

大学英語オーラル授業における形成的評価の試み

田中順子 (神戸大学国際文化学部助教授)

大学英語オーラル授業における形成的評価の試み

田中順子（神戸大学国際文化学部助教授）

1. はじめに

本稿では、形成的評価を当初取り入れることを目的としていなかった大学共通教育の外国語授業（英語オーラル）において、偶発的に授業改善の必要が認知され、教師自身が評価者として形成的評価を取り入れた結果、教師側と学習者側双方に有益な傾向が見られたことについて報告する。また、形成的評価の導入によってそのような有益な結果が生みだされた理由を、先行研究を参考にして分析を試みるものである。本稿では、特に形成的評価実施のプロセスと、研究結果のなかからリスニング能力との関係について報告する。

2. 形成的評価とは

2. 1 形成的評価の定義

形成的評価(Formative Evaluation, Formative Assessment)とは、教育プログラムの向上を意図して、そのプログラムの計画、立案、実施の過程において、意志決定の助けとなる情報を得るために行うものである(Flagg, 1990, p. 25)。授業においては、1年間あるいは半年間といった全授業期間中にわたって多くの機会を使って、教師が学生の学習過程を評価するものである(Boston, 2002)。そして、評価を通して得られた情報に基づき、教師と学生の双方にフィードバックを行い、授業を効果的にするための改善に生かしていくのである。この形成的評価は、すでに起こった学習について一連の授業の終了後に試験を実施する総括的評価(Summative Evaluation, Summative Assessment)とは対局に位置するものである(Boston, 2002)。ここでいう「評価」とは教師や学生が教授や学習における変化を目指して行う活動全てを広く指し、観察、クラス内ディスカッション、学生の宿題やテストの分析などをも含む(Black & William, 1998b)。評価の相互性や客観性、またその対象範疇は様々である。

形成的評価は、「中間的段階での暫定的な評価」であり、「ある目標達成のためのプロセスの中に組み込まれた評価」である(陣川, 1990, p. 11)。教師は、形成的評価をすることで、学生を通して自分の授業についての情報を得て自己評価を行い、「自分の教育活動を『自覚』し、活動の過程全体を『意識的に評価』し、その評価をもとに自主的にその活動を『改善』しようとし」(安彦, 1990, p. 231)、自分の授業の「見直し」を行う(陣川, 1990, p. 11)。教師が新しい教育アプローチを試みた場合には、形成的評価は特に有益であるといえる。新しい教育方法の導入においてはしばしば微調整を必要とするのだが、形成的評価の実施によって、必要な修正が必要とされるときにタイミング良くフィードバックが教師に与えられることになる(Horwitz, 1985, p. 83)。このように、評価で得られた情報が、学生のニーズに合うように教育や学習を柔軟に改変するために用いられた場合において、その評価が形成的であるといえる(Black & William, 1998b)。

3. 形成的評価のプロセス

では、形成的評価のプロセスとはどのようなものであろうか。唯一無二の標準的な手法が存在するわけではなく、評価者が評価対象の特定の教育プログラムに合わせて変えるのが通常である(Horwitz, 1985)。実践的な視座に基づいたプロセスの概略を以下に示す。

教育目標の具体化や問題や疑問点の生成

形成的評価の実施

授業法（過程・携帯・技術）の工夫と実践

(陣川, 1990, p. 9; Horwitz, 1985, pp. 83-84)

上に挙げた3項目以外に新たな項目を追加することも可能であろうし、上記の3項目をさらに細分化することも可能であろう。しかし、本稿では上記3項目からなる簡潔なプロセスで形成的評価をとらえることとする。それぞれについて以下に説明を加える。

3. 1 教育目標の具体化や問題や疑問点の生成

目標の具体化の実際について、陣川(1990)は次のような方法を提示している。生徒に指導する学力としての技能を「学習技能」とし、該当する項目を掲げ、それらの下位範疇に属する項目をリスト化するというものである。¹

3. 2 形成的評価の実施

評価方法としては、(1) テスト法(記入法、選択法)、(2) 自由記述法、(3) 短文法(2-3結合文)、(4) 記述法(キーワードの取り出し)、(5) Semantic Differentiation Method (SD法、意味微分法)、(6) 試行学習法、(7) アンケート法、(8) 参加観察法、(9) 構造化されたインタビューなどが掲げられる(Horwitz, 1985, pp. 85-87; 陣川, 1990, p. 11)。また、評価の方法を評価の性質や評価の対象の違いによっても分類することができる。評価が相互的なものかどうかによって相互評価法と自己評価法とに、評価の客観性を基準に絶対評価法と客観評価法とに、また評価の対象範疇によって認知面評価(できるかどうか、わかるかどうか、など)と情意面評価(例えば、「うれしい」などの感情)とに分類される(陣川, 1990, p. 11)。

形成的評価を実施したならばフィードバック(生徒を通して得られた情報)を手がかりに、自分の授業を見直すことが次のプロセスとなる。

3. 3 授業法の工夫と実践

形成的評価実施の次の段階として、形成的評価で得られた情報をもとに、すでに行ってきた授業方法に修正を加える。修正を加えるのは、授業の過程、形態や技術面といったいろいろな方面が考えられる。

補充・進化型A: 各個人の目標達成度によって、それぞれが異なった学習活動を行う、

補充・進化型B: 目標に達していない生徒を到達している生徒が援助していく学習の方法、

自主的学習: 学習者自身が課題を選択し、課題毎にグループをつくって学習を行う、

授業方法の変化: 一斉授業の形態をとるものの、その中の授業方法に変化をもたせ、小集団活動をさせたり、教育機器を使用したり、実験を行ったりする。

(陣川, 1990, p. 13)

¹ 学習技能の具体例として、陣川(1990)は文学的文章の読みの段階の指導をとりあげ、具体的な学習技能として、「読み通す、語句を知る、問題を持つ、吟味する、まとめる、読み味わう、感想をまとめる、振り返る、発展する」の9つを掲げている(p. 9)。

3. 4 形成的評価の主体

形成的評価の実施において重要な問題は、誰が実施するのかということである(Horwitz, 1985)。外部評価による教育プログラム評価や、同僚教師における相互評価の場合のように、授業担当者以外が評価者となることが考えられる。あるいは、授業担当者である教師自身が自分の授業の問題点を改善する目的で、いわばアクション・リサーチの一形態として形成的評価を行う場合もある。後にも述べるが、誰が評価の主体であるにしても、評価結果を教育プログラムに反映していくという性質上、評価対象となる授業の担当者、つまり授業者が、形成的評価の実施に同意していることが肝要になるのである。

4. 学習理論との関連

4. 1 学習理論上の意義

形成的評価は、教師の自己評価のみならず学生の自己評価をも意味するといえる(安彦, 1990)。形成的評価は教師への自己評価を意味するものと考えるのが一般的であろうが、形成的評価の実施によって教師の側のみならず学生の側にも自己評価が認められる。教師は形成的評価を自分で実施する際に、授業改善のために自分の授業方法に対するフィードバックを得て自己評価を行う。一方、学生(学習者)は、自分の学習についてのフィードバックを得て自己評価を行い、その結果学生は「自己教育力」(p. 235)を得る。

一方、学習者は教師が実施する評価を通じて自己評価を行い、現在の自分の能力で可能なことと可能でないことを知ることが出来るようになる。これは第二言語学習理論において“noticing the gap”、つまり学習者が自分の現在の知識体系と目標言語の体系との間にある隔たりに気づくこと、が学習のために必要だとされていることと共通している。ここでいう「自己教育力」にとって、自己評価の正確さ、つまり「自己評価能力」の的確さは、必要不可欠な要因とならずである。そのためには評価の基準が正確でなければならない。

4. 2 形成的評価の効果

Black & William (1998a)は、単行本や過去9年間分の160の学術雑誌に掲載された250編の論文を用いて、形成的評価の教育的効果についてのメタ分析を行った。その結果、形成的評価を実施することで学習が向上したという分析結果を得た。Black & William (1998b)によると形成的評価を実施することによる効果の大きさ(effect size)は0.4から0.7の間であった。この効果の大きさから判断すると形成的評価は学習に比較的效果があると言える。ただし、形成的評価導入による効果のみならず、教育の効果が現れるのには時間が通常かかるため、形成的評価を導入したことで正の効果がすぐに現れると期待してよいものではない。

5. 英語授業での形成的評価の導入

5. 1 導入までの経緯

本稿で報告する形成的評価は、共通科目としての英語オーラル科目において実施したものである。過去にも、アメリカのテレビニュース番組を題材としたテキストを用いてこの科目を実施しているのだが、以前から使用していたテキストに替えて、この授業の学期より同様の内容を扱った新しいテキストを用いることにした。そのため、学生にとってこのテキストを用いた授業の難易度がどうであるのかを筆者が過去のクラスと比較して評価することが困難であった。また、授業当初において学生の授業参加があまりなかったこともあって、以下のような問題があるのではないかと考えた。

- (1) テキストの難易度が高い
- (2) 授業方法が学生の能力に適合していない
- (3) 学習に積極的に取り組もうという意欲が学生にない

上記のような問題が存在するとすれば改善する必要があると考えたことから、筆者が形成的評価を授業者兼評価者として導入することにした。

5.2 実施対象クラスについての詳細

実施時期は2003年度前期である。科目は前期の英語オーラルで、学生にとっては入学後最初に受講した複数の英語授業のうちの1つであった。対象クラスの学生は文系学部Aに所属している1年生39名(女子25名、男子14名)である。学習者背景についてはアンケートを、授業開始日(2003年4月17日)に実施し、外国での居住歴、英語を媒体とした教育歴、英語能力試験受験歴、留学する意志、特に伸ばしたい英語能力などについて調査した。アンケートは、潜在的に英語能力が高くなる要因を持った学生や英語学習に高い動機付けをもった学生を割り出す目的で行われた。しかし、アンケート実施時に英語能力テストは実施しておらず、授業開始時において彼らの英語能力については不明であった。入学後すぐの授業であったため、同僚から対象クラスの成員の英語能力について情報を得ることも不可能であった。授業のテキストはアメリカのプライムタイムに放映されているテレビニュース番組を題材としたものであり、当該学期より新規に採択したものである。この授業での目標は、

1. 現代英語表現を知る、
2. 聞き取り能力を向上させる、
3. 大学生として知っておくべき社会・経済問題を知る、
4. 報道内容に対して自分の意見を持つ、
5. 自分の意見を表明する。

の5項目を設定していた。1と3については当該テキストを使用することで、学生は取りまざるを得なくなると考えた。

5.3 問題のあらわれ

5.3.1 授業方法と学生の能力の不適合と積極的な授業への取り組みの欠如

これは筆者個人の授業設計方法かもしれないが、何学部を担当するかによって、どのような授業活動が出来るかを推測する。例えば、その学部内の専門が英文学のように大学在学中に英語を必要とする可能性が高いものであれば、入学者も英語能力が高い者が多いであろうと予測するのである。しかし、これは非常に大雑把な判断基準であり、予測がはずれることも少なくない。実際には予測した学生の能力レベルや英語への関心と実際の学生の能力レベルや関心度合いがずれることがある。

予測にずれが生じる別の要因もある。学部の中には入学試験の際に専攻別に募集をするところもあるが、2年への進級前に専攻が決定するところもある。一方、2年への進級前に専攻が決まるのであっても、学部全体がおしなべて英語能力や関心が高い入学者によって構成されているところもある。このような理由で、筆者の経験では、授業を取りに来る人間の英語能力や英語への関心といった学習者背景が学部名以外にはわからないまま、前年度にテキストを選択し、授業方法を考えて学期が始まることも少なくないといえる。実際に、本稿で報告する授業の受講者は入学試験時ではなく、2年への進級前において専攻が決まる学部所属する者達であった。それを知らなかった筆者は、その学部学生の能力と関心について予測を誤ったのである。

5.3.2 学生との会談

上記の授業方法と学生の能力の不適合について発見する契機となったのは、学期開始から約1ヶ月のころに一人の学生が質問に来てカンファレンス(会談)をしたことからであった。その当時は、教科書の著者が提案する教科書の内容

の進め方が学生のレベルにあっていず、また学生との連携 (rapport) を築くのに時間がかかっており、学習者があまり積極的に授業に参加しないという問題も抱えていたところであった。学生の話から、当該クラスの学生の専攻は未分化であり、英語の能力にばらつきがあり、英語自体に興味を持っていない学生の率が当初の予想よりも多そうだ、ということが推測された。

5. 3. 3 テキストの難易度および授業方法

テキストは、同一授業名の授業間での重複を避け、学年次に難易度をあわせる調整作業が必要なために、前年度にテキストを選択している。そのため、授業開始後にクラスの学生の能力とテキストの難易度の不一致が起こっても変更することはできない。テキストの選択が失敗であったとしても、授業方法を変更して対処する必要がある。

5. 3. 4 リスニング能力の不足

その当時は当該クラスの学生のリスニング能力が低いのではないかと推測していた。公表されている直近の過去三年間の TOEFL スコアサマリーに (Educational Testing Service, n. d.) によれば、日本人受験者はリスニング、構造、リーディングの3つの分野のうちリスニングのスコアが最も低い。同様の傾向は、神戸大学の文系学部 B で実施した新入生対象の TOEFL-ITP でも見られた (Tanaka, 2002)。日本人英語学習者について、リスニングが不得意な傾向があるとすれば、学部 A の学生においてもリスニングが不得意である可能性があった。また、約1ヶ月間の授業での経験から、学部 A の学生についてはリスニング能力が低い可能性があると考えられた。

5. 4 形成的評価の実施と精神の実践

上に掲げた問題に対処するために、以下のような指標を設けて評価を行うこととした。

5. 4. 1 学生の背景を知る

授業開始時に学習者背景についての情報を知る必要があるのだが、授業開始時に行ったアンケートでは、海外在住経験や、英語会話の学習歴は判明しても、個々の学生の英語能力、特にリスニング能力を知ることはできない。そのため、英語能力を診断するために、SLEP (Secondary Level English Proficiency) テストのリスニング部分を実施し、学生のリスニング能力を測った。上記の「問題の現れ」の項で述べた一人の学生とのカンファレンスに加えて、複数名の学生とのカンファレンスを持った。その結果、形成的評価導入前に一人の学生とのカンファレンスをした時に得た情報と類似した情報を得た。これで「当該クラスの学生の専攻は未分化であり、英語の能力や学習動機にばらつきがある」という推測の新たな裏付けとなった。

5. 4. 2 中間授業アンケートの実施

カンファレンスによって様々な問題が存在することが予測されたため、それまで実施した授業についての中間アンケートを5月29日に実施した。内容は授業の進度、内容の難易度、内容への興味、視聴覚教材の英語の速さや難易度、行って欲しい授業活動の種類、などであった。これによって現状の授業に対する学生の反応と希望する授業についてのフィードバックを得た (アンケートデータについては今回報告しない)。

5. 5 授業法の工夫と実施

5. 5. 1 修正の着眼点

学生の英語学習への興味や能力にばらつきがある可能性が高いことが判明し、またそれまでの授業のどの面に改善が必要かわかった。その結果とリスニングテストの結果から授業改善を行うこととした。授業改善で目指したのは、

自発的に学習する意欲を出させる

学習意欲が継続されるようにする

の2点であった。この二つの目標を実現するために、次のような変更を行った。まず、英語テキストに載っているタスクをテキストの提示順序に従って行うことをやめた。代わりに、ビデオの内容を理解し、理解した内容を題材として後

の作業で使用するという、ビデオ視聴を意図的に努力して行う必要があるタスクを作ることにした。

上の2点を実現する上で、

自主的学習

(学習者自身が課題を選択し、課題毎にグループをつくって学習を行う)

授業方法の変化

(一斉授業の形態をとるものの、その中の授業方法に変化をもたせ、小集団活動をさせたり、教育機器を使用したり、実験を行ったりする)

(陣川, 1990)

を取り入れることにした。具体的には、以下の5点を具現化するような活動を取り入れることとした。

- (A) グループワークを導入し、協調的学習を図る。
- (B) グループワークでの成果を相互評価する。
- (C) 他の学生が自分たちをどのように評価しているかを知らせる。
- (D) 学生自身が意識的に自分を形成的に自己評価できるような手法の導入。
- (E) 一時点での能力が評価されるのではなく、学習が継続的にモニターされていることの明示、である。

上記(A)から(E)のすべての着眼点を含む活動としてはグループワークを、(D)と(E)の実現手段としては能力テストを取り入れた。学習が継続的にモニターされていること(E)を明示的にするために、グループワーク活動と能力テストについては1回目の実施後、同様のタスクがもう1回実施する旨と、1回目の成績と2回目の成績の伸びの大きさも評価の対象とすると学生に告げておいた。

5. 5. 2 グループワーク

形成的評価の実施後に取り扱った単元のうち、2つの単元の内容についてグループワークを取り入れた。それぞれの単元の内容は、現在の日本の状況にもあてはまり、かつ日本においても社会問題であるようなトピックについてであった。² 学生達はこれらの新しい単元に入る際に以下のような指示を与えられた。

ビデオニュースクリップの内容を理解するよう努めること

単元の内容に基づいて、後で与えられるトピックについてグループで討論すること

討論結果をグループの意見として発表原稿を作成し口頭発表すること

口頭発表は教師だけではなく他のグループ(ピア)からも評価されること

発表原稿は教師によって回収され評価されること

² ここでは現在も使用している教科書を使っている活動について述べているため、トピックの詳細な内容については記載しない。本稿がウェブ上に掲載された場合、詳細を記載すると次学期からの授業運営に支障を来す可能性があるためである。

作業の開始前に、学生達は2人あるいは3人から成るグループに分けられた。基本的に、自分の座席から遠くない人たち同士でグループを組ませた。グループ分けは座席の物理的な位置でおおむね決まったが、グループができない人たちについては教師がメンバーを指定してグループを作らせた。

すでに当該クラスで導入していたディクテーションのように1回で完結するタスクと、このようなグループワークの導入によって、学生が意欲的にかつ継続的に学習に取り組むことになるであろうと考えた。

5. 5. 3 能力テスト

先に述べたように、学習者背景を知るため、能力テストとして5月22日に SLEP テストのリスニング部分を実施した。さらに学期の終わりの7月24日にもう一度リスニングテストを行った。この二つのリスニングテストは、試験の実施直後に採点し、学生に素点スコアを知らせた。テストを2回実施する目的は、学生に自分のリスニング能力の実体を知って、意識的に学習を行っていくという、「自立的学習」あるいは「自己教育力」を伸ばしていくためである。

一方、教師の側においては、1回目のリスニングテストをプリテスト、2回目のリスニングテストをポストテストとして、形成的評価を通じた授業改善の結果がリスニング能力の向上にどう現れるかを検証し、今回の「形成的評価」の導入の試みについて「自己評価」しようと考えたためである。

5. 6 授業改善後の評価

学期末の授業終了時（7月24日）に、5月29日に実施した内容と同じ設問と、それにグループワークに関する設問を新たに加えたアンケートを実施した。これは形成的評価導入の後で、学生の授業評価がどのように変化したか、また学生が英語学習に積極的になったかどうかを知るためであった。³

6. 導入後の結果—リスニングテストスコアへの現れ

最初に述べたように、本稿では形成的評価の実施の具体的プロセスとリスニング能力との関連に重点をおいて報告する。本研究の対象クラスの構成員は39名であったが、そのうち英語ネイティブ・スピーカー1名と2回目のテストを受験しなかった3名を除いた、35名を分析対象とした。リスニングテストのスコアは、素点を試験マニュアル(SLEP Test Manual)に記載の変換表に基づいて尺度値に変換した。尺度値に変換する理由は、1回目と2回目に用いたテストの版が異なるため難易度が同一であるとはいえないため、比較可能にするために尺度値に変換する必要があったためである。35名の試験結果の概要は、以下の表1に示すとおりである。

³ アンケート結果については本稿では報告しない。

表1. リスニングテスト1回目と2回目の素点および尺度値と1回目と2回目と尺度値の差.

	L1	L1 scaled	L2	L2 scaled	d
被験者数(N)	38		35		35
平均値(M)	48.00	23.29	52.80	25.31	2.23
標準偏差 (SD)	4.32	1.92	4.68	1.92	2.24
最大値	56.00	27.00	61.00	29.00	6.00
最小値	38.00	19.00	44.00	22.00	-3.00

注. L1=1回目リスニングテスト素点, L2=2回目リスニングテスト素点.

L1 scaled=1回目リスニングテスト尺度値, L2 scaled=2回目リスニングテスト尺度値.

Rank L1 scaled=1回目リスニングテスト尺度値順位, Rank L2 scaled=2回目リスニングテスト尺度値順位.

d=L2 scaled - L1 scaled. Sum of the ranks=630.

1回目のリスニングテストの素点(L1)の平均値は $M=48.00$ ($N=38$)で、尺度値の平均値は $M=23.29$ であった。2回目のリスニングテストの素点(L2)の平均値は $M=52.80$ ($N=35$)で、尺度値の平均値は $M=25.31$ であった。1回目と2回目の尺度値の差(d)の平均値は2.23($N=35$)であった。授業初期と授業終了時のリスニング能力に変化があったかどうかを検証するために、 $\alpha = .05$ に設定して、ウィルコクソン符号順位和検定(Wilcoxon Matched-Pairs Signed-Ranks Test)で片側検定を行った。帰無仮説(H_0)と研究仮説(H_1)は以下のように設定した。

帰無仮説(H_0): このクラスの学生達の母集団では、授業初期と授業後のリスニングの尺度値の差の中央値はゼロに等しい。

対立仮説(H_1): このクラスの学生達の母集団では、授業初期と授業後のリスニングの尺度値の差の中央値はゼロよりも大きい。

$H_0: Md = 0,$

$H_1: Md > 0.$

$T=547.5$ で、 $T_{.95}=533$ であり、 $T=547.5 > 533$ であるため、帰無仮説は棄却された。⁴ よって、リスニングテスト2とリスニングテスト1の尺度値の中央値の差は0よりも有意に大きいと言えた。このことから、授業初めに受験したリスニングテスト1と授業終了時に受験したリスニングテスト2を比較すると、その尺度値に有意に正の増加が見られたと結論づけられる。

しかしながら、表1を見るとL2尺度値とL1尺度値の差であるd値の最小値は-3.00である。本稿には含めていないが、各被験者毎のデータを見ると、このように差が負であった者が4名、差がなかった者が5名おり、それぞれ $d=-3.00$ ($n=1$)、 -2.00 ($n=1$)、 -1.00 ($n=2$)、 0.00 ($n=5$)であった。前者4名はすべてL1尺度値では平均以上のスコアを取得しており、後者5名のうちの3名もL1尺度値で平均以上のスコアを取得していた。当該クラス全体ではスコアに正の増加がみられたものの、授業当初に比較的高いリスニング能力を保持していた学生に関しては、授業当初と授業終了時の間にスコアが

⁴ Marascuilo & Serlin (1988)の213-225ページに現れる計算式を参照。

減退する、あるいは変動がない傾向がみられた。

7. 考察

7. 1 スコアの推移の特徴と原因

結果の項で述べたように、形成的評価を取り入れた英語オーラルの授業当初のリスニングテスト(L1)のスコアと授業終了時のリスニングテスト(L2)のスコアの比較から、授業当初から授業終了時までの間にリスニングスコアに有意な伸びが見られた。先にも述べたが、TOEFLの結果から見られるように、リスニングが日本人学習者にとって最も不得意な分野であることから、リスニング能力が伸びたことは非常に望ましい結果であるといえる。不得意である分、逆に言えば最も伸ばせる余裕がある分野であり、リスニング能力を養成することによって、TOEFLなどの標準テストにおいてスコアの伸張が期待できる。

この形成的評価においては、ディクテーションなどのように聞いて文字に直すことで完結するタスクではなく、聞く、話す、社会・経済事象の内容を理解する、意見を持つ、討議する、書く、発表するといった様々な能力分野を統合したタスクを取り入れた。決してリスニング能力だけを訓練することに重点を置いていたわけではない。それにもかかわらず2ヶ月の短い期間に、尺度値で平均2.23ポイントの伸びが見られた。

何故このような伸びが見られたのかについて、ここでその理由を考えてみたい。第一点は、今回の授業の中で取り入れた協調的な学習の効果であろう。協調的な学習では、グループ内で個人が責任をもって役割を果たすために、学習者が自分で目標を設定してその目標に向けて努力をする。その基盤として視聴覚教材の内容を正しく理解しようと努めたため、リスニング能力が伸びたのではないかと考えられる。また、グループ内では能力の高い者が低い者の学習を助けて協調的に学習をするため、能力の高い者と低い者との間で相互的な対話を通して学習が生じたのではないかと推測される。また、グループ毎に発表してグループ同士で採点をさせる形式をとったため、他のグループと自分たちのグループとの差が明確にわかり、またその違いが相互評価の点数やコメントとしてフィードバックに現れてきたためだと考えられる。また、当該クラスの中に非常に音声言語に秀でた学生がいたため、その学生がモデルの役割を果たすことになって、他の学生の目標設定が明確になったのではないと思われる。このような協調的学習を通して、学生に自己教育力が生まれ、ひいてはそれがリスニング能力の伸張に反映されたのではないかと推測される。

しかしながら、結果の項でも述べたように、リスニング能力は入学時にもともと高い能力を持っている学生よりは入学時に低い能力を持っている学生が伸びた傾向が見られた。もともと高いリスニング能力をもっていた者ではなく低い能力をもった者が伸びたのは何故だろうか。先に述べた形成的評価のメタ分析によると、形成的評価の効果は成績の高い被験者よりもむしろ低い被験者に現れる傾向があるといわれている (Black & Wiliam, 1998b)。Black & Wiliam (1998b)によると形成的評価の中に見られるように、生徒にフィードバックを頻繁にすることによって、従来型の授業では授業についていけずに落ちこぼれていた生徒が、学校や学習を真剣に受け止めるようになるという。彼らの説明は高等教育機関ではなく初等・中等教育機関に在籍する者を想定しているため、この説明が大学生には必ずしも該当しない可能性がある。形成的評価が、学力遅滞のない、むしろ平均よりも優秀であると思われる成人被験者においてどれほどの学習の向上をもたらすのかは今後の研究をまっ必要がある。

では、逆にもともと高い能力をもった者のうちの半数以上(テストを2回とも受験した者で1回目の尺度値が平均以上であった者13名のうち7名)が、何故1回目と2回目の間で尺度値スコアに変動がないか、あるいは下がってしまったのであろうか。ひとつには、これらの高得点者は、能力の低い学生が協調的作業で学んだような自己教育力をすでに持っていたため、授業当初と授業終了時でスコアの増加がなかったのだらうと考えられる。もう一つの理由としては、もともと能力が低い学生には伸びる余地があり、高い学生には伸びる余地があまり残されていなかった(つまり天井に達していた)からかもしれない。例えば、尺度値27の学生の場合、このテストを標準化した際のサンプル(全受験者)のうち71%がこの学生よりも下のスコアをとっており、同様に尺度値26の場合は全体の67%の学生がこれよりも下のス

コアを取っていることになり、もともと比較的高いスコアだったといえる。⁵

結論としては、授業当初と授業終了時に実施したリスニングテストにおいて、全般的にはスコアが2.23ポイント有意に上昇した。これは主として授業当初にスコアが低かった者のスコアの上昇によるものであり、もともとスコアが高かった者の場合はスコアに変動がなかったか、むしろスコアが下降した。形成的評価の導入は能力が高い学生にとってはその効果があまり現れない(Black & Wiliam, 1998b)という報告があるものの、今回報告した結果が形成的評価の導入に起因するものであるかは不明である。本研究では、形成的評価を行ったクラスとその統制群の間で実験的に確かめたわけではない。そのため、今回報告した結果が形成的評価の導入によるものであるかについては言明ができない。学生の中に現れた学習効果に形成的評価が直接結びつくことがまだ統計的に実証できないまでも、教師の授業改善への効果(その改善を学習者が好意的にうけとめているかで測定)が出てくる可能性は考えられる。本研究に関して言えば、教師(筆者)が教室内で行う活動について意識的になった、という意味において十分に効果があったと言えよう。

7. 2 研究手法上の問題

ここで、形成的評価とその結果が持つ客観性について注意を喚起したい。往々にして質的なデータに基づいて行われる形成的評価は客観性の面で不十分であり、普遍性にかけて批判されている(Flagg, 1990)。その一方で形成的評価に客観性を求めることは不合理であり、「評価データが個人的で主観的であっても、そのデータが指導の改善に的確に役立つ」ば良く、形成的評価には何よりも教育性が求められる、という立場がある(安彦, 1990, pp. 232-3)。形成的評価が指導法の改善に用いられる場合に限定してもその手法の有効性をめぐって論議されているが、形成的評価の客観性がより問題となるのはカリキュラム評価に用いられた場合においてであろう。

もともと形成的評価は、「カリキュラム評価」の方法として提唱されたものである(Walker, 1992)。指導法の研究において有効性が疑問視される理由の一つが、上にも述べたように、質的データを評価に用いることから来る客観性の欠如であった。しかしながら、Ross(2003)が言うように、今日の言語教育プログラムにおいては、その成果を評価する上で重要な鍵となる指標が、どれだけ学生が参加できているか(student involvement)である(p. 7)。そのため、望んでいるような成果を言語プログラムが上げているかを評価するために、従来から使われてきた客観テストに加えて、ポートフォリオや、自己評価、ピア評価、発表や、討議などのような様々な評価方法が含まれるようになってきている。このことから判断すると、今日の言語プログラムで重要視すべきは指導法改善を目的とした質的評価方法の使用の是非を論議する段階はすでに通り越しているといえる。現時点において重要なのは、Ross(2003)が主張するように、様々な評価方法を取り入れた教育プログラムの成果が、いかにして学生の能力変化と年次を越えて長期的に関連づけられるかである(p. 7)。つまりは、すでにある程度うまく機能している言語プログラムに関しては、いかにして言語プログラムの通時的-一貫性(program coherence)(Newman, Smith, Allensworth, & Bryk, 2001, referred in Ross, 2003)を保つかが重要だということになってくる。

8. 今後の課題

今回は研究者と教師が同一人物であり、アクション・リサーチ的な意味を持っていた。形成的評価の実施において重要な問題は、誰が実施するのかということである(Horwitz, 1985)。形成的評価の導入については、形成期評価の実施決

⁵ 一回目のテストで尺度値が25から26であった者で、二回目も受験した者8名のうち、スコアが2ポイント上昇した者が2名、変化なしが3名、下降した者が3名であった。

定の主体との関連で何点か注意すべき点がある。形成的評価の実施には、評価者が教師自身であろうと第三者であろうと、教師が実施に同意し研究に参加乃至は協力することが不可欠である。教師が研究者兼教師の場合は、形成的評価の導入を決定するのは教師自身であるので問題はない。しかし、評価者が第三者であり、教師以外の者（機関を含む）が形成的評価の実施を意図する場合には、教師自身の実施への合意と協力を前もって取り付けておくことが必要である。なぜならば形成的評価には教師自身が被評価者として貢献することが必要不可欠であり、評価者としてではなくても教師が形成的評価に参加する必要があるためである。

このことは外部機関や教師の所属機関が形成的評価導入のイニシアチブをとることを全て否定するわけではないし、ましてや問題がある授業が垂れ流しのまま改善されないでよいということの意味するのではない。形成的評価は問題意識を持った教師が必要を感じたところで自発的に取り入れるのが最良の方法だと考えるためである。その意味において、すべての授業について形成的評価を一斉に取り入れるように教師全員に上意下達の形で指導しても意味がないといえる。形成的評価が教育の内容の改善、ひいてはカリキュラムの向上を目指すものである限り、評価の結果が教師によっていかされることを目的としているはずである。結果が改善にいかされなければ、その評価は形成的とは成り得ないのである(Black & Wiliam, 1998b)。

参考文献

- Black, P., and Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Black, P., and Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment [Electronic version]. *Phi Delta Kappan*, 80, 139-18.
- Boston, C. (2002). The concept of formative assessment (Report No. ED470206). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation. (ERIC Document No. ED470206) Retrieved November 24, 2003, from <http://www.ericfacility.net/ericdigests/ed470206.html>
- Educational Testing Service. (n. d.). 2000-2001 Edition TOEFL test and score data summary. Retrieved from <ftp://ftp.ets.org/pub/toefl/990611.pdf>
- Educational Testing Service. (n. d.). 2001-2002 Edition TOEFL test and score data summary. Retrieved from <ftp://ftp.ets.org/pub/toefl/10496.pdf>
- Educational Testing Service. (n. d.). 2002-2003 Edition TOEFL test and score data summary. Retrieved from ftp://ftp.ets.org/pub/toefl/10496_02_03.pdf
- Flagg, B. (1990). *Formative evaluation for educational technologies*. Hillsdale, NJ: Erlbaum.
- Horwitz, E. K. (1985). Formative evaluation of an experimental foreign-language class. *Canadian Modern Language Review*, 42, 83-90.
- Marascuilo, L. A., & Serlin, R. C. (1988). *Statistical methods for the behavioral sciences*. New York: Freeman.
- Ross, S. J. (2003). A diachronic coherence model for language program evaluation. *Language Learning*, 53, 1-33.
- Walker, D. F. (1992). Methodological issues in curriculum research. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 98-118). New York: Macmillan.
- 陣川桂三 (1990)「形成的評価からの授業作りを目指して」『中学校 新しい授業作りと形成的評価』梶田叡一編著 東京：東京書籍。

田中順子 (2002) 「入学時英語能力とその推移に関する一考察」『国際文化学研究』第18号 1-22.

安彦忠彦 (1990) 「形成的評価の新たな課題」『中学校 新しい授業作りと形成的評価』梶田勲一 編著 東京：東京書籍.

Formative Evaluation of an English Oral Communication Class for University Freshmen

Junko TANAKA (Associate Professor, Faculty of Cross-Cultural Studies, Kobe University)

The author carried out the formative evaluation as a teacher and researcher in a university Freshmen English class, whose aim was to develop oral communication skills using US television news programs as a material. When she found that some students were having difficulty in following the class, she decided to conduct the formative evaluation, which involved questionnaires, a new collaborative task that required a greater students involvement and peer-evaluation, and listening tests at the beginning and the end of the evaluation period. Positive changes were observed in the students and teacher.

As for the students, their listening proficiency, as measured by Secondary Level English Proficiency test, had a statistically significant gain of 2.23 points ($N=35$, $\alpha=.05$, $p < .05$) at the second testing. As for the teacher, she became more cognizant of her educational activities in the class. Since this study did not involve controlled experiments, it is difficult to know whether the gain in the listening test was due specifically to the formative evaluation, or whether a similar gain could have been obtained without the formative evaluation.

Despite such a methodological drawback, it can be said that the observed positive changes may be enough to warrant the use of formative evaluation for the purpose of improving educational program by the teacher.