

カリキュラムと評価の間で

— 短期大学における英語教育の場合 —

松村 伸一

<要旨>

カリキュラムと評価という言葉は、教育に携わるすべての者にとって馴染み深いが、その反面、それらが意味するところについては、かならずしも見解が一致しているとは限らない。いずれの概念も、教育行政と教育学の歴史的変化による影響を深く受けながら構成された複合的な概念だからである。

とはいえ、あまりにも専門的なとらえ方をしてしまえば、一般の教員には役に立たない秘教的な知識となってしまう。本稿の目的は、英語教育分野における、カリキュラム、評価、テストングに関する諸説を概観し、筆者と同様、英語教育論の専門的な訓練を受けていない語学教員が、これらの問題と関わる際に役立つパースペクティブを描くことである。

あわせて、上記の諸説を踏まえ、本学英文学科における英語教育カリキュラムと評価の現状を述べ、特に「英文講読Ⅰ」という科目で実施された試験問題について紹介する。それぞれの試験問題からは授業がどのように運営されていたかも透けて見える。試験問題を検討するための組織的な対応ができれば、学生のニーズに合ったカリキュラム改革にも役立つだろう。

キーワード：英語教育英語教育、カリキュラム、評価

Keywords: language learning/teaching, curriculum, course assessment

はじめに

少子化による大学受験者数の激減を背景として、大学・短大は未曾有の危機的な状況に置かれている。こうした中、大学短大における研究と教育の質が厳しく問われるようになってすでに久しい。大学短大で教鞭をとる英語教師の多くは、必ずしも英語教育の方法論を体系的専門的に学んだわけではないだろう。内外の大学院で英米文学や英語学の特定領域を学んだ者が、自分自身が専門と考える研究領域に関する講義や演習を担当したりしなかったりするかたわら、全国の大学や短期大学の教壇でいわゆる「語学科目」を担当している。私自身もその一人だ。こうした状況を、批判的に見る向きはかなり以前からある。たとえば「大学に入学したばかりの理科系学生にシェイクスピアを読ませる」というような類いの神話は、いまだに語られないわけではないだろう。初期近代英語で書かれたシェイクスピア作品を例として挙げるのは極端だとしても、文学を専門とする語学教員

が、学生の英語力や関心の在りかとは無関係に、自分の専門領域のテキストを取り上げているのではないか、そしてそれは教育上不適切ではないか、というような疑念は、いまだ根強い。

現在多少なりとも語学教育に関わりのある者であれば、そうした神話を言下に否定することだろう。語学教育がそのような牧歌的状况にある大学短大は、少なくとも今日では皆無と思われる。本来ならば、教養教育という立場から、たとえば理系の学生に文学作品を読ませるような方針をあえて立てることも考えてよいはずなのだが、たいがいは、学生のニーズに応えない授業設計と見なされるであろうことは、語学教員自身がよく認識しているものだ。管見の限りでは、担当者個々人が、学部や大学院で受けた学知とはやや次元を異にする体験知として自らの教育経験を咀嚼し、テキストの選び方、授業運営方法、テストの作り方や評価方法などなど、語学教育のあらゆる側面に関して、ある種職人的な誇りすら抱きつつ、語学科目を教えているというケースが、大半ではないかと思われる。とりわけ、学生による授業評価がほとんど義務化されつつある昨今、語学教育に向けられる、学生やその背後にある社会、そして他の専門領域を担当する同僚達からの目を一切意識しない語学教師は、あまりいないだろう。

本稿では、カリキュラムと評価という切り口から、短期大学における英語教育の現状と課題について考察したい。

まずカリキュラムと評価に関する英語教育分野での一般的な諸説を概観し、筆者と同様、英語教育論の専門的な訓練を受けていない語学教員が、これらの問題に取り組む際に役立つパースペクティブを描くことを試みる。あわせて、本学英文学科における英語教育カリキュラムと評価の現状を述べ、特に「英文講読Ⅰ」という科目で実施された試験問題について紹介したい。

1. カリキュラムとは何か ～ 概念の歴史と構成

カリキュラムという言葉は教育に携わる者にとってほとんど日常語と言っても良いはずだが、改めてその意味するところを説明しようとするとは決して容易ではない。それは、この語が担っている歴史と、その構成概念の複層性とに関連がある。

1-1. カリキュラム概念の歴史

カリキュラム (curriculum) とは、「走路」ないし「競争用二輪戦車^{チャリオット}」を意味するラテン語が語源である。それが教育用語として用いられるようになったのは、16・17世紀に大学の教育内容が国王や教会の厳しい統制を受けた時代のことであり、「権力の統制に反発した大学が、定められた教育内容のコースを強制的に走らされるという揶揄を込めて『カリキュラム』と呼んだ」¹⁾のが始まりである。『オックスフォード英語辞典』は、'a regular course of study or training, as at a school or university' と語義を示しつつ、初例としてグラスゴー大学公文書から1633年および1643年に書かれたラテン語の文言を挙げている。

宗教改革以後、市民革命以前、すなわちローマカトリック教会が一元的に「真理」を支配していた時代は終りを迎えたにもかかわらず、いまだ絶対王政が世俗世界を牛耳っていた時代に、知的自由を脅かされつつあった大学で生まれた言葉であるという事実は非常に興味深い。カリキュラムという語は、その誕生からすでに、公的統制と知的自由との政治的緊張関係をはらんでいたわけである。

カリキュラムという語は、その後主にスコットランドの大学で教科課程を表すのに用いられていた。やがて19世紀末アメリカ合衆国において、教育行政と学校との間の権限が分離される過程の中で、行政が規定する教科課程の大綱を‘course of study’（学習指導要領とほぼ同義）と呼び、「学校において教師と子供が創造する教育経験の総体」（p.4）を「カリキュラム」と呼ぶようになった。カリキュラムは、公的統制の側から、学校における教育の実質へと、再定義されたのである。

その後、20世紀初頭のアメリカにおける州・郡・市のカリキュラム改訂運動のなかでさらに概念が深化されていく。この運動は「大別して二つの系譜で示すことができる」（p.51）。一方の系譜は、1910年代に始まり30年代にピークを迎えた「合理化・効率化と『科学的経営』を求める運動」である。その背景には、大恐慌（1929年）以後の地域経済と教育の復興計画があった。もう一方の系譜は「私立小学校の主導の下で展開していた進歩主義的教育に原型をおくカリキュラム作成の運動」であった。これは各州における復興計画とも連動して、進歩主義的教育運動が公立学校にも普及する契機となった。その大きな牽引力となったのが、ジョン・デューイ（1859-1952）の教育理論と実践であった。これら二つの系譜のはざまに「カリキュラム開発」の考え方が形成される。その特徴としては「授業改善」「教師の参加」「教職の（再）研修」が挙げられる。（pp.51-53）

第二次世界大戦後、カリキュラム開発の理論化が進展する。ここでも二つの系譜ないしモデルが生まれた。一方は、創造的芸術活動をモデルとした、H・ラッグの「カリキュラム・デザイン」概念である²⁾（pp.53-57）。ラッグによれば、カリキュラム・デザインとは、自国とヨーロッパの文化に通じた感受性と教養を豊かに備えた教師が、「生活の諸技芸」（arts of living）を組織化し、生徒に与える手段を講じることであった。もう一方がラルフ・W・タイラーの「カリキュラム・プランニング」の概念で、行動科学に依拠しつつ、産業・工学モデルによってカリキュラム研究方法の基礎を築いた。タイラーは、カリキュラムと授業の過程を、段階的で循環的な過程としてとらえるとともに、各段階の行動目標（学習がどのように行動に反映されるべきか）を明確化することを求めた³⁾。タイラーの行動科学モデルを引き継ぎ、さらに発展させたのがH・タバであり、彼女は、スプートニク・ショックを背景に、学校教育の立ち遅れと過度な管理化を、授業改善を基礎とするカリキュラム開発によって克服しようとした⁴⁾。

タイラー、タバの行動主義的カリキュラム観は、科学的なカリキュラム研究開発の基礎を形成したが、1960年代末になると、それ自体あまりにも精緻化した科学的方法論に対して疑問が呈されるようになる。教育的価値から中立を保とうとする科学主義がカリキュラ

ムと授業を一面的で貧弱なものにしているという批判が、この時期に起こるのは、ほとんど必然的であろう。また、より根本的な問題として指摘された点は、心理学的知見に基づいた行動主義的カリキュラム観では、実際に教室で行なわれる授業がブラックボックスとして扱われてしまうことだった。事前テストと事後テストの結果を比較することによって、授業の効果を計るのが、行動主義的研究の主要な手法だったからである。こうしたカリキュラム研究への批判として生まれたのが、授業に「光」を照射するパートレットとハミルトンの「解明的評価」(illuminative evaluation)であった。これは参加観察法による文化人類学的アプローチであり、以前の研究が「測定と予測」に関心事としたのに対して、「記述と説明」に意を注ぐ手法である。テスト結果の比較という「数字遊び」(numbers game)から脱し、カリキュラムの質的研究へと道を開いたのである(pp.65-66)⁵⁾。

このようにたどってみると、カリキュラムをどう捉えるかは、それぞれの時代の政治社会状況や知的エートスときわめて深く関わっていることが分かるだろう。カリキュラム概念とは、あらかじめ明晰に定義された所与ではなく、教育に携わる者が主体的に構築するものであり、われわれ自身のイデオロギーを反映するものであることを、自戒を込めて思い起こしておきたい。

1-2. カリキュラム概念の構成

上述のようにカリキュラムは、公的に制度化された「教科内容の組織」と、教師と生徒学生が共同して教室の中で創出する「学習経験の総体」という二重の概念を含んでいる。

まず前者の下位概念としては、国の定める学習指導要領、それに則った検定教科書、学校が定める教育目標、またそれに則った科目表、時間割、指導計画などがある。教育学者や教育心理学者など、専門家の言説もそこには介入してくるだろう。カリキュラムのこうした一面の特質としては、体系性、規範性、先行性といった要素が挙げられる。体系性とは、さまざまな科目や人的組織が相互に関連し合いながら、全体として教育目標の実現を目指す機構を成しているということである。また、規範性とは、一定の強制力をもって、学校組織や個々の教職員、生徒学生らがそれに従うべきものとして存在しているということである。そして先行性とは、その「計画」としての性質上、教育や学習の実践より以前に存在しているということである。個々の教員や学生にとっては、それは《つねにすでに》そこにある所与の条件である。こうしたカリキュラム観は、カリキュラムを教育の現場とは離れたところに設定している。

後者のカリキュラム観は、より具体的には、たとえばジョン・デューイが子供中心主義の立場から唱えた「主題・経験・表現」(教師が主題を定め、子供がそれを経験し、表現する)、あるいは大量生産と企業経営をモデルとした「目標・達成・評価」、あるいはまた、批判的思考を備え、社会変革を志向する人物の育成を目指す「問題・思考・解決」といった学習・成長過程として、教育の現場に移し変えられる⁶⁾。この場合、カリキュラムは、まさに「走路」という原義のままに、つねに円環を描いて循環運動するプロセスとし

て特徴付けることができる。

特筆しておきたい点は、いずれのカリキュラム概念にも、評価 (evaluation, assessment) が組み込まれていることである。プラン、プログラムという意味でのカリキュラムの場合は、授業実践が定められた計画に適合しているかどうかがつねに評価される。一方、教育・学習プロセスの総体としてのカリキュラムの場合も、その最終段階で行なわれる評価が次の学習サイクルへと向かう^キ合図となる。計画にせよ目標にせよ、実証からのフィードバックを受けて、更新されなければならないのである。

では、このようなカリキュラム概念の再検討、特に「学習経験の総体」というカリキュラム観は、いったいわれわれに何を教えてくれるのだろうか。

教壇に立つ者であれば誰でも経験することであろうが、学生は教師が語る内容をそのまま、水差しから空のコップに水を注ぐように受け入れるものではない。学生の関心や予備知識などが複合的なフィルターとして働いているはずである。また、学生は教師が語る内容だけを受け入れるものでもない。教師の態度、周囲の学生達の挙措など、複合的で膨大な情報をも無意識であれ処理しているはずである。つまるところ、授業とは、教員が教えることと学生が学ぶこととの間のネゴシエーションの場である。教員と学生の両者を主体として捉えるならば、両者が自動的に一致することは、とうてい想定しえない。つねに潜在的に存在するこのミスコミュニケーションは、授業者としての教員のスキル、学生の側の意欲・志向性・前提知識、そして両者の情意的関係などに支障が発生した場合に、たちまち顕在化するだろう。とりわけ教員側としては、教員が何を教えるかばかりでなく、学生が何を学ぶかを考える必要がある。

しかしながら、この問題を、単に学生の意欲・知識不足や教員の意欲・技術不足に還元することもまたできない。カリキュラムというものを、「学生の学習体験の総体」と定義する以上、実際に学生が習得した知識を基準にして、科目構成と授業内容、すなわち通常の意味での「カリキュラム」を検討・評価することが重要であろう⁷⁾。

1-3. 英語運用能力とは何か

より具体的に英語教育カリキュラム、とりわけ短期大学におけるカリキュラムについて考えてみよう。まずわれわれはその目標について考えなければならない。われわれはふだん、英語教育の目標を「英語運用能力の向上」と考えがちだが、そもそも「英語運用能力」とはいったい何だろうか。

英語の語学力を体系的かつ分析的にとらえようとする試みについては、ロバート・ラドーの功績が大きい。ラドーは、発音・文法構造・語彙・文化的意味といった「言語要素」と、話す・聞く・読む・書く・翻訳という「技能」(skill) という二つの軸で、学力を把握しようとした。ただし、言語要素については、分析の便宜上、個別に挙げられているものの、実際の言語使用の場面では、四技能の中に統合されると説明している。ラドーの二次元的学力モデルはさらに精緻化され、構成要素間の動的関係も理論化された。やが

てR・L・クーパーが第三の軸として「変種」(variety)を加えた。この第三の軸は「社会領域」(social domains)などと呼ばれることもあり、話相手や状況によって言葉遣いが決定される社会言語学的領域である。たとえば、家庭、学校、同年齢の友人、地域社会など、シチュエーションに応じた言語使用ができるかという能力が対象となる。現在では三次元的に理解するのが一般的である。近年の英語教育で「場面・機能シラバス」が精力的に導入されているのは、こうした学力観の反映である。

とはいえ、たとえばリーディング力をさらにその構成要素へと分解するとすると、ほとんどは仮説の域を出ない概念しか存在しない。そこで「言語能力」(linguistic competences)を構造的に把握しようとする観点とはやや異なる観点として、言語使用者が実際にある言語を用いて何をすることができるかの能力を記述しようとする考え方が生まれた。それが「言語運用能力」(linguistic performance abilities)の概念である。J・B・キャロル⁸⁾に始まり、リヴァース⁹⁾、ヴァレットとディシク¹⁰⁾へと受け継がれていく。前述の「記述と説明」に基づくカリキュラム研究方法とほぼ同時期に、こうした考え方が生まれたのは偶然ではないだろう¹¹⁾。言語運用能力がその発展段階別に記述されると、評価の問題と直結することになるが、それについては後述する。

2. 評価とテスト

評価は教員と学生の双方にとって、カリキュラム以上に身近で切実な問題かもしれない。それゆえに今一度その基本的な考え方をおさらいしておく意義はあるだろう。

2-1. 評価論の基本概念

評価とは何かを考える上で、言語テストを例に取り上げてみよう。言語テストの領域では一般に、妥当性(validity)と信頼性(reliability)が評価に際して考慮すべき事項として挙げられる(実現性(practicability)、真生性(authenticity)などの概念を挙げる論者もいるが、前者の意味は明白だろうし、後者については妥当性と区別する必要を認めないので、ここでは触れない)。ただし、評価論という妥当性と信頼性は、日常語とは若干意味が異なるので注意したい¹²⁾。

妥当性とは、測定したい能力と、テストが実際に測定している能力とが合致しているかどうかである。たとえば、発音の力を測定したいのに、受験者に実際に発音させないテストがあれば、それは妥当性に欠けている。妥当性は、このように「測定したい能力」を前提とした概念なので、より正確には「構成概念妥当性」とも呼ばれる。構成概念(construct)というのは、「言語能力に関する理論の中でその存在を仮定されている、目には見えない能力(もしくは特性)」である。語彙力や文脈把握力などの能力は、実体として取り出すことができない能力で、それらに応じた(と想定される)設問アイテムを用意し、実証的に検証するより他ない。その場合に証拠となるのが、妥当性の下位概念である「内容妥当性」と「基準関連妥当性」である。内容妥当性とは、テストの内容が、測定し

ようとする対象の全体から偏りなく抽出されているかどうかに関わる妥当性である。これによって、構成概念の内容範囲が明示され、テストが測るべきものを測っていることが立証される。また、基準関連妥当性とは、別の信頼できるデータと照らし合わせて、結果がどの程度一致しているかによって、あるテストの妥当性を測ることである。これは、基準となるテストが当該テストとほぼ同時に実施されている場合には「併存妥当性」と呼ばれ、後で実施された基準テストと当該テストを比較する場合には「予測妥当性」と呼ばれる。

妥当性の概念は、採点の段階にも適用される。たとえば、リーディングのテストで内容についての質問を英語で書かせたりする場合に、スペリングや文法の誤りを減点対象とするのは、妥当性がない。リーディング力を問うているのに、別の能力で点数を左右させているからである。本来はこうした誤りが出ないように、設問形式を工夫することが望ましい。もっとも、コースで実施される教師の手作りテストの場合には、妥当性を厳密に検証することはほぼ不可能だろう。それでも妥当性を高めるために有効な手段としては、①「測定しようとしている構成概念について現在わかっているすべてのことを考慮に入れたテスト細目規定」を書いてみる、②「可能な限りにおいて常に直接テストングを使うこと」が挙げられる (p.37)。直接テストングについては2-4でも触れるが、上記の例で言えば、発音の力を測定するためには学生に発音させてみるようなテスト方式のことである。

一方、信頼性とは、言い換えればテスト結果の再現可能性である。たとえば、ある受験者がテストを受けたあと、あまり時間が経たないうちに、そのテストの内容はすべて忘れたうえで、もう一度その同じテストを受けたとして、これら2回のテストの点数がほぼ一致すれば、それは信頼性の高いテストである。ありえないことを仮定しているようだが、仮想的には今述べたのと同じ事態を作り出しうることは、信頼性係数計算の方法として後述する。信頼性の概念もまた、採点の段階に当てはめられる。異なる採点者が同じ答案に対して同じ点数をつけたならば、採点者信頼性は高いことになる。信頼性を高めるには、① 解答者が自由に記述・実践できる余地をできる限り少なくすること、② 設問（アイテム）が相互に独立していること（あるアイテムに正解することが別のアイテムに正解する条件とはならないこと）、③ 明確な指示を与えること、④ テスト形式をあらかじめ周知すること、⑤ 厳密な採点基準を設けること、などが挙げられる。

気をつけなければならないのは、妥当性と信頼性は時に矛盾することである。たとえば、採点信頼性を高めるために記号で解答する問題を多くすると、必然的に間接テストングとなり、妥当性が下がる。妥当性と信頼性を共存させるためには、直接テストング・アプローチを守り、構成概念（すなわち測定対象）を的確に分析し、明確な採点基準を用意することが求められる。

さらに客観的な目安がほしい場合には、点数リストが手元にあれば、妥当性と信頼性を数学的に計算する方法もある。現在では表計算ソフトで「相関係数」を算出することがき

わめて容易になっている。相関係数は、二組の数値がどの程度一致しているかを表す値で、最小値0は一致が全くないこと、最大値1は完全に一致していることを表し、通常はその間の値である。たとえば、非常にアイテム数の多いテストを実施したが、アイテム数をもっと減らし、実施時間を短縮してもテストの妥当性は保たれるかどうかを調べたいとする。その場合は、テストを受ける全学生から一定数のサンプルをランダムに抽出し、そのサンプルだけに長いテストを受けさせ、その結果を短縮版を受験した他の全学生の結果と比較すれば良い。また、あるテストの信頼性を測りたい場合には、折半法(split-half method)を用いれば、一回の実施でも測定可能である。一つのテストの点数を、たとえば奇数番と偶数番などでアイテムを二つに分けて計算し、その二組の点数の相関を調べることで、同じ学生が同時に二つの代替可能なテストを受けたものとみなすことができる。その値の意味を感覚的につかむための、ごく簡易な計算方法としては、相関係数を二乗してパーセンテージ表示した値を、大まかな一致度と考えてよい。どの程度の相関係数をもって満足すべき一致度とするかは、テストの目的によって異なる。ヒューズは、クラス分けテストなら0.7、コースの履修成績なら0.4程度で十分と述べている(pp.31-32)。

2-2. 評価の実施形態

評価というと、成績という形で示される教員による学生の評価がまず念頭に浮かぶが、これはカリキュラム全体から見ると、評価の一部分に過ぎない。たとえば、随時小テストを実施して返却し、学生自身に学習状況を伝えるケースも日常的にあるだろう。学習成果に関する評価という話題については、次節のテストの項で詳述することとして、ここではカリキュラムや授業の評価の形態について急いで概観しておきたい。

- ①教員自身による授業評価：コースデザイナーとして、どの程度授業目標を達成することができたかを、授業成果（学生のテストやレポート）から評価する。これはまた、以下のすべての評価においても貫かれるべき視座である。
- ②学生による授業評価：本学も含め、大半の大学短大で実施されている。本学の場合は、希望する教員のみ授業アンケートを実施しており、数値と自由記述の両側面から学生の満足度や改善すべき点などを知り、授業改善に役立てることができる。
- ③教員相互による授業評価：大学レベルでは和光大学の先駆的な試みが報告されている。「同僚性」(collegiality)の重要性が近年認識されつつあることを考えると、今後ひとつの潮流となる可能性もある。
- ④評価の評価：上記の評価は個別の授業改善に直結しているが、科目編成を含めたカリキュラム全体にまで、評価の帰結を反映させるためには、学校全体としての組織的取り組みが当然必要である。

考慮すべき事項として、語学科目は非常勤講師に多くを負っているという現状がある。今後もそれは続くだろう。一方、本学を含め、多くの大学短大では、語学科目に関してきわめて抽象的な授業概要を提示するに留まりがちである。「英語」という実態をつかみづ

らい科目名のもと、明確な指針を与えられないまま、教える側も教わる側も結局は、ルーティーン消化に陥ってしまう可能性は否めない。非常勤講師を大切な教育スタッフとして組織的に位置づけることは重要である。具体的には、① 依頼時に科目のねらいや目標を適切に伝え、必要であれば盛り込むべき授業内容を説明すること、② カリキュラムや授業のあり方について意見や不満を収集するなど、カリキュラムと評価の円環に参画してもらえる体制を整えることが挙げられる。

教育・学習過程における評価の役割はきわめて大きい。前述のような各種評価が適宜実施され、それぞれが相互に、また個々の授業とも、関連付けられた場合には、きわめて健全な授業運営が保障され、カリキュラム改善への展望が開けることになるだろう。

2-3. テスティングの考え方

とはいえ、授業担当者としてはやはり、自分の実施しているテストが有効なものであるかどうかが気になるところである。統計処理まで含めたテスト理論は、筆者のような素人の目にはあまりにも玄妙な研究領域にみえる。TOEFL などのような大規模テストの開発に関わるわけではないなら、ごく基本的な考え方のみ理解しておけば、十分に実用的だろう。逆の見方をすれば、日常的なテストでも満たすべき要件はあるし、また英語教師として語学科目を担当する以外でも、入試をはじめ、さまざまな形態のテストと関わる場面はある。ある程度はテストの考え方を知っておいて良い。

テストの基本的な考え方では、測定対象と測定目的（どんな能力を何のために測定するのか）を明確化することが、何よりも重要とされる。まず測定対象について言えば、たとえば「英語力」と一言で呼ばれる能力を、より詳細な構成要素へと、少なくともいったんは（後で統合的にとらえなおす必要はあるにしても）分析するのが手順である。本格的なテスト開発を進める場合には、測定すべき要素をリストアップした「細目規定」を作成することが必要となる¹³⁾。

一般の語学教員に手が届く範囲で考えると、テストが満たすべき条件としては、「測る側にとって関心のある能力を正確に測定した結果を一貫して与えてくれる」、「(テストが指導に影響する状況であれば) 指導に対して有益な影響を持つ」、「時間・費用の点で経済的である」(p. 8) といったことが挙げられる。

テストを実施する目的としては、「言語の習熟度を測る」、「授業の目標を学生がどの程度達成できたか確かめる」、「学生の得意な部分と弱い部分、知っていることと知らないことを診断する」、「ある言語プログラムのどの段階、どのパートが学生の能力に最もふさわしいかを見極め、学生のクラス分けに役立てる」といったものが考えられるだろう (p. 9)。これらの目的に沿ったテストについては、次節で触れる。

テストが教育・学習プロセスにどのような影響を及ぼすか、あるいはテスト結果をどのように利用するかも、重要な問題である。テストや評価が、学生を挫折させることを目的としていないのは当然であるが、それらの結果が次の学習段階に結びついていない場合、

学生の立場からは、ある種の「判決」と見えてしまうのはいたし方ない。可能な限り、テスト結果を次の学習段階への準備として位置づけることが望ましい。また、測定目的とは全く関係のないところで、テストの形式などが、学習姿勢に「波及効果」(backwash)を及ぼすのも、よく知られた現象である。波及効果には有益なものと有害なものがあるので、なるべく有益な波及効果をあたえるよう配慮することが望ましい。

2-4. さまざまなテストの種類

実際に試験問題を作成するための手引きとして、さまざまなテストの分類法を以下に紹介しておこう。小見出しでは、どんな観点からテスト・評価方法を見た場合の区分法であるかを示した。

■授業・学習プロセスとの影響関係

形成的評価 (formative assessment)：学習の進み具合を知り、その結果を以後の授業・学習計画に反映させるために実施する。学生が自分の学習成果を集めるポートフォリオを作成するのも役立つ。自己評価 (self-assessment) が奨励されることもあるかもしれない。

総括的評価 (summative assessment)：学期末・学年末に、学生の力が（個人として、もしくは集団全体として）どの程度伸びたかを測定するために実施される。

■評価の基準

集団参照的テスト (norm-referenced test)：一人の受験者の得点を他の受験者の得点と照らして評価するためにデザインされたテスト。受験者を集団の中で順位付けしたり、その偏差値を算出したりするのに利用する。入学試験などがこれに相当する。ただし、このやり方では、個々の受験生が英語を用いて何ができるかを知ることはできない。

基準参照的テスト (criterion-referenced test)：「目標参照的」と翻訳されることもある。受験者が英語を用いて実際に何をすることができるかを知るためにデザインされたテスト。何らかのタスクを満足できる水準で遂行できたか否かによって、受験者は可否に二分されたり、成績に応じた「帯域」(band) に分類される。実用英語技能検定 (英検) やイギリスの IELTS などがこの方式に分類できるだろう。

両者は、相対評価と絶対評価と呼ばれる評価方法に近いが、「わが国で絶対評価といった場合、客観的な実態把握に基づかないで、評価者が主観的に任意に定めた基準に基づく評価をいうかのような誤解がある」¹⁴⁾。つまり、基準参照的テストの場合には、客観的な基準があらかじめ定められ、受験者にも明示されていることが前提とならねばならない。

基準参照的テストには「(1) 実際に何ができるかに関して意味のある基準を設定し、この基準は受験者集団が変わっても変わらない。(2) 生徒はこの基準をクリアしようと努力する気になる」¹⁵⁾ という長所がある。

■テスト開発・作成の際のアプローチ

直接テストイング：測定したい技能を受験者に直接使わせて、何らかのパフォーマンス

を行なわせるテスト方法。「話す・書く」という産出的技能を測定する場合には、比較的实施しやすい。

間接テスト：測定したい技能を、別の何らかの手段によって、測定しようとするテスト方法。たとえば、韻を踏む単語のペアを選ばせることで発音の力を測ったり¹⁶⁾、文中のいくつかの下線部の中から、正用法としては誤りであるものを選ばせることで、ライティング能力を測ろうとするようなアプローチである。

■テストの目的

熟達度テスト (proficiency test)：受験者がどのような授業を受けたかとは無関係に、その言語の熟達度を測定する。

到達度テスト (achievement test)：個々の学生、クラス全体、そして授業自体が、どの程度目的を達成できたかを見極めるために行なわれる。授業日程のさなかに行なわれる形成的評価テストと、授業日程をすべて終えた後に行なわれる総括的評価テストが含まれる。ただし、この場合、テストの内容をどうすべきかについて、二つのアプローチ方法がある。一つはシラバス内容準拠アプローチ (syllabus-content approach) で、テストの内容は授業シラバスと実際に使用された教材に直接基づくべきだという立場である。もうひとつは、コース目標準拠アプローチである。これは「基準参照的テスト」のアプローチと同じで、テスト内容を、あくまで明示されたコース目標に準拠しているかどうかという観点から選択し、むしろ授業で使用した教材からは採用しない手法である。一般には前者を「公平」なテストとみなす語学教師が多いが、ヒューズは、本来、後者こそ妥当性の高いアプローチであるとして、こちらを強く推奨している。

診断テスト (diagnostic test)：四技能など、ごく大ざっぱに分類された能力を診断する場合には、既存の熟達度テストで代用可能である。しかし、ある学生がどの文法項目を理解していないかというレベルの測定になると、あらゆる文法項目を網羅した膨大な数のアイテムが必要となる。信頼性、実現性ともにきわめて低いため、純粋に診断を目的としたテストは、現実にはほとんど作成されていない。そうした中、興味深い試みの一つとして、欧州委員会のサポートで開発運営されている多言語教育プログラム DIALANG がある。インターネット上に公開されているサイト (<http://www.dialang.org/>) では、英語を含む14のヨーロッパ諸語について「読み・書き・聞く」の三技能と文法・語彙の力を診断するサービスを提供している。

レベル分けテスト：学生をある授業プログラムの中で最も適切な段階に配置するために役立つ情報を得るためのもの。ただし、個別の授業プログラムに準拠して作成されたテストでなければ、適切な役割を果たさない。原則として、教育機関において自作する必要があり、市販テストではその用を満たさないと考えるべきである。

■言語要素の扱い方

個別項目テスト：言語要素を一度に一つずつテストする方式。一連のアイテムがそれぞれ特定の文法項目を問うているような問題である。

統合的テスト：受験者が多くの言語要素を組み合わせて一つのタスクをやり遂げることを求める方式。講義を聞きながらメモを取ったり、まとまった文章を書いたりする形などが考えられる。のちほど説明するクローズ法のテストも、受験者が文法・語彙・文脈などさまざまな言語要素の知識を組み合わせて解答する必要があるので、統合的なテスト方式である。

■採点法

主観テスト、客観テストという言い方があるが、この区分はやや誤解を招きやすい。「客観的と言えるのは、テストの『採点』だけなのである。」¹⁷⁾ 通常は、記述式問題を主観テスト、記号解答式の問題を客観テストと呼ぶが、こうした解答形式そのものは、テストの妥当性にも信頼性にも関係がない。記述式の場合には、確かに採点者の判断する余地が大きくなるので信頼性は下がる可能性があるが、記述式の場合でも、さまざまな手段で信頼性を保つ手段はある。あまり意味のない区別であると言ってよいだろう。

■クローズ法 (cloze procedure)

リーディング力を測定するテスト形式としてかなり有効と目されているテスト形式。通常の言語情報には、余剰性 (redundancy) がある。メッセージの内容を理解するのに必要な量を越えた情報がわれわれには届いているということだ。だからわれわれは母語で他人の話を聞いたり、文章を読んだりするとき、雑音があつて言葉を聞き逃したり、未知の単語に出会ったとしても、内容を理解するのにあまり苦労しない。ちょうどゲシュタルト心理学のテストで地と図柄とが相補的に作用するように、減らされた余剰性 (reduced redundancy) は他の言語情報によって補われるのである。ところがこうした作用は、外国語になると困難を覚える。こうした現象を利用して、テキストから削除された余剰性をどの程度復元できるかによって言語能力を測定するのがクローズ法である。表面的には空所補充問題の形をとるが、文法や語彙の能力を測定するためのそれとは、全く性質が異なり、クローズ法は、文化読解力 (cultural literacy) まで含めた、統合的なテストである。具体的な作成方法は、パッセージの冒頭は「導入部」としてそのまま残し、その後はたとえば7語ずつ空欄にしていく。それだけである。100語程度のパッセージで10の空欄ができることになる。使い方としては、何のヒントもなくそのようなパッセージを出題することもあれば、かなりの語数がある長文の前や後に、100語程度の長さの要約文を掲げて、その要約文にクローズ法を適用することもある。復元不可能と思われる語にヒントを与えるべきか、採点時には正答にどの程度の幅をつけるべきか、など二次的な問題もあるが、ここでは詳述する紙幅はない。

すべての目的に適うテストは、原理的にありえない。ただし、今日のテスト理論を踏まえた最先端の試みはある。2001年に欧州評議会言語政策部門が公刊した『諸言語のための欧州共通の参照枠』は、現在考える限り、言語テストの理想に最も近いところにあるプランと言えるだろう。妥当性と信頼性を両立させつつ、直接テスト・アプ

ローチを用い、受験者への「良き波及効果」も持ちうる言語テストのあり方を追求する試みが、言語と文化の多元主義を最も切実に必要としている欧州連合において進められているのは興味深い。診断テストの項で紹介した DIALANG は、その一環として提供されているプログラムである。基準参照的テストは基準設定が成功の鍵であるが、『・・・参照枠』のそれは、きわめて水準が高いと言えるだろう¹⁸⁾。

3. 青山学院女子短期大学英文学科の英語科目とテスト

最後に、青山学院女子短期大学英文学科の英語運用能力養成科目の現状を紹介し、実際にどんな試験問題が出題されているかを切り口として、担当教員の学力観や授業のあり方について考える手がかりとしたい。

3-1. 本学英文学科英語教育カリキュラムの現状

まず青山学院女子短期大学英文学科のカリキュラムを簡潔に紹介しておきたい。本学英文学科は、英文学専攻と英語学専攻の二専攻からなり、一学年の定員は現在それぞれ180名、120名である。いずれの専攻の授業科目も、英語運用能力養成科目と専門科目からなり、専門科目としては、英文学専攻では一年次に英米文学史、二年次に英米文学研究などの文学関連科目が、また英語学専攻では一年次に英語学概論と英米史、二年次に英米文化研究など地域文化関連科目を置いている。英語運用能力の養成により直接関わる科目としては、外国人教員組織による統合的英語教育プログラムと、日本人教員による科目（英文講読（一・二年次）、英作文（一年次のみ））の二種類がある。専攻分離当初から長い間、外国人教員担当科目の数などに違いが設けられていたが、現在はそれらの科目も、授業内容・時間とも、両専攻に全く同じように配置されている。

名称こそ古色蒼然たる観を呈しているものの、英文講読は、英文学科ではいまだに、日本人教員が携わる重要な英語教育科目として位置づけられており、その内実についても各教員がそれぞれに工夫を凝らしてきた。一年次の「英文講読Ⅰ」では、2004年度より、①クラス編成を工夫することで、通常クラス（英文学専攻4クラス、英語学専攻3クラス）の2/3程度にクラスサイズを小さくして、②その10クラスすべてを専任教員が担当して、学習アドバイザー的な役割も担うこととし、さらに④授業の目標と内容を共通化する（具体的には、学生向けの冊子『講義内容』に掲載する科目の「ねらい」に同一の文を用い、「授業内容」についてもガイドラインを設けた）ことで、担当者が異なっても一定の目標を達成する足がかりとした。また、（少なくともわれわれにとっては）やや実験的な試みとして、複数のクラスで同一教材（市販テキスト）を用いて、随時担当者同士が各クラスの具体的な授業内容について情報交換することを申し合わせた。

実際に一年間の授業を終えた後の評価としては、これらの試みがすべて、期待したように機能したとは必ずしも言えず、反省点も多い。ここでは、この科目の後期試験として各担当者が出題した問題を紹介することで、当該科目の問題点や今後の課題を考える手がかり

りを供したい。筆者を含めた5クラス、全10クラス中半数が対象である。この場をお借りして、ご協力いただいた同僚諸兄姉には心から感謝申し上げたい。

3-2. 「英文講読Ⅰ」試験問題の形式

同じ名称の科目を担当している他の教員がどんな授業を行ない、どんな試験を実施しているのかは、関心はあるもののこれまでなかなかそのような機会がなかった。今回5クラス分のテスト問題を比較してみて、まず感じたことは、授業の目標、授業内容の骨子、そして教材まで共通化しても、試験問題はまことに多様であるということだった。一様であるべきだという含意は全くない。理想的な唯一のテスト形式という夢は、実現性の観点から、われわれには無縁であると考えてよい。それぞれの試験が異なる時間帯に実施されるため、同一問題を使うことができないという実際上の事情もある。それぞれのクラスで異なる教育・学習経験が積まれた、その反映として、試験問題を見てみたいのだ。以下になるべく客観的に各試験問題の形式をなぞりつつ、コメントを加えていこう。なお、5クラスのうち、2クラス（仮にA Bとする）がX社のテキストを、3クラス（仮にC D Eとする）がY社のテキストを使用した。

Aクラス：20の設問。指示文を含めてすべて英語でタイプされている（なお、他のクラスはすべて指示文は日本語）。テキスト1レッスン分の内容について設問が二つずつあり（10レッスン分が試験範囲）、一つ目が与えられた英文の正誤を記号で書き込ませる問題、もう一つが英問英答。ただし、後者は解答文の書き出し3～8語程度があらかじめ与えられている。両者の間に直接の関連性はない。テキストの内容は、学生が覚えていることが前提となる。

Bクラス：大問七つ。1)はテキストから抜粋された短文（10問）の空所に、与えられた語群（12語）から適語補充する問題。2)も同様の形式だが、与えられた語群は前置詞・副詞である。3)も短文の適語補充（5問）だが、これは語群に動詞が並び、必要なら活用させる記述問題。4)は、2文程度の文章の一部、約7語の整序問題。5)は、7つの文を並べ替えてパラグラフを作る問題（文の整序問題）。冒頭の一文は与えられている。6)では、異なるレッスンから2パラグラフが抜粋され、それらについて空所補充（記号で解答）や、文中の語句を二つの選択肢から一方を選ばせる問題（例：“a [A.near/B.nearby] hospital”）など（計15問程度）。7)は、3行程度の短い抜粋から15語程度の短文を英文和訳させている。

Cクラス：大問四つ。1)は、百語程度の英文を日本語で要約させる問題が2問（字数制限はなし）。2)は、四つの引用文（いずれも40語程度の短い文章）から一つを選び、その内容についての「考え」を日本語で述べさせる（解答欄をすべて埋める必要はない旨、注記あり）。3)は、10語前後の短文の一部に下線が引かれており、そこを「文法的に説明」する問題が5問。4)は、文中の語の派生語や反意語を書かせる語彙問題（15問）。

Dクラス：大問四つ。1)は、50語前後の引用文中1/3程度を英文和訳させる問題が7

問。2)は、1の引用文のうち3つについて、その言葉を言った(書いた)人物の名前を英語で答えさせている。3)は、80語程度の引用文の内容について、日本語の設問に日本語で答えさせる問題が2問。4)は、日本語で与えられた単語を英訳する語彙問題(15問)。

Eクラス：大問五つ。1)は適語補充。10語以下の短文16文について、それぞれ選択肢が四つずつ与えられており、そのうち一つを記号で選ぶ。2)は、イディオムの同義語を語群から選択する問題(6問)。イディオムは文中に下線部を引く形で示されている。語群選択肢は12語用意されている。3)は、約90語の引用文について、指示語の指示対象を答えさせたり、空所に埋めるべき前置詞を選択肢から選ばせたりする問題(7問)。倒置文を通常の語順に書き直させる問題も1問添えられている。4)では、15語程度のセンテンスについて主語や目的語を挙げさせて、統語分析を求めている(4問)。5)は、250語程度のまとまった文章を掲げ、その内容について英問英答問題(5問)。ただし解答文は四つの選択肢が与えられており、その中から記号で選ぶ形になっている。

これらの試験問題を踏まえつつ、それぞれの設問形式(test technique)についてコメントを加えておきたい。

多肢選択問題は、当て推量での解答が可能であること、妥当な錯乱肢の作成に労力がかかること、などなどから、作成者への負荷が高い問題形式とされている。今回検討した試験問題では、錯乱肢がよく配慮して作られている印象を持ったが、当て推量の可能性はどうしても残る。これは、「誤答は減点する」と設問の指示文でアナウンスすることで、避けることができるだろう。

空所補充問題は、リーディングのテスト・テクニックとして問題を作りやすく、かつ妥当でもありうる(クローズ法テストの項を参照)。ただし、語群から選択肢を選ばせるのは、アイテム同士が干渉する可能性があるため、問題がある。そうした形式の問題は、語彙習得のための練習課題としては有効かもしれないが、個々の語彙・語法・文法などの知識があるかどうかを測る試験問題としては不適切だ。一定数のアイテムに自信を持って解答すれば、残りのアイテムに関するヒントが出てくることになり、結果的に点数が高いものと低いものとに偏る可能性が高い。他の問題形式についても当てはまることだが、個々のアイテムはできるだけ独立していることが望ましい。

英文和訳問題については、訳す文が比較的短いこと(語彙・文法事項など問いの内容的妥当性)、採点基準が確固としていること(採点者信頼性)が考慮されていれば、有効な形式である。ただし、これらの条件を満たしていない場合は、テストの妥当性・信頼性ともに問題があり、日本語に置き換えないと英文を理解したと思えない習慣を身に付けさせてしまうなどの、悪しき波及効果を学生に及ぼす可能性も高い点には留意したい。

内容に関して日本語か英語で答える問題については、一般的に、設問自体になるべく多くの条件を付けることで、解答者の自由度を下げるのが望ましいとされる。解答の際にライティング力が重要なファクターとして介在してしまうこと、採点者信頼性を保つのが

困難になることが、その主な理由である。今回検討した試験問題では、ほぼワン・センテンスで解答できるように書き出し部が与えられるなど、それぞれに工夫されていたが、さらに短い解答を求めたほうがリーディング科目の試験問題としての妥当性は高いと思われる。解答が一つに絞れる英問英答形式や、空所補充形式をうまく取り入れることで、そうした難点は解決されるだろう。日本語での内容説明を求める問題は、さらに論争的かもしれない。採点基準の問題もあるが、それ以上に波及効果の点で問題があるだろう。こうしたテストに慣れた学生は、英文を読むたびにその内容を日本語に置き直す習慣を身につけ、そのために迅速な読みを身に付ける動機付けが弱くなる可能性があるからである。

冒頭にも述べたとおり、これらの試験問題の優劣をあげつらうことが目的ではない。テストという切り口から、各担当者が学生を評価する際の観点ばかりでなく、それぞれの授業のあり方や学力観などもうかがい知ることができることを確認しておきたい、というのが主眼である。各授業を通しての学習達成度を測定するのが学年末試験である以上、授業と対応したテストが実施されることを目指せばよい。

とはいえ、経験や知識の不足、あるいは授業形態のマンネリ化など、さまざまな事情で良い試験問題が作れないでいたり、作れなくなったりすることもあるだろう。今回は筆者の一方的なコメントという形をとったが、たとえば今後、同じ科目を担当する教員有志が、それぞれの試験問題や評価方法を持ち寄って、相互に啓発する機会を設けることはきわめて有意義であろう。もっとも、小中高の教員であれば、当然のこととしてこうした情報の交換と共有を行なっているはずである。テストの作成・採点という、学生にとってもきわめて重要なフェイズにおいて、教員の「同僚性」を高めていくことは、カリキュラムの研究と改善をはかる上でも大いに役立つことが期待される。

むすび

テストの理論と実践について触れたものの、たとえば40名程度の学生からなる授業について、過度にテストの精度を気にすることは、全く現実的ではない。学年末試験（最終到達度テスト）の点数について学生が重大な関心を寄せるのは、単位を取得できるかどうかだけであろう。また、出された成績が利用される機会自体ほとんどない。さらに語学科目は、学生個人にとっては多々ある授業のなかの一科目に過ぎず、たとえば海外留学を目指す際に学業平均値（grade point average）を算出するとしても、他の諸科目の成績と合算されてしまう。学期・学年末に教員が提出する成績というものを、純粹にその利用価値から考えるなら、まことに空しい営為とすら言ってよい。数少ない成績評価の利用価値とは、学生の学習方針に良い波及効果を及ぼすことと、より妥当性と信頼性の高い評価を下そうとする努力の中からより良いカリキュラム編成を生み出していくことである。カリキュラム改革にせよ、評価の精度向上にせよ、教育理論のような外的要因によって果たされることはない。それらは、実際に教場に立つ教師が、教場からのインパクトを反映しながら、進めるより他ないのである。

注

- 1) 佐藤学『カリキュラムの批評——公共性の再構築へ』（世織書房、1996）、p.4。なお以下のカリキュラム論は同書に大半を負っている。以下この節で、引用文の後などに、単に参照頁が示されている場合には、この本への言及である。
- 2) Harold Rugg, *Foundations for American Education*, 1947.
- 3) Ralph W. Tyler, *Basic Principles of Curriculum and Instruction*, 1949.
- 4) Hilda Taba, *Curriculum Development: Theory and Practice*, 1962.
- 5) M. Partlett and D. Hamilton, “Evaluation as Illumination: A New Approach to the Study of Innovatory Programmes”, 1972, in D. Hamilton (ed.), *Beyond the Numbers Game*, 1977, pp.6-22.
- 6) 佐藤、pp.9-11.
- 7) 近年、個別の授業に関する「コースデザイン」という考え方と作業が重視されつつある。授業の教育目標を明確化し、教員と学生はその達成に協力し合わなければならない、そのためには特に、教員側の教育意欲・教育スキルの向上が必要とされている。2004年度「特色ある大学教育支援プログラム」に採択された名古屋大学の取組み「教員の自発的な授業改善の促進・支援—授業支援ツールを活用した授業デザイン力の形成—」は、その成果の一端がウェブ上でも公開されており、参考になる（「成長するティップス先生——名古屋大学版ティーチングティップス——」<http://www.cshe.nagoya-u.ac.jp/tips/>）。
- 8) John Bissell Carroll, “The Psychology of Language Testing” in Alan Davies (ed.), *Language Testing Symposium: A Psycholinguistic Approach*. Oxford University Press, 1968.
- 9) W. M. Rivers, *Speaking in Many Tongues: Essays in Foreign Language Teaching*. Newbury House, 1972.
- 10) R. M. Valette and R. S. Disick, *Modern Language Performance Objectives and Individualization: A Handbook*. Harcourt, Brace, Jovanovich, 1972.
- 11) 以上の英語力観の歴史的変遷は、垣田直巳監修、青木昭六編集「英語教育モノグラフ・シリーズ」『英語の評価論』（大修館書店、1985）の第2章「英語学力の本質」I節（田中正道分担執筆）によった。
- 12) 以下は主に、アーサー・ヒューズ『英語のテストはこう作る』第2版、静哲人訳（研究社、2003）に拠りつつ、『英語の評価論』第一章II節（萬谷隆一分担執筆）を参考にした。以下この節で、引用文の後などに、単に参照頁が示されている場合には、ヒューズの著書への言及である。
- 13) ヒューズ、前掲書、第7章「テスト開発の手順」を参照。以下本節で出てくる頁数も同書への言及。
- 14) 『英語の評価論』、p.110.
- 15) ヒューズ、前掲書、p.23.
- 16) R. Lado, *Language Testing: The Construction and Use of Foreign Language Tests*. Longman, Green, Co., 1961. 同書に挙げられている有名な（ないし悪名高い）例。
- 17) J・B・ヒートン『コミュニカティヴ・テストング——英語テストの作り方』語学教育研究所テスト研究グループ訳、土屋澄男、齋藤誠毅監修（研究社、1992）、p.30.
- 18) 原著 *Common European Framework of Reference for Languages: Learning, Teaching,*

Assessment は、ケンブリッジ大学出版局から出版されているほか、http://culture2.coe.int/portfolio/documents_intro/common_framework.html に直接アクセスするか、欧州評議会（Council of Europe）のホームページ（<http://www.coe.int/>）から “Language Policy Division” のページへたどっていくかすれば、pdf ファイルを入手可能である。