

第3章 eラーニングにおける評価技法

学習目標:カークパトリックの4段階とは何かを、例を挙げて説明できる。

目標明確化の3要素とは何かを、例を挙げて説明できる。

出入口を明確にするための3つのテストの用法を説明できる。

学習課題の5分類とそれぞれの評価方法を説明できる。

形成的評価の目的と技法のあらましを説明できる。

本章の概要

代表的なIT系WB Tコース17事例を調査した結果、学習目標の設定やレディネスの明示が極めて低い結果を示していた。目標を研修受講者に明示することが評価技法の第1である。

カークパトリックの4段階とは、反応、学習、行動、業績である。40年以上も企業内教育の評価技法の標準として用いられてきた。

目標明確化の3要素とは、目標行動、評価条件、合格基準である。目標を明確化しなければ、到達したかどうかを確かめることもできないし、目標が妥当なものかどうかあいまいになる。

出入口を明確にするために用いるテストは、前提テスト、事前テスト、事後テストある。この3つを組み合わせることで、誰に何を教えるのかを明確にして、資格・必要性・効果を明らかにしていくことができる。

学習目標の分類枠としては、ブルームの目標分類学とガニエの学習課題の5分類が広く知られている。5分類とは、言語情報、知的技能、認知的方略、運動技能、態度である。到達度の測定方法が異なるのみならず、到達を支援する最善の方法が異なるという観点から学習目標の性質にもとづいて課題を分類したものである。

形成的評価とは、教材が完成する前に、教材の品質向上を意図して試行データを集める工程である。形成的評価は、とくに独習型の教材開発においては重要であり、様々な技法が蓄積されている。

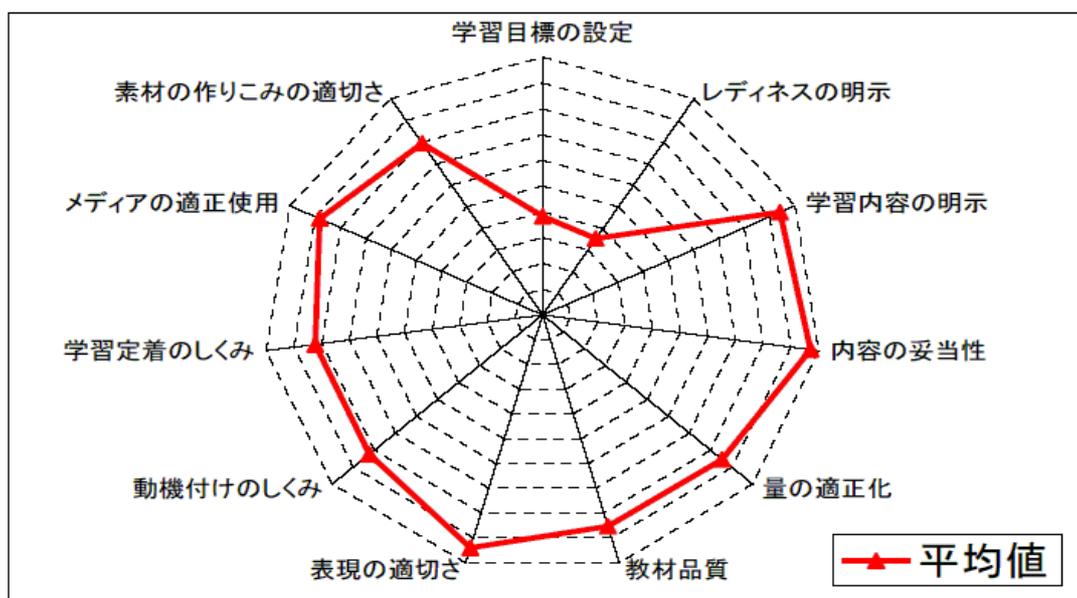
第1節 学習目標を知らせることの功罪

評価の視点を整理した研究として、CAITが機械システム振興協会からの受託研究として実施した「遠隔学習システムの利用促進に関する調査研究」がある（機械システム振興協会、2001：2002年の調査報告書もあり、2001年度調査との比較結果も報告している）。この調査では、代表的なIT系WBTコース17事例を選択し、内容の知識をもつ調査員があらかじめ設定した調査項目に沿って評価を行った結果、学習内容・評価に関して図表3-1のような結果を得ている。とくに、学習目標の設定やレディネスの明示が極めて低い結果を示していることが読み取れる。

この調査にあたっては、コース開始後の記述のみを調査対象としたため、コース選択時の参考のためにWBTサイトに記述されている学習目標やレディネスがたとえあったとしてもそれは数字には反映されていない。それが結果の数値が低い原因の一つであると説明しているが、それを割り引いたとしても、良好な結果とはいえない。また、同調査報告書は、「国外で制作されたものを日本語化したものに、学習目標やレディネスを明示しているものが多い傾向がある（p.17）」としている点も注目に値する。

学習目標を知らせることは、評価の第一歩である。様々な評価技法を本章では紹介する。

図表3-1：IT系WBTコースの学習内容に関する評価結果



出典：機械システム振興協会（2001）「遠隔学習システムの利用促進に関する調査研究報告書」（CAIT）p.17

第2節 コークパトリックの4段階：企業内教育評価の40年選手

3-2-1：なぜ教育評価法が求められるのか

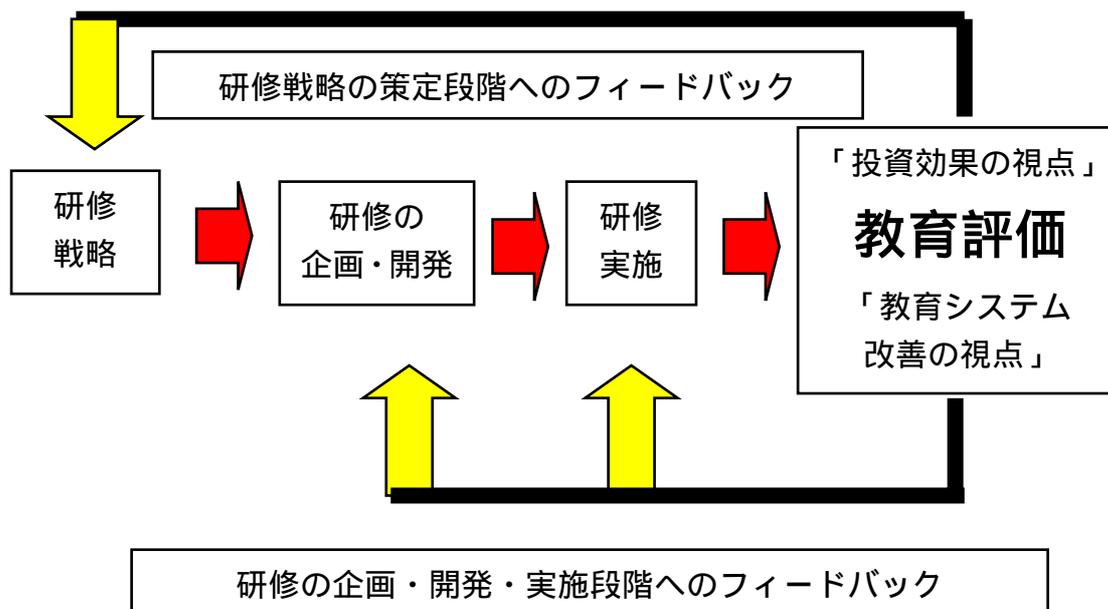
情報化社会になってビジネスはますますソフト化、ソリューションビジネス化が加速している。さらに業務環境、顧客ニーズ、技術の変化の速さが増し、教育への重要性は増している。しかし日本では社員教育への多大な投資に対して、その投資効果は明らかではない。その結果教育の満足度、業務遂行への役立ち度が不明で教育をコストと考える文化から、景気や業績が苦しくなるとコストカットの対象となり、教育縮小の憂き目にあっている。

教育の評価をすることによって企業教育の企業への貢献度を出来る限り明らかにして、トップの正しい認識を得ることにより安易にコストカットに甘んじることのないようにしなくては、企業の将来を誤らせる要因にもなる。また教育の評価を実施して、教育のPDCAをまわして、運営法や仕組みを改善していかねばならない。

* P D C A = P l a n ・ D o ・ C h e c k ・ A c t i o n

教育評価の目的を図表3-2に整理しておく。研修は研修の企画・開発・実施段階へのフィードバックと研修戦略の策定段階へのフィードバックがあり前者を形成的評価、後者を統括的評価ということがある。

図表3-2：企業内教育における教育評価の目的



教育評価の視点は「投資効果の視点」と「教育システム改善の視点」に分けられる。「教育システム改善の視点」は教育の仕組みや教育プログラムなどの改善に役立てる評価である。カリキュラム、教材、インストラクター、学習環境に関するデータを収集し、改善点あるいは強化すべき点を明らかにする。「投資効果の視点」は経営の視点から当該プログラム継続の価値を判断するものである。投資効果である教育目標の達成度を把握し、投資（研修）継続の経営判断を行う。

3-2-2：カークパトリックの4段階評価モデル

カークパトリックの4段階評価モデルとは、図表3-3に示す、反応 [Reaction]、学習 [Learning]、行動 [Behavior]、業績 [Results] の4段階で研修の評価を行うことを提案したモデルである。

図表3-3：カークパトリックの4段階評価モデル

レベル1；Reaction（反応）	満足
レベル2；Learning（学習）	理解
レベル3；Behavior（行動）	実務での活用
レベル4；Results（業績）	業績への貢献

カークパトリックはアメリカの経営学者で1959年にこのモデルを発表した。40年以上経た今でもカークパトリックの評価モデルは共有の認識として変わることなく使われている。カークパトリックは今でもASTDなどで毎年必ず講演をし、多くの人を集めている。それだけ親しまれ、常識となった評価法である。カークパトリックの評価はよくレベルといわれるがステップと表現することもある。これはレベル1からやがてレベル4に行くまでのステップという意味も含まれているという。

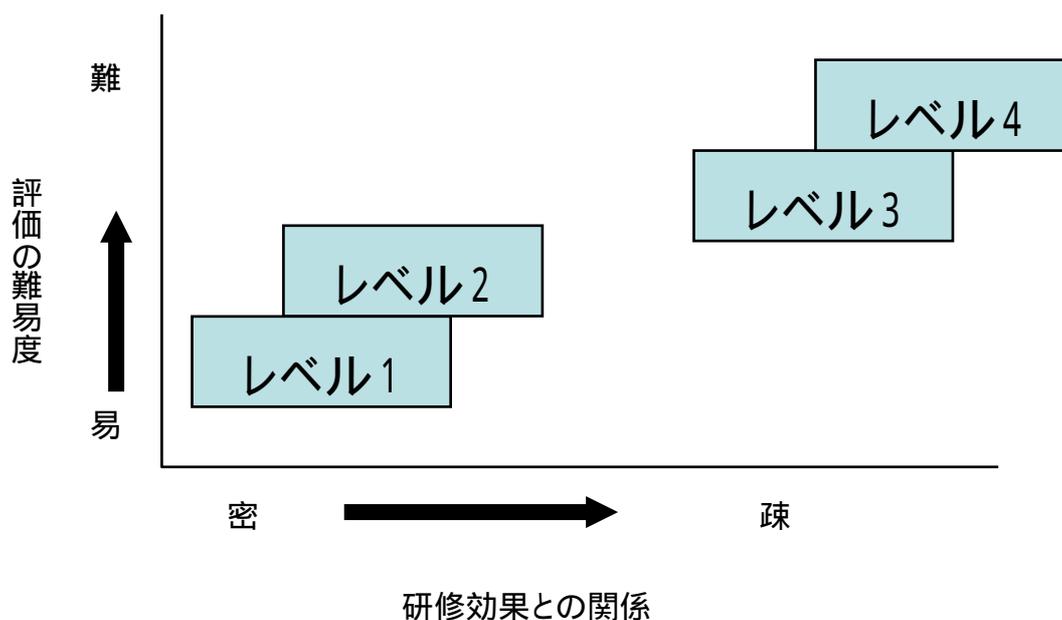
カークパトリックの評価法の背景にある、アメリカでの評価・研修のトレンドをまとめると、（1）受講者はお客様であるであるという認識・（2）事業の戦略、目標、方針に答え得る研修の提供・（3）社内研修部門も社外研修部門との競争・（4）顧客満足度を高めるため教育評価に力を入れ研修内容改善に努めている・（5）評価システムを電子化し速やかな評価を実施している・（6）評価は当然である、という特徴がある。

インストラクショナル・デザインでは受講者はお客様であるという認識ですべてを捉えている。教育する側は教える一段高いところにいる存在であるという概念は無い。教育の目的は事業の戦略、目標、方針に応え得る研修の提供にあり、そのために評価をし、改善をしていくための方法として評価を使っている。社内研修部門はインストラクショナル・デザイナーなど専門職により、運営されており、質的にもコスト的にも社外の研修部門との競争しているという認識である。そのために顧客満足度を高めるため教育評価に力を入れ研修内容改善に努めている。

評価のためには多くの学習者と連絡を取り、データをまとめなければならない。そのために評価システムを電子化し速やかな評価を実施している。日本では教育の評価というと、人格や能力全体の評価のイメージがあり、評価はなじまないがアメリカでは教育は投資であるという考えから、評価は当然であると受け入れられている。

アメリカにおける教育評価の実施率を見ると、レベル1評価は、92～95%出実施されているのに対して、レベル2評価は30～34%、レベル3は11～14%、レベル4評価に至っては2～3%という低い実施率に留まっている（ASTD提供、単位企業）。図表3-4に、評価レベルと難易度・研修効果との関係を図示する。レベル1と2は教育終了直後に評価を行い、レベル3と4は教育終了後3～6ヶ月後に行われるという。アメリカでもその実施率はおおよそ上記データの通りであるが、日本では調査してもまだレベル3と4を実施しているという報告は聞くことができない。

図表3-4：評価レベルと難易度・研修効果との関係



3-2-3：教育評価手法の実施例

レベル1（学習者の反応をつかむ）

レベル1の満足度調査には、受講者、インストラクター、M g rへのアンケートや、苦情システムによる情報収集、あるいは電子化されたアンケートなどがある。Q & A方式を基本とし、具体的には、感想文を書かせる。客観性を高めたいときには、3～10の選択肢にマークすることにより、定量化し、フリーコメントで補足させる。

コースがいくつかのセッションに分かれているときには出来る限り、セッションの直後にアンケートすることが望まれる。また、各セッションが有機的つながりを持っているときにはコース終了時に全体的視点からまとめて個々の評価をすることも良い。

レベル2（知識やスキルの習得状況を測る）

レベル2（理解度）の評価方法には、テストの合格率、ロールプレイング、習得スキルのチェックシート、前後テスト、電子メールによる前後のスキル比較などがある。テストは出来るだけ簡便な方法で行うことが望ましい。

テストの目的は受講者の知識・技能レベルの評価ではなく、コースのアセスメントと改善のためのデータ収集であることをアピールする必要がある。テスト内容はパイロットコースを活用して、信頼性を高めることが必要である。

レベル3（業務への活用度合いを把握する）

レベル3（実務での活用度）評価には、上司による作業観察、第3者によるスキル観察、受講者・上司・部下へのアンケートやインタビュー、研修前後の行動比較、対照群との行動比較、電子メールによる顧客や社員の満足度調査などの方法がある。

習得したものが実務に反映されることが教育における投資効果（ROI； Return On Investment）把握の前提条件とすると、レベル3（行動）評価は非常に重要なチェックポイントである。被験者としては本人だけではなく、上司などの観察者も考慮すべきである。さらに、調査の中立性を保つ意味からも面接調査は社外の調査機関の面接者を立てる場合がある。

レベル4（業績への貢献度を求めて）

レベル4（業績への貢献）の評価には、研修前後の作業時間比較、研修前後の実績データ比較、研修前後の収入比較（コンサルタント会社）、対照群との収入比較などがある。技能教育の場合には、貢献度を比較的に定量化しやすい。また教育の成果とその結果の相関がつけにくいこともあり、調査を困難にしている。

調査のテクニックとしては結果に至る重要なステップの改善状況から結果を推定する方法も取られる。例えば生保の営業マンは初めての顧客に訪問する際の訪問承諾の電話が契約の成否に大きな影響を与えると判っている場合、電話での訪問アポ取り教育で承諾取り付け率が2倍になればその営業マンの生産性は2倍になるというような成果の推測である。

3-2-4：教育評価の導入にあたって

米国の環境では、評価への違和感がない、あるいは、研修コースの設計・開発にIDを採用することで評価があるのが前提になっている。それに対して、日本の環境は、教育の評価を人物評価と同一視しがちで、難しいものと考えており、評価結果をストレートに受け入れないことに注意が必要である。

教育の評価が日本で行われない理由は2つある。ひとつは教育での評価が全人評価のようにとる文化が日本ではあったことである。少し前まで幹部候補者を集め研修でふるいにかけるというようなことをやっていたことが教育評価への足かせになっている。さらに現場でも教育部門に全面的信頼がなく、例え仕事に直結する教育を行っても素直に受け入れない土壌がある。これらの文化を覆していくためにも、業務に直結した仕事の遂行能力を養成する教育の実践の実績を積み社内の信頼をひとつずつ積み重ねて行くより他、手段はないのであろう。

教育の評価は理屈の上では新しい教育システムの導入と同時であるのであるが、前述した文化的歴史もあるので、まず業務に貢献する教育を提案し、その実績を積んでから、少しずつ評価法を導入するというのが、日本での現実解なのであろう。インストラクショナル・デザインのよいことを実践しても素直にその良さを理解されずに、誤解に基づく、抵抗に遭遇することは十分想定できる。

教育評価を円滑に導入するためには、教育は業務遂行能力の向上に寄与してこそ価値があるという共通認識が必要である。設計・開発・実施に体系的なアプローチを採用し、評価結果を積極的に活用できる仕組みを作る。そして、評価法の定着には時間が掛かるので、ある程度の猶予時間が必要だと思われる。



コラム：A T & Tにみる教育評価の理念



A T & Tを例に、教育評価の理念を見てみよう。人材育成の責任の所在については、研修部門と事業部に分ち合われ、それぞれが専門化した役割を担う体制が整っている。研修に対する考え方としては、研修と実務の境がなくなっている。研修は事業に影響を与えたときに価値あるものとなる。教育評価の目的は、レベル2、3、4のデータを収集し事業への影響度合いを確定することと考えられている。

企業内教育の進化した会社では教育部門は現場の事業部内に所属し、その会社全体の方向づけや教育手法の統一化は教育部門を横串にした、ミーティングなどで行っている例が多い。ある程度の規模では、教育を業務に寄与させるためには事業部内にあるのが教育目標を策定するには良い方法である。さらに巨大企業などでソフィスケイトされたシステムにより、会社全体を纏めながらも、個人中心の教育、情報提供、コンピテンシー・マネジメント機能ができるのであれば、H C M（ヒューマン・キャピタル・マネジメント）として一部門で管理することも可能かもしれない。

企業内教育の評価は日本ではコンピテンシー・マネジメントと並んでその実践に最も苦勞のする仕組みであろう。情報化社会になり、ユーザーのニーズ変化、技術、サービス内容も激しく変化し、その変化に対応するための学習組織の構築はこれからの企業にとって必須の責務である。人が糖尿病の恐れがあっても、運動もせず、飽食する人は多いように、時代への最適化に挑戦する企業はほんの一部かもしれない。しかし変化はゆっくりしかも速くやってくる。備えの確かな企業や学校が生き残ることだけは確実である。

（第2節 執筆担当：小松 秀園）

第3節 目標明確化の3要素：目標行動・評価条件・合格基準

目標明確化の3要素とは、目標行動、評価条件、合格基準である(鈴木、2002：第3章を参照、Webサイトに練習問題あり)。システムのアプローチで最も重要なのは、何を目標にするかを誰もが分かる形で記述することだとされ、目標の明確な記述方法がいろいろと提案されてきた。カークパトリックの評価の第2段階(学習)の成否を確かめるとともに、学習の焦点化を図る意味で、学習開始時に学習者に公表しておく(cf.ガニエの9教授事象のうち第2事象)ことにも使われる。

とくに評価条件について、考えておこう。暗記することが勉強だと思っていると、何も見ないでテストするのが当然と考える。しかし、「見ながらでも出来ればよい」と考えることで、不要な暗記を強要する必要がなくなる。ジョブエイド(E P S S)の発想につながる。

次にあげる2つの目標のうち、どちらがより明確に書かれているであろうか(鈴木、1987)。

- A. 教授目標を明確化する方法について理解する。
- B. 教授目標を明確化するための3つのポイントのリストをみながら、与えられた未知の目標がそれに従って書かれていない場合にすべて直すことができる。
ただし3つのポイントとは、目標の行動化、条件の記述、基準の記述とする。

目標Aと目標Bを比較すると、AよりはBの方が、何をどの程度まで教えたいのかが明らかであり、読み手に意図が確実に伝わりやすい。つまり、AよりはBの方がより明確な目標といえる。明確な目標とそうでないものを見分けるコツは、「その目標にたどり着いたかどうかをどうやって確かめたらよいのか」を考えてみることである。目標Aに到達したかどうかを確かめるために、まさか学習者に「理解しましたか」と尋ねるわけにもいくまい。「理解した」かどうかを確かめる方法は多様にあり、人それぞれに考え方も異なるであろう。一方で、目標Bの場合は、学習者がこれまでに見たことのない不明確な目標を幾つか与えて、それを明確な目標に書き直せるかどうかを確かめればよいのである。つまり、明確な目標があれば評価方法が簡単に導き出せるものである。

3-3-1：行動で目標を表す

目標明確化の第1のポイントは、学習者の「行動で」目標を表すことである。上の例で学んでほしいことは「教授目標を明確化する方法について」であり、その方法についての理解を深めて欲しいことは確かである。しかし、「理解を深めた」ということは学習者の内的な(脳の記憶の)状態が何らかの形で変わったということであり、外からその変化は観察できない。「・・・を理解する」「・・・を知る」「・・・に気づく」というような目標は、学んでほしいことをそのまま記述している反面、うまく教えられたかどうかをどうやって確かめたらよ

いのが明確であるとはいえない。

目標Bの場合はどうであろうか。ここでは、目標明確化の方法について理解したかどうかを、その方法を実際に応用して目標を書き直すという学習者の「行動で」具体的に示している。この場合の「書き直す」という行動を、目標行動という。いうまでもなく、目標Bで学んでほしいことは「書き直す」行為そのものではない。書き直す方法だけを意味もわからず機械的に覚えたとしても、それは、目標明確化についての「理解」を示すものではないからである。「書き直す」という行動が「理解が深まった」ことを具体的に、外からわかる形で表すための一つの指標として用いられているということに注意したい。

3-3-2：評価の条件を示す

目標明確化の第2のポイントは、目標行動が評価される条件を明らかに示すことである。すなわち、目標Bにおける条件は、「3つのポイントのリストをみながら」と「未知の」という部分である。この目標では、学習者が3つのポイントが何であるかを暗記しておく必要がないことは明らかである。評価される時には試験問題の一部としてリストが示されるのであるから、それぞれのポイントが何を意味しているかを知っていればよいのである。また、いままでに例えば例題や練習問題として学習者が見たことのある目標は評価には使わないということが、「未知の」という条件から読み取れる。これは、書き直したものを丸暗記して正解することを防ぐためのものである。応用力を試すためには、新しい例に適用できるかどうかで評価する必要があるからである。条件にはこの他に「電卓を使って」や「辞書を見ないで」のように、学習者が目標行動を行うときに何を使ってよいのか、あるいはどのような制限があるのかを示すものがある。

3-3-3：合格基準を示す

目標明確化の第3のポイントとして、目標が達成されたかどうかを判断する基準を記述することがある。目標Bの場合は、「すべて」という部分がそれにあたる。この場合、例えば「与えられた5つの目標の中で4つ以上は」と置き換えることも可能である。その他の基準として、「1分以内で泳ぐ」のような速さや「誤差5%以内で測定する」のような正確さを明らかにするものを目標に含める場合がある。

目標を明確なものにしていくためには、何を学ばせたいのかをそのまま示すAのような目標からスタートしてもよい。上にあげた明確化のための3つのポイントをあてはめて、目標行動を選択し、条件や基準をつけ加えることで、徐々に明確な目標にすることができる。この作業の手助けとして有効なことの一つに、授業のできばえを評価するためのテスト問題を作ることがあげられる。評価の問題を作成することで何を教えようとしているのかを明らかにして、その具体的な評価の方法から逆に明確な教授目標を導き出すことも考えてよい。



コラム:経済学的視点からも目標の明確化が求められる



『教育を経済学で考える』の著者である小塩氏は、何を持って教育成果とみなすかを経済学の立場から問題にしている。気になる記述が多々あったうちから、目標を明確にすべき、という主張をいくつか引用させていただく。

「基礎学力の向上」「豊かな子どもの心の育成」「生きる力」などなどの教育目標が考えられている。「しかし、そうした目標がどこまで達成できたか、それをきちんと評価する仕組みはなかなか存在しない。」(小塩、2003、p.107-108)

「教育がその「生きる力」を子どもたちにどこまで身につけさせることができたか、どうやって検証するのだろうか」と疑問に思えてくる。最近では、政策評価の重要性が盛んに叫ばれるようになってきているが、政府がその「生きる力」を子どもに身に付けさせることを目標とするなら、その目標の達成度が教育行政の評価にとって最大のポイントとなるはずである。いったいどうするのだろうか。」(p.109)

「あまりに崇高な目標は、実は誰も相手にしていない。教育目標は、成果が具体的な数字で明らかになるほど望ましい。」(p.110)と指摘し、具体的な数字で示そうとする動き(「読売教育メール」から引用：世田谷区の事例)を歓迎している。たとえば、

「深く考える子」「図書室の本の貸出冊数を二倍にする」

「4月から10月にかけてつねに10種類以上の花が咲いているようにする」

「卒業時における進路の結果に対する満足度が五段階評価で四以上の生徒が90%を越えるようにする」

「筆者は、教育成果は、すべて数字で評価すべきもの、あるいは評価できるものだと思う。[小学校の恩師の一言、喧嘩したほろ苦い思い出など]かといって、教育成果をまったく測定不能なものとみなしてしまい、外部の評価を拒否することははたして望ましいことだろうか。教育も、批判を受けない聖域であってはならない。教育の中にも、きちんと検証できる成果はあるはずである。それさえも拒否するのは、独善的だという謗りを免れないだろう。」(p.111)

経済学からみた教育効果を検証する理由は2つある。(1)教育にはコスト(授業料、税金)がかかっている。コストに見合っただけの成果を教育機関がきちんと挙げているかどうかは、限られた資源をいかに効率的に配分するかという、経済学にとっても重要な関心事項。(2)公平性の観点から、教育成果がそれぞれの子どもにどのように異なって発揮されているかにも注目する。とくに義務教育の場合、格差の大きさは政府の関与のあり方に疑問を投げる。

出典：小塩隆士(2003)『教育を経済学で考える』日本評論社

第4節 出入口を明確にするために用いる3つのテスト

出入口を明確にするために用いるテストは、事前テスト、事後テスト、前提テストである。この3つを組み合わせることで、誰に何を教えるのかを明確にして、資格・必要性・効果を明らかにしていくことができる（鈴木、2002：第3章を参照）。

学習目標の明確化は「出口」の明確化である。つまり「何を教えるのか」「何を学ぶのか」を知ることになる。この判定は、「事後テスト」で行う。つまり、事後テストは、学習目標に到達したかどうかを判定するためのテストであり、学習成果を確認するためのものである。不合格であれば、必要に応じて再学習をして、もう一度事後テストに挑戦する。

学習する内容のうち、少しでも知っている、あるいはできる可能性があるものが含まれている場合、学習前に事前テストと前提テストを実施する。事前テストは、事後テストと同じ出口レベルの評価で、これに合格すれば「研修の必要なし」となり、次のレベルに飛び級する。事前テストに不合格であれば、無事に「研修の必要あり」となり、研修を開始する。

一方の前提（レディネス）テストは、研修では扱わない基礎事項を既に習得しているかどうかの準備状況を確認するための評価となる。これに合格すれば「資格あり」で学習がスタートでき、不合格であれば、「資格なし」と判断され、前段階の研修をやり直してから「出直してください」となる。

事前テストと前提テストは組み合わせて実施することも可能で、その場合は、どのレベルからスタートするかを見極める「診断テスト」として用いられることになる。この二つのテストはいわば「入口」の明確化になり、「誰が学ぶのか」「誰に教えるのか」を規定する。レベルが低すぎる人も、そして高すぎる人も入れないように、門番の役割を果たしている。

一斉指導で、レベルの差が問題になる。どんな研修でも難しすぎてついてこれない人と、易しすぎて退屈な人が出る。これは必要悪だと考える人もいれば、何とか個人差に対応しなければ優秀なインストラクタではないと考える人もいる。しかし、出入口の管理をしっかりやれば、受講者の知識・スキルレベルをそろえることは可能である。研修内容をしっかり分析して小口に分け、出入口をしっかり管理することで、無駄・無理・ムラがなく、しかも誰もが充実した時間を過ごすことが可能になる。

これに気づいたアメリカでは、完全習得学習（マスタリーラーニング）が採用された。功罪があるといわれるが、理にかなった、実にID的なアプローチだと筆者は思う。

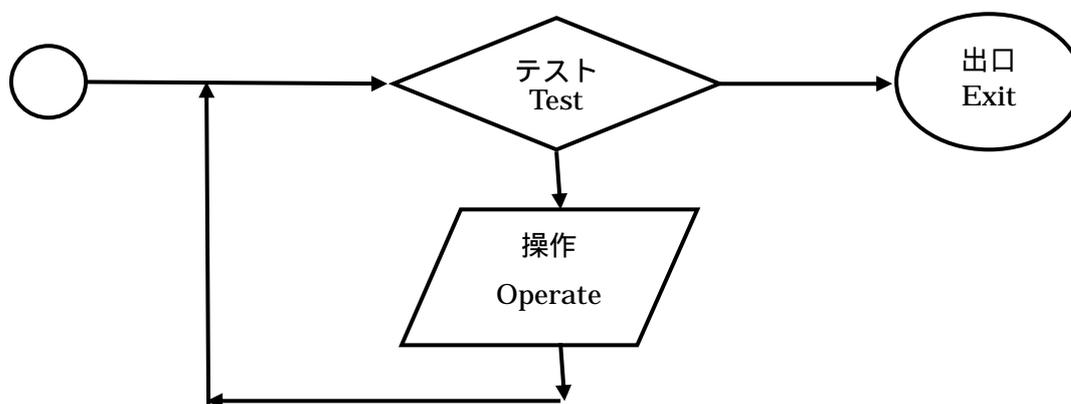


コラム：TOTEモデル(Test - Operate - Test - Exit)



TOTEモデルとは、図表3 - 5 に示されているように、ある作業を行うのに、まず、すでに目標が達成されているかどうかをチェックし、すでに達成されている場合は作業をしないで抜け出す。目標が達成されていないことが判明したら、その目標に向けてある一定の量だけ作業を行い、再び目標が達成できたかをチェックする。達成できれば抜け出し、できていなければ作業に戻り、チェック - 作業 - チェックを繰り返す。つまり、TEST (チェック) - OPERATE (作業) - TEST (チェック) - EXIT (抜け出す) の頭文字を取って、TOTEモデルという。

図表3 - 5 : TOTEモデル



TOTEモデルとは、ある特定のゴールを目指して進む時に常にゴールに達したかどうかをチェックしながら作業を進めることを図式化したモデルである。エアコンの温度制御などはこのモデルに基づいて動作している。とても簡単なモデルであるが、eラーニングの進め方を考える時に、最も大切な考え方を示している。

このモデルは、事前テストをやって、学習する必要があるかどうかを見極めた後に、必要があるものだけを学習するというIDの考え方のもとにもなっている。事後テストをやってダメならばまた学習を繰り返す(Operate)ことになり、合格すれば出て行かれる(Exit)と考えるのもこのモデルがベースである。でたらめに作業を開始しないで、必要があった場合にのみ作業を開始するという点、学習目標を最初に示せば、「これはもう知っているからやらないで他の教材をやる」といった選択が可能になること、あるいは、受講者に評定(点数)を付けるためのチェック(TEST)ではなく、できぐあいを見守り、次の手段を講ずるためのチェックであることなど、示唆に富んでいる。

第5節 学習課題の性質と適切な評価方法：ブルームとガニエ

学習目標の性質と評価方法をマッチさせることを目指して、様々な目標分類枠の提案がこれまでになされている。最も著名なのはブルームらによる学習目標分類学(タクソノミー)である。ブルームの目標分類学は、評価方法の整理のための枠組みとして、目標のレベル分けを提案したものであり、ブルームの著作が翻訳され、初期の教育工学研究の基礎として内外の教育関係者に強い影響を与えてきた(詳細は、梶田、1992などを参照)。とりわけ、目標としては高レベルのものを掲げておきながら、テストとなると暗記による断片的知識のみでも解答可能な問題が出される、という矛盾を解消し、高レベルの段階を測定するためにふさわしいテスト問題を例示したことが功績であった。

ブルームらは、教授目標をまず次の3つの領域に分類した。

- (1) 認知領域 (Cognitive Domain) : 知識の再生や知的技能の発達についての目標
- (2) 情意領域 (Affective Domain) : 興味・態度・価値観の変容、適応力などの目標
- (3) 精神運動技能領域 (Psychomotor Domain) : 運動技能や操作技能に関する目標

さらに、それぞれの領域をより基礎的な目標からより高次のものへと階層的に分類するためのクラス分けを設定した。認知的領域は、最も基礎的な目標のクラスから、(1)すでに学習したことを思い出し必要に応じて利用できる情報としての「知識」、(2)伝えられる情報の意味をとらえて利用できる力としての「理解」、(3)すでに学んだことを新しい課題場面や具体的状況に適用する力としての「応用」、(4)問題を構成要素に分解・再構成し、問題の全体的な構造を明らかにする力としての「分析」、(5)部分をまとめて新しい全体をつくり出す力としての「統合」、そして、(6)価値や意味を判断する力としての「評価」の6段階に分かれている。

情意領域は、「どう思うか」あるいは「どう感じるか」といった目標を扱う領域である。価値の内面化の程度というものさしで、基礎的なクラスから、「受け入れる」「反応する」「価値づける」「組織化する」「個性化する」という5段階を設けている。精神運動技能領域には、体育で扱う様々な運動技能や文字を書いたりタイプライター - を操作したりする技能が含まれる。この領域に関するブルームらの段階分けはまだなく、いくつかの段階分けが試みられてはいるが定説はない。これらを応用して、各教科・内容領域で、目標の分類やレベル分けが盛んに行われた。

一方、ID理論の生みの親として著名なガニエは、学習成果を図表3-6に示す通りに分類した(注: Webサイトに練習問題あり)。

図表3 - 6 : ガニエの学習成果の5分類と外国語学習の例(鈴木、1989)

学習成果分類	内容と外国語学習の例
1)言語情報 (verbal information)	学習の結果「述べること」ができるようになる事実、ラベルなどの学習。例として、すでに母国語で習得済の概念(例えば犬)に対応する外国語の単語(犬 = dog)を記憶すること。「犬は英語で何ですか」と問われたときに、「犬は dog です。」と述べること(口述するか記述する)ができるようになるのは、言語情報を習得したこととみなす。
2) 知的技能 (intellectual skills)	学習の結果、これまでに遭遇したことのない新しい例にその学習の成果が「適用」できるようになるような認知的な学習。知的技能は、さらに次の下位領域に分類される。
2a) 弁別学習 (discrimination)	学習の結果、2つのものが「異なる」かどうかを見分ける(聞き分ける)ことができるようになるような学習。例として、r、l、rと連続して発音された時、何番目が他と違うかが言えるようになること。この場合、どれがrでどれがlかはわからなくとも、違うものが指摘できれば、2つ違いが学習できた、つまり弁別学習が成立したとする。
2b) 概念学習 (concrete & defined concepts)	学習の結果、グループ(概念)に属する事例とそうでない事例を分類することができるようになるような学習。例として、母国語にはない文法上の概念(例えば単複同形とそうでないもの)を区別するとか、母国語と外国語の概念のずれを学習するような場合。seat という単語を「言語情報」として学習する場合には、「座るもの」と述べることであればよいが、「概念」として学習したかどうかを試すためには、「映画館の椅子は seat か」「公園のベンチは seat か」... という問いに正しく答えられなければならない。
2c) ル - ル学習 (rules; higher-order rules)	学習の結果、規則(ル - ル)を適用することができるようになるような学習。例として、疑問文に書き換える、主語や時制に合わせて動詞を変化させるなどの文法規則をどの文にも応用できれば、ル - ル学習が成立したとする。また、これまでに学習したル - ルを多く利用して、たとえばある状況で相手に自分の意思を伝えるという問題解決をなしとげるなどの、複数ル - ルの一連の適用能力を特に高次のル - ル学習(higher-order rules)という。
3) 認知的方略 (cognitive strategies)	学習の方法の学習。これまでの学習体験から得た、あるいは外からのヒントによって学んだ学習を効果的にする手立てを必要な時に使えるようになったとき、認知的方略を学習したという。例としては、単語を語呂合わせで覚える方法などがある。
4) 態度 (attitudes)	学習者が個人的な選択の機会に遭遇したときに、その選択を左右するような働きかけを内側から与えるような状態をつくる学習。例えば、外国人と積極的に接する態度を学習すれば、外国人と接触できるような機会を自分で選ぶ行動にできるようになる。英語を学ぶ態度を身につければ、他のことより英語の学習に時間を割くようになる。その行動を支えるものとして、外国人と接すること、あるいは英語を学習することに対する肯定的な態度を身につけているとみなす。
5) 運動技能 (motor skills)	学習の結果、(認知的なものだけでなく)筋肉の運動を伴う技能を駆使できるようになるような学習。例としては、口の筋肉の運動を伴う発音や、指の動きを伴うアルファベットの運筆技能などがある。

注: 鈴木(1989)の本文を表形式にまとめた。

学習成果の5分類とは、言語情報、知的技能、認知的方略、運動技能、態度である。これらの分類枠は、到達度の測定方法が異なるのみならず、到達を支援する最善の方法が異なるというIDの観点から学習目標の性質にもとづいて課題を分類したものである（第7章に後述）。

認知領域の学習成果は、三種類に大別されている。知的技能は、分類方法や計算方法などの約束事を学び、それを未知の例に適用する力（手続き的知識）の習得を指す。言語情報は、名前や年号などの与えられた情報を再び記述する力（宣言的知識）の習得を指す。認知的方略は、自分の学習過程をより効果的にするための力（学習技能）の習得を指す。このうちの知的技能には、ガニエの長年の研究成果から下位分類が設けられている。

情意領域には、態度の習得がある。態度とは、例えば「人種差別」や「数学を学ぶこと」などあらゆるものごとや状況等に対する肯定的あるいは否定的な感情であり、学習成果の一つとみなしている。また、運動領域では体（からだ全体、あるいは一部）を動かして一定の課題を成し遂げられるようになることが運動技能として取り上げられている。体育実技で取り上げられる学習課題の他にも、キーボードのタッチタイピングや外国語学習での発音方法なども、運動技能に含まれる。以上をまとめると図表3-7のようになる。

図表3-7：ガニエの5つの学習成果と出入口の明確化

学習成果	言語情報	知的技能	認知的方略	運動技能	態度
成果の性質	指定されたものを覚える 宣言的知識 再生的学習	規則を未知の事例に適用する力 手続き的知識	自分の学習過程を効果的にする力 学習技能	筋肉を使って体を動かす/ コントロールする力	ある物事や状況を選ぼう/ 避けようとする気持ち
学習成果の分類を示す行為動詞	記述する	区別する 確認する 分類する 例証する 生成する	採用する	実行する	選択する
成果の評価	あらかじめ提示された情報の再認または再生 全項目を対象とするか項目の無作為抽出を行う	未知の例に適用させる：規則自体の再生ではない 課題の全タイプから出題し適用できる範囲を確認する	学習の結果より過程に適用される 学習過程の観察や自己描写レポートなどを用いる	実演させる：やり方の知識と実現する力は違う リストを活用し正確さ、速さ、スムーズさをチェック	行動の観察または行動意図の表明 場を設定する。一般論でなく個人的な選択行動を扱う

出典：鈴木克明（1995）「放送利用からの授業デザイナー入門」日本放送教育協会、p.??
（表 III-2 の一部）



コラム:ポートフォリオで自己評価と自己アピール能力を育てる



ポートフォリオ [Portfolio] とは、書類入れやファイルを意味する言葉である。日本の小中学校における総合的な学習の評価方法として、近年注目されている外来語である。ポートフォリオ評価は、たとえば「学習活動において児童生徒が作成した作文、レポート、作品、テスト、活動の様子が分かる写真やVTRなどをファイルに入れて保存する方法」(グロワード, 1999, p. 8)と定義されている。

ポートフォリオ評価は、単なる記録ではなく評価なので、学習の過程で創出されたものすべてを保存するのではないとの考え方が一般的である。すなわち、残す意味があるものを選んで子ども自身の目の前でファイルすることを通して、1)子どもが達成したことが何であるかを子ども自身に明確に伝え、2)どうしてそれが高く評価されることなのかをわからせ、3)子どもの達成感や自尊心、あるいは自己効力感を高め、そして4)次の課題が何であることを示して自分の学習活動をコントロールするためのメタ認知を育てることを意図するものである。何を残して学習成果を最大限にアピールするか、という意味で、証券ポートフォリオ(連動しない証券を組み合わせることで、変動に強い証券セットをつくること)と底通する用法である。

イギリスでは、16才以上の生徒は、自分自身でポートフォリオを作成し、一般職業資格取得に値するだけの学習をした証拠を示すことが義務づけられているという(グロワード, 1999)。自分の学習成果をまとめて整理し、それをもとに「これだけの成果をあげました」と自己アピールできる子どもを育てる。中学校あたりから、証拠を揃えて自己主張をするという訓練ができれば、たくましい子どもが育つだろう。

また、アメリカのポートフォリオ学習を紹介している小田(1999)は、「日本の総合的な学習では、評価の観点を明示しつつ進めている学校は少ない(p.8)」と指摘する。アメリカでは、学びが始まる前に評価観点を示した上で調べ学習などを行わせていると言う。評価の観点を事前に明示することは、ポートフォリオ学習の本質であり、「自分の変化について自分で気がつくための学習(p.75)」だとする。フィードバックによる自己評価と修正が組み込まれているのである。

たとえば、ミネソタ州の事例「マーケットリサーチ」では、観察ノートと図表と小論文の提出が義務づけられている。各提出物に対して、観察ノートではわかりやすさと重要なポイントが含まれているかどうか、また図表では、表示の正確さ、パターンの発見、そしてそのパターンの示し方などが評価基準であると示される。生徒はまず自分自身で「優れている・満足のいく状態・改善を要する」の3段階でチェックし(これをループリックと呼ぶ)、教師による評価と比較しな

がら自己の学習を振り返り，自己評価力をつけていく。

日本では、「ポートフォリオでは、学習の結果ばかりではなくプロセスも評価する」といわれることが多い。しかし、これは間違いであると筆者は思う。プロセスを評価するとは、言外に「結果は思わしくなかった子の頑張りを加味する」との親切心があるようであるが、それはかえってありがた迷惑ではなからうか。頑張りを評価するとなれば、能力が高くて頑張らなくてもできてしまった子の評価は低くなるのだろうか。プロセスをていねいに見てあげるためにもポートフォリオは使うことは可能だが、上記の趣旨から考えると、「あなたは自分の学習の成果をどの証拠を持ってアピールするのですか」という厳しい問いを突きつけているように感じられる。それは、成果の積み重ねを問うているわけであり、プロセスをうんぬんしているわけではない。作品を上手に組み合わせて、自分の学習の成果をアピールできる能力がどの程度ついたか、という自己評価力（もちろんこれも学習成果でありプロセスではない）をも含めた評価である。ガニエの分類にあてはめると、それは認知的方略の育成と評価を狙ったものだと解釈できよう。

日本では、評価といえば、先生から下されるもので、それは自分の価値をも決めかねないという威力を持ったものとして、受け取られてきた。そんな歴史を背景に持ちながらも、今、新しい評価の形が模索されている。自分を高めるための評価、自分の弱点を知って修正するための評価、自己アピールできるようになるための評価などと急に言われても、戸惑うばかりだろう。子ども時代をずっと旧来の評価観で過ごしてきた今の大人にとって、評価の意義を見直すことはなかなか困難なチャレンジであろうが、学習過程における評価の役割とその威力を伝えていくのもID者の重要な責務である。その自覚のもとに、評価を如何に実行し、受講者にとってのパワフルな道具として再認識させる努力が求められている。

【参考文献】

グロワ - ド（鈴木秀幸訳）（1999）『教師と子供のポートフォリオ評価（総合的学習・科学編）』
論創社

小田勝己（1999）『総合的な学習に適したポートフォリオ学習と評価』学事出版

第6節 eラーニングで用いる教材の形成的評価

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、研修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試用する工程を指す(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、体系的なIDプロセスの所産として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、IDモデルに基づいて理論的見地からデザイン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工程である。

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。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的な教材の改善に役立てられる。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。IDプロセスの一ステップとして教材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。現在のIDプロセス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(詳細は、鈴木、1987 を参照のこと)。

形成的評価の中心をなす手法は、学習者からのデータ収集に関するものである。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図表 3 - 8 のようになる(鈴木、1987)。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。

図表3 - 8：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	データの種類		
		試用以前	試用中	試用後
外部協力者	指導者	・教材の適切性	・教材の管理	・教材に対する意見、感想及び提案
	学習者	・前提行動のテストの結果	・教材に組み込まれたテストの結果	・事後テストの結果
		・事前テストの結果	・教材とその内容に対する意見	・教材に対する意見、感想及び提案
		・事前・事後テストの指示と項目の明確性		・完了までの経過時期
チーム内	SME	・内容の正確性最新度及び語彙レベルや例題の妥当性		
	ID専門家	・学習指導原理の適切な応用		・教材改善の適切性
		・用語の適切性		

出典：鈴木（1987）の表1の一部術語を変えて簡略化した。

注：SME = Subject Matter Expert（内容領域の専門家）

1対1評価

学習者からのデータを収集する第一段階は1対1の形成的評価[one-to-one formative evaluation]と呼ばれる。教材の設計開発担当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。設計開発担当者が学習者と共に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対象者集団の中からやる気のある者、中ぐらいの者、そしてあまり熱心でない者の順に1対1で教材を使用して学習する。異なった反応が上・中・下の協力者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を考える。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。

小集団評価

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担当者が学習の過程に介在したが、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、教材の対象者集団を反映させ、また結果を数量的に検討するために8人から20人の協力者を募る。したがって、対象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協力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

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、教材の管理や他の研修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。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、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。実施初年度は、ある意味で「実地トライアウト」の要素を含む。稼働させながら、問題に対応し、2年目の運用に向けて徐々に品質を高めていくことになるからである。

形成的評価のこれらの工程は、システム工学が提唱する 版と 版の試験運用や、ユーザビリティテストなどに用いられている手法と共通点が多い。完成品の品質を確保するためには、どの分野でも提唱されている工程であるが、一方で、品質管理が徹底していない教育分野では、疎かにされがちな工程でもある。

評価結果を改善につなげる

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。また、鈴木（2002）では、印刷教材を例に、改善へのアイデアをまとめるためのチェックリストを提案している（注：Web上のツールとしても試せる）。様々なメディアを駆使したプロジェクトにおける形成的評価の実際については、Flagg（1989）が紹介している。

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。『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。』（Dick & Carey, 1985, p.223）としている。

教材の改善には具体的に何が含まれているかをイメージする手がかりとして、デバート（Debert, 1979）の改善のリストを図表3-9に示す。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イメージできよう。また、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。

図表3-9：教材改善方略のリスト（デバートによる）

付加	<ul style="list-style-type: none"> 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インストラクタへの訓練を加える。 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。
簡易化	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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. 教材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。

出典：鈴木（1987）の表2を一部の術語を入れ替えて再掲した。

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言及している。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テジーを再点検し、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。

 コラム: 評価機関(有識者)の評価はあてにならない? 

6つのWBTパッケージを比較した30人のボランティアが出したレイティングに開きがあったことを報告した実証研究を見つけた。そのレイティングが有識者によるレイティング(www.lguide.com)と正反対の結果であったことから、有識者の評価をあてにせずに、形成的評価で実証的なデータをとることの必要性を主張している。評価機関の評価はあてにならないのだろうか?以下に、研究の概要をメモしておく。

比較対象: マイクロソフト社の「アクセス」を学習するための市販ソフト6つ(価格帯は20-99ドル)。見つけた市販ソフト10個のうち、1つはCD-ROM提供のためWBTでないので除外、1つは高価すぎて(179ドル)除外(買わなかった)、もう一つは、発売元がすぐに倒産してサービス停止で除外、最後の一つは違う会社が売り出した同じ内容のソフトだったので除外。それで6つが残った。

実験参加者: 新聞広告等で集めた21歳以上の大人30人。47%男性, 53%二十代, 37%初心者(自己評価)。Windowsパソコンが使える, アクセスの使用経験が3時間未満で, プログラマ・ウェブ開発者として働いてない人(事前審査で確認)。2時間の実験協力に対して40ドルを支給した。

実験手順: 6つのソフトのうちの2つのみを体験, 比較させた。6種類×それ以外の5種類で合計30通りの組み合わせを, 一人ずつ。順序効果を相殺して割り当てた。ソフト体験45分+事後テスト+アンケートの順に2セット。ソフト体験は, 「このソフトを自宅で学習しているつもりで, 45分の中でできるだけたくさん見るように」毎回教示した。事後テストは多肢選択式の知識問題を6つのコース全てをチェックした助手が作成した20問を10問ずつの2セットに分けて実施。順序効果は相殺。

レイティング(結果):

(1) どちらのソフトを買うか(値段が同じだとしたら)? 選ばれた数(最大値10) A9, B7, C5, D5, E3, F1 (AはBとの比較で1回だけ負け; FはEとの比較で1回だけ勝ち): AB好まれたソフト, EF好まれなかった, CDその中間【図1】

(2) 2つのソフトの差はどの程度か(1差は小さい~7差は大きい) 30人の平均=5.3

(3) 2つのソフトのどの側面をどの程度に評価するか(全体, つかいやささ, 読みやすさ, 魅力, 目標の明確さ, 教え方, エンターテイメント性, ナビゲーション) レイティングの低いソ

フト (E, F) はエンターテインメント性が極端に低かった【表1】

レイティングと価格: 反比例 (価格 = A \$20, B \$27, C \$55, D \$98, E \$75, F \$99)【表5】

安いソフトほど好まれた

レイティングと成績: 比例 (平均点 = A 8.4, B 7.8, C 6.7, D 7.3, E 7.5, F 6.3)【表1】 10点満点 遊びばかりで学びがないのが好まれたわけではない。むしろその逆。好まれたソフトでは成績も良い。[実験手順から, 良い成績が収められたから好みになったとも考えられる]

ソフトの特徴 (差はどうして出たか): アンケート結果と特徴の分析

グラフィックを多用し, 構造的には単純明解なものが好まれた

ソフトAは一方向的に漫画キャラクターが語りかける方式, オプションの多くは未使用。ソフトBはAとは逆に双方向性を多用。クイズや練習コーナーが好まれた。双方向性については, 好みが別れる。一方向が必ずしもダメではない。[初心者特に]

逆に, 文字が多く, 多種多様な機能が満載のソフトは好まれなかった

ソフトFのオプション (10分の使用法説明が付属): 何種類もの形式の練習, クイックチェック, レファレンス, 学習者コミュニティ, チューター, PDF形式の捕足資料, 書店, ソフト購入ページ, ヘルプページ, スレッド型BBS, チャットコーナー。もっとも多くの機能ともしっかり洗練されたID手法が用いられていたが, 使われずに無視されることもある。45分だけの試用で協力費が支給された実験参加者の場合のみにあてはまるのか, より長期間でより動機づけられた学習者の場合もそうなのかは今後の研究が必要としている。

有識者レイティングとの差: 両方が揃ったDを除く5つのソフトのレイティングがまったく逆【図5】 有識者は難しい文言を好むとする説 (bafflegab phenomenon) の事例かもしれない。IDプロセスで強調されている形成的評価が重要であることだけは確かだとしている。

有識者評価と試用実験結果が逆という話は、以前からも研究されていて聞いたことがあるが、eラーニングの世界でもあてはまるのかもしれない。実験の組み立ても良くできていて、研究のやり方も参考になる。

【紹介した研究】

Hassett, J., Ingram, A., Hassett, M., & Marino, E. (2003). What Do Learners Like? Ratings of Off-the-Shelf Web-Based Training Courses. *International Journal on E-Learning*, 2 (1), 50-60. [Available Online]: <http://www.ace.org/DL/index.cfm/fuseaction/View/paperID/11553>

(おわり)

【参考文献】

- 梶田叡一 (1992) 『教育評価 (第2版)』 有斐閣
- 機械システム振興協会 (2001) 『遠隔学習システムの利用促進に関する調査研究報告書』(C A I T) [Available Online] <http://www.jipdec.jp/chosa/caittext/wbt/indexh12.htm>
- 鈴木克明 (1989) 「テレビ番組による外国語教育を補うドリル型C A Iの構築について」 『放送教育研究』 17, 21-37 [Available Online] <http://www.iwate-pu.ac.jp/home/ksuzuki/resume/journals/1989b.html>
- 鈴木克明 (2002) 『教材設計マニュアル』 北大路書房
- 鈴木克明 (1989) 「12. 教授目標明確化の方法を、例をあげて説明せよ」 沼野一男・平沢茂編著 『教育の方法・技術』 学文社
- 鈴木克明 (1987) 「C A I教材の設計開発における形成的評価の技法について」 『視聴覚教育研究』 17, 1-15
- Dick, W., & Carey, L. (1985). *The systematic design of instruction* (2nd Ed.). Scott, Foresman, and Company.
- Flagg, B.N. (Ed.). (1989). *Formative Evaluation for Educational Technologies*. Lawrence Erlbaum Associates.

	章末レポート課題 (第3章)	
--	-------------------	--

次に挙げる2つの課題のうち、1つ以上について考えてみましょう。

- 1) この章(第3章)を読んで疑問に思ったことやコメント・意見・感想などをまとめてみましょう。なお、この章の記述に関連するこれまでの経験談や付け加える情報・調べてみたこととその結果(情報源の名称を付けること)などがあれば、それも含めて考察すると理解が深まるでしょう。
- 2) 日本の企業内教育の評価について、本章の第2節を主に参考にして、あなたが知っているeラーニング事例について分析しなさい。日本では企業内教育の評価があまり行なわれない背景や、これから評価が必要とされる理由、あるいは学習者はお客様であるという考え方について、何がどこまであてはまると思うかに言及すること。なお、eラーニング事例に限定せずに、これまで受けてきた学校や会社での被教育体験、あるいは自分が行っている教育活動について分析してみるのもよいでしょう。

レポート閲覧・交換上の注意

閲覧方法：「eラーニングファンダメンタル」学習支援Webサイトの中に、「章末レポート交換用掲示板」があります。これまでの書き込みは誰でも閲覧できます。
Webサイトトップページ(<http://www.et.soft.iwate-pu.ac.jp/eLF/>)から本章が属する「目的と評価」を選択すると、第3章用の掲示板があります。

交換方法：「交換用掲示板」への書き込みは、ユーザー登録を済ませると可能になります。
ユーザー登録には、本名および電子メールアドレスが必要ですが、投稿に際しては、本名を名乗らずに、ニックネームでの登録・情報交換ができます。

留意事項：掲示板の閲覧は本書の読者以外も可能であることに留意し、公開できないような内容は書かないでください。また、個人名や特定団体名称などの使用や誹謗中傷にあたる恐れがある記述にも注意してください。削除・変更の権限はWebサイト管理者が有し、必要に応じてユーザー登録の取り消しも行います。

採点基準：eLCからの修了証を目指してブレンディング講習を受講される方への提出期限・提出方法・採点基準などは別にお知らせします。

受講者の反応

(レポート課題1：第3章への感想・コメントなど)

ダメなテスト問題 urakumi さん (2003年09月15日)

eラーニングコンテンツの開発に1年ほど従事してきたが、その中でいくつかのテスト問題に出会ってきた。多くの場合 SME にあたる人が作成したものだ。きちんと考えられたテスト問題もあるが、中にはダメなものもあった。例えば、次のようなものである。

- (1) きちんと勉強した人が100点を取ることを認めない。背景に「自分が教えている内容は、非常に高度なものであり、簡単にわかってたまるか」という意識がある。ひっかけ問題を出すケースも見られる。
- (2) 常識で答えられる、あるいは選択肢のつくりが雑で明らかに正解/不正解がわかる。テスト問題の作成技術が欠けている。
- (3) 問題数が少なすぎる。数時間の学習内容でテストが2~3問しかない。
- (4) 教えていないことをテストに出す。(1)で述べたような意識が背景にある場合もあるが、なかには教えていないことまで付加情報として提示することが受講者へのサービスだと思っている人もいる。
- (5) 同じ問題を何度も使いまわす。知的技能の課題であるにもかかわらず、まったく同じ問題を練習問題にも修了テストにも出す。
- (6) テストを教材のおまけだと思っている。教材を作るので力尽き、締め切りぎりぎりになって適当なテストを作ってくる人もいる。

偉そうに書き連ねたが、実はこのようなケースに出会ったとき私は何もなかった。ただ、誤字脱字だけを直してそのまま開発を進めた。言い訳になるが、理由は次の2つだ。1つは、SME = 講師の場合が多く、自分が教えている内容に絶対的な自信を持っている。したがって、それをあえて覆すことがためらわれるのである。もう1つは、組織の壁である。研修部門、開発部門と組織が縦割りにされており、役割分担が明確になるあまり、それぞれの範疇の仕事しかしなくなっている。たとえ、作り直しを依頼しても、納期に間に合わない。

eラーニングの開発において、ID者の存在は必要不可欠だと言われている。すべてのプロセスにおいて、ID者が存在することが望ましいのはいうまでもないが、このテスト問題の作成は評価方法・目標設定と切り離せない関係にある。(1)~(6)の例を考えても、ここに、ID者の活躍の場は十分にある。一方、ID者にとっても、テストに関する技術の習得は特に大切だということもまたいえるのではないだろうか。この章を読んで強くそう思った。

減点主義 ひろりさん (2003年09月17日)

教育評価が日本で行われない理由のひとつに「教育での評価が全人評価のようにとる文化が日本ではあったこと」(p.7)と述べられているが、私はもうひとつ「減点主義」というのが根強くあると思う。評価されるということは、すなわち、xを付けられ、減点されることなのである。人は減点されて、あなたはここができない、だからダメだといわれたら誰でも身構える。授業で間違えることを恐れるがゆえに発言しない理由がここにある。何もいわなければ、恥もかかないし、減点もされない。したがって、黙っているのが一番。評価もなるべくされたくない。面子をつぶすだけの評価は誰でも受けたくないものである。このような状況で評価結果をストレートに受け入れないのは当然である。

私の授業では、加点主義で評価し、その内訳は何が何%...というように最初に明示するが、学生はその加点主義という意味がわからず、しばらくは戸惑っている。「間違えて結構、それも学習の過程だから」といってもポカンとしている。そのうちに、とにかく努力をすれば加点されるらしいということがわかってやる気を出してくる。その様子を見ていると、日本の減点主義がいかに学習を縛っているか、その根深さを感じる。

受講者の反応

(レポート課題1：第3章への感想・コメントなど)

部下のことを考えて「まずまずの評価」をした 秋さん(2003年09月16日)

かつて「インストラクショナルデザイン入門」という本を読んだとき、教育・学習過程を徹底的に細分化し、論理的に構成していく方法に驚かされました。この章を読んでそのことを思い出しました。日本の教育風土をeラーニングの側から見ると、評価が非論理的であることが見えてくるように思えます。総合的評価、全人的評価という言葉は多くの場合、論理や分析を軽視したベテランの「勘」や「情」によるものでしょう。かつて部下の評価をさせられたことがあります。評価項目が「業務遂行能力はあるか、ないか、普通か」といった設問で、考えれば「あの点はよいが、別の点はだめだ」ということで、評価が出来ず、結局、部下のことも考えて、まずまずの評価をすることになりました。個別の事項について評価することは、「金に細かい」といった言い方に似たマイナスイメージがあり、大雑把に「がんばってる」「やる気が見えない」といった評価基準ばかりのようです。

学校の先生の評価も同様で、具体的に子供たちのどんな能力をどういう方法でどこまで伸ばすかという目的意識(ということは教育者の能力ということですが)は乏しい。その結果、テストの点数イコール全人的評価になってしまい、知識量を測る単一評価になっている。しかしオンライン学習を進めている学校では、教材自体が細分化され、項目ごとに結果が点数化されるため、児童生徒が自分のつまづきや課題を理解し、学習効果が上がった事例もある。明確、論理的に項目化された評価システムは必要なことと痛感しています。

目標明確化の3要素、学習課題の5分類など方法は、学習者が自ら学ぶ道を見つけるためにも、先生が的確な方法を築くためにも、広めるべきものではないかと思いました。

絵の授業のポートフォリオ学習 Kazusさん(2003年09月16日)

息子は小学校に上がる前の半年間、米国の保育園に通っていた。帰国して日本の小学校に上がった最初の参観日。教室の壁の生徒の絵の中に、いくら探しても息子の絵が見つからなかった。必死になって探していると、担任の若い女性教師がやってきて言った。「殴り書きだったので、張り出すのをやめました。」一瞬、耳を疑った。米国の先生は、その一筆一筆の意味をこどもたちに丁寧に問うて、その発言をもとにタイトルなどをつけて張り出してくれた。「殴りかき」などと評価されたことは一度もない。学校では殴り書きの息子も、自宅で私と一緒に絵を描くときは、いきいきと楽しそうだ。私は、単純に、米国の保育園の先生がしてくれていたとおりに息子に接しているだけだが、よく考えると、これこそ、米国流ポートフォリオ学習だったのだと思う。

ポートフォリオは使い方次第 wombatさん(2003年09月16日)

ポートフォリオについては、プロセス評価を水増しして救済措置として使ってしまうこともあり、これは確かに、成績優秀者から見ると不公平感があるかもしれない。しかし、この点はメリットにもなりうる。特にトコロテン方式に合格者を出さなければならない機関(例えば大教室での必修科目で、脱落者が多いと翌年の再履修者が増えて困る場合など)で、一定数の学生を救うことが機関からも期待されるケースである。

しかし、ポートフォリオは、とにもかくにも、本人が最後には形に残るもの(作文などの制作物)を作りあげ、友人からのコメントをもらうなど、重層的な評価ができる点こそを評価すべきだと思う。デジタルなものが出来れば、単なる通知表を最後に渡すことに比べ、貴重な財産となるであろうし、ある者にとっては自己実現を実感できる材料にも、なりうる。教師側が、成績をつける手段としての価値は、その次に来ることであろう。

受講者の反応

（レポート課題1：第3章への感想・コメントなど）

レベル3評価をやっています（事例提供） ライナスさん（2003年09月16日）

当社でも、カークパトリックの4段階評価モデルを常に意識し、機会を捉えてはデータ収集を試みている。筆者は日本におけるレベル3、4の実施報告がないと理解されているが、当社では電子回路設計技術の基礎レベル研修コースに対し、レベル3の評価を実施した経験があるので以下に紹介する。

この研修コースは毎年実施される公開コースであり、カリキュラムは初級者を対象としている。5日間コースは講義と実回路の製作評価演習で構成されている。受講者には、“大学では理論を学んできたが、実際に回路に触れることでより理解が深まった”などのコメント、即ちレベル1（反応）評価としては十分な満足度を得ている。この研修コースの受講者数十名に対し、6ヵ月後の実務における活用度、即ちレベル3（行動）評価を実施した。形態は、本人と直属上司に対するアンケートである。詳細は提示できないが、結論として、“実務ですぐ活用している”ケースでは研修効果を非常に高く、“実務でまだ活用していない”ケースでは相対的に評価値が低かった。即ち、実務における“即戦力”スキルや知識と“遅効性”スキルや知識では、3～6ヶ月後程度の短期経過時の評価は完璧ではないと言える。

現在、ITSSが制定されたのを受け、IT分野ではスキル体系が綿密に設定され、これに基づく教育研修と達成レベルが明確な定量的レベルで提示されるようになった。しかし、変化に即応して新たな戦略を打ち立てる力ことが、今、企業において求められている。これをどうやって評価するか、特にレベル3、4評価こそが正に的確な指標であり、その方法論を明らかにしたい。

因果関係を見つけること自体が貢献になる テンテンさん（2003年09月16日）

1. 経営変革を促進し、業績を向上させる。
2. そのための、教育目標を具体的に設定する。
3. 教育目標を達成するための、より具体的な学習課題を設定する。

この1. 2. 3. の因果関係が明確になれば、それ自体がタイヘン素晴らしい貢献だと思う。これが明確になるということは、「こういう学習をすれば、教育目標がクリアーされ、社員の能力が向上し、行動が変わり、生産性が高まり、必ず経営変革が成功し、業績が上がる」という因果関係を明確にするということであり、これは、効果的な「経営の勘所・ツボ」をおさえることになるからである。

東洋医学でも、ヤブ医者とは、やたらにあちこちのツボをおさえて、努力のわりには却って効果が出ず、患者も痛い思いをするばかりだが、名医ともなると、ほんの少数のツボ中のツボをおさえるから、簡単で効果が出て、患者も痛い思いをほとんどしない。このような、ツボ中のツボとしての教育目標、学習課題の設定をする能力の開発がID者には求められる。このようなID者の能力開発、そのような能力を備えたIDを育成することが、日本のe-ラーニング発展、日本の教育（学校・企業・社会）の発展のツボかもしれない。

バランス・スコア・カード（BSC）の応用で、教育と業績・経営変革の因果関係の仮説を立てながら、教育目標・学習課題の「当たる確率」を高めていくことも考えられると思う。

受講者の反応

(レポート課題2：eラーニング事例の紹介と考察)

つたない取り組み例からの考察 shidaさん(2003年09月15日)

(レベル1)受講後に、アンケート形式で、答えてもらっている。集合研修の場合も、最終日にまとめて記述してもらっている。以前は「満足」という書き方をしていたが、最近では「役だったかどうか」という聞き方が主流。でも、中には、「役立ったかどうかは、職場に戻ってからわかるので、今は不明」という、「その通りですね」といわざるを得ない回答をされるケースもある。

ここでは、受講者満足を大事に考えているが、本来の趣旨からいえば、受講者がその場では満足でなくても(たとえば、つらかった、というような評価でも)成果があがれば、近い将来は満足に変化する、ということも考えられるので、読み取りが難しい面もある。また、集合研修の場合、評価は悪くつけづらい(特に社内講師の場合)面もあるので、立ち会った研修担当者が感じたことも重要になっている。その点、eラーニングでは、より率直な評価が得られるように感じている。顔が見えないから、なのだろうか?

(レベル2)集合研修では実施していなかったが、eラーニングには、テスト機能があるので、組み入れている。章立てした、各章の最後に、「それでは、テスト1を受講してください」とナビを入れて、テストを受けて、内容が理解できたら、次の章に進むように、とガイドしている。つまり、テストは何度でも受講可能で、わからなかったら、テキストに戻るように、としている。これは、受講者にとっては復習になるが、テキストそのものの改善をすすめるにあたっては、平均点が高く出てしまうので、改善点が見えにくい、というマイナスがあるかもしれない。

(レベル3、レベル4)現在は実施していない。能力開発が、経営陣と研修スタッフと職場が三位一体となって進んでいけば、「日々の仕事、人事(異動など)、研修」が好循環で進んでいくので、たとえば、評価面談(査定のための面接)の中で、上司と能力開発についてもじっくり話し合う、という土壌が生まれていくと考えられる。レベル1では、受講者をお客様として想定しているが、本当のお客様は、受講者の向こうにある職場、取引先、本当のお客様である。そのことを見逃さないためにも、レベル3、4はとても大事な取り組みだと考えられるが、日々の業務の中で、フォローにまでパワーを使いきれないジレンマがあるのが現状。「それだけのコストと時間があつたら、ほかのことをしよう」ということになってしまう。レベル3以上の評価をスムーズに導入するためには、「日々の仕事、人事(異動など)、研修」のサイクルで見えていく仕組みづくりと同時に進めることが必要だと思う。そうすれば、より率直な評価が可能になるのではないかと、思う。

評価を嫌う企業の不思議 よしりんさん(2003年09月17日)

今まで携わったeラーニング導入コンサルティングの事例を振り返ると、「お客様である企業」から求められる評価は、レベル1「満足」とレベル2「理解」がほぼ100%。実務への活用、および貢献についての評価の要望は、残念ながらごく僅かです。こちらもデータをきちんと収集したいので、わりと熱心にお勧めするのですが、「いえ、うちはまだ、そこまでは・・・」と逆に断られてしまう場合が多いのです。これは企業の教育御担当の方々が、まだまだ「学習者に評判の良い教育」や、「教育することそのものが目標である教育」で満足していらっしゃる、余分な追跡調査で思いがけないアラが発覚するのが怖いのでは、あるいは面倒なのでは、と思うのは空振りの多いコンサルタントの僻みでしょうか?ただ、僅かですが、業績への貢献のみならず、eラーニング導入で組織全体のパフォーマンスのスタイルや、the way of thinkingがどのように変革したのか、レポートして次のステップに役立てたい、と言われるうれしい場合も出てきました。

受講者の反応

(レポート課題2：eラーニング事例の紹介と考察)

(前頁からの続き)

今後は、企業のほかのあらゆるパフォーマンス同様、教育というパフォーマンスについても、導入の仕様書、導入経過のレポート、導入の妥当性説明と導入の評価結果が必要となると思います。経済状況厳しき折から、効果(=利益)の見込めないパフォーマンスは切られるのが当然で、教育とて例外ではありません。益のない教育に注ぎ込むよりは、職場環境の改善に投じたほうが生きた「投資」である、と捉えられても仕方ないですものね。

評価と「お客様」扱い kitさん(2003年09月16日)

私が携わった内定者に対しての入社前教育のe-Learningの例を述べたい。約100人の内定者に対し、入社前に自社や自社商品に関する教育を行った。内容は自社や自社商品に関する約50のキーワードに関するもので、目的は「入社後の新人研修で落ちこぼれないこと」、目標は「新人研修を理解するのに不可欠(=出題された)キーワードの意味を答えられる」ことであった。そして内定者には「新人研修初日の講義は皆さんがこれらのキーワードを知っている前提で進みます」という教示をした。評価は以下のとおり行った。評価の主な目的は「内定者e-Learningを翌年実施するか否かの意思決定」であった。

<レベル1：e-learning終了時のアンケート>詳細は述べないが、全般に好評であった。

<レベル2：直後テスト・修了状況と新人研修初日の小テスト>このe-Learningでは「全て大事なキーワードなので、100%覚えるまで修了としない」とした(その代わりリトライ可とした)が、全員修了した。また、新人研修初日の講義前に小テスト(e-Learningでのテストから5問抜粋)を行ったが、全員が4問以上に正解していた。

<レベル3：新人の初日のアンケート、および登板した講師の印象のインタビュー>新人アンケートでは、「事前のe-Learningは役に立ったか?」の質問に対し、80%の新人がyes。登板講師の印象は、「e-learningを行わなかった前年に比して、受講生の反応がよく講義がしやすかった」となった。

<レベル4>対象者が新人であることもあり行わなかった。ただし「新人研修全体の理解への寄与」を知るために、新人研修が終了する5月末に行ったアンケートの中で「入社前のe-Learningが新人研修全体の理解に役立った」と回答した新人が75%、「まあまあ役立った」が15%いた。

上記の結果、翌年も継続することが組織的に決定された。さて、このケースでは修了率100%という、提供側としても望外の結果を得られた。その要因のひとつとして「社内の学習者はおお客様である」という問題に対する我々のスタンスを挙げたい。

我々は「快適に学習していただくために『お客様』としての配慮をするが、学習を完遂するためには『お客様』扱いせずあくまで『社員』として厳しく接する」というスタンスを持っている。逆に言えば、社員として厳しく接するが、学習の快適さに関しては「相手が社員だし、まあ、いいか」といった妥協をしないように心がけている。内定者のe-Learningにおいても、快適さを考える上では『お客様としてどうか』という観点で考え、工夫し、チェックしていったが、反面、動機付け(口の悪い同僚は「恫喝づけじゃないか?」と揶揄したが・・・)としては進捗が思わしくない学習者に「学習しないと大変なことになる」「困るのは自分だよ」といったようなプレッシャーを掛けていった。

いささか古い(伝統的な?)考え方もかもしれないが、このような「あたりの厳しさ」も学習者への愛情であり、その愛情は社内の教育では不可欠ではないかと考えている。そして「お客様」扱いしきってしまうと、この愛情も薄いものになってしまうのではないかと私は危惧している。この考えが正しいか否かは、今後、自分たちが実施する教育を評価する中で考えていきたい。