイントとなる。

II-2. Assessment (査定)

Assessment (査定) とは、学習の質と量を見立て (内容の吟味)、その後の対処行動を決定することである。そのときに、ただ一つの視点よりも複数の視点から児童生徒を捉えることによって、より正確な対処行動が決定できるのは言うまでもない。そもそも、児童生徒の達成状況は地図上の一点ではなく、幅広く複数の広がりを持つことから、一つの目標だけを追うことはかえって学習効率を悪くすることが考えられる。したがって、児童生徒の個人的特性を考慮する複数のものさしを用いて評価を行うことによって、よりよい学習支援が期待できる。そこでその多面的なものさしについて以下に列挙する。

第一に、多面的知能観が挙げられる。多面的知能観 (Multiple Intelligence) とは、人間の知的能力は一次元的なものではなく、多次的であるという考え方である。すなわち、例えば、算数の学力が低いからといって、その児童の全体の知的能力が低いということにはならないという考え方である。この Gardner, H. (1983) の考え方に基づいて、小田 (1999a) は次のように表現している。

- ① 視覚的・空間的把握力 (絵画, 作図, デザインなど)
- ② 論理的・数学的把握力 (比較して結論を得る能力など)
- ③ 言語的能力 (作品を考察し, 創作する能力)
- ④ 身体的能力 (運動, 行動力など)
- ⑤ 音楽的能力 (演奏, 批評)
- ⑥ 対人折衝能力 (電話のかけ方, 勉強会の運営能力など)
- ⑦ 内省的な能力 (自己評価能力)
- ⑧ 自然観察能力

これらの分類のいくつかは、教科設定の根拠の一つになっていると思われるが、⑥対人折衝能力や⑧自然観察能力に関しては、新設された「総合的な学習の時間」にも対応すると考えられる。すなわち、新学習指導要領において学習の目標とされるのは、教科の知識の習得だけではなく、それと同等に、より幅広いコミュニケーション能力や問題解決能力の習得も含まれている。

また、この考え方は、最近注目されているポートフォリオを利用した総合的な学習の理論的根拠

にもなっている。ポートフォリオ (portfolio) とは、学習の履歴を表す道具のことであり、そこに学習の過程で集めた文献や写真、レポートなどを集積させていく。そして、その集積したものを対象として、自己評価活動が行われ、学習の軌道修正や新たな展開に役立てていく。特に、多面的知能観との関連では、⑦内省的能力が注目され、ポートフォリオをもとにした評価活動が期待されると思われる。すなわち、学習のプロセスにおいて、次の行動を決定するとき、ポートフォリオに収納された情報を見直すことによって、「あれが足りない」「これに関してはどうなっているのだろう」と、自分の知識や理解の構造を振り返ることができ、具体的な対処行動を選択しやすくなる。したがって、ポートフォリオは、結論のプロセスが分かり、Assessmentを促進させるための有効な道具と言えよう。

次に、「気質」という考え方と Assessment との関係について取り上げる。気質 (temperament) とは、個人の示す情動反応の特徴 (坂野, 1999) のことである。Thomas, A. et al. (1963) は、子どもの気質について、「活動性」、「生理的機能の規則性」、「初めての刺激に対する接近・回避傾向」、「順応性」、「反応の強さ」、「反応の閾値」、「機嫌」、「気のまぎれやすさ」、「注意の幅と持続性」の9つの行動カテゴリーに分類した。これらは生得的な行動特徴であるため、例えば、学習環境に慣れるのに時間がかかる児童生徒に対しては、すぐに評価結果を求めずじっくりと待つ、あるいは、環境への慣れを早めるような手立てを考案する、といった配慮や評価視点が必要である。

気質を学習環境との相性という視点で捉えることもできる。例えば、教授法に関して、講義形式において学習効果が上がる児童生徒と、演習形式において学習効果が上がる児童生徒に分けることができる。このように、児童生徒の個人的特性によって、有効な教育方法が異なる現象を適正処遇交互作用 (aptitude-treatment interaction, ATI) という。市川 (1995) によれば、この概念はクロンバック (Cronbach, L.J.) によって提唱され、個人的特性を情報にした、教育効果の「最適化」と表現されている。すなわち、児童生徒にとっての最適な学習環境がそれぞれ異なるということであり、例えば、気のまぎれやすい児童生徒に対し

では、同一の課題にじっくりと取り組むよりも、次々と課題を替える方法が適しており、学習効果も期待できるかもしれない。

さらに、コンピテンスも正確な Assessment のためには考慮されるべき考え方であろう。コンピテンス (competence) とは、人にすでに備わっている潜在的な能力と、環境に能動的に働きかけて自らの「有能さ」を追求しようとする動機づけを、一体として捉える力動的な概念をさす (橋本, 1999)。言い換えると、有能感と言える。したがって、児童生徒のコンピテンスに応じた適切な評価環境を整備することで、児童生徒自身が「これならがんばれそうだ」「やってみよう」と感じ、学習をより促進することが期待できる。

以上のように、児童生徒の個人的特性が考慮される一方で、教師による学級や学校、教育活動全体に対する評価である教育評価モデルも存在する。例えば、教育評価モデル (model of educational evaluation) とは、Stake, R.E. (1970) によって示された確かな評価資料を得るための構造モデルである。それは、記述資料と判断資料の情報から構成されている。

図1に示されているように、記述資料は、教師の意図領域と実際の観察領域とから成り、それらは、さらに、先行条件、教育的交渉、及びその成果に分かれる。教師の意図したことは、必ずしもそのまま実現されず、意図領域 (①②③) と観察領域 (④⑤⑥) にはしばしばズレ (uncongruence) がある。例えば、「①学級の全員が欠席せず、かつ予習をしてこよう意図したが、④実際には数名の欠席者と、予習をしてこない者がいた」「②教師があらかじめ考えていた指導案が、⑤なんらかの事情で80%ぐらいしかできなかった」ということになる。

一方で、判断資料は、標準領域 (⑦⑧⑨) と観察領域 (⑩⑪⑫) とから成り、それらは、さらに、記述資料の場合と同様に、先行条件、教育的交渉、及びその成果に分かれる。標準領域は、教師も含めて、広く学校当局、保護者、地域社会、国家などの教育の提供者が、教育に期待するものである。例えば、「⑦教師は学級の全員がある予習をしてこようことを期待し、⑧校長は魅力ある授業を期待し、⑨国家は一定水準まで学力が向上する成果を期待する」と仮定する。ところが、観察領域にお

	記述資料		判断資料	
	教師の意図の記述	実際の観察結果の記述	期待される標準	観察結果の判断
先行条件	①全員が欠席せず、予習をしてくる	④数名が欠席し、予習をしてこない	⑦予習は全てやってくる	⑩数名が予習をしてこないの、期待通りでなかった
教育的交渉	②指導案を準備	⑤指導案の80%しかできなかった	⑧魅力ある授業が期待される	⑪指導案の80%しかできなかったが、魅力ある授業だった
成果	③全員が練習問題を解ける	⑥数名が練習問題を解けない	⑨一定水準（指導により理解する）まで学力が向上する	⑫全員が練習問題を理解し、期待水準まで達した

図1 教育評価モデルの例

いて、「④生徒の80%しか予習を行わず、しかし⑤授業は魅力的で、⑥国家の期待水準を越える学力の向上が見られた」ということであれば、判断基準は、「⑩先行条件としての予習は、教師の期待通りには行われなかった。⑪しかし、教育的交渉としての授業の魅力は、校長の期待に添い、かつ、⑫その成果としての学力の向上は、国家の期待水準を越えた。」となる。

この教育評価モデルには、必要に応じて問題となっている項目を絞ることや、問題発見のための視点として活用することが考えられている。

以上のことから、Assessmentという視点から考えてみると、対処行動の決定という診断性と予測性が重要となるが、それは一つの視点からではなく、より多くの視点から検討されるべきである。さらに、児童生徒の意欲に基づく最適な学習環境設定と、それに伴う教師環境の整備についても、日常的な論議が求められる。

II-3. Evaluation（評価：重要性の判断を伝達して共有する）

Evaluation（評価：重要性の判断を伝達して共有する）とは、当事者が相互主体的（Inter-subjective）に価値判断を行う相互交渉を意味している。すなわち、Evaluationとしての評価とは、お互いが評価者であり、各々の自己評価に相手の他者評価をクロスさせる相互性が主張された

考え方である。

Evaluationを用いた代表例としては、教師の教育責任としての評価、すなわち、保護者への説明責任・結果責任（accountability；アカウンタビリティ）という考え方が挙げられよう。小野（2000）によれば、アカウンタビリティは、教師自身が児童生徒の学力をどう捉え、学習成果をどのように評価するか、ひいては学校の教育成果を教師がどう評価するか（学校評価）という問題と切り離して論ずることはできない。したがって、アカウンタビリティとは、児童生徒だけではなく、教師自身あるいは学校全体が、他者評価とともに、自己評価・自己点検を行い、現状維持でよいのか、何をどのように改善しなければならないのか、といった次なる行動について、情報を分かち合い（シェアリング）、児童生徒とともに相互主体的に決定することである。

次に、「総合的な学習の時間」におけるEvaluationの例を挙げる。総合的な学習を行う教師が、年に何回か、子どもと親と「会議（conference）」をもつ機会を設け、その場でポートフォリオを活用しながら、学習の評価を行う方法について、高浦（1998）は報告している。そこでは、子どもと学習の過程を振り返り（共有し）、次に努力すべきめあてを明確にすることが期待されている。

II-4. まとめ

ここでは欠席日数を例に挙げて、Measurement, Assessment, Evaluationの3つの視点から評価を考えてみる。Measurementの視点では、欠席が何日だったという数値が評価となる。そして、欠席がどのようにカウントされているのか、という欠席の定義が重要である。例えば、入院や通院のために欠席した場合、通常の欠席として扱われるのか、あるいは、どこまで授業時間数としてカウントされるのかを明確にしておく必要がある。さらに、授業開始後何分までに入室した場合は出席として扱われるのかということも考えられるであろう。そして、それらの基準に対して、時と場合あるいは教師によってばらつきがあった場合、信頼性は低いということになる。また、必ずしも出席日数と学力が正比例関数になっていない、ということであれば、学力評価として出席を考える妥当性は低いと言えよう。

ところが、Assessmentの視点では、あと何回欠席すると進級が困難であるとか、このままでいけば単位が取得できないということであれば、その分プリントの課題を出す、放課後の補習を計画するなどのように対処方法も含めて評価行為となる。そこからさらに、なぜ欠席が多くなってしまったのかについて両者で話し合い、その原因が、友人関係の悩みであることが理解されたとする。実際、教師は、欠席にどう対処するかというAssessmentにとどまらず、出欠の問題を超えて、教師・児童生徒間の信頼関係や児童生徒の気持ちをいかに読み取るかなどの教師自身の自己評価にもつなげていくことになるだろう。これらのように、欠席を他人ごととして捉えるのではなく、相互主体的に捉えて、共に根本的解決を目指していく評価がEvaluationである。そうすると、欠席日数のような記録のための評価というのは、単に、Measurementだけにとどまらず、むしろ、それをもとにして、いかにEvaluationとしての評価へ運んでいけるかが、教育評価という視点からは重要なのではないだろうか。

この前提条件として、両者の信頼関係が求められるのは言うまでもない。したがって、日常的に、相手との会話や行動から、思いや考えを聞き取り、対処しサポートしていくといった姿勢を持つことがEvaluationにおいて重要であると言えよう。

III. 教育評価とは

次に、これまでのMeasurement, Assessment, Evaluationの評価視点をもとに、教育評価のポイントについて考えていきたい。本論では、教育評価の基本的考え方として、①自己評価、②内容基準による評価、③非数量的評価を重視する立場に立ちたいと思う(図2)。以下では、これら3つの立場について詳しく述べていきたい。

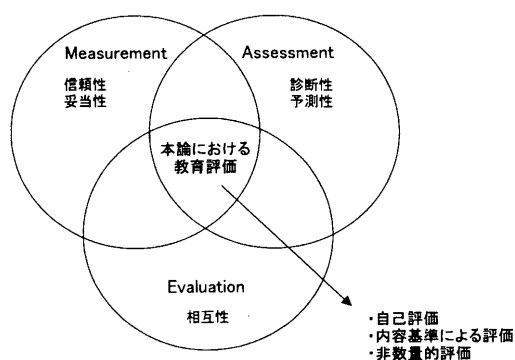


図2 本論における教育評価の基本的考え方

III-1. 自己評価

本論における教育評価の基本的考え方は、評価も自分自身の学習の一部であるという前提から、自己評価を基本としている。この背景には、教師の取り組み方によって、普通の小学生・中学生・高校生でも自分なりの理論・仮説を発見したり、他者の考えを正當に批判できたり、情報を分析・解釈できるという考え方がある(Meier,D, 1995)。そして、自己評価のねらいについては、次のようにまとめられる。

- ① 自己評価能力を高める
- ② 相互に啓発する
- ③ 計画・方法の修正をする
- ④ 学習の成果を明らかにする(成就感)
- ⑤ 課題を明らかにする(発展性)

さらに、これらの自己評価のねらいは、シェアリング(当事者相互の振り返り)を行うことで、より効果的に機能する(図3)。

シェアリング(当事者相互の振り返り)とは、優れた他者評価を介した自己評価のことであり、振る舞い(挑戦、協力、情動を含む)や視覚化の作業、表現(文章表現、身体表現)を対象として

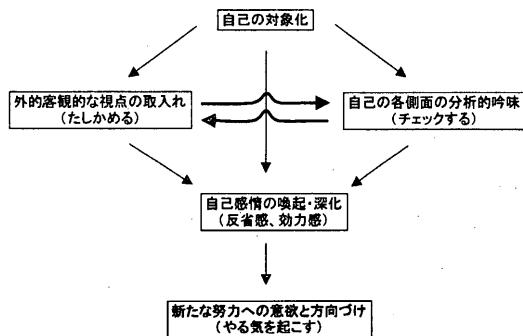


図3 自己評価的活動の心理的過程
(梶田, 1983)

行われるものである。例えば、ブリーフィング（ミニ報告会）、ポストイット（壁新聞）、プレゼンテーション（発表会）、振り返りカード（記録・過程記述作業）などであり、これらを通して、教師を含め、児童生徒同士がお互いに主体となって振り返ることである。

ここで、教師の自己評価について触れておきたい。小田（1999a）は「教員の自己評価システム」として、以下の項目例を紹介している。

- ① その生徒の学び方の「質」をどのように評価しますか？
- ② その生徒の「考え方の質」をどう特徴づけますか？
- ③ その生徒は授業でどのような質問をしますか？
- ④ その生徒はもっと深く探求してみようとチャレンジしたくなるようなプロジェクト学習を自分で選びましたか？実例を挙げてください。
- ⑤ その生徒が自分で高い価値を認めているのは、何でしょう。

これらの項目への対応は、いずれも深い教科内容への情熱と理解度が前提となる。そして、教師の振り返りの深度が、児童生徒各自の把握の深度へと結びついていくと仮定されている。すなわち、児童生徒との対話などから得られる資料も参考にしなければ、上述の項目への対応は難しい。そして、自分の仕事を確認する習慣を身に付けることを目指していく。そのための環境として以下のことが挙げられる。

- ① スムーズな「学習」よりも、失敗や困難、

困る、戸惑う、途方にくれるような、つかかかる「学習」の計画（鈴木, 2000）

- ② 評価の具体的参照点（項目）の設定（モチベーションの維持に必要）
- ③ キーワードを利用したフィードバックシステム（テーマへの接近度を随時評価修正）
- ④ 明確な目標設定のために実行される、ウェビングとチャックダウン（掘り下げ質問）

次に、自己評価の観点についても考察する必要がある。梶田（1983）は、次の5点を示している。

- ① 授業・活動への参加状況
- ② 向上・成長の状況
- ③ 学習に関する習慣・態度
- ④ 対人関係のあり方
- ⑤ 自分自身の全体的あり方

これらの具体的な観点や基準については、児童生徒自らが設定する、または、教師との対話（大まかな枠組みは準備）のなかで決めることが重要であると思われる。なぜなら、一方的に決められた観点や基準を与えることは、児童生徒自らによる目標設定の機会を失わせ、しかも、児童生徒自らの目標と一致するとも限らないからである。

また、実際の自己評価においては、学習の過程において、出来たこと、向上したことに着目していくことも重要であると思われる。梶田（1983）は、いずれの観点に対しても、結果だけではなく、決意表明的コメント（例えば、「これからは、～の点に特に努力したい」）をそえることの重要性を述べている。すなわち、自己評価は出来た出来なないに印をつけることではなく、時間のかかる発見行為であり、このことをしっかり認識する必要があると思われる。

また、上述の自己評価と、教師、児童生徒相互、特別指導・協力者、保護者、学校を支える地域の住民など、多様な目からの評価とを比べることで、自己評価をより正確に具体的に確認することができる。秋田大学教育文化学部附属中学校（2002）では、「総合的な学習の時間」において、①生徒、②指導者、③保護者の3者による評価を実践している。①生徒による評価では、「総合 DOVE」と呼ばれるポートフォリオや「振り返りカード」をもとに自己評価を行い、各種発表会では、相互評価、参観者からの評価を交換し、それらを自己評価に利用した。②指導者による評価では、教師が

行う評価の目的を「指導・支援のために」「カリキュラム開発のために」「保護者への説明責任のために」「記録・証明のために」の4点に整理して実施し、③保護者による評価では、指導協力を行った保護者からのコメントや、各種発表会での評価、三者面談やアンケートなどを用いている。

Ⅲ-2. 内容基準による評価

本論における教育評価の方法は内容基準による評価である。絶対評価と呼ばれることもあるが、この用語はもともと教師「絶対」主義を語源としており（田中，2002），誤解を招きやすいため本論では用いないこととする。ここでは、内容基準による評価について、相対評価と比較しながら、それぞれの背景や、長所、短所を取り上げ、考察していきたい。

鈴木（2002）によると、相対評価（集団準拠評価）の基礎原理は、心理測定学（psychometrics）であり、人間の生得的な能力を測定することが目的とされている。そして例えば、知能の測定を選抜場面に用いることで、生徒が将来、学業で成功するか否かを予測されていたと説明している。相対評価の長所としては、定員が決まっている選抜、組分け、集団一斉指導への適合性を明示できることや、教師の主観的判断による偏りが少なく、得られた数値に信頼性があることなどが挙げられる。一方、短所としては、教師行動の自己評価に向かないこと、達成度や進歩の判断が難しいこと、学習行動の改善が必ずしも成績評価の向上に結びつかないことなどが挙げられる。

内容基準による評価の長所としては、授業の改

善や個人指導方針の決定に役立つことが挙げられよう。短所としては、評価基準の設定が難しいこと、到達目標を行動の言葉で明示しにくい、人格や態度の評価には向かないことなどが挙げられる。

また、鈴木（2002）は、内容基準による評価としてのクライテリオン準拠評価を説明し、さらにそれを①ドメイン準拠評価と、②スタンダード準拠評価の2つに分類して、それぞれの理論的背景について記述している（図4）。なお、クライテリオン準拠評価（criterion-referenced assessment）とは、一定の基準（criterion, クライテリオン）と比較して児童生徒の学習を評価する方法である。

①ドメイン準拠評価（domain-referenced assessment）は、行動主義および要素主義に立ち、明確な行動で基準を用いて判断する「ブロック積み」の学習観を持っている。例えば、計算能力は、足し算、引き算、掛け算、割り算の順に、積み上げることによって獲得されるという考え方である。この場合、「掛け算ができないのは、足し算や引き算の理解が不十分であるから」という判断ができる。しかし、体育の「さかあがり」ができないということに対しては、積み上げ式の考え方を導入しにくいかもしれない。また、掛け算ができないということは、計算能力全体に対してどのようなことが言えるのかについてもわかりにくい。したがって、複雑な技能も客観的に判断できるという長所がある一方、短所としては、評価可能な学習領域が少ないこと、個別の評価と全体との関係がわかりにくいことなどが挙げられる。

	相対評価	クライテリオン準拠評価	
		ドメイン準拠評価	スタンダード準拠評価
評価基準	集団	正解数	事例との照合
背景	心理測定学	行動主義	構成主義
長所	<ul style="list-style-type: none"> 選抜、組分け、集団一斉指導への適合 信頼性が高い 	<ul style="list-style-type: none"> 授業の改善、個人指導方針に適合 到達度が分る（納得できる） 	
短所	<ul style="list-style-type: none"> 教師行動の自己評価に向かない 達成度や進歩の判断が難しい 学習行動の向上が成績向上に結びつかない 	<ul style="list-style-type: none"> 評価基準の設定が難しい 到達目標を行動の言葉で明示しにくい 人格や態度の評価には向かない 	

図4 各評価法の比較（鈴木（2002）を一部加筆修正）

②スタンダード準拠評価 (standard-referenced assessment) は、構成主義の学習理論を基礎としている。すなわち、学習とは、量の増加というより、本人の知識構造の組替え、つまり、新しく学んだ知識や概念によって、既存の知識を組み替えていくことを指す。例えば、「同じ数値を繰り返して足し算する場合は、掛け算が便利である」ということを理解すること自体が評価の対象となる。そして、さらにこの問題に対するアプローチの仕方に変化が見られた場合も、学習の重要な進歩として評価される。そして、スタンダード準拠評価の大きな特徴は、評価基準を実際の事例を用いて説明することである。すなわち、その事例をもとに、該当レベルに当てはまるかどうかを判断される。しかし、その判断には一貫性が求められるために、教師の研修の必要性も指摘されている。

以上のような内容基準による評価を用いる場合には、観点や操作的定義に基づく基準及び評定関数を、児童生徒と保護者に事前に公表する事が必要であると思われる。なぜなら、本論の立場は、相互主体的な評価である Evaluation を前提としているからである。さらに、自己評価をどの程度利用するのか、観点の重みづけと評定関数の設定をどうするのかについての議論も必要であろう。

ところで、人は全ての学習観点に興味を持ち、スキルアップしようとするのであろうか。必ずしもそうと言えないのであれば、どのように個性を大切し、どのような児童生徒像を描こうとしているのかという、個性尊重と新しい児童生徒観もこれからの課題であると言えよう。

さらに、内容基準による評価の代表値をどのように設定するのか、あるいは、観点別評価をどのように統合するのかについての議論も必要である。領域限定の知識理解度に準拠した評価である「ドメイン準拠評価」、パフォーマンス・レベルに準拠した評価である「スタンダード準拠評価」の2種類の観点別評価が、独立に判断可能で、かつ、ブロックのように積み重ねることができる、あるいは、平均値を求めるなどの演算ができるという根拠は乏しい。それでは、全体印象・最重点項目など、代表的なレベルや理解度の観点別評価のひとつを主観的に代表として選ぶべきであらうか。それとも領域内で最も優れた評価を代表値として

選ぶべきであらうか。どちらであるにしても、要素構成主義・最重要項目主義は、教育評価の全てではない。むしろ、それは伝え方(コミュニケーション)の問題でもある。教師は、児童生徒の人間形成によりよい形になるように、記述された評価の表現方法を使い分けるべきである。

Ⅲ-3. 非数量的評価

個人ベースの改善という内容基準による評価の長所をより引き出すためには、非数量的評価が求められよう。教師が行う非数量的評価の具体例として、支持的コメントと「学習の質」の評価を取り上げたい。

支持的コメントとは、教師が児童生徒に対して、学習成果の大小に関係なくがんばりを認め、次の課題を発見させるようなコメントのことである(佐々木・森, 2000)。言葉掛け評価と合わせてこれから検討していく必要があると思われる。言葉掛け評価(暗黙の評価)とは、①さりげなく、②曖昧(受け取った側が解釈できる)で、③その場限りの評価のことを指す。例えば、「そう!それから?」といった興味を示すことや、「この色いいいね」、「この文章好きだなあ」といった言葉掛けが例として挙げられる(海保, 2001)。と同時に、教師には、児童生徒に対して、本人の熟成期間として「待つ」姿勢が求められるであろう。必ずしも人はいつも全てにやる気満々であるとは限らず、評価(指導)を急ぐことは、本人の学習ペースを崩し、かえって学習意欲の低下に結びつく危険性が考えられるからである。この学習意欲については、評価されたときの感情が大きく影響する。市川(1995)は、成績を伝えられたときの感情について、奈須(1990)をもとに図5のように整理している。学習意欲との関連では、例えば、成功場面における統制感・向上心である「今度もがんばろう」、「がんばったかいがあった」は、自己効力感や積極性と結びつくであろうし、誇り・承認への欲求の「自慢したい」も学習意欲向上の重要な要因となる。一方、失敗場面における無能感・あきらめとしての「がんばってもだめなんだ」、「自分は頭が悪い」というのは、失敗の原因を自分の能力においており、学習意欲低下が懸念される。ここでは、能力不足ではなく、努力不足であることを伝えて意欲低下を抑えなければならない。以上のように、教師が成績を伝えたときには児童

成功場面	喜び	うれしい／楽しい／バンザイ／よかった
	統制感・向上心	今度もがんばろう／がんばったかいがあった
	誇り・承認への欲求	自慢したい／親にほめられるぞ／通信簿が楽しみ
	驚き	本当かなあ、信じられない／意外だ
失敗場面	後悔	がんばればもっとやれたのに／どうしてももっとがんばれなかったのか
	落胆・罰の予感	親になんていおう／おこられる／もうだめだ／どうしよう
	くやしき	くやしい／ちくしょう／くそー
	無能感・あきらめ	やっぱり／がんばってもだめなんだ／自分は頭が悪い

図5 成績を伝えられたときの感情の例（市川，1995）

生徒は多様な感情を抱く。したがって、単に、テストを返却して終わりではなく、そこから児童生徒の感情を察知し、それに応じた適切な事後指導を行うという、Evaluationとしての評価が求められることは言うまでもない。

「学習の質」の評価については、達成のレベル、学習の質のレベル（意欲，関心，集中力，学習環境との相性，多面的知能，気質），個人の伸びのレベルなど，多角的な表現による評価が必要であると思われる。重要なことは、単に、全ての細分化された項目を評価することではなく、その評価の伝達が、本人の意欲や自己発見，肯定的自己評価などへと結びつくようにフィードバックすることである。そこで、相互発見のための実際の評点の具体例を挙げる。小田（1999b）が紹介しているアメリカ・ミネソタ州で使用されている評価用紙を参考にすると、観点別，提出物別にいずれも他者・自己評価兼用として，“改善を要する・良くしたい所がある”（N = Needs Improvement），“優れている・すばらしい”（E = Excellent），“満足のいく状態・これでOK”（S = Satisfactory）の3種類による記述方法がある。ともすると、1，2，3のような数量的評価は、序列意識を生じさせるかもしれない。しかし、Needs Improvement，Excellent，Satisfactoryを用いた評価（以下、NES評価と略す）は、各評点が独立して意味を持つ名義尺度による質的評価である。すなわち、E = 3点，S = 2点，N = 1点，といった間隔尺度でもなく，EはSより優れている（E > S），または，SよりもNは劣っている（S > N）

といった順序尺度でもない。したがって、全ての項目がNであっても最低点という意味にはならず、またそれ以前に、全項目に記入する必要もない。必要な項目に必要なメッセージを貼り付け、しかも、自分自身と他者（仲間，教師など）の評価が併記されることによって、改善点や自分が気づいていなかった点の確認が可能となり、NES評価は、的を絞った集中的な学習の進化に役立つと思われる。

評点の記述方法は、その対象も様々であり、自己評価のみならず、相互評価，改善度評価，過程評価，作品評価など多くの視点が考えられる。しかも各々の視点に応じて、柔軟かつ適切な使用が求められている。さらに、Evaluationにおける相互主体的なコミュニケーションの促進を考慮すると、情報機器やCD-Rなどを用いた評価情報の管理についても合わせて検討しなければならない。例えば、児童生徒と教師の両者がアクセスしやすく、容易にコメントを書き込めるようなシステムを構築することによって、日常的な評価活動が期待できるであろう。

IV. おわりに

小学生でも自己評価や教師評価が出来る。したがって、教師が評価者で、児童生徒は被評価者というだけの構図は成り立たない。特に、内容的基準による教育評価では相互主体性を重視していることから、評価活動というのは児童生徒と教師との人間関係の問題にも関連すると思われる。また、児童生徒がいつも全てにやる気満々であることは

考えにくい。したがって、相互主体的な関係において留意したい点は、教師には、「人はその人なりに良くなろうとしている」という認識を持つ事、または、認める姿勢を持った、愛のある評価を行うことではないかと思われる。そして、Measurementの持つ信頼性と妥当性、Assessmentの持つ診断性と予測性、Evaluationの持つ相互性をどのように組み合わせて、自己評価、内容基準による評価、非数量的評価をどのように行うかが重要となるであろう(図2)。

新学習指導要領の実施に伴い、より多視点での評価が求められるようになった。しかし、具体的な手続き等はまだ不十分であり、だからといって、安易に行えるものでもない。評価の問題はこれまで以上に重要視されなければならない、かつ、慎重に議論されなければならない。

V. 文献・資料

文献

- 1) 秋田大学教育文化学部附属中学校 (2002)「研究報告第69号」.
- 2) Gardner,H.(1983) Frames of mind: The theory of multiple intelligences. New York: Books.
- 3) 橋本憲尚 (1999) 中島義明・安藤清志・子安増生・坂野雄二・繁樹算男・立花政夫・箱田裕司編「心理学辞典」有斐閣.
- 4) 市川伸一 (1995)「学習と教育の心理学」岩波書店.
- 5) 梶田叡一 (1983)「教育評価」有斐閣.
- 6) 海保博之 (2001)「指導と評価第47巻6月号『連載生きる力を育てる評価⑧生きる力の評価—認知面—』」図書文化.
- 7) 児島邦宏 (1998)「教育の流れを変える総合的学習」ぎょうせい.
- 8) Meier,D.(1995) The Power of Their Ideals: Lessons for American from a Small School in Harlem.
- 9) 奈須正裕 (1990)「学業達成場面における原因帰属、感情、学習行動の関係」教育心理学研究, 38, 17-25.
- 10) 小田勝己 (1999a)「ポートフォリオ学習と評価」学事出版.
- 11) 小田勝己 (1999b)「指導と評価第45巻3月号『総合的な学習の評価方法—ポートフォリオ評価のやり方—』」図書文化.
- 12) 小野由美子 (2000) 岡東壽隆・林孝・曾余田浩史編「学校経営重要用語300の基礎知識」明治図書.
- 13) 坂野雄二 (1999) 中島義明・安藤清志・子安増生・坂野雄二・繁樹算男・立花政夫・箱田裕司編「心理学辞典」有斐閣.
- 14) 佐々木典彰・森和彦 (2000)「総合的な学習へのポートフォリオ評価の導入に関する一考察」秋田大学教育文化学部教育実践研究紀要, 22, 83-92.
- 15) Stake,R.E. (1970) Objectives, priorities, and other judgment data. Review-of-Educational-Research,40(2),181-212.
- 16) Stevens,S.S. (1951) Mathematics, Measurement, and psychophysics. In S.S.stevens(Ed.), Handbook of experimental psychology. New York: Wiley.
- 17) 鈴木秀幸 (2002)「指導と評価第48巻8月号『教育評価再入門 (5) 評価の結果の解釈 (2) —評価の背景となる学習観との関連—』」図書文化.
- 18) 鈴木敏恵 (2000) 鈴木敏恵編著「ポートフォリオで評価革命!」学事出版.
- 19) 高浦勝義 (1998)「総合学習の理論・実践・評価」黎明書房.
- 20) 田中耕治 (2002)「指導と評価第48巻11月号『<参考資料>をどう受けとめ、どう生かすか—目標に準拠した評価を定着させるために—』」図書文化.
- 21) Thomas,A., Chess,S., Birch,H.G., Hertzog,M. & Korn,S. (1963) Behavioral individuality in early childhood. New York University Press.

資料

- 1) 東奥日報 (2002) 夕刊2002年9月13日号第9面.