

CAI教材の設計開発における 形成的評価の技法について

フロリダ州立大学 鈴木 克明

I. はじめに

スクリバン (Scriven, 1967) によって形成的評価の概念が明確化されて以来、プログラム学習教材の実証的な効果の診断手段として、さらに授業のシステムの設計開発の手続きの中で、形成的評価の研究が重ねられてきた。我が国においてはブルーム (梶田他訳, 1973) による業績が教育評価の立場から広く紹介され、主に教師が前面に立って行う集団学習での応用が研究されてきている (藤田, 1982)。近年の独立型の教材、特にCAI教材の開発普及に伴い、教材のシステムの設計開発における形成的評価の概念ならびに実践的手法の重要性が増していると思われる。本稿では、定められた学習目標の達成を促すという意味で「ひとり歩きのできる教材」を形づくる手段としての形成的評価について概観し、これまでに提案された具体的な技法をCAI教材の設計開発へ応用することを検討する。

II. 形成的評価とは

形成的評価は、効果的な教材づくりを実証主義的にすすめるために重要なステップである。「インストラクターが授業の効果と効率を高めるためにデータを収集する過程」(ディックとケーリー [Dick & Carey], 1985, p. 198) と定義され、授業で用いる教材が完成する前に、教材の改善を目的として学習者等が実験的に教材を使用する過程を指す (AECT, 1977, p. 265)。形成的評価とそれに伴う改善のサイクルを経た教材は、その学習指導効果が実際の学習者からのデータを用いて向上された教材とされ、システムの設計開発過程の所産として学習指導に用いられることになる。つまり、学習課題の構造分析や学習指導理論に基づく教授ストラテジーの選択・応用など、理論的見地から設計された教材が果たして実際に効果的に使用できるかを確かめ、教材の改善に役立てるデータ収集の作業が形成的評価と呼ばれる過程である。

初期の授業改善への試みは1920年代にさかのぼるとされているが (ケンパー [Cambre], 1981)、形成的評価という用語は1967年にスクリバンによって用いられた。スクリバンは、クロンバック (Cronbach, 1963) の教育評価の拡張の提唱を受け、教育評価の総括的 (summative) な役割と形成的 (formative) な役割とを二分することで明確化をはかった。総括的な役割をもつ教育評価

の結果は教材作成機関の外へ流れ、教材の利用法や教材への認識を向上するために使用される。一方、形成的な評価の結果は内部情報として未完成な教材の改善に役立てられる。当時の教育評価が総括的な役割に多く用いられていた中で、クロンバックとスクリバンによる形成的評価の概念の提唱は重要な意味をもつものとされている。つまり、形成的評価においては、評価の対象は学習者でないばかりか、『完成された』教材でもないことに注意する必要がある。授業のシステムの開発の過程の一ステップとして授業の改善のためにデータを収集するわけである。

形成的評価の概念が導入されて以来の20年間で方法論の体系化が進められてきた。学習者のデータの重要性を強調するスキナー派のプログラム学習教材の開発過程では、開発テスト (developmental testing) という名のもとに形成的評価の技法が研究された (マークル [Markle], 1967)。ベーカーら (Baker, 1974; Backer & Alkin, 1973) は、プログラム学習教材に用いる形成的評価のモデルと教材の改善の手続きを示すガイドラインを提案し、形成的評価のデータを教材の改善に直接結びつけた。プログラム学習での形成的評価は我が国にも影響を及ぼし、例えば沼野 (1976) はトライアウト (tryout) の概念を組み込んだ授業の設計手順を提唱している。また、学習者検証の重要性は教材のシステムの設計開発にも早くから認識され、たとえばバークマン (Burkman, 1971) の手がけたISISやISCSと呼ばれる中等教育レベルの新しい理科カリキュラムの開発においては、自習教材の執筆中に夏季休暇の中学生の協力を得て形成的評価が行われた。

一方セサミ・ストリートなどの教育テレビ番組の制作においては、パーマーら (Palmer, 1974, 1981) によって形成的研究 (formative research) が体系化され、制作中の番組の効果向上を達成しながら基本的な番組制作変数の研究に役立てる『折衷的でかつ実用的な』方法論がとられてきた。さらに近年のイギリスにおけるオープンユニバーシティのテレビ放送教材の制作過程においては、大規模な番組の学習指導効果に関する形成的評価が行われており、その方法論がネイゼンソンとヘンダーソン (Nathenson & Henderson, 1980) によって発表された。また、10年程前からコムスキ (Komoski, 1974) らによって提唱・推進されてきた州指定教材の学習者検証と改善 (Learner Verification and Revision; LVR) もアメリカ合衆国のいくつかの州で立法化された。特にフロリダ州では州指定教材 (教科書を含む) のリストに採用され購入の補助対象となるための条件として、LVRが実施されたことを示すLVR報告書の提出が出版社に義務づけられるなど、LVRの方法論が具体化してきた (ベドロス [Vedros], 1986)。これらの方法論の体系化は、異った用語を用いて異ったコンテキストの中で進められてきたが、いずれも教材の改善のためのデータ収集であり、共通している点が多い。現在の授業のシステムの設計開発モデルに用いられる形成的評価の手法は、これらの研究所産を基盤として形づくられてきた。

Ⅲ. なぜ形成的評価か

図1は現在アメリカ合衆国において広く知られているディックとケーリー（1985）による授業のシステムの設計開発モデルである。授業目的と授業開始時に学習者ができること（前提行動）が確認され、そのギャップを効果的にうめるために授業設計が行われる。授業設計は学習指導理論（たとえばガニエ [Gagné], 1985) に基づく仮説的な過程であり、最も効果的とされている学習目標の順序だてや各目標に対する学習指導ストラテジーを規定する。次に授業設計案に基づいて教材を開発し、その教材が実際に効果的かどうかを形成的評価の過程を通して確かめる。形成的評価から得られたデータは教材の改善、あるいは理論的な前提の再検討に用いられ、学習効果がさらに確認される。最終的には、教材の改善の必要性がなくなった時に、総括的評価が行われ、システムの過程を終結することになる。

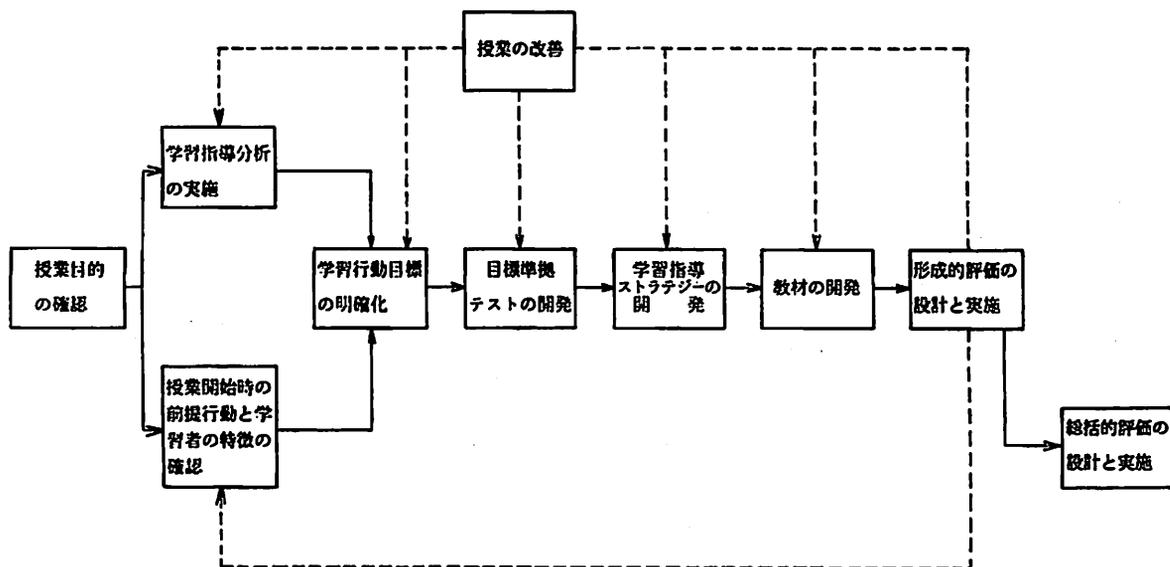


図1. ディックとケーリーの授業のシステムの設計モデル

注: Dick & Carey (1985, p. 2-3) より。

授業のシステムの設計開発にあたって形成的評価を行う第一の意義は教材の効果向上である。学習指導理論の発展によって教材設計の理論やモデルがより効果的な教材設計を助けられるようになってきたが、理論のみに基づく教材設計は未だ完全ではない。理論のみに基づく教材の設計開発の効果が100%確立されれば学習者から試験的にデータを収集する形成的評価のプロセスは不要となるわけであるが、人間の学習の過程や記憶構造、動機づけなどがすべて明らかになるまでは学習者からのデータが必要である。

ところで、形成的評価は特にCAI教材等の独立型の教材のために重要である。教師が介在、あ

るいは伝達する授業においては、学習指導の場面での臨機応変な処置がとられるが、独立型教材を用いての授業は学習者が他からの援助なしに学習を進めることが前提となっている。従って、独立型教材の開発には教材が使用される以前に効果を確認することが不可欠であり、形成的評価と教材の改善を行うことが特に有効となる。

第二には、形成的評価を行うことで、学習指導の理論やモデルの妥当性を検証し、不完全な部分を補うことが挙げられる。つまり教材の形成的評価を通して、教材設計の基礎として用いられた理論的根拠をも形成的に評価するわけである。たとえば、Aの技能を学習するためにはBとCの概念とDの知識が必須であるという学習ヒエラルキー分析に従って設計された教材の形成的評価に於いて、多くの学習者がDのテストで低い点をとったにもかかわらずAのテストではマスタリーに達したとする。テストが妥当なものであるとすれば、Dの知識がAの技能習得に不可欠であるとした分析に誤りがあったことになり、分析の見直しが必要となる。このように形成的評価はあくまでも教材の改善のためのデータ収集を目的とするが、同時により理論的な見地からの洞察が得られることもしばしばある。

第三に、形成的評価を行うことで教材の利用者層が具体的に定められ、実際に使用可能な、作る価値のある教材づくりをすることができる。つまり、学習指導の現実からの遊離を防ぐ働きをもっているといえる。たとえば、小学5年生用の教材を作成する際に、4年生の内容の復習をどの程度盛り込むかという問題がある。学習指導分析（たとえば前出の学習ヒエラルキ分析）によって目標達成に必須な下位行動が明らかにされ、4年生の内容のうちの何を再確認させる必要があるかが決定される。ここで、下位行動のうちAとBはすでに学習済みであるから再確認のためのまとめを1ページ含めば十分とし、CとDは学習が未だ十分に行われていないので（あるいは4年生の学習内容ではないので）説明と練習のページを含むなどの判断がなされ、教材が開発される。しかしながら、これらの判断はあくまで仮定されたものであり、形成的評価の際に、たとえばAとAを下位行動にもつ目標につまづきが見られ、Cは簡単すぎてつまらないとの感想が多く実際に正答率も高かったなど、予測に反する結果が出るがしばしばある。形成的評価の結果を受けてAの説明や練習のページを増やしたりCの項の一部をカットするなどの教材の改善を行い、学習者の実体に見合った教材にしていくことができる。

学習者の実体に見合った教材という点に関しては、実証的な教育研究を解釈したり、あるいは自ら実施する場合にも重要な視点であり、形成的評価を行った教材を用いた実験であったかどうか教育研究の解釈に意味をもつとされてきている。例えばブリッグス（Briggs, 1984）は教育研究は4つの『文化』に分類されるとし、学習指導に応用できる情報を直接に得られるのは実験のデザインが良いだけでなく授業設計の原理に従う『第4の文化』に含まれる研究からだけだとしている。『第4の文化』の特徴としては、学習課題の分類と目標、テスト、教材の整合性が確かめられることと学習内容が現実のカリキュラムを反映していることの他に、教材がシステムの設計開発

され、形成的評価の手続きを踏んでいることが挙げられている。最近の研究誌にみられる実証的研究には、『用いられた教材は形成的評価された…』というような記述が多くみられるようになってきており、教育研究者にとっても形成的評価の技法はその重要性を増しているといえよう。

第4に、授業を設計開発する者にとって、体験的な訓練になることが挙げられる。学習者要因を考慮しながら授業の設計開発を進める中で、実際に学習者から得られる情報ほど確かな手がかりはない。データを解釈し、教材を改善し、また学習者からのデータを収集する。このサイクルを通して、どの学習指導ストラテジーが有効か、事後テストの不振は何に起因しているかなど、教材の設計開発の鍵になる判断をする技能が経験を通して養われていく。これはちょうど学校教師が教室での授業の体験を通していわゆるベテランの技量を習得していくのに似ている。科学的な学習指導理論を補う意味で重要な体験に基づく教材の設計技術を習得するためにも形成的評価を実施することが有効である。

最後に、教材の形成的評価が後の教材の採用を促進するための方略として行われることもある。一つには形成的評価の過程を経ることで教材の効果向上が試みられたということであり、さらには、利用者の意向を取り入れた教材の改善を行い教材と利用者の距離を縮めることを狙うわけである。たとえば前出のバークマン（1971）のISISプロジェクトにおいては、実験室での第一次形成的評価の結果に基づいて改善されたトライアウト用教材を用いて、アメリカ全国に点在する大都市の実験校での第二次形成的評価を行っている。この第二次形成的評価は教材の効果をより実験的な状況の中でテストし改善案を探ることと同時に、イノベーションである新しい教材のデモンストレーションとしての効果も予測通りに得ている。このような大規模の教材開発プロジェクトでは、イノベーションの普及方略としての形成的評価の側面も見逃せない。

形成的評価は、授業のシステムの設計開発の過程の中で不可欠な要素であり、形成的評価を含まないモデルは不完全なものといえる。理論的な裏づけがどうであれ、学習の促進に役立たないのでは単なる机上の空論になってしまう。形成的評価には前述のように多様な効用がみられるが、学習者からのデータを授業の改善に役立てるために収集するという第一義の目的を忘れてはならない。

IV. 形成的評価の実際

授業のシステムの設計開発のモデルの中で、形成的評価の手法についての詳しい記述を含むものにディックとケーリー（1985）によるものがある。これは、ディックの形成的評価体系化の業績（たとえば Dick, 1977）に基づくものであり、今日において最も完成された方法論といえる。以下に形成的評価の実際をディックとケーリーのモデルに従って要約する。

形成的評価の中心をなすものは、学習者からのデータ収集である。学習者からのデータには事前・事後テストや教材の中に組み込まれたつまづきを発見するための質問項目の他に、教材に関する感想や提案を求めるアンケート、並びに学習時間などの記録も含まれる。その他の情報源としては教

材を使用する生徒を担当する教師や該当教科の専門家、あるいは教材設計を専門とする者などが挙げられ、各情報源から得られるデータの種類と収集の時期は表1のようになる (Vergara, 1985)。これらのうちから形成的評価の目的が教材の効果を高めることにあることを念頭において、改善に役立ちそうな情報を収集するわけである。

表1. 形成的評価に用いるデータの種類、情報源ならびに収集の時期

情報源 のタイプ	情報源	データの種類		
		トライアウト以前	トライアウト中	トライアウト以後
外部	教師	<ul style="list-style-type: none"> 教材の適切性 (D) 	<ul style="list-style-type: none"> 教材の管理 (B) (D) 	<ul style="list-style-type: none"> 教材に対する意見、感想及び提案 (B) (D)
	生徒 (学習者)	<ul style="list-style-type: none"> 前提行動テストの結果 (B) (D) 事前テストの結果 (B) (D) 事前、事後テストの指示と項目の明確性 (D) 	<ul style="list-style-type: none"> 教材に組み込まれたテストの結果 (B)(D)(N) 教材とその内容に対する意見 (N) 	<ul style="list-style-type: none"> 事後テストの結果 (A) 教材に対する意見、感想及び提案 (A) 完了までの経過時期 (D) (N)
内部	教科専門家	<ul style="list-style-type: none"> 内容の正確性最新度及び語彙レベルや例題の妥当性 (B) (D) 		
	教材設計 専門家	<ul style="list-style-type: none"> 学習指導原理の適切な応用 (M) (B) (D) 用語の適切性 (B) 		<ul style="list-style-type: none"> 教材改善の適切性 (A)
(M) マークル	(D) ディックとケーリー	(A) 右記のすべて		
(B) ベーカー	(N) ネイゼンソンとヘンダーソン			

注: Vergara (1985) の表2 (p.24) を一部改訂したもの。
かっこ内の文字はどのモデルに含まれているかを示す。

学習者からのデータを収集するトライアウトの第一段階は1対1の形成的評価 (one-to-one formative evaluation) と呼ばれる。教材の設計開発者が学習者一人ずつを相手に教材を用いて学習を進め、教材やテストからわかりきった誤りを取りのぞき、学習者の教材に対する反応をみることを目的とする。この段階においては、学習者が教材を批判的にとらえ、学習を進める中で不明瞭な点などを自由に表現できる雰囲気をつくることが重要である。設計開発者が学習者と共に1ページ

ずつ学習を進め、計画に従って適所で質問をするなどその雰囲気をつくる努力が要求される。

1対1の形成的評価には教材の成否に最も影響を及ぼすと思われる尺度の上・中・下の3名の学習者を少なくとも含むことが提唱されている。たとえば、やる気のあるなしが学習効果に最も関係があるとすれば学習者集団の中からやる気のある生徒、中ぐらいの生徒、そしてあまり熱心でない生徒の順に1対1で教材を使用して学習する。異った反応が上・中・下の生徒から得られた際には、枝分かれや任意の項目を設けるなどの方略が考えられる。1対1の段階には、教材のみならず、テストやアンケートなども同じような手続きで検討し、明らかなミスを取りのぞき、次の段階に臨む。

形成的評価の第二段階は小集団形成的評価(small group formative evaluation)と呼ばれる。1対1の段階では設計開発者が学習の過程に介入したが、小集団の段階では学習者が独立して学習を進めた場合の問題を探る。それと同時に、1対1の結果改善した箇所が効果的かどうかをも検討する。小集団の評価には、教材の利用者集団を反映させ、また結果を数量的に検討するために8人から20人の学習者を用いる。従って、利用者集団が多様な学習者層を含む場合には特に形成的評価の参加者の選出には注意を要する。小集団の評価の結果は学習目標別に、あるいは学習者の特徴によって分類され、教材の改善が検討、実施される。この段階の評価を行うことで小規模の教材開発の場合は効果の検証が十分為される場合が多いが、特に大規模な開発の際には次の段階に進む。

実地トライアウト(field tryout)と呼ばれる形成的評価の第三段階は、教材の管理や他の授業との関連などの実際的な条件が加わった中での評価である。我が国の場合では、学級集団を単位としての教材利用が最も一般に行われていることから、実地トライアウトは学級集団を用いて、担当の教師によって行われることになろう。実地トライアウトを1対1や小集団評価なしで行う場合も状況によってはありうるが(例えばネイゼンソンとヘンダーソン(1980)の教育放送番組の形成的評価)、実地トライアウト以前に1対1、小集団評価を経て段階的に教材の改善を行う方がより効果的であるとされている。

実地トライアウトが行われる理由として、教材普及のための方略上のものが挙げられる場合があるが、トライアウトの結果が教材の改善に用いられる限り、形成的評価としての役割を果たしているといえる。教材やテストそのものは小集団の段階でその効果が規準に達して改善の必要性がないとしても、教材使用者のためのマニュアルやガイドブック等の教材管理の側面を含めたトータルシステムとしての評価が行われることもある。あくまでも広義の教材の開発過程の一環としてのフィードバックがある以上、形成的である。一方、新しく開発された教材の効果をテストするために行われる試験的な利用で、その結果が教材の改善でなく教材採用の決断に用いられる場合は、その評価は総括的な役割を果たしている。つまり、両者は互いに相反的ではなく、一つの実地トライアウトが両方の役割を担うこともありうるわけである。

形成的評価の3つのステップについては以上の通りであり、その手法はディックとケーリー(1985)に詳しい。しかし一方で、形成的評価で得たデータをどのように解釈して教材の改善に結びつける

かは不明瞭な点が多い。ディックとケーリーは教材の改善を行う方法を形成的評価の次の章をさいて説明しており、有用なアイデアが数多くみられるが、その冒頭において教材の改善に関するモデルはいまだにわずかであることを指摘している。つまり、『結局、データを最も合理的と思われるように解釈し、そしてデータが示していると思われる方向に、あるいは我々の学習過程に関する知識に基づいて改善するのである。』(p. 223)としている。形成的評価が教材の改善を目的とする以上、形成的評価を計画する際には得られたデータをいかに解釈、活用するかを考察することが不可欠である。

まず、教材の改善には具体的に何が含まれているかを考察するのが有効であろう。デバート (Debert, 1979) は効果的な教材改善を訓練するための手法を提案した際に、表 2 に示されているような改善のリストを示した。アルファベットの数と同じ26の改善が含まれているが、これですべてというわけではない。教材の改善と一言でいっても多様な対応の方法があることがわかる。一方、ネイゼンソンとヘンダーソン (1980) は教材の改善を加える (Add)、削る (Delete)、動かす (Move)、修正する (Modify) の4つのカテゴリーに分類することが便利であるとしている。

表 2. 教材改善方略のリスト

-
- | | |
|------------|--------------------------|
| <u>付加</u> | A. 前提技能と知識に関する説明を加える。 |
| | B. 学習者に教材の使い方の訓練を加える。 |
| | C. 教材を使って指導する教師への訓練を加える。 |
| | D. 予告 (先行オーガナイザー) を加える。 |
| | E. イラストを加える。 |
| | F. 作業の補助を加える。 |
| | G. 例を加える。 |
| | H. 活動を加える。 |
| | I. フィードバックを加える。 |
| | J. 転移のための練習を加える。 |
| | K. テスト項目を加える。 |
| | L. 動機づけを加える。 |
| | M. 多様性を加える。 |
| <u>簡易化</u> | N. 複雑さのレベルを下げる。 |
| | O. 用語を簡単にする。 |
| | P. より小さい単元を使う。 |
| | Q. より大きい単元に教材を合わせる。 |
| | R. 順序を変更する。 |
| | S. 関連性の薄い情報を削除する。 |
| | T. 関連性の薄い活動を削除する。 |

- その他 U. 例をより関連のあるものにする。
V. 教授メディアを変える。
W. 授業設計のフォーマットを変える。
X. 形成的評価に用いた学習者を変える。
Y. プロジェクトを破棄する。
Z. 変更をしない。
-

注：Debert (1977) による。

次に、開発された教材の効果を妨げる原因には何がありうるかをささえ、それらに関するデータを収集する計画を立てる必要がある。いうまでもなく、事後テストでの各学習目標に関する達成度は教材改善の必要性の有無を示す最も重要なデータであり、学習目標が達成されれば教材の効果が確認されたとみなされ、それ以上の改善が行われるとは限らない。しかし、ネイゼンソンとヘンダーソン (1980) は、事後テストの結果がよい場合でもその他のデータを解釈し、より質の高い学習経験を提供するための教材の改善を行うことが大切であるとしている。ネイゼンソンとヘンダーソンはテストでの達成度の他に収集するデータを5つに分類し、教材の明瞭度、教材と学習者のレベルの一致度、学習者に要求されている行動と実際の行動の一致度、学習者の動機づけや態度のレベル、学習に要した時間に関する情報を集めることを提唱している。

さらに別の視点から、ディックとケーリー (1985) は、たとえテストの結果が悪くてもすぐに教材そのものの修正に取りかかるべきでないと指摘している。まずテストそのもの、次に学習課題の分析、さらに用いられた学習指導ストラジーを再点検し、それら相互間に整合性が保たれ、また学習者とのコミュニケーションが正しく行われていたかを確認し、その後に教材そのものの修正を考えるべきであるとしている。たとえば事後テストに悪い項目が含まれていれば、たとえ教材そのものが適当であっても得られたデータは思わしくないものとなる。また、教材の修正を行うにあたっては、それにかかるコストが修正による効果の上昇に見合うものかも考慮しながら改善の優先順位を決めることも必要な場合があるとしている。

V. CAI教材の設計開発への応用

授業のシステムの設計開発のモデルの多くがプログラム学習教材の設計開発技法の影響を強く受けて発展してきたため、モデルの応用は、教師が前面にたつ学級集団での授業の設計に比較して、CAI教材等の独立型メディアを用いた授業の方がスムーズに行われ易い。ガニエやウェージャーら (Gagné, Wager & Rojas, 1981; Wager & Gagne, in press) はCAI教材の設計にプログラム学習の技法が多く用いられていることを指摘し、さらに学習指導理論を活かしたCAIコースウェアの設計開発についての具体例を提供している。また、CAI教材は基本的には他のメディアを用

いた教材と同様な過程を経て設計開発されるべきであるとする一方、CAI特有のメディア属性（グラフィックスと文字を用いた授業内容の精巧度、問いかけと解答に応じたフィードバックの提供能力など）を活用することでコンピュータをより効果的に授業に使うことができるとしている。

CAI教材の形成的評価は、たとえばディックとケーリー（1985）によるモデルに従って、他のメディアを用いた授業の場合と同様に行うことができる。つまり、教材がプログラムされた後に数名の学習者を用いての1対1の形成的評価と改善、さらに小集団の評価や実地トライアウトと、形成的評価の3つの基本的な段階に従って行うわけである。個別学習が前提となるCAI教材の場合は、小集団の評価といっても一人一人がコンピュータに向かって個別に学習を進めることは1対1の評価と同じである。しかし、教材の設計開発者が学習に介在せずに行われる所が異なる。また、CAI教材に付随する使用者のアンケートや事後テストなどの印刷物も同様に形成的評価の過程を経るべきである。

一方で、CAI特有のメディア属性によって、形成的評価の過程に特別な注意を要する側面がある。第一に、CAI教材の構造を柔軟なものにし、学習者個々に合わせた授業状況を用意するために用いられる枝分かれ、誤答の種類に応じたフィードバック、あるいは任意のヒントなどの評価が困難なことが挙げられる。もし形成的評価に参加した学習者が特定の誤りを犯さなかったり、任意のヒントを選択しなかったりすれば、その部分は一度も学習者の目に触れないことになり形成的評価のデータは収集できない。逆に任意のヒントを強制した場合には学習の流れが設計されたそれとは異なるものになり、効果の測定が不正確になってしまう。

この問題に関しては解決策がいくつか考えられている。ゴラス（Golas, 1983）は、1対1の段階では枝分かれの機能を使わず、教材のすべての部分をひとつおき評価することを提唱している。この立場は、1対1の段階では枝分かれ機能の効果の検証よりも教材の誤りや不明瞭な点を探ることが優先されるべきだとするものである。また、小集団の段階では学習者をいくつかのグループに分けて、その効果の差異を調べる方法がある。たとえば鈴木、李とスミス（Suzuki, Li, & Smith, in press）はメニュー構造をもつCAI教材の形成的評価の際に、2つのグループに異なる学習順序を指示してヒントの使用頻度や学習効果の差異を調べ、優位な選択順序をアドバイスの形で教材のメニュー画面に加えた。このように、コンピュータの記録機能を利用して事後テストと任意選択との関係を調べることを形成的評価に取り入れることも可能である。

次に、形成的評価を設計開発のどの段階で、どの程度完成に近いCAI教材を用いて行うかという問題がある。たとえばCAIが授業内容を精巧に提示することができるという理由で選択された場合、その効果を予測するためには精巧な画面（高密度グラフィックスやアニメーションなど）を用いて形成的評価を行うのが最も安全である。しかし一方で、形成的評価以前に多額の開発費用が投入されている場合、教材の改善のためにこれまでに開発された部分を修正するのが実際的には困難である。

教材の完成度と形成的評価の時期の問題はプリント教材以外の開発コストの高いメディアを用いた教材開発全般に共通するもので、形成的評価の研究対象となってきた。例えばテレビを用いる教材の形成的評価では台本を用いる方法（グッドマン [Goodman], 1984）や未完成のフィルム等を用いての比較研究（バートンとアバーサ [Burton & Aversa], 1979）などが行われており、開発初期からの形成的評価が一般的に提唱されている。CAI教材においても、例えば前出のゴース（1983）は紙上の画面ディスプレイ案を用いての1対1の形成的評価を取り入れるべきだとしている。

CAI教材の柔軟な構造と早期からの形成的評価の効用を合わせて考えた際、開発のチャンクを小規模におさえ累積的に教材を完成させていく方略も効果的かつ能率的であると思われる（鈴木、出版準備中）。教師がコースジェネレーター等を用いて自作のCAI教材を開発する時などのように、形成的評価に用いる学習者の獲得が容易な場合には特に有効である。例えば、まずテストの部分を作成し、形成的評価を行う（第一サイクル）。次にドリル演習型の教材をテスト項目にフィードバックを加えることで作成し、形成的評価を実施する（第二サイクル）。最後にドリル演習型のプログラムに、導入・説明のフレームを加えて個人指導型（tutorial）の教材に拡張し、形成的評価を行うという具合である（第三サイクル）。この方略には、各サイクルでの形成的評価には完成に近い教材が用いられると同時に、小規模な開発の積み重ねなので教材の改善も比較的容易に行うことができるという利点がある。また、開発のチャンクを小さくすることで、実用的な教材が短期間のうちに完成することは現場に直結した教材開発に特に有効であろう。これは一つの案にすぎないが、いずれにしてもCAI教材の設計開発には形成的評価の過程をできるだけ取り入れる方向での工夫が求められよう。

VI. おわりに

本稿では独立型教材のシステム的设计開発における形成的評価について概観し、特にCAI教材の設計開発への応用に関しての考察を試みた。現在授業のシステム的设计開発に広く用いられている形成的評価の手法はCAI教材にも応用できるが、CAI特有のメディア属性に配慮する必要があることを示した。コンピュータのハードウェアの普及に伴い、多数のCAI教材が市販されるようになってきた。また、コースジェネレータなどのコンピュータ技術の発展によってCAI教材の開発が簡便になり、教師の手による開発も促進されるであろう。形成的評価が広く実践され、CAI教材の質の向上に貢献することが強く望まれる。

謝 辞

本稿の執筆にあたり、フロリダ州立大学ディック（Dick）教授に貴重なコメントをいただいた。また、本稿の出版に際して前回と同様に国際基督教大学石本菅生教授に援助いただいた。ここに感

謝の意を表したい。

参考文献

- AECT. (1977). *Educational technology: Definition and glossary of terms* (Vol. 1). Washington, DC: Association for Educational Communications and Technology.
- Baker, E.L. (1974). *Evaluating instructional programs*. Los Angeles: National Institute for Educational Researcher Training Task Force, UCLA.
- Baker, E.L., & Alkin, M.C. (1973). ERIC / AVCR Annual Review Paper: Formative evaluation of instructional development. *AV Communication Review*, 21, 389-418.
- Briggs, L.J. (1984). Comment: Trying to straddle four research cultures. *Educational Technology*, 24, (8), 33-34.
- Burkman, E. (1971). *Intermediate science curriculum study: Final report*. (Contract No. OEC 2 - 6 - 061762 - 1745). Tallahassee, FL: Florida State University.
- Burton, J.K., & Aversa, F.M. (1979). Formative evaluation information from scripts, scratch tracks, and rough cuts: A comparison. *Educational Communications and Technology Journal*, 27, 191-194.
- Cambre, M.A. (1981). Historical overview of formative evaluation of instructional media products. *Educational Communications and Technology Journal*, 29, 3 - 25.
- Cronbach, L.S. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672-683.
- Debert, R. (September, 1979). Revisionary tactics. *NSPI Journal*, 18-21.
- Dick, W. (1977). Formative evaluation. In L.J. Briggs (Ed.), *Instructional design: Principles and application*. Englewood Cliffs, NJ: Educational Technology Publication.
- Dick, W., & Carey, L. (1985). *The systematic design of instruction* (2nd Ed.). Glenview, IL: Scott, Foresman, and Company.
- 藤田恵暉 (1982). 「授業改善のための評価」大内・中野 (編) 『授業に生かす教育工学シリーズ第一巻・授業の設計と実施』図書文化.
- Gagné, R.M. (1985). *The conditions of learning and theory of instruction* (4th Ed.). New York: Holt, Reinhalt, and Winston. (邦訳) 金子・平野 (訳) (1982) 『学習の条件 (第三版)』学芸図書.

- Gagné, R.M., Wager, W., & Rojas, A. (1981). Planning and authoring computer assisted instruction lessons. *Educational Technology*, 21(9), 17-26.
- Golas, K.C. (1983). Formative evaluation of computerassisted instruction. *Educational Technology*, 23(1), 26-28.
- Goodman, I.R. (1984). Evaluating scripts for instructional media programs: A structured script evaluation procedure, *Educational Technology*, 24(3), 25-30.
- 梶田一他 (訳) (1973). 『教育評価法ハンドブック』第一法規.
- Komoski, P.K. (1974). ERIC / AVCR annual review paper: An imbalance of product quantity and instructional quality: The imperative of empiricism. *AV Communication Review*, 22, 357-386.
- Markle, S.M. (1967). Empirical testing of programs. In P.C. Lange (Ed.), *Programmed instruction*. Chicago: The sixty-six Yearbook of the National Society for the Study of Education, Part II, NSSE.
- Nathenson, M.B., & Henderson, E.S. (1980). *Using student feedback to improve learning materials*. London: Croom Helm.
- 沼野一男. (1976). 『授業の設計入門—ソフトウェアの教授工学』国土社.
- Parmer, E. (1974). Formative research in the production of television for children. In D.R. Olsen (Ed.), *Media and symbols: The forms of expression, communication, and education*. Chicago: University of Chicago Press.
- Parmer, E. (1981). Shaping persuasive messages with formative research. In R.E. Rice, & W.J. Paisley (Eds.), *Public communication campaigns*. Beverly Hills, CA: Sage Publications.
- Scriven, M. (1967). Goals of evaluation versus roles of evaluation: Formative and summative evaluation. In R.W. Tyler, R.M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. AERA Monographs on Curriculum Evaluation, 1. Chicago: Rand McNally.
- Suzuki, K. (in press). A short-cycle approach to CAI development: Three-stage authoring for practitioners. *Educational Technology*.
- Suzuki, K., & Li, R., & Smith, R. (in press). *The systematic development of a computer courseware: The Reiser-Gagné model of instructional media selection* (CET Technical Report). Tallahassee, FL: The Center for Educational Technology, Florida State University.
- Vedros, R.G. (April, 1986). *LVR Florida style: One state's response to improv-*

ing instructional materials. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Vergara, A. (1985). *The effects of selected formative evaluation procedures on the effectiveness of instructional materials.* Unpublished doctoral dissertation, Florida State University, Tallahassee, FL.

Wager, W., & Gagné, R.M. (in press). Designing computer-aided instruction. In D.H. Jonassen (Ed.), *Instructional designs for microcomputer courseware.* Hillsdale, NJ: Lawrence Erlbaum, Publisher.

Techniques of Formative Evaluation for CAI Courseware Design and Development

Katsuaki Suzuki
(Florida State University)

This paper provides an overview of formative evaluation in the context of the systematic design and development of stand-alone instructional materials with a special emphasis on the application of various techniques of formative evaluation to CAI courseware development. The historical advancement of formative evaluation concepts and techniques is first described from the introduction of the term by Scriven in 1967 to later works by Markle, Baker, Burkman, Palmer, Komoski, Nathenson and Henderson, and Dick and Carey. Next, the reasons for conducting formative evaluation are listed: (1) to improve the effectiveness of the materials using student data, (2) to contribute to the formation of instructional theories and models, (3) to keep the materials consistent with the entry behavior and characteristics of the target audience, (4) to provide a training opportunity for instructional designers, and (5) to facilitate later adoption of the materials.

After the three-stage procedure of formative evaluation is briefly introduced using the model by Dick and Carey (1985), potential problems with instructional materials and methods of possible revisions are provided in order to link formative evaluation data and consequent revisions. It is pointed out that although the general model of formative evaluation can be applied to CAI development, deliberate attention should be paid to the unique capabilities of the computer as an instructional medium. The discussion includes: (1) timing for conducting formative evaluation in relation to cost effectiveness of the production as well as the prediction of the effectiveness of the final product, and (2) techniques for evaluating CAI courseware that has adaptive structure to individual learner differences. Finally, an alternative 3-step approach to develop CAI courseware is proposed for the practitioners with limited resources.