2　測定と評価
――学習者の動機づけと教師の指導力向上となるテストを

教師は学習者の英語力をテストして評価する。同時に学習者の理解度を確認し，教え方を振り返る機会となるので，テストは授業と表裏一体の関係にある。例えば，オーラルコミュニケーション重視の授業を実践しておきながら，読み書きの筆記テストだけで聞く話す力を測定しなければ正しい評価とはならない。また，ビジネス英語の授業目標に「英文の履歴書が作成できる」や「プレゼンテーションができる」が含まれていれば，履歴書を書くテストやプレゼンテーションをするテストを実施すべきである。JACET教育問題研究会（編）(2017）によると，「テスト結果を利用して学習者の成長につなげるように支援することが評価を行う教師の務め」である。学習者は不足している能力を発見し，さらに授業者は自らの指導方法を反省し改善していくことが理想である。

客観テストと主観テスト

出題形式が記号による選択式の問題であれば採点者が誰であっても客観的に採点できる。一方，自由英作文などの記述式の問題やディベートなどの表現力を問う評価は主観的になりやすく学習者に不公平感が生じてしまう恐れがある。主観や印象で評価しがちな態度や技能をより客観的に評価するためには，評価項目を細分化したり，採点者を複数にしたりする。

パフォーマンス評価とポートフォリオ評価

パフォーマンス評価（**performance evaluation**）とは英語の実技テストであり，ディスカッション，ショッピング，クライアントとのミーティングといった課題に対して，学習到達状況の評価基準を観点と尺度の表にまとめた**ルーブリック**（**rubric**）などを用いて総合的に評価する。表の形式は，文法・発音・アイコンタクトなどの各項目に対してそれぞれの尺度における典型的な発表の様子を説明した記述からなっている。**ポートフォリオ評価**（**portfolio evaluation**）は学習者が自由英作文などを長期にわたって収録し，学習者の内省・成長への気づきを促し，教師からの指導や評価を受けたときの記録などを系統的に蓄積していくものである。いずれの評価法も学習者に事前に十分理解させておく必要がある。

総括的評価と形成的評価

総括的評価（**summative evaluation**）とは，学期末に中間考査，期末考査などの

長期間の学習成果の測定結果を重視して学期中全体の学習成績を通知表に5段階評定などとして示すものである。**形成的評価（formative evaluation）**とは，狭い範囲の学習内容についての理解度をテストすることや授業態度を観察するといった短期間に行う評価である。よって日々の授業では頻繁な形成的評価におけるコミュニケーションが教師と学習者に介在し，学習者はどこでつまずいているのかを把握する機会となり，教師は指導法を振り返り改善するための材料となる。

規範参照評価と基準参照評価

規範参照評価（norm-referenced test）には，能力テスト（熟達度テスト）とクラス分けテストがある。民間試験の英検，TOEICや大学入学試験のように，全般的な言語能力を測り，受験者全体の中での相対的な位置を調べることができる。同じ試験問題でも受験者集団が変われば評価も変わる。受験者の得点は正規分布曲線を描くので集団内での順位付けができる。一方**基準参照評価（criterion-referenced test）**には，到達度テストと診断テストがある。学習者が目標にどの程度到達しているかを測定するテストであり，学習者間の学力差を比較することが目的ではない。テスト範囲を決めて行う小テストや期末考査などは授業の理解度を測定できる。全員が到達点に達していれば全員が満点の評価を得ることもある。よって，教師は学習者の現状をより適切に把握しつつ自らの授業改善を短いサイクルで行うことができる。

信頼性と妥当性

信頼性（reliability）とは，測定しようとする能力を測定できるだけの安定した精度のテストであるかどうかである。昨日実施したテストを同じ学習者に今日実施しても，結果が変わらなければ信頼性がある。信頼性を確保するためには，テストの中身だけでなく，環境要因も無視してはならない。試験会場の室温，時間帯，学習者の健康状態や疲労度などがテスト結果に誤差を生じさせる。安定した測定値を得るためには，テストは1回きりではなく，繰り返し形成的に実施することが望ましい。**妥当性（validity）**とは，テストで得た得点が運用能力と相関するか，テストで測定したいことが正しく測定できているかどうかの指標である。言語の機能（提案する，不平を言う，謝罪するなどを表現すること）の測定は，筆記試験だけでは不十分なので，例えばペアワークやロールプレイなどのパフォーマンステストも加えると学習成果をより正確に測定できるため妥当性があるといえる。

（市島　清貴）