

# 教育評価の鳥瞰図 ～ 評価コストを軽減するための一つの試案

Review of Current Studies of Educational Assessment: A Draft Proposal to Reduce the Cost of Assessment.

平真木夫 (Taira Makio) \*

**要旨:** Educational assessment is one of the most annoying tasks for the teachers at all age levels. The misunderstanding and misusing of tests/measurements is prevailing for them and it can enlarge the costs of evaluation. Lack of understanding of scale levels and externality of criterions/standards are the major reasons for this misunderstanding. To reduce the cost of assessment we need better understanding of tests/measurements, especially need to know the limitations of each tests/measurements. The author reviewed the current studies of educational assessment, and shows the applicable conditions for each tests/measurements. A draft bird's-eye view of each tests/measurements is proposed to enhance the better understanding of applicable conditions for them.

**キーワード:** educational assessment/evaluation (教育評価)、bird's-eye view of tests/measurement (評価手法の鳥瞰図)、scale levels (尺度水準)、externality of criterion/standards (基準・規準の外在性)

## 1. 緒言 - 正論と現状

「教育における評価は、教育活動の点検と修正」、「教育的価値とそれを具現化した教科目標との即応関係」、「テストは教育行為の領収書」といった意見がある。それぞれは教育評価を学習活動に生かすための考え方として正しいといえるであろう。

また、「指導と評価の一致」という特に近年よく耳にするスローガンの背景には、(1) 言葉で伝えられないものは教育内容として明示できない、(2) 明示できないものは教育の達成を評価できず成績評価が不可能になる、というロジックが存在する。幼稚園から大学まで含めて、全てのカリキュラムは、明示的で評価が可能な知識・技術の教授によって構成される必要がある、と言い換えることができるであろう。

しかしながら、教育現場においてこれらの意見やロジックがどれだけ正しく理解されているかという、かなり危ういといわざるをえない。たとえば、後述する各評価方法が正しく理解されていない場合には、それぞれの評価の適用条件を無視した過剰な利用が生じることなどが予想される。各評価方法の制約条件を知らないままに活用すると誤った評価を行ってしまう可能性が高く、全ての成績評価を数値化しないと問題があるような思い込みが形成されているように思われる。また、評価について過剰な対応をしまい、不必要な負担感が増してしまう可能性もあるだろう。

本論文では、まず始めに教育評価に関する重要な概念を整理し、それぞれの手法の利点と欠点について考察する。そして、最後に、それぞれの評価手法の相関関係を図示し、学力評価の使い分けについて1つの方向性を示したい。

## 2. 生徒による評価の受容

本節では、評価を行うことの原点として、テストを行うメリットとデメリットについて考え、評価が生徒たちに及ぼす代表的な影響としてテスト不安について考察する。

### 2.1 テストを行うメリットとデメリット

テストを行うことのメリットとしては、まず第一に、テストの予告が学習の動機づけとなることを挙げることが

\* 宮城教育大学教育学部学校教育講座、e-mail: m-taira@staff.miyakyo-u.ac.jp、<http://www.miyakyo-u.ac.jp/school/taira/>

できるであろう。また、テストを受けることによって、知識の理解度を確認でき、答案を書くこと自体が学習となる。更に、答案が返却されるときに、自分の長所・短所が確認でき、その後の学習活動に役立てることもできる(辰野, 2001)。他にも、同一のテスト問題を他学級に実施することによって、当該学級の特徴を知ることができたり、テストの作成を通じて、教師がどんな学習内容を重視しているかを表現することもできるであろう(大野木, 1994)。すなわち、正しく意図通りに試験が実施され、また、試験結果も生徒たちに正しく受容されれば、テストを実施することは何ら問題が生じないといえるであろう。

しかしながら、現実的にはテストの実施に関しては様々な問題が指摘されている。ここでは主にテストを作成する段階での問題として以下の3点を指摘してみたい(池田, 1992)。

1. テストは時間的に制約されたものであり、時間内に最高のパフォーマンスを示すことができるとは限らない。
2. どのような課題が選定されるかによって、結果が大きく左右される。
3. 評価ポイントの置き方、採点方法によって、得点が大きく異なる。

これらの問題はテスト作成に関するリテラシと呼ぶべきような能力の欠如によって生じるといえるであろう(大野木, 1994)。たとえば、テストの作成方針として、大まかに「大問主義」と「細目積み上げ主義」の2つが存在する。大問主義とは、100点満点の試験の中で問題が1つないしは数個しか存在しない種類の試験であり、細目積み上げ主義とは、観点ごとにまとめられた複数の問題群から試験が構成されているような試験である。大学入試における小論文がまさに前者の大問主義の典型的な試験であり、細目積み上げ主義による試験としてはセンター試験などが挙げられる。

いかなるテストでも時間的な制約があることは回避できないが、どちらの問題作成方法が評価漏れの問題を回避できるかは明らかであろう。特に、大学入試における小論文の場合は、多くの場合は観点の置き方で成績が大きく変化し、極端な場合には同一採点者の中でも成績判定が揺らぐことがある。たとえば、渡辺ほか(1988)は、小論文の採点結果について、次のような興味深い特徴を報告している。

1. 総合評価の評定者間の一貫性は低い
2. 総合評価の安定性は評定者によってかなり異なり、評定者内の一貫性は高いとはいえない
3. 分析的評価のための観点は、内容面と表現面とに分類できる
4. 分析的な評価と総合的な評価は同程度の信頼性しかない
5. 文字の丁寧さに関する評定は評定者間で一貫している
6. 総合評価は論文を読んだときの印象と強い関連がある

つまり、採点基準が明確に構造化され、そしてその基準が採点者どうしの間で確実に共有されていない限り、論文型のテストの客観性は測定論的にいうとかなり低いといえるであろう(大野木, 1994)。ペーパー試験に限らず試験全般においてもっとも重要なことは、評価基準をきちんと作成し、基準が揺らがないことであるが、上記の例から示されるように現実的には多くの問題が存在する。テストを行う目的と、実施形態を正しく把握した上で評価することの重要性が示唆されるであろう。

## 2.2 テスト不安と学習方略

天野(1993)によれば、1877年の時点でテストに伴う教育上の問題が既に報告されていたことが明らかにされている。その当時に問題とされていた事柄とは、発達が十分でない児童が競争心に駆られること、試験前に過度の勉強をすることは心身の発達を妨げるおそれがあること、試験のために勉強するという悪い習慣を形成することなどである(天野正輝、「教育評価史研究」Pp.49-51)。これらは現在も未解決の問題といえるであろう。

心理学ではこのような問題をテスト不安 (test anxiety) として研究対象としてきた。その定義は以下の通りである。

「テスト場面に代表される評価的状况において喚起される不安反応で、課題遂行や成績に対し抑制的または促進的な影響を与える」(奈須正裕、「教育工学事典」、Pp. 387)

一般に、低不安の生徒は、評価場面で挑戦することによって成績が一層良くなり、テストを完全にやり遂げることによって不安を解消しようとする。それに対して、高不安の生徒は、テストと無関係な反応をして成績が悪くなることが報告されている (Naveh-Benjamin & Lin, 1994)。

前者の成績上位者たちが抱える問題はさておき、まず始めに後者の高不安で成績下位の生徒たちの問題を考えてみよう。彼らはテスト不安以外にも、学習性無力感 (learned helplessness) が形成されていることが指摘されている (鎌原ら, 1983)。学習性無力感とは、「強制的、不可避的な不快経験を繰り返し体験することによって、何をしても環境に対して影響を及ぼすことができないことを、誤って学習してしまう現象」と定義できる (Seligman & Maier, 1967)。つまり、本来は学習によって解決できる場面であるにも関わらず、いかなる状況でも学習できないことを誤って学習してしまい、その結果、解決困難な状態が継続される現象と言い換えることができるであろう。

この学習性無力感とテスト不安とが組み合わさると、どのような状況に陥るのだろうか。たとえば、成績下位群の生徒にとって、強制的、不可避的な不快経験とは「期末テスト」、「受験」などの避けて通れない試験のための勉強であり、環境に対して影響を及ぼすことができない状況とは「勉強したことが試験の結果に反映できなかった」状況が相当している。このようテストに伴う不快な経験を繰り返すうちに、自分は学習能力の低い無力な人間であるという、「学習できないことを学習」してしまうことになる。その結果、たとえば、自分の能力の低さが露呈されることを嫌う場合には、テストに際してそもそも勉強をしないという回避策が採られたりするようになる (『わかる』ということの意味」、Pp.84-85. 佐伯, 1995)。

また、テスト不安が低いと考えられる層の生徒たちも、このような無力感と無縁であるわけではない。たとえば、平 (2001) は宮城教育大学の1年生を対象として調査を行ったが、大学に入学して2ヶ月後の時点で、センター試験の主要科目 (英数国) を解く力が約半分まで目減りしていることが示された。これは、論証する能力が重視される数学のような科目ですら、暗記科目ととらえられているためである。他の教科も、どう理解するかではなく、機械的な暗記に重点をおいていたことがうかがえる。暗記をベースにした学習方法を採用した場合、学ぶのを止めると急速に忘れていく。このような学力残存状態はその結果生じたものと言えるであろう。

しかし、このような学習方法は、大学受験を目的とした勉強という意味では極めて理にかなっている。だからこそ、受験勉強で学んだ内容をすっかり忘れてしまうという事実に対しても、学習者本人が疑問を抱いたり、不満を感じたりすることがほとんどない。「受験勉強が終わったのだから、すっかり忘れてしまうのは当たり前だ」という姿勢といえよう。つまり、テストに対して苦手意識を持たないと推測される層の大学生たちにも、学習性無力感と似たある種の諦めが存在すると考えられる。

## 2.3 評価基準、学習目標、学習活動の対応関係

天野 (1993) が報告しているように近代の教育制度の成立当初から、テスト不安の高低に関わらず、教室における学習活動が評価によって大きく規定されているといえよう。ただし、評価を行う人間の意図がそのまま正確に学習活動に反映される保証はない。

たとえば、学習目標と学習の動機づけに関する研究から、教師は理解を教授目的と考えているのに対して、生徒は成績を目的とした学習動機が形成されている結果が報告されている (Nicholls, 1984)。つまり、教師が設定している学習目標と、生徒たちが理解している学習目標の間に違いが生じる可能性があるということである。また、生徒たちは、特定の学習目標 (e.g., 試験をパスする) を達成するための学習方略のコストに敏感であることも示されている (Nolen, 1996)。たとえば、上述したセンター試験の数学のように、出題される単元を理解するためのコストが高いと感じられた場合には、生徒たちはすぐに忘却してしまうことが分かっている、公式などの暗記で乗り切る

ことを選択することが多い。何故ならば、試験に合格することができれば、暗記主体で乗り切っても理解主体で乗り切っても差がないからである。

また、一般に試験というものは、問題作成者が想定した学習スタイルと異なった方法で解答が可能である。たとえば、有坂(2002)は、選択肢だけで構成されている問題であれば次のような方法で正解を「発見する」ことが可能であると受験生たちに指南する。(1)他の選択肢となんらかの関係をたくさん持っているものが正解である、(2)正解の選択肢から誤りの選択肢を作るということを踏まえた上で、正解かどうか吟味せよ。有坂のアドバイスは受験テクニックとしても極端なものであるが、このような抜け道が存在する限り、生徒はより低い労力で試験を乗り切ることができることが分かるであろう。

同様のことは学習活動の内容と学習の目標の混同についてもいえる。たとえば、虹の絵を描く活動を行う場合には、虹を書くことではなく絵の具を混ぜて色を作り出す技術を習得することが学習目標となるはずである。また、授業で台形の面積を求める活動を行った場合も、授業の目標は台形の公式を覚えることではなくて、どのような図形の面積でも三角形や四角形に変形して求められることが活動目標となるはずである。もちろん、授業を行う教師としては、学習目標と見た目の学習活動との違いを把握して授業を行っているはずであるが、「活動あって学びなし」という表現が存在するように、生徒が学習目標を正しく把握している保証はない。生徒たちが考える学習目標は、親や社会など学校外の環境から影響を受けることも示されているが、評価活動を通じた教師の教授方針に強く影響を受けることが明らかにされている(Nolen, 1994)。教師がその単元でどのような理解を理想としているかを生徒たちに正確に伝えるためにも、適切な評価基準を作成することが重要といえるであろう。

### 3. 評価方法と用語の整理

評価方法は、評価を実施する時期と目的に応じて使い分ける必要がある。以下は実施時期に応じて評価手法を分類したものである(表1)。次節ではそれぞれの評価方法の特徴について紹介していく。

表1. 評価の実施時期と手法の対応関係

実施時期	評価目的	評価方法
単元導入時	診断的評価	絶対評価、相対評価
中間	形成的評価	絶対評価、個人内評価
単元終了時	総括的評価	絶対評価、相対評価、個人内評価

学校側の説明責任を求めるような文脈で、学校ごとの学力格差をみるためにテストを実施することも今後は十分にありうる。したがって、評価を行う目的の中には上記の通り行政的なものも含まれるはずである。

しかし、平成14年度の時点では、学校評価を目的とした評価テストが実施されることはまだ一般的でなく、評価は学期や単元を1つの区切りとして行われることが大多数といえるであろう。評価を行う目的を学習指導に限定して分類すると、表1の通り、診断的評価(diagnostic evaluation)、形成的評価(formative evaluation)、総括的評価(summative evaluation)にそれぞれ分けられる。

診断的評価とは、学期始めや単元導入時における生徒たちの学力や知識量を測るための評価である。評価手法としては、後述する絶対評価(absolute interpretation)、相対評価(relative interpretation)が主な手法である。

形成的評価とは、学習の途中で学力の移り変わりを分析する評価である。具体的には、学力の移り変わりを調べる小テストなどを繰り返して行い、生徒の学力の変化をモニターしようというのが形成的評価であり、結果を評価する場合には絶対評価と個人内評価(self-evaluation)を行う。指導と評価を一致させるために小刻みにモニターするが、学力の形成過程を把握し、軌道修正に役立つ情報をフィードバックすることが目的であるので、この時期に行われる評価は原則的に学期末の成績に含めない。

総括的評価とは、学期末や単元終了時に生徒たちがどこまで到達したかを分析するために行われる評価である。いわゆる期末試験が典型的な例であるが、その評価の仕方には、絶対評価、相対評価、個人内評価といった具合に複数の方法が存在する。

なお、表1では、評価方法を絶対評価、相対評価、個人内評価の3つに分類したが、後述するとおり様々なサブタイプが存在する。また、現実的には教育評価を行う目的も様々である。たとえば、東(2001)は、(1)教育行政の資料としての評価、(2)学校の管理・運営の資料としての評価、(3)教師の学習指導の資料としての評価、(4)子どもに情報を与えるための評価、(5)保護者の参考にするための評価、(6)子どもの処遇決定のための評価(資格認定、振り分け、選抜)、(7)カリキュラム改善のための評価など、様々な目的を挙げている。

体重を量るのに物差しを使うことがないように、利用目的に応じてどのような評価方法がよいかを判断していく必要がある。特に、一般の心理検査と同様に、1つの評価方法で知ることができる側面はかなり限られていることを正しく理解し、それぞれをテストバッテリー(test battery)のように組み合わせて、生徒の学力を多面的に評価することが重要といえるであろう(臨床場面におけるアセスメントでは、質問紙による性格検査以外にも、知能検査、投影法、動作型の心理検査などを並行して行うことが一般的である。このように、複数の心理検査を組み合わせてアセスメントを行うことが多いが、この組み合わせをテストバッテリーと呼ぶ)。

#### 4. 相対評価の注意点 ~ 準拠集団と情報量 ~

いわゆる評価の類の中でもっとも最初に想起されやすい評価手法が相対評価に基づくものであろう。そして、おそらく相対評価はもっとも評判の悪い評価手法である。たとえば、筆書が担当している大学の講義では初回の講義で評価活動に関するアンケートを必ず実施しているが、学生たちの中には、相対評価を必要悪ですらなく、純粹に不要で、生徒をスポイルするような評価であると回答する者もいる。相対評価に関するこのような指摘はある一面で正しいものの、相対評価の技法を正しく運用できれば有益な評価情報を得ることもできる。

たとえば、準拠集団(母集団)から無作為に抽出して選ばれた生徒たちの成績と比較して、当該学級の各生徒個人の成績がどうであったかを判断することが、本来の相対評価である。しかしながら、現実的には、その準拠する集団の特徴にしたがって、ノルム準拠の評価(norm-referenced test/measurement)とコホート準拠の評価(cohort-referenced test/measurement)の2つの方法が存在する。つまり、生徒たちの相対的な学力位置を測る母集団(準拠集団)の設定の仕方によって相対評価を分類できるということである。

前者のノルム準拠評価は、国、県といった比較的大規模な集団に準拠して、各生徒の相対的な学力を計測するものである。この場合、問題の難易度を計測するために抽出される被験者の人数は、どんなに少なくとも100人を下回ることはまずあり得ない。市販されている試験の場合には、1000人以上の生徒たちの協力の下に難易度評定が行われている。また、この評定は特定の学力層の生徒だけで行われるのではなく、様々な学力の生徒が含まれるような工夫が施されている(いわゆる、無作為標本抽出の手續きに基づいて、各問題の難易度が標準化されている)。

それに対して、コホート準拠評価では、クラスや学校など、比較的小さい集団が準拠集団である。一般にコホートというと世代などの集団を意味しているが、ここではそれよりももっと小さい集団を意味している。つまり、各生徒の学力は各学校、各学級に依拠して相対的に評価されることになる。

上記の議論から、相対評価として通知票に記載されていた成績評価は、後者のコホート準拠評価であったことが分かるであろう。たとえば、5段階評価を行う場合には、クラスの中で学力が正規分布することを想定し、1は下位10%の生徒、2は下位10%~30%まで、3は下位30%~70%まで、4は下位70%~90%まで、5は上位10%といった具合に割り振っていく(1クラス40名の場合には、それぞれ4名、8名、16名、8名、4名が機械的に割り振られる)。このような成績評価は試験の点数などが求まれば機械的に割り当てることができるので、教師にとって使い勝手がよく、また、なぜその評価であったかを説明することも一見すると簡単にみえる。

しかしながら、このコホート準拠評価には幾つか致命的な欠陥がある。技術的な欠点としては、30名とか40名

といった小規模の集団では、学力が正規分布しづらいことを指摘できる。たとえば、テストの得点で言えば、クラスの中で上位群と下位群とにきれいに別れているような双峰性の分布状況が生じても何ら不思議ではない。また、各クラスの平均得点が異なっているような場合にも機械的に評価が割り振られるので、評価と試験の得点との対応関係を論理的に説明することが難しい。たとえば、クラスの中で上位20%であるから4という評価がつけられたと説明することはできるが、他のクラスに対してその4の価値が低かったり高かったりする可能性がある。つまり、クラスが違ったら本来は4ではなくて3であったかもしれないといった具合に、成績評価がクラス分けによって決められると非難されるおそれもあるだろう。

したがって、相対評価と一口に言っても、前提となるその適用条件が大きく異なっていることが以上の議論から分かるであろう。一般に、学級を準拠集団とした相対評価は、国や県を準拠集団とした相対評価よりも論理的な整合性が十分ではなく、後述するような情動的な価値も低いと言わざるを得ない。

#### 4.1 相対評価のメリットとデメリット

母集団の中における相対的な順位に基づいて生徒の学力を評価することが、相対評価の定義である。極端に言えば、大規模標本調査を前提とした問題の標準化だけが関心事であり、問題の内容的な妥当性に関してはあまり拘らないといえるであろう。すなわち、相対評価においては、学習目標の達成度を判断する基準が必ずしも用意されているわけではないので、当該分野の習熟度を評価することは難しいということである。同様の理由で、相対評価は単なる相対的な位置の記述でしかなく、全ての子どもの学力保障といった教育理念とは無関係であるという指摘もある(中原, 1996)。

たとえば、学年の進行にともなって同一世代の中で学力の差が拡大していくような場合には、結果の解釈に注意が必要である。ある生徒が複数の学年にまたがって同じような学力順位にいたとき、行政的な判断として教育成果を評価する場合には、上がりもしなかったが下がりもしなかったといった程度に評価されるかもしれない。しかし、実際には、その生徒の学力順位が下位グループに属するものであれば、実際の習熟度で考えた場合にはよりできなくなっている可能性もあるだろう(Brookhart, 2002)。すなわち、世代の中の相対的な順位に変化はなくても、学力そのものは低下していると表現することができるかもしれないということである<sup>1</sup>。

また、これはコホート準拠評価において顕著な問題であるが、クラス全体のテスト得点が全体的に高くなった場合には、クラスの中における相対的な学力順位が変化しない限り、成績評価も変化しないという問題を指摘できる。すなわち、コホート準拠評価は、個人内の学力変化を把握するには適していないといえるであろう。特に、生徒の努力と成績評価との対応関係が見えにくいという問題がある(Krampen, 1988)。ただし、ノルム準拠評価の場合には、クラスや学校を単位とした変化とは無関係であるため、個人内の学力変化を把握することもある程度可能である。つまり、個人的な努力によってもたらされた学力変化に関する情報を知ることができる点で、ノルム準拠評価はコホート準拠評価よりも情動的な価値が高いといえるであろう。

次に、これはノルム準拠評価にもコホート準拠評価にも当てはまるが、評価を行う基準が順位だけであるために、生徒たちの学習目標を不適切なものに変えてしまうおそれもある。たとえば、後述する絶対評価のように学習内容に関する習熟度が評価基準であれば、生徒たちは習熟度を高める方向で学習活動を行うはずであるが、校内における順位が評価基準であればそれを目標にした学習活動を行うと考えられる。極端な場合には、有坂(2002)が紹介するような受験テクニックを使っても、順位が高くなれば成績評価が高くなるため、評価の高さと学習内容の習熟度との乖離が生じる可能性があるだろう。

習熟度と成績が乖離してしまうという可能性は、藤澤(2002)が指摘しているような、職員集団における教師の

<sup>1</sup> このような歪みは、成績の得点順位がパーセンタイルで表示されているときにも存在するので注意が必要である。たとえば、得点が平均値から離れるほど、得点(z得点等)の差とパーセンタイルで示される順位との差の対応関係が悪くなる傾向にある(パーセンタイルが50から60へ上がるよりも、30から40へ上がる方が得点の向上が著しい)。また、中央値の左右で得点の分布が非対称であることもある。

世間体の問題と組み合わせられると、より現実性を帯びてくるようになる。たとえば、期末試験の直前の授業などでは、教科書の中の特定の問題が試験に出るか否かが教師によって告げられたりする。教師によるこのような予告は生徒が学習しようとする項目を少なく制約する自殺的な行為であり、「テストに合わせて教える・学ぶ」というネガティブな循環を促進させることになる。しかしながら、自分の担当しているクラスの成績が、他のクラスと比較して低くなることを回避したい場合には、このような予告を行わざるを得ないような圧力を教師は強く感じるであろう（「ごまかし勉強・下」藤澤、2002年、Pp.120）。もちろん、教師に対するこのような圧力は絶対評価を行った場合にも存在しうるが、相対評価の方が学習目標のすり替えがより生じやすいと考えられる。

相対評価に関するこれらの批判を教育関係者は真摯に受け止めるべきであり、生徒たちを安易に相対評価することは厳にして謹むべきであるといえるであろう。しかしながら、このような批判があることを理解した上で、相対評価を生徒の相対的な学力を測る目的に限定して利用すれば、何ら問題がないとも言い換えることができるであろう。

たとえば、自己の職業適性などを考えるときには、当該分野における習熟度に関するデータだけでは不十分で、他者と比較してその分野に適性があるか否かの情報も重要な判断材料になるだろう。もちろん、自分の興味関心だけにしがって判断することも可能ではあるが、当該分野に進む場面で競争原理がはたらいている限り、他者との比較を一切行わずに判断することはあまり現実的な方策とはいえない。東（2002）が指摘するように、生徒や保護者が進路選択を判断する資料として、相対評価は重要な意味を持つようになるといえるであろう。

ただし、このように進路選択の判断材料として相対評価を考えた場合にも、ノルム準拠評価の方がコホート準拠評価よりも優れているといえる。卑近な例を挙げれば、たとえば、過去の受験データなどから、所属学校で上位10位以内に入っていれば、どこそこ大学に合格できるといった情報も意味があるかもしれない。しかしながら、所属学校の全体的な学力レベルが変化した場合には、その予想の信頼性は低くなってしまふ。一般的には校内の相対的な位置情報よりも、予備校などが実施する模擬試験の結果に基づいた方がより正確な予測ができるといえるであろう。

## 5．絶対評価と目標準拠評価

絶対評価とは、学習対象とされる領域における習熟度で生徒を評価することである。その意味で、絶対評価は全て到達度目標に準拠した評価といえる。ただし、相対評価と絶対評価とは完全に別の課題によって評価されると誤解されているきらいもあるが、相対評価に用いられる課題を目標に準拠した形で作成することも可能である。何故ならば、相対評価の試験で重要なことはあくまでも問題が適切に標準化されていることであり、絶対評価で用いられる問題がこのような手続きにしたがって標準化されていても全く支障がないからである。むしろ、理想的な試験問題とは、問題の難易度が標準化の作業を通じて明らかにされていて、なおかつ個々の問題が当該領域の学問体系に対して整合性をもったものといえるであろう。ただし、S-P表の分析などで重視されているように、よい試験問題というものは、きちんと理解できている生徒とそうでない生徒の弁別能力が高いものであり、学力に応じて問題の正答率がなだらかに単調増加している必要はない（佐藤、1982）。

本節では絶対評価をこのように定めた上で、評価のサブタイプと運用上の注意について考察したい。

### 5．1 尺度水準の問題

絶対評価のサブタイプについて言及する前に、まず始めに変量・変数の型（variate / variable type）と尺度水準（scale level）の違いについて説明しておく必要がある。

基本的に、変量は、連続変量（実数）、離散変量（整数）、類別変量（論理値）の3つの型に分類される。このうち、前者2つは量的変量（量的データ）であり、最後の類別変量は質的変量（質的データ）となっている。そして、それぞれの変量の型は、観測されるデータがもっている距離情報の多寡にしたがって順序づけることが可能である



(尺度の水準を定めることが可能という)。

ここでいう尺度とは、具体的には試験問題の採点方法などを意味している。たとえば、センター試験では原点(0点)が存在し、離散変量ではあるが原則的に各得点間の差は等距離で、なおかつ比率関係が保たれていることが想定されている(50点の生徒の正当数は10点の生徒の正当数のおおよそ5倍)。このように得点間の距離に関する情報が比率に関しても保持されているような尺度を、統計学では比率尺度(ratio scale)とよぶ。

しかしながら、実際の試験問題では、50点の生徒の正当数が10点の生徒の正当数の5倍であるということも、また、学力が5倍であるということも、必ずしも明確に保証されているわけではない。つまり、原点(0点)は存在するものの、得点間の距離が等しい保証はないということである。たとえば、100点の生徒と80点の生徒の実力差と、20点の生徒と0点の生徒の実力差が質的に等しいとは言い難いであろう。このような場合は、尺度空間がユークリッド距離にしたがっておらず、大ざっぱな得点順位の違いだけが意味を持つといえる。このような尺度は統計学で言うところの順序尺度(ordinal scale)とよぶことができるであろう。

その他、量的変量には、比率に関する情報はないが得点間の等距離性が保証されている間隔尺度(interval scale)というものが存在するが、尺度水準が高いほど距離に関する情報量も多くなるという特徴がある(比率尺度>間隔尺度>順序尺度)。また、一般に尺度水準が低いほど適用可能な範囲は広いが、逆に推定の精度は低くなるという特徴も指摘できる(100点満点の尺度から優・良・可・不可という尺度に変換することは可能であるが、逆向きの変換は不可能であるということ)。

ここまで述べてきた量的変量の特徴は、大まかに言えば、四則演算の中の和と差の計算と、平均値を求めることが可能である点にある。もちろん、厳密に言えば、順序尺度に対して平均値を計算することは意味をなさないが、ある程度の幅をもった変数であれば(a,b,cといった三段階ではなくて1~7といったより多段階の変数)シミュレーションの結果から間隔尺度に近似した結果が得られることが示されている(萩生田・繁樹,1996)。しかし、ここではそのようなテクニカルな処理の違いではなくて、質的データと量的データとの根本的な違いについて考えていきたい。

上述したとおり、量的変量の特徴は平均値などの演算の対象とすることにあることを既に指摘したが、逆に言えば、質的データに対しては平均値を求めてはならないといえる(尺度水準としては最も低い位置にある名義尺度:nominal scale)。たとえば、質的データの例としては、性差や所属クラブ等を挙げることができるが、このようなデータに対して平均値を求めることは全く意味をもたないことがよく分かるであろう。

もちろん、既存の学力検査場面では、このような性差や所属クラブの違いが評価の対象となることはあまりなかったが、しかし、学習指導要領の改訂にともなって興味や関心といった態度の違いも評価の対象に含まれるようになってきている。部活動に参加しているか否かや、ボランティアに参加しているか否かなど、価値観の相剋が生じやすい活動にまで評価対象が広げられつつあるのが現状であり、これらの活動を評価する場合には、適切な尺度を特に注意して選択していかなければならないであろう。また、総合学習などでは、高度な問題解決能力を身につけたか否かといった、厳密な数量化が困難な学習目標が掲げられているのが一般的である。このようなケースでは、質的データに限りなく近い順序尺度を適用すべきであろう。

## 5.2 絶対評価の分類

絶対評価は前節で述べた尺度水準の違いに応じて2つのタイプに分類することができる。平均値を求めることに意味があるような、量的変量として評価を扱うことができる場合にはクライテリオン準拠評価(criterion-referenced test/measurement)を利用し、数量化が難しい場合にはスタンダード準拠評価(standard-referenced test/measurement)を利用する<sup>2</sup>。

<sup>2</sup>一般に、クライテリオン(criterion)は規準と訳され、スタンダード(standard)は基準と訳される。前者はカッティング・ポイントのような規準という意味合いが、後者は典型例という意味合いが含まれている。



クライテリオン準拠評価では、まず第一に、評価すべき学習領域が明確に規定されていることが必要であり、当該領域に関する学力を規定する体系が明らかにされていなければならない。つまり、習熟度や到達度に関して明確な規準が存在したり、各学習段階における知識の包含関係が明らかになっている場合などに適応することができる評価手法といえるであろう。

たとえば、台形の面積を求めることができるためには、3角形の面積を計算できるようになっていることが必要であり、更に、一般的な面積の定義も理解できている必要がある。そして、台形の公式を単純に当てはめることができる程度では単元の習熟度としては不十分であり、公式を適用できない場合などについても正しく理解できているかの確認も行う必要があるだろう。このように、各領域の体系と照らし合わせて、どこまで理解できていればその後の発展的な学習に進むことができるかの判断が可能であるときに、クライテリオン準拠評価を適応することが可能といえるであろう。また、そのような明確な習熟度の規準が存在すれば、各段階に応じた得点評価や順位づけも可能となり、カットイング・スコアなどを設定することもできるようになる。

次に、クライテリオン準拠評価では原則的に量的変量を扱うのに対して、スタンダード準拠評価では上述したとおり質的変量として評価を処理することになる。前節でも一部指摘したが、学習活動への取り組み方や関心といった情意面も評価の対象として組み込まれるようになっており、もっとも難しい評価方法といえるであろう。たとえば、「地域社会に興味・関心を持ち、社会活動に積極的に参加する」ことを、数量化して評価することはかなり難しい作業であることは想像に難くない。

このように新規に設立された学習領域以外にも、既存の一般教科においても情意面の評価が重視されるようになってきているので注意が必要である。たとえば、国立教育政策研究所は2002年に公開した「評価基準・評価方法等の研究開発（報告）」の国語に関する資料の中で、「国語への関心・意欲・態度」、「話す・聞く能力」、「書く能力」、「読む能力」、「言語についての知識・理解・技能」の5つを、評価を行うときの観点として設定している。また、たとえば日本標準教育研究所（2002）は、【関心・意欲・態度】として「物語の内容に関心を持ち、進んで読み、学習のまとめをしようとする」ことを、国語の教科書の単元別に示した「到達目標と評価基準」の中で具体的に例示している。その他にも、各授業を単位とした指導目標とその評価も厳密な数量化を行うことは難しいであろう。

このように、数量的にきちんと定義することが難しい学力を評価する場面では、評価基準を示した言語表現と、評価基準に対応した具体的な学習事例を組み合わせることで評価することになる。これがスタンダード準拠評価の一般的な方法となっている。たとえば、授業の目標（読むこと）として、「 氏の行動と生き方を一通り読みとることができる」という目標が掲げられていたとした場合には、この表現を基準として、それよりも深く読解できたと判断した場合にはA、それよりも浅いと判断した場合にはCと判定する程度に留めたりする（北尾, 2002）。

### 5.3 絶対評価・目標準拠評価の注意点

まず始めに指摘できることは、目標の具体化と評価を行う期間の問題である。これまで、中学校などでは評価というと中間テストや期末テストなど比較的長いスパンを単位として評価が行われてきたが、絶対評価へと移行していく今後は単元レベルで評価を考えていくべきである（北尾 2002）。

たとえば、極端な状況として1つの学年期間を単位として考えた場合には、評価の時に設定する目標は学習指導要領に示されているような方向目標とならざるをえない。もちろん、そのような抽象的な目標を具体的な観点に細分化することも可能であるかもしれないが、そのような大がかりな分析をするよりも、各単元のように具体的な学習項目を設定しやすいレベルで評価し、最終的に学期末に評価を集計しなおした方が効率がよい。また、評価に用いる規準や基準が校外に公開される可能性も今後は十分にある。そのような事態に備えて、評価基準とする到達目標は、暗黙の了解事項として処理するのではなくて、言葉で具体的に説明することが容易な単元レベルで記述していくべきであろう。

次に注意すべきポイントは、尺度水準が高いものから低いものへの変換は可能であるが、逆は不可能であるという原則をきちんと守ることである。これはたとえば、A、B、Cといった順序尺度で評価した結果を、あとから無理

に5段階評価へと変換してはならないということである。ただし、近年の通知票の記載書式としては、観点が細分化された反面、個々の観点の評定段階は大まかな区分けが採用される傾向にある。そういう意味では、比率尺度のように得点間の距離情報の多い評価データを、距離情報の少ない尺度へと変換する記載方式が増えていき、実際にはそれほど深刻な問題にはならないと予想される。

むしろ、尺度変換で問題とすべき場面は、純粋なスタンダード準拠評価から得られた質的データを順序尺度以上に引き上げるような場面が相当している。たとえば、ボランティア活動への参加・不参加といった事項についての評価は、本来は特別な価値判断が含まれないと思われる。しかしながら、小林(1996)が報告しているように、1994年度の時点で、茨城、福島、東京などでは、学校外における特別活動への参加状況を点数化して評価する方式を打ち出している。その他、先述したとおり、指導要領の中の観点別指導事項として【関心・意欲・態度】が含まれるようになっている。

学習活動の情意面を重視する方策は、高学力・低意欲という日本型の高学力への反省が根本にある。この問題は特に理数系科目において顕著な特徴であるが、「分数ができない大学生」という昨今の学力調査の結果や、科学技術に対する成人の無関心や知識不足とも関係している(文部科学省編、「平成13年版科学技術白書」、「第6節 科学技術に対する国民の理解」の第1-2-47 図「OECD加盟国民の科学技術への関心の比較指数」参照)。そのような意味では、情意面を評価の対象に含めることにも一理はあるかもしれないが、これに過剰に適応してしまうと、戦前に採用されていた忠誠度評価型の態度主義(主観的な絶対評価)や、子どもの人格の丸ごとの管理へとつながる可能性もあるだろう。恐らく、情意面に関する評価は、評価全体の中での占める割合を軽微なものにしたり、情意面の良さに関する評定項目を他の項目から独立させて記載することが望ましいと考えられる。たとえば、辰野(2001)が提唱するように、相対評価と絶対評価の成績を合計して記載するのではなくて、それぞれを併記する2重システムのような書式を採用することも有効な手段といえるであろう。

## 6. 個人内評価

近代の教育制度の中で行われてきた試験では、形式が構造化されていない自由面接方式が採用されていた(辰野, 2001)。そのような試験の評定は恣意的で主観的に過ぎるとの批判が当然のように生じ、試験の厳密性と客観性を高めるために客観テストが開発され、論述式、選択式、空所補完など様々な技術が生み出されてきた。これらの技術開発と並行して、統計学的な進歩にもなって相対評価が開発され、また、過度の相対評価への反省から学力の形成的な側面を重視した絶対評価が提唱されるようになってきている。この一連の流れの中で、形成的な側面を更に重視した評価として、近年は個人内評価(self-assessment)が盛んに行われようになりつつある。

既に定義したように、形成的評価とは学習途中における学力の移り変わりを分析することを目的としているが、個人内評価とはその変化に対する学習者本人の評価活動となっている。このような学習者本人が行う評価活動は既に各個人の中である程度自発的に行われてきた行為であるが、教師の側が生徒たちに働きかけて意識的に実行させることに特徴がある。また、総合学習への取り組みで多く使われるようになってきたポートフォリオを用いた評価(portfolio assessment)などでは、このような自己評価能力の評価を行うことが目的とされることもある。

歴史的には、B.Bloomの完全習得学習理論(mastery learning theory)の台頭にもなって、形成的評価の重要性が指摘されてきたが、安易な形成的評価には問題点も指摘されている。たとえば、形成的な評価を行うためだけに、形成的なテストを頻発することは一般的にナンセンスといえるであろう。実際に、小テストや確認テストを毎時間行う理由は、生徒たちの理解度を逐一把握すること以外に、知識の定着といったドリル的な役割が期待されているからであろう。この問題と同様に、生徒の理解状態を把握するためだけに、何度も個人内評価を促すことは一般にコストが高い上にそれほど高い効果も得られないと予想される。

## 6.1 重要な特徴と注意点

個人内評価活動に関する評価については後述するとして、ここでは個人内評価に関するいくつか重要な特徴と注意点について整理する。

まず始めに整理しておかなければならないことは、個人内評価は絶対評価や目標準拠評価とは異なるということである。もちろん、規準・基準となる目標が何らかの形で個人の中で存在しなければ評価することは不可能である。しかしながら、これらの目標が既存の学問体系などと一致していることは一般に希であろう。たとえば、先述したとおり一般に学習活動と学習目標とは異なったものであるが（e.g., 虹を描く 混色技術を習得する）、生徒が設定する学習目標と教師が設定する学習目標との間に齟齬が存在しても不思議ではない。

また、望ましい学習方法を教師が生徒に示すことは可能であっても、望ましい知識状態を初めから押し付けることには無理がある。たとえば、台形の面積を求める公式を教える場合には、公式の導出過程を理解できることと公式の適用条件を理解できることが目標として設定される。このような単元で、生徒の理解がそれぞれ不十分であった場合には、これらを独力で具体的な理解目標として設定することは難しいはずである。極端な場合には、公式の導出方法を機械的に暗記させたり、練習問題を定型的なパターンとして暗記させることにも繋がりがねない。

しかし、たとえば、「数学においては、当該単元で紹介された公式の導出過程と適応条件をきちんと理解した場合において、正しく理解できたと判断する」という一般的な目標設定は可能であろう。つまり、このような単元に依存しない中立的な評価基準（学習スタイル）を生徒たちに身につけさせることは十分に可能であると思われるが、規準として特定の知識を押し付けることは理解を抜きにした暗記へとつながる可能性もあると思われる。

ただし、ここまでは何らかの形で目標を設定した例について言及してきたが、個人内評価の場合には、明確な目標設定を行わずに、単純に過去の自分自身が判断基準となることもありうる。たとえば、台形の面積を求める単元であれば、前回の授業では公式の導出方法まで理解できたが、今日の授業でその適応条件まで理解できたと判断するような場合が当てはまるであろう。恐らく、単元の具体的な内容に即した理解状態に対する自己評価とは、このような「当社比」の評価しか行うことができないと思われる。

何故ならば、認識論の立場から学習をとらえると、知識の広がり自体に関して終端条件を設定することは困難であり、学習目標は常に更新し続けられる性質をもっているからである。たとえば、台形の面積を求める単元が終了したとしても、面積の後には体積の計算を学習目標として設定可能であり、体積の後にはより一般化した概念として積分を学習目標として設定することが可能である。一般に学習者は、教師のように当該分野に関する鳥瞰図を習得しておらず、もっていた場合にも、教師による適切な補正が必要とされる。そのような意味では、学習者個人が形成している単元に関する具体的な到達目標は、個人内評価ではなく、本来は教師による絶対評価の対象として扱うべきであろう。

さて、ここまでは一般教科における個人内評価について考察してきたが、次に、総合学習の文脈における個人内評価についても言及してみたい。総合学習では、特定の教科・単元を設定しない問題解決型の学習が行われることが多く、特定の正解が存在しないような社会的問題がテーマとして選ばれる傾向にある。このような問題を学習対象とした場合には、各生徒が設定した到達目標の適切さを判断することは教師であっても困難な作業であり、必要に応じて専門家の協力を要請するなどの対応がとられたりする。

このような問題解決型の調べ学習の文脈では、当該領域に関する生徒の知識度（学力）と目標設定の対応関係に特に注意すべきだろう。たとえば、平（2002）は調べ学習全般が抱える問題点として次の4点を挙げている。（1）生徒の自主性に依存した目標設定になりやすいこと（過度の放任主義）（2）到達度の評価が難しいこと（メタ認知の難しさ）（3）体系的・構造的な知識が獲得される保証がないこと（這い回る経験主義）（4）賢い生徒はますます賢くなり、そうでない生徒は伸び悩むこと（格差の拡大と再生産）。

たとえば、検索サイトを用いた調べ学習が教育現場では流行しているが、しかし、その学習効果がどれほど期待できるかはあまり明らかでない。むしろ、平（2000, 2002）や Taira（2001）が報告しているように、既にある程度の知識がある状況においてのみ、検索サイトが学習場面で有効に機能すると考えるべきであることも示唆されている。

たとえば平(2002)は、WWWと検索サイトを用いた実験から、WWW全体の中に占める情報量の多寡とは関係なく、被験者本人にとって最も身近で分かり易いページ、問題として理解できるページから閲覧し始める傾向にあったこと、そして、その後の閲覧行動も類似の選択方針にしたがっていたことを報告している。つまり、WWWの検索と閲覧行動は、既存の知識構造を広げるような方向ではなく、検索前に形成されていた興味関心に基づいて、既存の知識構造を強化・補完していくような方向に機能する可能性があるといことである。

これらの実験結果を別の言葉で表現し直すと、(1)到達目標のレベルとその達成度は学習に取りかかる前段階の知識量に依存しており、また、(2)達成度に関する個人内評価は実際の到達度の絶対的な高さとは無関係に行われる可能性が高い、と言い換えることができるであろう。たとえば、台形の面積の例で挙げた認識論でいえば、台形の面積までの知識しかない生徒が掲げる到達目標は体積の求積が精一杯の目標設定であろうが、体積の求積まで理解している生徒は積分の概念を目標として掲げることができる可能性がある。このとき、学習によって前者の生徒が体積まで理解できたのに対して、後者の生徒が積分の理解に失敗した場合には、理解の絶対的な到達度と達成度に関する主観的な評価はそれぞれ全く異なった値となるであろう。

## 6.2 個人内評価の評価方法

総合学習と関連した調べ学習などでは、問題を設定する能力や、知識の自力構成力、修正力、発表能力、交流力、共同する力など、様々な能力が必要であると考えられているが(小田,2001)、本節では個人内評価の評価方法について主に考察したい。

個人内評価といえども評価の対象とする場合には何らかの基準や規準が必要とされる。平成14年の時点で個人内評価が積極的に利用されているのは主に総合学習に限られているが、本来は全ての教科で利用されるべきものである。べきであるという表現よりも、生徒たちは何らかの形で既に日常的に実践しているはずであるが、自己流に基づいて暗黙に行われている自己評価がどのようなものであるかを自覚することは様々な局面で有益な結果をもたらすと思われる。また、これから述べるように、生徒たちが暗黙に行っている自己評価は必ずしも合理的な評価ではない。そのため、必要に応じて教師がアドバイスしていく必要がある。

たとえば、個人内評価、自己評価に関する研究は、心理学ではメタ認知(metacognition)に関する研究として古くから行われてきた。メタ認知とは、MosesとBaird(1999)によれば、1890年のJamesの研究にまで遡ることができるが、一般的な定義としては、「記憶や、思考、注意などの認知的活動をモニターするような上位の認知的活動」となるだろう。別の言葉で更に言い換えると、「自分自身の思考についての思考」と表現できる(Flavell,1977)。つまり、特定領域に固有な認知的な活動ではなくて、問題解決や視知覚など人間の認知活動全般に影響をもつようなプロセスがメタ認知といえる。

学習と関連した場面で例を挙げれば、記憶方略の切り替えや、意思決定手法の切り替え、問題空間の探索方法の切り替えなど、「どの場面でどのような方略を利用すると効率的な学習や問題解決ができるか」に関する知識と、判断が相当している。より具体的には、何かを記憶しなければならないような状況において、丸暗記する方略と、既存の概念と関連づける記憶方略(精緻化リハーサル:elaborative rehearsal)のどちらの記憶方略が効率的であるかを判断し実行に移す過程が、記憶場面におけるメタ認知的活動となっている。

一般に、6歳をこえる頃から様々な方略(思考の道具立て)を用いることができるようになるが、自分の状態を的確に判断し、それに応じた方略を採ることができるようになるのは、小学校の高学年に入ってからであると考えられている(清水,1995)。たとえば、新規の概念を長期間正確に保持・再生できるように記憶するときには、既に熟知している領域の概念と関連づけて記憶する精緻化方略を選択することが望ましい。教科書などでも、原子核の構造を学習するときには太陽系のアナロジーが使われている。しかし、小学校一年生の児童であれば、このように関連づけることができる知識自体が少ないので、思考の道具として精緻化方略を利用できたとしても、実際に使える場面は限られているといえるであろう。

ただし、関連づけることができる周辺知識を十分に蓄積していると考えられる大学生でも、単なる丸暗記も精緻

化方略も、記憶の残り易さはどちらも同程度であると判断するという実験結果が報告されている (Schaughnessy, 1981)。しかし、Nolen (1996)や、センター試験への取り組み方と学力の残存状態について説明したとおり、学習目標と学習方略とはある程度対応関係が成立していることも経験的に把握されている。つまり、現時点では、学習者が各学習方略の機能についてどの程度熟知しているかについては、心理学的にはまだ明らかにされていないといえる。

しかしながら、上記のようにまだ不明な点が多いものの、学習目標に対して生徒が選択した学習方略がどの程度適切であるかを生徒に問うことと、その判断の適切さを問うことは可能であると思われる。この場合には、目標自体の適切さも評価の対象に含まれるはずであるが、個人内評価の適切さを評価する場合には、上述したようなメタ認知の適切さを判断基準として利用することになるであろう。

ここまでの議論は個人内評価をミクロな視点で評価する方法であったが、次に紹介する評価は個人内評価をこれよりもマクロな視点でとらえる方法となっている。もちろん、個人内評価に関する研究が本格的に始められたのは近年になってからの話であるので、本論文で紹介する発達段階モデル (Towler & Broadfoot, 1992) も確定的なものではない。しかしながら、何故そのようなマクロなモデルが必要とされるかという理由は簡単に説明できるであろう。それは、自己評価能力をどのように発達させるかについての全体的な見通しがなければ、生徒の自己評価能力をどう伸ばすかの指導の方向性が見えてこないからである (鈴木, 2002)。すなわち、評価とは到達状態と目標状態との差分を教育的価値に基づいて判断する行為である以上、何らかの目標状態を設定する必要があり、そのための物差しとしてこれから紹介するモデルが使えるのではないかということである。

Towler と Broadfoot (1992) が提唱する発達段階モデルは以下の4つの段階 (phases) から構成されている。

1. 知識段階 (knowledge phase) - あのときはどうだったかななどの報告ができる
2. 分析・理解段階 (analysis/understanding phase) - どしてそうだったかななどが説明ができる
3. 評価の段階 (evaluation phase) - 学習の状況について判断したり、何が達成されたかを説明できる
4. 総合の段階 (synthesis phase) - 何を学習したかを、長期的な学習の文脈 (将来の学習活動・ビジョン) の中で評価したり位置付けることができる

たとえば、上記の発達段階モデルで、心理学の卒業研究がどのように自己評価されるかを説明してみよう。

知識段階とは調査なり実験の結果を単純に説明できる段階が相当しており、良かったとか悪かったとかその程度の自己評価が行われる。次の分析・理解段階は、どうしてそのような結果が生じたかを分析的に説明できる段階であり、当該領域の知識の理解度について自己評価が行われる。評価の段階では、その研究プロジェクト自体が成功したか失敗したかの判断が行われ、比較的客観的な根拠に基づいた自己評価が行われる。そして、総合の段階では、結果に対する評価を受けてプランを修正したり拡張する作業が行われ、長期的な研究プランに対して現在ほどの程度に位置するかの自己評価が行われる。これらの発達段階の過程で、中心性のある課題を自力で発見する能力や、知識を自力で構成する能力、計画を修正する能力、他者へ説明する能力、批判を咀嚼する能力、良い研究に対する審美眼などが培われることになる。

### 6.3 個人内評価を行うことのメリットと注意点

通常の評価場面で成績が低い生徒は、おそらく相対評価でも絶対評価でも評価が低くならざるをえないであろう。しかし、個人内評価を積極的に進めることによって、各生徒なりの進歩を実感できるようになる可能性がある。そして、個人の進歩を認めることが、自己効力感と自尊感情を高め、結果的に、学習意欲を高めると考えられる (小田, 2001)。また、相対評価や絶対評価で測定された学力と、自分の進路観とを結びつけることができる可能性もあるだろう。

たとえば、得意科目と大学進学などの進路選択は現状でも一致することが多いといえるであろうが、果たしてこ

れで本当に問題がないのかという議論の余地があるであろう。一例を挙げてみれば、たとえば、理数系が得意でその他の科目の偏差値が高いからといって、それだけで果たして医学部に進んでよいものか、その後の適性などを考えると大きな問題があるように思われる。もちろん、これらの進路決定において、相対評価や絶対評価から得られる情報は判断材料として大いに活用されるべきである。しかし、前節で紹介したような、自己評価の発達段階モデルにおける総合の段階を抜きにして進路を決定したならば、高校までの学業と大学入学後の学業とのギャップに苦しむ学生も出てくるであろう。実際に、平(2001)が報告したとおり、多くの学生は大学受験を通して学ばれた知識が忘れ去られても、それほど明確な疑問を感じていない。つまり、ギャップが生じることは当然であると割り切っているのであるが、割り切ることができない場合には、退学などの道を選ぶしか残されていないのではないだろうか。

以上の指摘は個人内評価を行うことのメリットであるが、既に繰り返し指摘したように、何を進歩とするかという判断基準の曖昧さが残されている。特に、指導を行う教師の側に、自己評価の発達に関する明確なイメージが存在しないと評価を行うことすらできなくなると考えられるが、現時点では、教育学や心理学の分野でも未開発の研究分野であり、あまり明確な指針を提示できない状況にある。たとえば、様々な学力や能力が総合学習を通じて学ばれると考えられているが、具体的な概念規定は研究者によって千差万別である。そして、その結果、たとえば、オープンエンドな学習環境では、生徒が何をしても評価すべきだという誤解が生じてくるように見受けられる。すなわち、このような極端な誤解は、課題内容の多様さと、自己評価能力などの評価基準の曖昧さが混同されているために生じているように思われる。

また、自己評価をすること自体が目的化してしまうと、評価活動が単なるルーチンワークになってしまう可能性もあるだろう。特に、観点別に数値を用いて評定する自己評価活動の場合には、作業が簡単なために評価を毎時間行うことも可能である。このような場合には、作業内容を振り返るのではなくて、作業量を想起するだけで自己評価をこなすことも十分に可能である。実際に、教師が設定した観点に対して生徒が  $\times$  をつけていくような場合には、一つの単元の期間中に延々と同じ評価が続くことが多い。

このような自己評価活動には、ほとんど意味がないことは明らかであろう。極端な場合には、教師が本来すべき評価作業を、生徒に押し付けて手抜きをしているように見えてしまう可能性もある。観点別の自己評価は学期末のとりまとめが容易であるが、生徒が目的や活動内容をきちんと理解したうえで評価するようなスタイルになっていないと、教師が意図したような効果が得られにくいと考えられる。形成的な自己評価という観点を大切にするのであれば、生徒の側に立ったような評価方法とペースで自己評価を促すべきであろう。現実的には、自己評価活動が独立して存在するのではなくて、予習・復習を効果的に行うのに役立つような形式と内容になっている必要があるのではないだろうか。

## 7. 学習評価テストバッチリの必要性

現在、学習評価に関して教育現場を悩ませている問題として、評価の過剰を指摘することができる。個別的には様々な難しさがあり得るであろうが、一般的には、(1) 序列化の過剰、(2) 評定対象の過剰、(3) 作業記録の過剰が指摘されている(国民教育文化総合研究所教育評価研究委員会, 1998)。

(1)の問題は、いわゆる選抜尺度の過度の精密化の問題であり、既に考察したとおり、データの尺度水準と照らし合わせて過剰なデータ処理を行っている可能性を示唆している。(2)の問題は、関心・意欲・態度など、評定が困難な対象も評価の対象に含まれていることである。そして、(3)の問題がもっとも深刻だと思われるが、通信簿、指導要録、内申書など、教師は膨大な評価を強いられている反面、それが有効に活用されることは一般的に希であるという問題である(たとえば、転校や学級の組み替えを行う時には、各生徒の指導要録などが引き継がれるはずであるが、これらを参照する教師はそれほど多くない)。

つまり、情報公開の流れにともなって、過度に精密な評価を行わなければならないような雰囲気は形成されてお

り、更に指導要領の改訂にともなって評価が難しい情意面までも評価対象が広がったものの、それらの評価結果が実際の指導に活かされる機会は多くない、といえるであろう。すなわち、負担感に加えて徒労感も増しているといえるであろう。その他、教師がなすべき仕事としては、これらの評価活動以外にも、地域との連携を推進することや部活動などの特別活動の指導などが存在し、教師の日常は多忙を極めている。

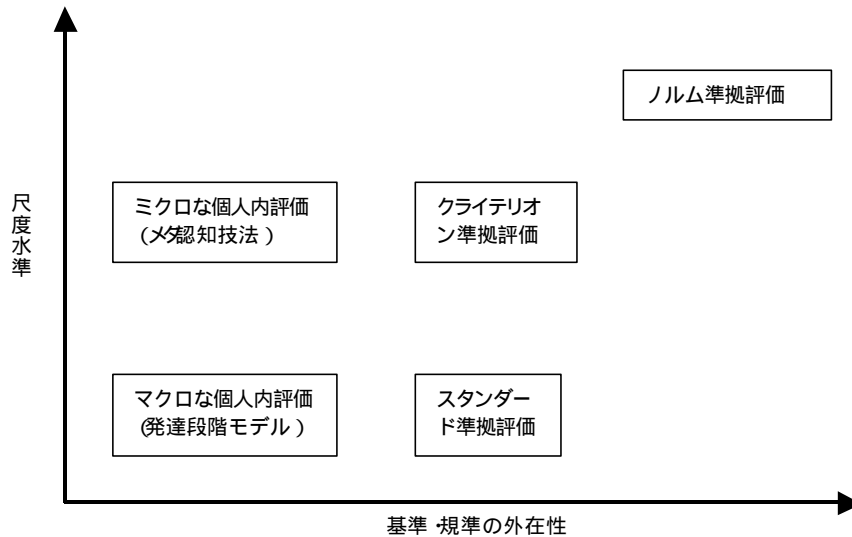


図1. 各種評価技法の見取り図

さて、上記のような教師の多忙感を救済することを願って、最後に各種学力評価技法の見取り図を提示したい(図1)。もちろん、このような見取り図を提示することによって教師の多忙感が軽減されるという保証は全くないが、少なくとも、収集する資料から「どのような評価情報を得るか」を明確にする手助けになると思われる。たとえば、総合学習におけるポートフォリオを作成するような場面では、漫然と資料が蓄積される一方で、それをどのように整理すればよいかの見当がつかないことが多い。そして、そのために、ポートフォリオは使い勝手が悪いと批判を受けることがある(高浦, 2001)。しかしながら、図1に示したような鳥瞰図が頭の中に入っていれば、これらの雑多な情報を整理してまとめることも比較的容易になると思われる。

たとえば、心理検査の査定報告書は、シュナイドマンの図のような各種の心理検査の関係を図示した見取り図を念頭において作成されることが多い。このような見取り図を用いるメリットの1つとして、査定情報間の齟齬を解消するときなどに重要な働きをもっていることを指摘できる(田中, 1996)。たとえば、知的能力に関する情報として、一般の知能検査(e.g., ウェクスラー尺度など)で高い知的水準が示されたのに対して、ロールシャッハテストの所見からは低い知的水準が示された場合には、知的能力に関して異なった査定情報が得られたことになる。このとき、2つのテストが構成概念として同一の知的水準を計測しているならば深刻な矛盾を引き起こして解釈が困難になる。しかしながら、それぞれのテストで測定されている知的水準が全く異なっているため、統一的な解釈を行うことも可能となる。

つまり、複数の評価項目を設定して査定するような状況で整合性をもった解釈を行うためには、図1のようなテスト間の位置関係を把握していることは必要不可欠といえるであろう。逆に、このような見取り図がない場合には、多くの査定情報が得られたとしても、高浦が指摘するようにまさに雑多な情報にしか見えないはずである。今後は、各評価方法の個別的な研究以外にも、各評価方法どうしの対応関係に関する研究も深めていく必要があるといえるであろう。



## 文献

- 天野正輝 (1993). 「教育評価史研究」 東信堂
- 有坂誠人 (2002). 「有坂誠人の現代文速解例の方法 試験で点がとれる」 学習研究社
- 東洋 (2001). 「子どもの能力と教育評価」 東京大学出版会
- Brookhart, S.M. (2002). What will teachers know about assessment, and how will that improve instruction? in R.W. Lissitz and W. D. Schafer (ed.) Assessment in Educational Reform. Chap.1 Pp.2-17.
- Flavell, J.H. (1977). Cognitive Development. Prentice-Hall.
- 藤澤伸介 (2002). 「ごまかし勉強」(上・下) 新曜社
- 萩生田伸子・繁榊算男 (1996). 「順序付きカテゴリカルデータへの因子分析の適用に関するいくつかの注意点」 心理学研究 Vol.67, 1-8.
- 池田央 (1992). 「テストの科学 - 試験に関わる全ての人に」 日本文化科学社
- 鎌原雅彦・亀谷秀樹・樋口一辰 (1983). 「人間の学習性無力感に関する研究」 教育心理学研究 Vol.31, Pp.80-95.
- 北尾倫彦 (2002). 「目標準拠評価の手順と課題」 指導と評価 Vol.48. 8-12.
- 国立教育政策研究所編 (2002). 「評価基準の作成、評価方法の工夫改善のための参考資料 - 評価基準・評価方法等の研究開発(報告) - 」
- 国民教育文化総合研究所教育評価研究委員会 (1998). 「教育評価研究委員会報告」  
(<http://www.jtu-net.or.jp/%82%88%82%90/Soken/pdf/hyouka.PDF> last accessed: 30/09/2002 )
- Krampen, G. (1988). Competence and control orientations as predictors of test anxiety in students: Longitudinal results. Anxiety Research, 1, 185-197.
- 小林洋文 (1996). 「行政による入試改革は成功しているか」 全国到達度評価研究会編 「子どものための入試改革」 法政出版 Pp.55-73.
- Moses, L.J. & Baird, J.A. (1999). Metacognition. In Wilson, R.A. and Keil, F.C. (Eds.) The MIT Encyclopedia of the Cognitive Science. Pp.533-535.
- 文部科学省 (2001). 「平成 13 年版科学技術白書」(<http://www.wpi.mext.go.jp/kag2001/index.html> Last accessed: 29/09/2002 )
- 奈須正裕 (2002). 「学習性無力感」 日本教育工学会編 「教育工学事典」, Pp. 387-389. 実教出版
- 中原克己 (1996). 「学校教育の改革と入試制度」 全国到達度評価研究会編 「子どものための入試改革」 法政出版 Pp.167-183.
- Naveh-Benjamin, M., Lin, Y-G. (1994). Measuring and improving students' disciplinary knowledge structures. In Pintrich, P. R. and Brown, D. R. (Eds) Student motivation, cognition, and learning: Essays in honor of Wilbert J. McKeachie. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc. Pp. 51-78.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. Psychological Review, Vol.91, Pp.328-346.
- 日本標準教育研究所 (2002). 「教科書単元別 到達目標と評価基準・国語・光村 3・4 年」 日本標準
- Nolen, S.B. (1995). Effects of a visible author in statistical texts. Journal of Educational Psychology, Vol. 87. Pp.47-65.
- Nolen, S.B. (1996). Why study? How reasons for learning influence strategy selection? Educational Psychology Review, Vol.8, Pp.335-355.
- 小田勝巳 (2001). 「ポートフォリオで学力形成」 学事出版
- 大野木裕明 (1994). 「テストの心理学」 ナカニシヤ出版
- 佐伯胖 (1995). 「わかるといふことの意味」 岩波書店
- 佐藤隆博 (1982). 「S-P 表の見方と学習診断のための利用法」 佐藤隆博編著 「S-P 表の活用 - 中学校編」Pp.8-29. 明

## 治図書

- Seligman, M., & Maier, S. F. (1967). Failure to escape traumatic shock. *Journal of Experimental Psychology*, Vol.74, 1-9.
- Shaughnessy, J.J. (1981). Memory monitoring accuracy and motivation of rehearsal strategies. *Journal of Verbal Learning and Verbal Behavior*, Vol.20, 216-230.
- 清水寛之 (1995). 「記憶力」 高野陽太郎 (編) 「認知心理学 2 記憶」 第 8 章 Pp.169-187. 東京大学出版会
- 鈴木秀幸 (2002). 「自己評価能力をどう育成するか - イギリスでの自己評価の研究と実践から」 指導と評価 Vol.48 21-24.
- 平真木夫 (2000). 「教育の情報化の動向を探る」 認知科学会テクニカルレポート JCSS-TR-34, 1-22.
- Taira, M. (2001). The role of domain knowledge and PC literacy of novice users. *Proceedings of ICCS'2001*, 658-662.
- 平真木夫 (2001). 「類推の思考における位置づけ」, 宮城教育大学紀要, Vol.35, 313-327.
- 平真木夫 (2002). 「検索前の知識状態から推測される検索方略と知識の広がりについて」 第 19 回日本認知科学会 発表論文集 24-25.
- 高浦勝義 (2001). 「ポートフォリオ評価としての生活科の評価の取り組み」 総合学習 Vol.5 66-67.
- 田中富士夫 (1996). 「査定情報の総合と伝達」 田中富士夫編著 「臨床心理学概説」 第 6 章 Pp.82-91.
- 辰野千尋 (2001). 「改訂増補 教育評価基本ハンドブック」 図書文化
- Towler, L. & Broadfoot, P. (1992). Self-assessment in the primary school. *Educational Review*, Vol.44, 137-151.
- 渡辺洋・平由実子・井上俊哉 (1988). 「小論文評価データの解析」 東京大学教育学部紀要 Vol.28, 143-164.