

補 足 資 料 (文献紹介)

## 評 価 と は 何 か

EVALUATION: WHAT IS IT?

David Nevo

## Program Evaluation

---

A PRACTITIONER'S GUIDE FOR  
TRAINERS AND EDUCATORS

ROBERT O. BRINKERHOFF  
DALE M. BRETHER  
TERRY HLUCHYJ  
JERI RIDINGS NOWAKOWSKI

Kluwer-Nijhoff Publishing  
Boston The Hague Dordrecht Lancaster  
a member of the Kluwer Academic Publishers Group, 1983

## 評価：それは何か？

デビット・ネボ

“評価”という用語は、いろいろな意味に用いられている。また、評価の仕方もさまざまである。相異なる目的で使用される場合もあれば、適切な評価を決める基準が異なる場合さえありうる。ある人がどのような意味で評価という言葉を用いているのかを正確に知るためには、多くのことを調べる必要がある。

評価が論じられるときに、その評価という言葉がどのような意味で用いられているかを要約する際に役立つと思われる、10項目の質問をつぎに記述する。この10項の質問を用いれば、ある人が評価について論じたり、あなたが評価することを求められたときに、その人がどのような意味で評価という用語を用いているかを知ることができるであろう。さらに、あなた自身が「評価」という言葉をどんな意味で用いているか、あるいは次の機会に評価という言葉を用いる場合、どのように定義づけるかを明確にする上でも役立つと思われる。

### 1. 評価をどのように定義するか。

評価の特徴は何か。それをどのように知ることができるか。“測定”や“調査”とどのように違うのか。

### 2. 何のための評価か。

なぜ、評価するのか。評価の目的は何か。どのような役割をもっているのか。意思決定のためのものか。責任を果していることを説明するためのものか。資格認定や証明発行のためのものか。人々に動機づけをあたえるためのものか。プログラムの変更や改善のためのものか。さらには、そのほかの理由によるものか。

### 3. 評価の対象は何か。

評価できるもの、あるいは評価すべきものは何か。評価しようとしている“事象”は、学生か、教師か、プロジェクトか、プログラムか、機関か、あるいはそれ以外のものなのか。

### 4. 対象のどの側面、どの次元を評価しようとしているのか。

評価される事柄に関して、どのような質問をすればよいのか。どのようなタイプの情報を収集すればよいのか。評価しようとしているのは、対象のいかなる側面であるか～手段、方法か、影響・結果か、過程・実施についてか、スタッフとクライアントとのやりとりか、目標・計画についてか、コストや利益についてか、要求についてか、組織的特徴か、それら以外のものか。

### 5. 対象を判断する際の基準は何か。

判断したことをどのように解釈しようとしているのか。収集した情報に、いかに価値的な意味づけをしようとしているか。対象の“よい”あるいは“わるい”をどのように決めればよいのか。それを決めるときに優先すべきことは、目標の達成の度合か、プランの実施された度合か、ニーズにどれだけ合致しているかということか、社会的目標・理想の達成か、別個の対象との比較か、

規則・ガイドラインの堅守の度合か、聴衆・観客の期待との合致か、それら以外のことなのか。

6. 誰のための評価か。

クライアントは誰か。評価するための対象層は誰か。誰の情報ニーズに役立てられるのか。～自分自身のためか、学生のためか、教職員のためか、資金を出している機関のためか、一般大衆のためか、あるいはその他の人々のためか。

7. 評価をするにあたっての手段・手続は何か。

どのように評価に着手し、どのように次の作業に移っていくか。評価プロジェクトの主な段階はどのようなものか。評価の実施にあたって“最良”の順序はあるのか。

8. 評価にあたって、どのような調査方法を用いればよいのか。

情報収集としてどのような方法をとるか。評価にあたって、どのような調査の設計をすればよいか。評価の“最良”の方法は、テストやアンケートか、専門家による討論会か、実験的なデザインか、調査と相関分析か、人類学的手法・ケーススタディか、“司法的”裁判か、自然なアプローチか、あるいはその他のアプローチか。

9. 誰が評価をおこなうのか。

どのような評価者をあてるべきか。評価者が保有すべき技術は何か。評価者の権限や責任は何か。評価をおこなうのは、プロの評価者か、内部の評価者か外部の評価者か、評価の専門家か評価の対象分野の専門家か、通常の教職員か、それら以外の人々か。

10. 評価の判断の基準は何か。

よい評価とはどのようなものか、それはどのようにしたらわかるのか。よい評価の特徴はどのようなものか。“評価の評価”はどのようにすればよいのか。評価のありかたは、実際的で役に立つのか、正確で信頼性のある情報が提供できるか、現実的であるか、慎重で精密なことか、合法的・倫理的におこなわれることか、客観的・科学的であることか、あるいはもっとほかのことなのか。

この節の初めに述べたように、この10項の質問はいろいろな場合に使用できる。しかし、この10項の質問を実際に役立たせるには、つぎの二つのやり方がある。

(イ) 評価とは何かについて、あなた自身の考え方を組織化するために用いること。

(ロ) 他者が評価について論じているときに、どのような意味でその言葉を用いているのかを理解するために用いること。

\* 評価に関する文献からの、10項の質問に対する解答

評価に関する主なるアプローチを知っておけば、評価やそれに関連した問題に対して、理解を深める意味で有効と思われる。

1. 評価をいかに定義するか。

諸々の文献には、いろいろな評価の定義がみられる。Tyler (1950) は“教育目標が実際にどれくらい達成されたかを定めるプロセスである。”と定義している。これはタイラーが使い始めた定義であり、よく知られたものである。もう一つの定義は“意志決定のための情報を提供する。”というものである。これは Cronback (1963), Stufflebeam (1971), Alkin (1969) など指導的な教育学者によって唱えられたものである。これも広く知られている定義である。

“食いちがい (Discrepancy) の評価”の創始者である、Provus (1971) は“食いちがいが存在するかどうかを決定するための、ある尺度をもとにした達成の比較”と評価を定義している。最近では、評価の定義を“メリットや価値の評価”とすることで、かなりの合意が得られつつある。(Scriven, 1967; Glass, 1969; Stufflebeam, 1974) 教育評価に関する12団体の代表者17名からなる、評価の尺度についての合同委員会では、評価の定義を“ある目標の価値またはメリットを体系的に調査すること”としている。

このような評価の価値判断の側面を強調した定義は、潜在的に被評価者たちの間に強い懸念をよび起こし、評価反対論者に評価への反抗を引き起すことにもなりかねない。被評価者やクライアントにとっては、評価を価値判断を抜いて定義する方が、受け入れられやすいであろう。しかし、評価の重要であり、かつ必要不可欠な側面である価値判断を無視したのでは、評価への積極的な態度の育成はありえないであろう。評価への積極的な態度を伸長するためには、教育の諸分野における評価の建設的な役割を強調する、ということが考えられる。

2. 何のための評価か。

「形成的評価」(formative evaluation)と「総括的評価」(summative evaluation)とを区別したのはScriven(1967)である。このような区別の重要性を認識したのは、彼がはじめてではないが、この二つの評価の役割、機能について述べたところに意味がある。さらに、Stufflebeam(1971)はこの二つの機能について言及し、“意志決定に役立てるための、行動前に実施する評価”と“アカウントビリティ(どれだけ責任を果しているかを判定すること)に役立てるための、行動が終わった後の評価”とに分けている。このように、評価は二つの機能を果すことができる。前者の機能では、現在、行なわれている行動(あるいはプログラム、人物、製品など)の改善・発展のために評価が用いられる。総括的機能においては、評価はアカウントビリティ、証書授与、選抜や進級などのために用いられる。

評価についての文献にはあまり出てこないが、評価の第三の機能について述べておこう。それは、評価の「心理的、または社会・政治的機能」(psychological or socio-political function)である。評価は形成的目的だけに行われるのではなく、あるいはアカウントビリティやその他の総括的目的のみのために用いられるのではない。そのような事例が数多くみられるのも確かである。つまり、評価の目的がある事柄を広く知ってもらうためであったり、評価される人々がのぞましい行動をするように動機づけることであったり、宣伝のためであったりする。このような目的に評価を使用(あるいは誤用)することに対して、個人的感情がどうであろうとそれを無視することはできない。

確かに、評価の機能に“正しい”ものと“誤ったもの”があるわけではない。評価はいくつもの機能に役立てることが可能なのである。このような三つの機能以外にも、評価にはさらに多くの機能があると思われる。相異なる機能は相異なる評価法によって取りあつかわねばならない。

### 3. 評価の対象は何か。

あらゆる訓練の实在物を評価の対象とすることができる。その中でも学生や教職員は教育の分野ではいつも評価の対象となってきた。最近、プロジェクト、プログラム、教材、教育機関なども教育評価の対象となっている。評価に関する文献を調べなおしてみると、つぎの二つのことがいえる。

(a) ほとんど何でも、評価の対象となりうることであって、それを学生や教職員の評価に限定すべきではない。

(b) 評価の設計を進める上で、評価対象を明確に叙述することが大切である。

「何を」を評価しようとしているのか、明確に決めておくことが極めて重要である。それによって、収集すべき情報の種類が何であり、それをどのように分析すればよいかを決めることができる。つまり、評価の焦点をしぼるのに役立つ。また、対象を明確にすることによって利益関係のある者など、影響を受けている者の間の、価値観の対立や政治的脅迫の解決にも役立つことがある。

### 4. 対象のどのような側面、どのような次元を評価しようとしているか。

評価対象が選ばれたならば、次に対象の諸側面のうち、どの側面を評価しようとしているのか、決定しなければならない。初期の研究で評価は主に結果に焦点をあてる傾向があった。その場合、教育対象の評価とその機能の質的結果(つまり、成績)とを同じ意味に用いていた。しかし、最近の評価モデルでは、評価しようとする変数の範囲を拡大しようとする試みもみられる。(Stake, 1967; Stufflebeam, 1969; 1974; Stufflebeam, et al., 1971; Alkin, 1969; Provus, 1971)

例えば、Stufflebeam の CIPP モデルでは、評価対象の4つの側面に焦点をあてている。

(1) その目標、(2) そのデザイン、(3) その実施プロセス、(4) その結果。

このアプローチによると、教育プロジェクトの完全な評価をするには、  
(a) その目標のメリット、(b) その計画の質、(c) そのプランが実行された割合、(d) その結果の価値。

この4つの評価をしなければならないというのである。

「結果の評価」(outcome evaluation)を「総括的評価」(summative evaluation)と混同したり、「プロセス評価」(process evaluation)を「形成的評価」(formative evaluation)と混同したりしてはならない。「結果の評価」とは、結果に焦点をあてた評価であり、「プロセス評価」とは結果よりもプロセスに焦点をあてた評価のことである。この二つとも「総括的」評価にも「形成的」評価にもなりうるのである。

#### 5. 対象を判断する際の基準は何か。

教育評価において、評価対象やその対象のある側面を判断する際に用いる基準の選定がかなり難しい。“評価は目標が達成されたかどうかを決定することに限定すべきだ”という立場もある。その立場をとれば基準を部分的に無視することができるので評価作業は容易なものとなる。つまり、目標の内在的価値を判断することはせず、単に“目標への達成度”を評価の基準として用いる。それでは、無意味な目標であったり、達成されるだけの価値のない目標であった場合にはどうするのか。そのようなものであるにもかかわらず、評価の基準として用いるというのはおかしいと思われる。

けれども、“目標の達成”が評価の基準の一つの基礎であることは疑う余地がない。この他に、評価基準の基礎として次のようなものが考えられる。

- 実際のクライアントや潜在的クライアントのニーズ
- 理想的あるいは社会的価値
- 専門家など適当な集団が定めた、既知の標準のセット
- その他の対象と比較しての効用性や品質

特定の対象についての評価に用いる基準はその対象の独自のコンテキストやその評価のもっている機能とのかかわりで決定しなければならない。この点は評価専門家の誰れもが認めているところである。多くのケースで、数多くの基準からあるものを選定する権限を評価者自身ではもっていない。また、その権限をもつべきでもない。しかし、そのような選択が評価者により行なわれようと、他者によって行なわれようと、なされた選択に対してそれをうまく意味づけるのは評価者である。

#### 6. 誰のための評価か。

評価がすべてについて役立つには、まず特定のクライアント、聴衆に対して役立つものでなければならない。評価研究をみると、誰が“最も適切な”聴衆であるかを書いていない場合が多い。これは重要な問題であり、次の三点の主張がみられる。

- (1) 一つの評価のクライアント、聴衆は複数でありうる。
- (2) 評価の聴衆が違えば、評価のニーズも違ってくる。
- (3) 評価の初期段階に、評価の聴衆は誰か、彼らの評価へのニーズは何か、を明確にする必要がある。

評価へのニーズの違いは、いろいろな面にあらわれてくる。つまり、収集すべき情報の種類、データ分析のレベル、評価の結果の報告形式などにあらわれる。また、評価ニーズのすべてに応えられない場合もありうる。その際には、どの評価ニーズに応えようとするのかを明確にする必要がある。

#### 7. 評価をするにあたっての手段・手続は何か。

評価実施のプロセスは、評価理論のとらえ方によって違ってくる。目標が達成されたかどうかを決定する活動を評価とみる理論に立てば、つぎのような評価プロセスとなろう。

- (1) 行動目標として、目標を記述する。
- (2) 測定用具を開発する。
- (3) データを収集する。
- (4) 収集した事実を解釈する。
- (5) 報告書を作成する。

意志決定のために情報の提供をするものとしての理論にたてば、評価プロセスはつぎのようになる。

- (1) 意志決定者の情報へのニーズの確定。
- (2) 適切な情報の収集。
- (3) 意志決定者への評価情報の提示。

さらに、スタッフ（教職員）の理解、関与、知識を高めることをねらいとした評価であればつぎのようなサイクルをたどる。

- (1) 問題領域の確定。
- (2) スタッフの期待や価値観の意味づけ。
- (3) 実行にあたっての情報の収集。
- (4) 不一致があるということについて報告書をスタッフに提供し、その矛盾の解決に役立てる。

評価の実施にあたって、いかなるプロセスをとるのが最良であるのか、一致した見解はみられ

ない。しかし、評価の実施にあたって評価者と聴衆との十分な相互作用が必要であり、それらを通じて評価ニーズを確定し、結論が得られたらそれについて意見を交換することが重要である。この点については評価専門家の見解はおおむね一致している。“評価はデータの収集・分析という技術的な活動に限られたものではない”のである。

#### 8. 評価にあたって、どのような調査方法を用いればよいのか。

評価研究に用いられる、いろいろな調査方法の有効性について議論は多いが、(Provus, 1971, Stufflebeam, et al 1971) さまざまな調査方法が近年、導入されるようになってきている。

伝統的な実験的・準実験的な方法(CampbellとStanley, 1963)に加えて、自然的な方法(GubaとLincoln, 1981; Patton, 1980)、裁判法(Wolf 1975)、システム分析などの方法が認められるようになってきている。

ある方法がすぐれていることを強調する方法論者もいる。しかし、一般的には折衷的アプローチが評価の方法論としてよいとされている。このアプローチはどの評価の目的に対してひとつの方法が最良であるという見解をとらず、特定の評価の目的に適合する方法を探すべきであるというものである。

#### 9. 誰が評価を行なうのか。

専門集団になるために、“良い”評価者の特徴は何であり、そのための評価者の教育をどのようにやればよいか、評価者は熱心に考えてきた。有能であり信頼される評価者になろうとすれば、多くの特徴を備えていなければならない。例えば、測定や調査の方法に関する技術力、社会的なコンテクトや評価対象の内容に対する理解、人間関係の処理技術、人間としての誠実さ、客観性、組織の中での権限・責任をわきまえているといった特徴をもっていることが必要であろう。

これらのすべての資格を備えた人を見出すのはむずかしい。ゆえに、チームを組み、評価を行なう必要がある。さらに、特定の評価の仕事に最もふさわしい特徴を備えた人を選ぶ必要もあるわけである。

また、誰が評価を行なうかを決める際に、つぎの二つの区分が重要とされている。第1に、内部評価者と外部評価者との区分である。プロジェクトの場合、内部評価者とは、そのプロジェクトに雇用され、プロジェクトの管理者側に直接報告を行なう者をさす。内部評価者のもっている客観性、外部への信頼性はもちろん、外部評価者の場合とは異っている。それに対して、外部評価者はプロジェクトが直接に雇った者ではなく、プロジェクトからかなりの独自性をもっている。

第2に、プロの評価者とアマの評価者との違いである。この区分はScriven(1967)が述べているもので、評価者の質に関する価値判断をするものではなくて、訓練や専門性に関する重点の置き方の違いをいったものである。アマの評価者とは評価を主たる専門にしていない者をさし、その人々が評価に加わったとしても評価の仕事の一部を行なうに過ぎない。それに対して、プロの評価者とは、集中的に評価に関する教育を受けた者をさし、その人々の主たる責任は評価の遂



行それ自体である。持論、アマの評価者の評価技術は、プロの評価者に劣る。しかし、アマの評価者はプロジェクトに固有の評価ニーズを心得ていることもあり、評価しようとするプロジェクトのメンバーと、より親密な関係をつくりあげられる利点がある。

これら2つの区分は互に独立したものであり、内部アマ評価者、外部アマ評価者、内部プロ評価者などがありうる。

#### 10. 評価の判断の基準は何か。

教育活動についての評価のための基準を開発する試みが近年、いくつか行なわれている。

Boruch と Cordray (1980) はそのような基準を6組、分析してそれらの間にはかなりの重複や類似性があると報告している。最もよく検討されている、総合的基準は教育評価基準に関する共同委員会(1981)が発表した基準である。これがおそらく最も広くコンセンサスを得ているものと思われる。この基準を設定したのは Stufflebeam 博士を議長とする17名から成る委員会であり、メンバーは教育評価に関する12の団体の代表者である。この委員会が作成した基準は30項目であり、4つの群に分類されている。その4群はつぎのごとくである。

- Ⓐ 有用性の基準（評価が実際の情報ニーズに役立っているかどうかをみる。）
- Ⓑ 実行可能性の基準（評価が現実的で適格なものかどうかをみる。）
- Ⓒ 適切さの基準（評価が合法的・倫理的であるかどうかをみる。）
- Ⓓ 正確さの基準（評価が技術的に十分な情報を伝えているかどうかをみる。）

表1がその30項目の基準リストである。

表1

##### A. 有用性の基準

1. 聴衆の確定
2. 評定者の信頼性
3. 情報の範囲と選択
4. 評価の解釈
5. 報告書の明晰さ
6. 報告書の普及度
7. 報告書の適時性
8. 評価の影響力

##### B. 実行可能性の基準

1. 実際の手続き
2. 政治的実行可能性
3. コスト効果性

##### C. 適切さの基準

1. 公式の義務
2. 利害の衝突
3. 充分でかつ素直な表現
4. 公衆の知る権利
5. 人権
6. 人的相互作用
7. バランスのよい報告
8. 財政上の責任

##### D. 正確さの基準

1. 対象の確定
2. コンテキストの分析
3. 記述された目標と手続き
4. 防御できる情報源
5. 妥当性のある測定
6. 信頼性のある測定
7. 体系的データ制御
8. 量的情報の分析
9. 質的情報の分析
10. 適正な結論
11. 客観的な報告

これらの基準のすべてを充す評価があると考えるわけではない。これらの基準のうち、どれが相対的に重要なものであるかについては決まっているわけでもない。しかし、このような総合的基準やその開発の原理は、評価の発展上、また評価の質の評定上、考慮すべき重要な問題である。このような意味でこのリストは貴重なよりどころとなる。

Cronbach (1980) は“評価の遂行に関連する基準は、評価の結果に関する基準ほど重要ではない”と考えている。このような見解は、プログラムの改善に有効な評価が“最もよい”評価であることを示唆しているのである。

#### \* 教育訓練をめぐる評価のとらえ方について

この序章では、他者およびあなた自身が評価をどのように考えているか、を整理する手がかりについて検討した。評価に関する相異なる見解や定義を示したが、これらを見ると今までに論述された意見や潜在的な意見がかなり広範囲にわたっていることがわかった。しかし、そうであるからと言ってわれわれが同意できる統一的な評価アプローチがないという意味ではない。それは評価概念の解釈と他者が教育訓練や開発プログラムの評価を実施するのを助ける経験とを結びつけた見解である。

われわれの見解が“正しく”また最良のものと考えているわけではない。しかし、この見解は論理的にすぐれており、かなり有用なものであると思っている。

一般に、評価を教育訓練や職業的準備の一部として欠くべからざるものと思っている。教育訓練をうまく実施するには、教育訓練のニーズ、目標、デザイン、実施面の活動、直接的効果、長期的効果などについて評価する必要がある。われわれは評価を機能的にとらえている。つまり、ある目的～教育訓練に役立てる、という目的で評価が行なわれる。特に、評価は役に立つものでなければならない。評価は教育訓練プログラムの重要な側面について適切な決定をするのに用いられる。ここでは、このような題材に照らして10項の質問にどのように解答すればよいか、を記述しよう。

##### 1. 評価をどのように定義するか。

評価とは、教育訓練のメリットや価値を測るために職業的開発や訓練プログラムの諸側面に関する体系的な調査である。

##### 2. 何のための評価か。

評価は、教育訓練プログラムの開発・実施・繰返しの必要性に活用する。それは1つ、あるいは複数の目的にそって利用される。例えば、特定のプログラムの改善（形成的評価）、アカウントビリティや選抜（総括的評価）、知識への動機づけ・知識の習得やスタッフなどの支持を得る

こと（心理的評価）などのために役立てる。

3. 評価の対象は何か。

あらゆるものを評価の対象としうるが、これらの題材においては評価対象は教育訓練プログラム、あるいはその要素の一部となる。

4. 対象のどの側面、どの次元を評価しようとしているか。

評価を教育訓練プログラムに活用する際、評価の焦点は目標、ニーズ、訓練の設計、実施、処理、訓練の効果などにあてられるべきである。

5. ある対象を判断する際の基準。

教育訓練の評価に際しては次のような基準を考慮すべきである。

- (a) ニーズ、目標、価値に込んでいるかどうか。
  - (b) 利用できる資源や機会を最適に利用しているかどうか。
  - (c) 効果的な訓練実践を着実にこなしているかどうか。
  - (d) 意図した目標、またはその他の目標を達成したかどうか。
- いくつかの基準が併せて用いられる場合も多い。

6. 誰のための評価か。

評価は直接的な関係者や間接的な人々の情報ニーズに役立てる。評価に際して、これらの関係者を慎重に見極めるとともに、彼らのニーズや興味を決定すべきである。

7. 評価の実行にあたって、どのような手段、手続きをとるか。

評価はつぎの7つの機能に関する決定や活動を含むべきである。

- (1) 評価の焦点化
- (2) 評価の設計
- (3) 情報の収集
- (4) 情報の分析
- (5) 評価により得られた情報の報告
- (6) 評価の管理
- (7) 評価の評価

これらはこの順序に行なう必要はない。一つの機能が繰返されることもよくある。これらの7つの機能を行なうためのガイドラインやオプションは「ソースブック」に書いてある。

8. 評価に際して、どのような調査方法を用いればよいのか。

折衷的なアプローチをとるのがよいと思われる。評価に際して、用いられる調査方法は行動科

学やそれに関連する分野のものがよい。そのような方法は特定の評価の目的にうまく適合するからである。技術上の現水準を考えると、ア・プリオリにある特定の調査方法のみを用いることはない。

9. 誰が評価を行なうのか。

つぎのような立場の個人、あるいはチームが評価を行なうのがのぞましい。

- (a) 調査の方法論やデータ分析の技術に関して広範な能力がある者。
- (b) その評価プロジェクトが行なわれる社会的コンテクスト、あるいはその評価プロジェクトに関する固有の内容を理解している者。
- (c) 人間関係を適切に維持し、評価にかかわりのある個人や集団と親密な関係をつくりあげていく能力をもつ者。
- (d) 以上のすべての能力を総合するための概念的なフレームワークをもっている者。

10. 評価の判断の基準は何か。

評価にあたって、つぎのような基準を充足し、かつ適切なバランスをたもつことがのぞまれる。

- (a) 有用性（有効であり、実際的事であること。）
- (b) 正確さ（技術的に十分であること。）
- (c) 実行可能性（現実的で慎重なものであること。）
- (d) 適切さ（合法的、倫理的に実施されること。）

実施する価値のない評価をいかにうまく実施してみても意味のないことである。

## REFERENCES

- Alkin, M.C. "Evaluation Theory Development." *Evaluation Comment* 2, 1969: 2-7.
- Alkin, M.C. quoted in *Educational Research & Development Report*, Volume 3, Number 1, Winter 1980, pp. 8-12.
- Boruch, F.R. and Cordray, D.S. *An Appraisal of Educational Program Evaluations: Federal, State and Local Agencies*. Evanston, IL: Northwestern University, 1980.
- Brinkerhoff, R.O. Evaluation of Inservice Programs. *Teacher Education and Special Education*. Vol. III, No. 3, Summer, 1980. pp. 27-38.
- Campbell, D.T. and Stanley, J.C. "Experimental and Quasi-Experimental Designs for Research of Teaching." In N.L. Gage (Ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally, 1963.
- Cronbach, L.J. "Course Improvement through Evaluation." *Teachers College Record* 64, May 1963: 672-683.
- Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.E., and Weiner, S.S. *Toward Reform of Program Evaluation*. San Francisco: Jossey-Bass, 1980.
- Dornbusch, S.M. and Scott, W.R. *Evaluation and the Exercise of Authority*. San Francisco: Jossey-Bass, 1975.
- Glass, G.V. *The Growth of Evaluation Methodology*. Research paper no. 27. Boulder, CO: Laboratory of Educational Research, University of Colorado, 1969 (mimeo).
- Guba, E.G. and Lincoln, Y.S. *Effective Evaluation*. San Francisco: Jossey-Bass, 1981.
- Joint Committee on Standards for Educational Evaluation, *Standards for Evaluations of Educational Programs, Projects and Materials*. New York: McGraw-Hill, 1981.
- Nevo, D. "The Evaluation of a Multi-Dimensional Project." In A. Lewy *et al.*, *Decision Oriented Evaluation in Education: The Case of Israel*. Philadelphia: International Science Services, 1981.
- Patton, M.Q. *Qualitative Evaluation Methods*. Beverly Hills, CA: Sage Publications, 1980.
- Provus, M.M. *Discrepancy Evaluation*. Berkeley, CA: McCutchan, 1971.
- Scriven, M. "The Methodology of Evaluation" in R.E. Stake (Ed.), *Curriculum Evaluation*, AERA Monograph Series on Evaluation, No. 1. Chicago: Rand McNally, 1967.
- Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O., and Provus, M.M. *Educational Evaluation and Decision-Making*. Itasca, IL: Peacock, 1971.
- Stufflebeam, D.L. *Meta-Evaluation*. Occasional Paper Series, The Evaluation Center, Western Michigan University, December 1974.
- Tyler, R.W. *Basic Principles of Curriculum and Instruction*. Chicago, IL: University of Chicago Press, 1950.
- Wolf, R.L. "The Use of Judicial Evaluation Methods in the Formation of Educational Policy." *Educational Evaluation and Policy Analysis*, 1, 1979: 19-28.
- Worthen, B.R. and Sanders, J.R. *Educational Evaluation: Theory and Practice*. Belmont, CA: Wadsworth Publishing Co., 1973.