

どうする 観点別・絶対評価

根岸雅史

(東京外国語大学助教授)

今年度より、中学校において、絶対評価が始まった。これまでに行われてきた「観点別評価」と合体する形での開始である。この絶対評価をめぐるのは、様々な混乱が起きている。この混乱は、「観点別・絶対評価」の実施方法に始まり、内申書絡みの「観点別・絶対評価」の是非論に至るまで、広範にわたっている。日本においては、評価の観点を示された評価基準(のりじゅん)をもとに、各学校において「生徒がどこまでできるようになっているべきか」を示す評価基準(もとじゅん)を作成することになっているが、このためのモデルとなるような評価基準が示されていないことによる混乱も大きい。

1. マトリックスのどこをテストによるか

いわゆる「評価計画」を立てる場合、「内容のまとまりごとの評価基準」のマトリックスの中のどこをどのような方法で評価するかを決定する必要がある。その評価方法は、ペーパーテストに始まり、観察、インタビュー、スピーチ、テープ・ビデオ、学習ノート、ワークシート、レポート、パフォーマンス、ポートフォリオ、自己評価・相互評価、質問紙などの方法があるわけであるが、評価にあたっては、その特性を生かした選択が必要である。本稿では、紙幅の制約もあり、その選択に関して、いくつかの点にしぼって論じる。

1. 「コミュニケーションへの関心・意欲・態度」

「コミュニケーションへの関心・意欲・態度」の評価は、かつてはペーパーテストによる方法も試みられた時期もあったが、最近では、この観点は教室内の活動の観察によっていることが多いのではない。この観点の評価は、「言語活動への取り組みの

積極性や意欲」および「コミュニケーションを続ける工夫」に関わっている。前者は、教室内で課された様々な活動に対する取り組みの様子を観察することになる。それに対して、後者は、主に発表技能の中でとらえることになるのではない。もちろん、「聞く」や「読む」といった受容技能においても見ることはできるが、「工夫」を必要とする困難な状況が本当に存在するかどうか、外からは見えにくいために、これを受容技能で調べるには、ある程度実験的な設定をテストなどの中で行う必要がある(例えば、未知語を文章中に入れておいて、あきらめずに読み続けられるかを見る、など)。

2. 「表現の能力」としての「話すこと」と「音読」

「表現の能力」のうち、「話すこと」と「音読」については、教室内の活動の観察によることが多いが、テストによる評価も不可能ではない。これらのテストは、学期中に1時間を割いて、順に実施するというのが一般的であろう。この場合は、受験者以外の生徒はワークシートなどをやらせておくとよい。

3. 「言語や文化についての知識・理解」

このうち、「『言語』についての知識・理解」の評価については、何を評価するかについて議論が割れているようである。いわゆる文法・語彙テストを用いるべきでないとする主張と文法・語彙テストを用いてもよいとする主張とがある。確かに、実際に運用に至らない文法や語彙知識を見ることは意味がないという議論も成り立つかもしれないが、日本の学習環境を考えた場合、どうして運用に至らないのかを文法や語彙の観点から知ることには、診断的な意味合いはあるのではない。

2. 観点別・絶対評価のポイント

以下に、観点別・絶対評価のための10のポイントを挙げて、順に論じる。

1. 指導目標を立てる

定期試験などの achievement test は、本来特定の指導目標の到達を見るものであったが、これまでは必ずしも目標に準拠していない定期試験も少なくなかった。ところが、絶対評価では、指導目標を立てることは必須のプロセスである。つまり、ある期間の指導において、「～ができるようにする」というような具体的な目標設定が必要なのである。

では、具体的には、どのような目標設定が必要であろうか。絶対評価のための指導目標の設定といった場合、文法や機能、語彙、表現に関する指導目標の設定は比較的容易であろう。それに対して、「理解の能力」や「表現の能力」の評価においては、「英文を正確に読みとる」などという漠然とした目標設定になっていたりする。しかしながら、これではどの課を指導しても同じ目標が立ってしまう。実際には、中学の教科書は基本的に文法シラバスで構成されているために、もう少し言語寄りの記述が必要であろう。例えば、「受動態を含む文を聞き取って理解することができる」というような具合である。ただし、こうした目標にしたがって作ったテスト問題を見た場合、確かに英文には受動態が含まれているが、実際の解答はキー・ワードの聞き取りだけでできてしまう問題が少なくない。こうなってしまうと、このテスト問題からは、上の目標が達せられているのかは見えてこない。

2. 到達目標を立てる

指導目標が立ったら、今度は到達目標を立てる必要がある。到達目標とは、指導目標のどこまでが到達していればよしとするかの記述である。例えば、on, above, under, below などの前置詞の使い分けについて、6割方できていればよしとするのか、9割以上の出来でないと目標に到達したと判断しないのか、ということである。テストによる絶対評価の場合、到達目標は、指導目標に対応するそれぞれ

の大問ごとの正答率を持って定義することになるので、実際はこの線引きのための点数の決定が到達目標の設定となる。もちろん、多くの場合、テスト・データは連続的で、この連続データのどこかで線を引くわけであるが、この際には、習得しているかどうかを分ける意味のある違いを教師がいわば主観的に判断することになる。さらに、その際には、テスト項目の難易度も考慮に入れる必要がある。つまり、TFのようなテスト問題であれば、「サイコロをころがしても」50パーセントの正解率はあるわけであるから、このような問題の線引きでは、60や70パーセント以上などとしなければ、意味がない。

3. 評価は複眼的に行う

テストで見える英語力は、英語力の総体の一部であり、全てではない。したがって、評価にあたっては、テスト以外の場面における生徒の言語使用の観察データなども利用するべきである。その際、テストはその得意とする部分の測定に用い、それ以外については代替的アセスメント (alternative assessment) 方法を用いるというような複眼的評価が望ましいだろう。

4. テスティング・ポイントは明確に

これまでに英語の定期試験を見てきた経験では、テスティング・ポイントが明確でないテストがきわめて多い。つまり、どのような能力や知識をそのテスト項目で測定しようとしているのかが見えてこない問題が多いということである。このようなテストは、観点別・絶対評価においては、致命的な問題点を抱えているといえることができる。絶対評価では、それぞれの指導目標について、到達基準に達しているかないかを判断しなければならないため、そのテストにおいて、それぞれの大問が何の能力や知識を測定しているかが明らかになっていなければならない。これを確実にする1つの方法としては、大問ごとにタイトルを付けるという方法がある。これにより、教師はテスティング・ポイントを明確にせざるを得ないだけでなく、生徒に対してもどのような能力を測定しようとしているのかが明らかになる。

5. 総合問題はやめる

典型的な英語の定期試験問題の中に、いわゆる「総合問題」がある。これは、あるまとまった文章があり、その文章をもとに発音問題や文法問題、語彙問題、和訳問題、内容理解問題などが含まれている、いわゆる「ごった煮」の問題である。この種の総合問題では、そこで何点とれたといっても、元々が「ごった煮」であるから、この問題の出来から、観点ごとに絶対評価を行うことは不可能である。また、仮にその総合問題のサブカテゴリーをもとにデータを作成したにしても、それぞれは項目数が極端に少なく、信頼に足るものではない。さらに、総合問題の出来を「理解の能力」として使うことがあるようだが、こうした扱いの妥当性のなさは、その問題構成を見れば明らかである。

さらに、この総合問題の波及効果も問題である。実は、このようなバラバラな言語知識を実際に言語使用の中で求められることはまずない。総合問題に備えさせる中で、いわば「特殊な」言語処理プロセスを英語学習者に課すことになってしまう。さらに、この総合問題は、いわゆる「英語」ができるとは、あのような断片的な知識を問う問題を解くことができることだと学習者に広く思い込ませてしまっているのである。これらの点からも、「総合問題」を採用しなければならない積極的な理由は見つからない。

6. 直接的測定を行う

絶対評価においては、学習者の「何がどのくらいできるか」について具体的に知る必要がある。したがって、当然その測定は直接的測定にならざるを得ない。例えば、「発音の能力の測定」に際しては、「紙と鉛筆による間接的測定（実際の発音を観察していないという意味において間接的）」によって、実際の「発音の能力」を知ることはできない。発音の能力は、生徒に発音させなければわからない。これは、テストの妥当性に関わる重要な問題である。

7. 信頼性を高める

従来定期試験では、総合点をもとに評定を出すケースが多かったと聞いている。こうしたケースでは、定期試験が30問のテスト項目から構成されて

いたとすると、この30問の解答結果をベースに評定を導いていた。それに対して、同じような問題構成のテストを用いて、観点別・絶対評価を行うために、仮に「コミュニケーションへの関心・意欲・態度」以外の3つの観点をこのテストで見ようとする、それぞれの観点到使える項目数は平均で10項目となる。一般に、テストの信頼性は問題の項目数が多ければ多いほど高くなるから、10項目で同じような信頼性を達成するには、よほど精度の高いテスト項目から構成しておく必要がある。

また、観点別・絶対評価では、教室内活動の観察が強調される傾向にあるが、この観察は、よほどシステムティックに行わない限り、信頼性の点では、テストに及ばないということは認識しておく必要がある。 「コミュニケーションへの関心・意欲・態度」について観察で行う場合に、活動に参加していなければC、参加していればBというような基準で評価するのであれば、参加していない生徒だけをチェックしても、ある意味で全員の到達度を見ることにはなるかもしれない。しかし、表現の能力の評価として、ペアワークなどで生徒が話しているのを観察するとなると、せいぜい数組のペアワークをわずかの時間「見る（基準に達しているかを判断する）」くらいであろう。このような観察に高い信頼性が期待できないのは、明らかである。ましてや、理解の能力を観察によることは、もっと難しい。

8. テストの合計点を出さない

日本のテストの伝統からすれば、テストの合計点が不要であるということは、なかなか納得のいかないことかもしれない。しかし、よくよく考えてみれば、合計点というのは他者との比較には役に立っても、実は様々な異なった能力の指標を合計してしまっているために、絶対評価においては何の意味も持たない。言い換えれば、合計点は「総合的な英語力」の指標であり、あることができるかできないかの指標にはなり得ない。

9. 「ごった煮」採点をしない

定期試験の採点は、これまでどのように行われてきたであろうか。一般的には、まず、答案の「丸

付け」を終えると、大問ごとの合計点を計算し、それを解答用紙の端に書き留め、次に、大問の点数を合計して、合計点の欄に記入する。そして、名簿に生徒1人ずつの点数を書き込み、答案を返却するといったところではないだろうか。実は、この「当たり前」のプロセスが、絶対評価においては大きな問題を引き起こすのである。それはなぜかという、このプロセスの中で、生徒のそれぞれの観点ごとの情報が、解答用紙とともに教師の手元から消えてしまうからである。

こうした問題をなくすためには、少なくとも合計点は解答用紙の上で計算しないことである。まず、採点を終えた時点で、大問ごとの合計点を出すことは、従来と同じであるが、次に「大問ごとの合計点」をそれぞれ名簿に記録する。この大問情報を、観点別の評価資料とするのである。もちろん、このような作業はパソコンの表計算ソフトを用いることでかなり効率的に行うことができる（下の表、参照）。こうした採点法を採用することにより、生徒個人個人の出来がわかるだけでなく、大問を縦に見ることで、指導目標がクラスとして達成しているのかもわかる。その結果をもとに、その後の指導の調整が可能となる。

表) パソコンの表計算ソフトによる得点の記録例

	大問1	大問2	大問3	合計
藤田 義人	8	5	7	20
山田 文	6	4	8	18
クラス平均	7	4	7.5	18.5

10. 説明責任を持つ

最近では、どこでも説明責任 (accountability) が問われるようになってきており、これは評価においても同様である。評価における説明責任とは、評定やスコアがどのようなプロセスから出てきて、どのような根拠に基づいているか、そして、それが何を意味するかなどを評価に関わる全ての人々 (教育委員会、校長、同僚、保護者、生徒など) に対して、それらが求められたときに、説明する義務である。これは何も絶対評価にのみ生じる義務ではないが、絶対評価においては、指導目標や到達目標とは何かか

ら始まり、その目標への到達をどのように判断したのか、さらに、それらの結果をどう総括して、評定としたかなどについて、全ての人々に納得のいく形で説明することはかなりの困難を伴うであろう。説明責任の持てるテスト作りのためには、上記に述べたように、指導目標と到達目標を公開し、その上で、どのような方法でそれを測定するかを明らかにしなければならない。

3. 観点の重み付けの是非

評定を出す際のそれぞれの観点の扱いに関しては、せっかく導入された観点であるから、全ての観点を平等に扱うべきだとする主張がある。それに対して、指導目標にふさわしい重み付けがなされるべきだとする主張もあり、現場は混乱している。しかしながら、平成14年2月に国立教育政策研究所教育課程研究センターから出された『評価規準の作成、評価方法の工夫改善のための参考資料 (中学校) — 評価規準、評価方法等の研究開発 (報告)』では、評定の出し方に関して、「評定に総括する際の観点別学習状況の評価結果についての観点ごとの重み付けが異なることがある」と明記され、「重み付け」を認めたととれる。例えば、「言語や文化についての知識・理解」の評価として、英語の手紙の様式 (日付や名前などの書き方) というような言語文化的な知識を見るときした場合、これと手紙本体の文章を書く「表現の能力」を同じ重み付けとしたのでは、無理があるように思われるのであれば、「表現の能力」重視の重み付けが望ましいということになる。

4. 評価の先に

現状では、いかに絶対評価を実施するかということばかりに目が奪われているが、本来は絶対評価を行った結果、目標に到達しなかった生徒がどれくらいいたのか、そして、彼らに目標を達成させるためにはどうするのが、むしろ問題である。絶対評価というシステムが、こうした本来の貢献をできる日が一日も早く来てほしいものである。