

第1章

「エバリエーション」の成立

田中耕治 『教育評価』 岩波書店

ある意図的な教育(学習)活動の成果を判断する営為またはその方法を規定する概念として、従来から使用されていた「試験」「考査」「測定」とは区別する意図を持って、「教育評価」という概念が使用されるようになるのは、第二次世界大戦直後のことである。それは、英語である「エバリエーション(evaluation)」が邦訳されたものであった。「教育評価」という学術用語は、戦後生まれの言葉なのである¹⁾。

その後、「教育評価」という用語は、教育(学習)効果に対する判断行為一般をさす概念として教育世界の中に定着し、今日に至っている。しかし、この事態は言葉が日常化した結果というよりも、「教育評価」概念が、その当初内包していた固有の意味を失い、「試験」「考査」「測定」という概念と混同・融合されていく経緯と重なっている。ここには、教育評価のひとつの立場である、「相対評価」の興隆が大きく作用した。そして再び、「エバリエーション」としての教育評価の概念が復権するのは、日本では1970年代になつてからである。

この章では、まずは「教育評価」の本来の意味を探るべく、「エバリエーション」を提唱したタイラー(Tyler, R. W., 1902-1994)の考え方をみてみよう。その上で、次章では邦訳された「教育評価」概念が、その反対物にまで転化していく理由を探るとともに、再び「エバリエーション」として復権する歩みとその到達点を明らかにしてみたい。このような基礎作業によって、教育評価のさまざまな立場を規定するとともに、それらの関係が構造的に浮き彫りになってくるだろう。

第1節 「メジヤメント」運動とその批判

「エバリエーション」という概念は、1920年代に隆盛をきわめていた「メ

「ジャゼメント」運動(measurement movement)に対する批判を背景として、1930年代にタイラーによって確立されたものである。そこで、まず批判対象となった「メジヤゼメント(教育測定)」の特徴とその考え方に対する当時の代表的な批判を紹介したい。

アメリカにおいて、20世紀を前後する頃から勢いを増すことになった「メジヤゼメント」運動は、すべて近代の産物として、世界的にも大きな影響を与えることになった。この時期のアメリカは、法人資本主義の進行と階級・階層の移動と分裂、東欧系を中心とする新移民の激増、学歴社会の展開とリわけ高校進学率の上昇のなかで、誰もが納得のいく人材配分の公開の装置を必要としていた。そこでは、血統、財産、門地、年功に左右されない、「誰がいつやつても同じ結果が出る」という、「メジヤゼメント」運動が提案した「テスト」法に期待が集まっていた。第二次世界大戦後の日本において、教育評価論の主流を占めることになった「絶対評価」の考え方は、その源流を「メジヤゼメント」運動に発している。そこで、まず「メジヤゼメント」運動を構成することになる二つの系譜(「アチーゼメント・テストの系譜」「精神検査法の系譜」)を取り出すことによって、その特徴を指摘してみよう。

(1) アチーゼメント・テストの系譜

「メジヤゼメント」運動の問題意識をよく表現するものとして、ソーンダイク(Thorndike, E. L., 1874-1949)の「もし物が存在するのなら、それは量として存在する。もし、それが量として存在するならば、それは測定できる。(“If a thing exists, it exists in some amount, if it exists in some amount, it can be measured.”)(1914年)という言葉がとくに有名である。この言葉には、ニュートン物理学以来、空想や直観を排して科学の究極の洗練さを保証してきた「測定」への信頼と、その測定方法を教育世界に転用することで教育の科学化をはかるうとする意図が込められていた。

このような立場からは、「絶対評価」に基づく伝統的な判定方法に対する批判が生まれることになる(コラム1-1参照)。「絶対評価」とは評価をする者の絶対性を規準にするものであって、たとえば西洋中世の大学における学位授与や伝統芸能に継承されている資格認定に由来するものである。したがって、「絶

—コラム1-1 「絶対評価」の交通整理

「絶対評価」という言葉は現在少なくとも三つの意味で使われている。そのひとつは、評価をする者の絶対性を規準にするものであって、その場合には往々にして評価者の主観的で恣意的な判断が支配的になる。このような評価は戦前の「審査」の時代に横行していたことから戦前型の「絶対評価」と呼んだり、やや一般的に「認定評価」とも命名されている。

「絶対評価」の二つ目の意味は、「個人内評価」と同じ意味で使われる。ここでは、その子どもの持つ個性を絶対視するという意味である。

「絶対評価」の三つ目の意味は、2001年の指導要領改訂で採用された「目標に準拠した評価」である。ただ、「絶対評価」の語感からは戦前型の「絶対評価」を連想しやすいことから、「絶対評価」という言葉の使用には慎重であるべきで、本書ではこの混同を避けるために「絶対評価」はひとつ目の「認定評価」の意味に限定して使用する。

対評価」の絶対性は評価者の見識や属身的な権威に依存するものであり、そこに歴史的な存在理由を持っていた。しかしながら、このような絶対性を崩壊させた近代社会は、その時代の申し子である「メジヤゼメント」運動によって、「絶対評価」における評価者の主観的で恣意的な判断のありようを遠慮なく暴き出すことになる。スターチ(Starck, D.)たちが報告した(1913年)、論文体検査の恣意性や、正確に採点可能と思われた1枚の平面幾何学の答案結果が最低28点から最高92点までの変動幅であったことはよく知られた話である²⁾。

このような「絶対評価」の主観性・恣意性を克服し、アチーゼメント・テストの結果解釈をより客観的にするために標準テスト(standardized test, より限定して standardized achievement test)が開発されることになる。標準テストとは事前に標本集団に実施して得た集団標準(このプロセスを標準化と呼称)に照らして、検査結果を解釈するテストのことであり、規模の大きい母集団(国レベルや州レベル)における子どもたちの位置を知るには有効とされた。ソーンダイクの教え子であるストーン(Stone, C. W.)やカーティス(Courts, S. A.)の開発した算数の標準テスト(各1908年、1909年)、ソーンダイク自身が開発したスペリングの尺度を用いた標準テスト(1910年)などがその代表的なものである。

なお、その際ソーングイクの学習論を特徴づけていた「結合説(connectionism)」が、テスト作成に大きな影響を与えることになった。「結合説」とは、すべての学習過程は等質であって、洞察や推理といった高次の精神活動も神経結合の量的差異にすぎないとするもの(刺激=反応学習説: stimulus-response theory)である。したがって、量化が可能で容易である事実・情報や「知識」の測定結果によって、高次の精神活動の様相も類推可能であるという主張に根拠を与えることになった。この結果、測定が容易な客観テストの形式(再生法、完成法、真偽法、多岐選択法など)で出される多量の問題群を解く「正確さ」や「速さ」が、重要な評価の指標となる。

(2) 精神検査法の系譜

「マジヤメント」運動を構成するもうひとつの系譜は、精神検査法(mental testing)の開発をめざす動きであって、ターマン(Terman, I. M., 1877-1956)はその代表的な人物である。ターマンによれば、少年期に出会った「骨相学(phrenology)」こそ人間の個人差に興味を持つきっかけとなったとされる。骨相学とは、精神の諸機能が局在する脳皮質の形と、頭蓋骨の形との相関が高いことを根拠として、逆に頭蓋骨の輪郭から精神機能の特性を推論しようとする、発表時には純粋な医学理論として登場したものである⁹⁾。

しかし、19世紀前半のアメリカに移入される項には「人相学(physiognomy)」の一種として通俗化して大流行をみるとともに、南北戦争を境にしてその非科学性のために衰退の運命を辿ることになった。にもかかわらず、人体上の可視できる特性から精神の個人差を推論しようとする発想こそ、ターマンをして精神検査法に導いたものである。個人差への着眼という、すぐれて近代個人主義の所産とされる思想的発露が、民衆の次元では骨相学の流行として浸透し、「マジヤメント」運動を準備していくという運関は興味深い。

さて、精神検査法は、「メンタル・テスト」という呼称の創始者であるキャテル(Cattell, J. M., 1860-1944)を経て、ターマンによるビネー知能検査法の改訂によってほぼ完成に達した。ここで大切なことは、この精神検査法のあゆみに対するゴールトン(Galton, F., 1822-1911)の影響であって、その思想は「マジヤメント」運動の人間観、子ども観、発達観に強く刻印されることになった。ゴー

ルトンは母方の従兄弟にあたるダーウイン(Darwin, C., 1809-1882)の「種の起源」(1859年)に触発されて、人類遺伝学の研究を始めた人物である⁴⁾。

19世紀後半の思想形成に新基軸を構築することになるダーウインの進化論は、生物界に普遍的に存在している種のなかの個体差が、どの種でもその種の生存可能な個体数以上の個体数を生産する(ブルサス説)ので、個体間に生存競争を生じさせる原因となると指摘する。その結果、少しでも生存上に有利な変異を持った個体が、最適者生存(または自然選択の結果)として生き残り、その有利な個体差を遺伝によって次代に伝えることによって、やがて新しい種の登場をみることになる。つまり、種の不変性というドグマが否定されるときにも、個体間の変異(variation)こそ生物進化の前提と考えられたのである。

しかしながら、このダーウインの主張はあくまでも生物学説として提起されたもので、人類との関係や人間社会への適用については、彼自身きわめて慎重であった⁵⁾。ところが、まさにこの点がダーウインニズムの是非をめぐって鋭く問われたのであり、スペンサー(Spencer, H., 1820-1903)による社会ダーウインニズムの主張は19世紀後半のアメリカ社会を席卷する勢いを示した。社会ダーウインニズムとは、この人間社会においても最適者生存(the survival of the fittest. この用語はスベンサーの造語)の法則が貫徹して、最適者(勝者)が生存競争を通じて社会的に優位な立場を占めるもので、しかもこの事態こそ人間進化の健全な過程であるから、国家による制限や社会の変革によって競争に干渉してはならないと主張した。この社会ダーウインニズムは、当時レッセ・フェール(自由放任主義)下で法人資本化を強かに推進していたアメリカの企業家にとつて、自らの立場を正当化する思想としておおいに歓迎されることになる⁶⁾。

ゴールトンもまた、社会ダーウインニズムの影響のもとで、人間の個人差や人種差について関心を持つようになる。その著「遺伝的天才」(1869年)⁷⁾において、ケトラー(Quetelet, L. A. J., 1796-1874)が人体にも適用できるとした「鐘型の頻度の法則」(正規分布)は心的能力にも適応可能であり、その「名声」で判別できる「拔群」(天才)は4000人につき1人であると述べている。また、人種間においても、明白な優劣の差がある(たとえば、「ネグロ人種はヨーロッパ人種より劣る」と指摘する。さらに一歩を進めて、文明社会では自然淘汰が弱まることから、人類の生物学的改良のためには、人為選択(出生の統制)が必要であると

主張する。これが、アメリカの心理学界にも影響を与えることになった優生学(eugenics)の起源となった考え方である。

ターソンは、このゴールトンの考え方に基づいて、ビネー(Binet, A., 1857-1911)の知能検査法(コラム1-2)を改訂して、「メジヤメント」運動の成果とされる「スタンフォード=ビネー知能尺度」を提案する。しかしながら、ターソンとビネーは知能程度に差のある子どもたちを識別するための客観的で信頼のおけるテスト法を開発するという点では一致しているが、両者には発達・教育観においてかなりの差があった⁹⁾。ターソンの主張はある意味で明快であった。子どもたちの知能は生得的に決定されているのであるから、それを正確に診断できる知能テストの結果に従って、子どもたちを分類し、それぞれの知能レベルに見合う教育活動(具体的には能力別編成(tracking))を実施すべきである。しかるに、現状の学校では、子どもたちの知能差を無視して、平均化しようとする無駄で非効率的な教育活動を行っている。その結果、優秀児にも劣等児にも学校は苦痛なものになっている。

しかも、IQ(精神年齢÷暦年齢×100、ターソンが知能検査法に導入)の恒常性から判断して、子どもたちの教育可能性の予測は早期に可能であり、小学校1年生でもかなり正確に判別できる。また、この知能レベルは、親の社会階層と子どものIQが対応していることから、遺伝するものと考えられた。したがって、ターソンによれば、今日の民主主義とは素質の平等を唱えることではなく、知能開発のための機会の平等を与えることであり、とりわけ将来の国の命運を担うIQの高い優秀児を早期に発見して、彼らにこそ手厚い教育(高等レベルの教育)を保障すべきであると主張した。その際、「数量化の使用」と称されるゴールトンが提唱した、知能分布は「正規分布」するという考え方が、ターソンに多大の影響を与えた。その結果、「正規分布」するように作成された知能テストこそ科学性や客観性を保証すると理解され、ややシニカルな表現であれば、このような作為に基づいて構成された知能テストによって、まさしく「個人差」が「発明」されたのである。

1917年にアメリカが第一次世界大戦に参戦すると、新兵の人材配置を効率的に処理するための委員会(会長ヤークス(Yerkes, R. W., 1876-1956))が招集される。ここに至って、「メジヤメント」運動を構成していた二つの系譜は合流し

—コラム1-2 ビネーの知能検査法—

ビネーが知能検査法を開発する契機となったのは、1904年にフランスの文部省が知的障害のある子どもたちの状態を報告する委員会を招集し、その委員になったことに始まる。そこでビネーは、知的障害児を診断し分類する従来の方法とくに医学的方法(骨相学に由来する頭蓋測定法や人体測定法など)の主観性を手厳しく批判して、知能の測定尺度による分類を1905年に提案する。さらに、1908年の改訂において、知的水準の発達の事実に着目して、正常児を標準とする年齢尺度法を提起する。そして、没年となった1911年には1908年版にさらに改訂を加えている。

ビネーの場合には、知能検査法で測定された知能水準はあくまでも現在の状態であり、将来の発達に対する予測可能性を持つものとは考えていない。ここには「知能」の持つ複雑性や多面性(理解力、工夫力、方向づけ、批判力)を現行の知能検査では十分に把握できないとする慎重な態度とともに、それ以上に環境や教育の条件によっては知能水準も可塑性を持つとする確信があったからである。あくまでも診断は治療のためにこそ必要とされたのである。

参考文献：T. H. ウルソフ著、宇津木保訳、「ビネーの生涯」誠信書房、1979年(原著1973年)参照。

て、「軍隊テスト(Army Alpha と the Beta Test)」を作成する。ここでは、大量の新兵を効率的に分類するために、採点が容易で時間短縮をもたらすものとして、従来の「個別式」に代わって「集団式」かつ「客観テスト」が採用されることになった。

また、生得的な知能水準に学業も強く相関するという精神検査法の立場が一般化して、アチーブメント・テストの分野にも「ガウス曲線」である「正規分布曲線(ノーマル・カーブ)」を客観性の規準とする相対テスト法が普及していき、この関係から、教育測定家たちの関心は、テストの内容的妥当性(そのテストで何が測定されるか)よりも、テストの信頼性を高めるための統計的な尺度法の開発(テスト、ゼロ・ポイントなど)に集中していき、このようにして「軍隊テスト」の作成によって「メジヤメント」運動の地位が高まり、産業界や教育現場にテストが普及し、以後10年間はとくに映画の題名をもじって「テスト狂時代」と称されるようになる。第二次世界大戦後の日本に導入されることになる「5段階相対評価(5は7%, 4は24%, 3は38%, 2は24%, 1は7%)」は、この「メジヤメント」運動にたしかな淵源を持っている。

(3) 「メジヤメント」運動に対する批判

以上の特徴を持つ「メジヤメント」運動がまさに絶頂期にあった1920年代に、そのパラダイムに対して、本格的な批判が開始された⁹⁾。その代表的なひとりには『教育的価値』(1911年)の著者であり、本質主義者(essentialist)として著名なバグリー(Bagley, W. C., 1874-1946)の批判であり、もうひとりには進歩主義者(progressivist)の代表であったデューイ(Dewey, J., 1859-1952)の批判であった。この両者の批判は、「相対評価」の問題性を確認するためにきわめて重要である¹⁰⁾。

バグリーの批判は、「メジヤメント」運動に浸透していた社会ダーウニズムに向けられた。すなわち、もし生得的知能によって教育の可能性が制限されるならば、学校は教育機関であるよりも生得的知能の量を証明し、優秀児を選抜する機関にすぎなくなる。しかし、民主主義社会における公教育は、庶民の子どもたちに、人類文化の遺産を普及することによって、主権者としての統治能力を形成する役割を担う。したがって、リーダーの質はまさに庶民の知性のレベルに依存するのであって、その逆ではないと強調される。

もちろん、子どもたちの学習上の「個人差」は存在する。しかし、問題の核心は、この「個人差」の意味をどう解釈するかにある。それは絶対的なものではなく、「個人差」を考慮した教授方法の改良を通じて、共通の教養を獲得させることは可能である。その際、あらためてその「個人差」を発見する道具としてのテストの役割を再評価すべきである¹¹⁾。テストや測定それ自体が悪いのではなく、テストや測定の宿命論的解釈(知能テストで測定されるのは教育程度の差であるのに生得的知能の差と考えること)こそ問われなくてはならないと主張したのである。

他方、デューイは、「知能差→教育差→社会階層差」という「メジヤメント」運動が想定しているシェーマを合理的ケースト制であると批判した上で、その運動内部にあった矛盾を次のように抉剔する。すなわち、言うところの「個人差」とは、統計的基準(標準化)によって量的に分類されたものであって、決して固有の質と内面の自由を根本的徴表とする個性(individuality)の差をとらえてはいない。しかも、その「個人差」の規準となっているのは、生理的な精

神能力ではなく、現在の社会の成功者の知性を反映しているにすぎないと指摘する。したがって、個性の分析のためには、量的基準もひとつの方法ではあるが、絶対的なものではないと強調する。

その後、デューイは、「存在→量→測定」という「メジヤメント」運動のもうひとつのシェーマに対して、質的側面を軽視するばかりか、存在するもの(既存の子どもの状態や現行の教育課程や教科書)を肯定する保守的な理論であると論破する。デューイにとつては、教育とは存在する事実を基礎にしつつも、それを変革する当為の営みであり、彼はまさにその営みのただ中で子どもを把握しなければならぬと考えたのである。

この両者に共通する「メジヤメント」運動への批判内容は、「個人差」は生まれながらの恒常的なものではなく、それはとりわけ教育活動との関数関係として生じ、したがって可変的なものであると考えるところにある。この「メジヤメント」運動への思想的批判こそ、次にみるタイラーによる「エバリ」ユエーション」の提案を準備することになる。

第2節 「エバリ」ユエーション」の主張

チャーターズ(Charlers, W. W., 1875-1952)の下でカリキュラム研究に従事し、ジャッド(Udd, C. H., 1873-1946)の影響を受けたタイラーが、「エバリ」ユエーション」論の揺籃の地であるオハイオ州立大学に地任するのは、1929年のことである。タイラーが地任した頃のオハイオ州立大学は拡張期を迎えており、教育研究の充実を期そうとしていた。テスト部門(the Division of Accomplishment Testing)に配属されたタイラーは、まず植物学と動物学の基礎コースを教授していたスタッフと協力して、大学における教育実践の改善に取り組むようになる¹²⁾。

(1) タイラーの問題意識

まず、タイラーは当時の大学生に、大学教育に関する興味深いインタビューを行っている。その結果、60%の学生が、学生の任務は、教師が重要であると指摘した「事実」や「情報」を暗記することであると答えている。そのよう

に考える理由は、試験ではもっぱら教科書や講義に出てくる「事実」や「情報」を暗記し、想起することが要求されるからである。他方、教師たちは、「事実」や「情報」なしには思考が成立しないことから、思考の活性化のためにはまず「事実」や「情報」の暗記が必要であり、そのことが学習への刺激にもなると考えていた。

このような状況に対して、タイラーの問題・批判意識は次のようなものであった。まず、試験が持つ教育実践上の規定性への着目である。たしかに、以上のケースでは、試験が否定的な「拘束服」の役割を演じてはいる。しかし、だからといって、試験廃止論を唱えるのではなく、その試験が実際に教育実践に与える効果を認めたいうえで、その効果を教育的価値の実現の方向でコントロールしようとする。この問題意識が、「test」や「examination」や「measurement」に代わる「evaluation」を意図的に使用し、「メジャメントからエバリュエーション」へとパラダイムを転換させたものである。

その際、タイラーは当時の「メジャメント」運動に生じていた、測定条件に統計学的手法を導入することによって測定行為の自己目的化をもたらす傾向を批判して、テストや試験の目的規定を「その教科の特定の目標に生徒が到達する度合いを測定する道具」であることを確認する。とりわけ、測定原理の転換を求め、「妥当性」とは「その教科の本質を測定しているのか」ということであり、「目標の測定こそ妥当性の本質的な規程である」と指摘する。まさしく評価の規程は、価値的な判断規程となる「教育目標」であると把握した。

さらに、この教育目標自体の拡張が自覚化される。先のアンソートにあったように、試験を「事実」「情報」に限定する心理学上の背景には、先に指摘したようにソーングの「結合説」がある。再論すれば、この理論は、すべての学習過程を等質とみなし、洞察や推理といった高次の精神活動も神経結合(S-R結合)の量的差異にすぎないと説明する。したがって、量化的可能であり、容易である「事実」や「情報」に関する知識の測定の結果によって、思考力や理解力といった高次の精神活動の様相も類推可能であるという主張に論拠を与えていた。

これに対して、タイラーに影響を与えたジャツドは、事実や情報の「記憶力」にすべての精神能力が還元されるのではなく、原理の適用や推論といった

精神活動は「記憶力」よりも高次であって、それ独自の機能と論理を持っていると批判した。後にジャツドとともに著した本の中で、タイラーは、「事実の暗記は、より高次の精神活動の発達を押し止めるということになりやすい。もし、原理の適用や推論といった高次の精神活動を教育しようとするならば、それに適合する学習条件が必要である」と述べている。もちろん、これらの高次の精神活動(科学の方法といった認知面だけでなく、関心・態度といった情意面も含む)を目標の射程に入れるということは、目標論と測定論に新たな課題を課すことになった。

(2) 「エバリュエーション」の構想

以上の問題・批判意識にたつて、タイラーは次のような新たなテスト作成の具体的な提案を行っている。

- I コース目標の定式化
- II 各々の目標を生徒の行動との連関で規定すること
- III 生徒が各々の目標の存否を示すであろう状況を収集すること
- IV 状況を生徒に提示すること
- V 各々の目標に照らして生徒の反応を評価すること
- VI 評価の客観性の決定
- VII 必要な時に客観性を改良すること
- VIII 信頼性の決定
- IX 必要な時に信頼性を改良すること
- X 必要な時にもっと実践的な測定方法を開発すること

この構想には、「メジャメント」に代わる新しいパラダイムとしての「エバリュエーション」の特徴がよく現れている。とりわけ重要な二つの点を指摘しておきたい。まず、ステツプIで目標の内容面を規定した後に、ステツプIIで後に「行動目標」と呼称されるようになる提案を行っている。この「行動目標」の提案は教育評価の研究と実践に重要な影響を与えたものであり、その真意を明確にするために、少し引用が長くながるが、以下に示しておこう。すなわち、「各々の目標は、そのコースが生徒に発達することを助けるべき行動の種類を明確にする用語で定義されなければならない。すなわち、記述は、その目

標に達した人に期待できる反応を叙述することによって、その目標の意味を説明するようではなければならない。このことは、その目標が達成された時、どのように伝えることができるかを明確にするのに役立つ。なぜなら、目標に達した人は、この分析で特定化された行動によって特徴づけられるだろうから。行動は、ここでは広義に使われており、生徒の適切な反応たとえば精神的 (mental)、肉体的 (physical)、情緒的 (emotional) 等の種類を意味している¹⁴⁾。

このように、教育目標に対して「行動の分析」をとった理由を、タイラーは次のように例証している。たとえば、動物学でカエルの解剖を教授する場合、従来であれば「教科書分析」によって実験の記述を暗記させ、その結果をテストしていた。しかし、「事実から推理を引き出す能力」といった場合、該当する実験の詳細を暗記することが重要なのではなく、自分で実際に実験を説明できることが期待されている。すなわち、実験を観察し、実験の結果を読み、道理ある推理によって仮説を構成することを実際の行動で示すことが肝要なのである。したがって、教育目標は「生徒がこの内容をなすために期待される反応を規定する」必要がある。このように教育目標を内容面と行動面の二次元で記述することこそ、「目標に準拠した評価」の原型となった提案である(表1-1参照)。

そして、このようにして設定された目標の実現の度合いを測定する方法について、タイラーはさまざまな方法を提案している。たとえば、「推論」を測定する方法として、5段階の尺度法を考案したプロセスを紹介したり、客観テストである「真偽法」「多肢選択法」「完成法」を補う方法として「Master-list Tests」の方法(表1-2参照)を紹介したりしている。そして、「マジヤメント」運動を批判して、もし目標が子どもたちに実現しなかった場合には、「治療的授業 (remedial instruction)」を実施することを要求している。評価行為とは子どもたちの「個人差」をあばくことではなく、あくまでも授業実践のための「診断的道具 (a diagnostic instrument)」を提供するものと考えたのである。

以上のことを整理してみると、「エバリュエーション」概念の特徴が浮き彫りになってくる。まとめてみると、①評価の規準は、教育目標である、②教育目標は、高次の精神活動を含む重要な目標群を含むべきである、③教育目標は、生徒に期待される行動で記述すべきである、④目標実現の度合いを知るための

表1-1 タイラーによる二次元の目標記述

目標の内容の側面	目標の行動的側面						
	1. 重要な原理の理解	2. 信頼できる情報源への精通	3. プラータを解釈する能力	4. 原理を適用する能力	5. 研究成果を報告する能力	6. 幅広い興味	7. 社会的態度
A. 人体の諸機能							
1. 栄養	×		×	×	×	×	×
2. 消化	×	×	×	×	×	×	×
3. 血液循環	×	×	×	×	×	×	×
4. 呼吸	×	×	×	×	×	×	×
5. 生殖	×	×	×	×	×	×	×
B. 動植物資源の利用							
1. エネルギー関連	×		×	×	×	×	×
2. 動植物の成育を条件つづける環境要因	×	×	×	×	×	×	×
3. 遺伝と遺伝学	×	×	×	×	×	×	×
4. 土地利用	×	×	×	×	×	×	×
C. 進化と発達	×	×	×	×	×	×	×

出典：ワルツ W. タイラー著、金子孫市監訳、「現代カリキュラム研究の基礎」日本教育経営協会、1978年(原著1949年)、61頁。

多様な評価方法が工夫されるべきである、⑤もし、目標に未到達の子どもがいた場合には、治療的授業が実施されるべきである、⑥以上のことは、カリキュラムや授業実践の改善につながる、⑦以上のことは、実践家と研究者の協力によって行われることが望ましい、となる。これらの提唱は、「8年研究」によって精緻化され、後に「タイラー原理」として定式化されるシカゴ大学でのシラバスに結実することになる¹⁵⁾。

(3) 「8年研究」とタイラー原理

PEA (進歩主義教育協会：Progressive Education Association) が主催した「8年研究 (the Eight-Year Study) [(1933-1941年)] は、その委員長アイキン (Aikin, W. M.) のもとに、1930年から活動を開始する。この「8年研究」の委員会によれば、従来の中等教育カリキュラムの改革が所期の目的を達成できない最大の理由は、

表1-2 タイラーのテスト開発例

第1欄	第2欄	第3欄	歴史的事件, その他
1, 2, 4	5	3	ガリレオ

No.3の自然科学の発展に注目して下さい。ここでは、ガリレオは、この要因に偉大な影響を及ぼす活動をしましたから第3番目の欄に記入されています。さらに、No.5の探検時代の幕開けでは、ガリレオは中程度の直接関与をしていますから、この要因は第2番目の欄に記入されています。そのほかの要因については、ガリレオが影響を及ぼしたものではありませんから、第1番目の欄に記入されています。同じ方法で、次の表にあげられたそれぞれの人物、事件、制度について、適当な欄に要因の番号を記入して評価して下さい。

第1欄	第2欄	第3欄	歴史的事件, 人物 あるいは制度
直接の影響なし	中程度の直接影響	大きな直接影響	
			ウェストフリア条約後の平和
			ナントの勅令
			東インド会社
			コレット(Colet)
			イギリス議会

ヨーロッパの発展における主要な要因

1. 政治における絶対主義の発展
2. カトリック教会の改革
3. 自然科学の発展
4. 国民主義の台頭
5. 探検時代の幕開け

出典：Madaus, G. F. and Stufflebeam, D. (eds.): *Educational Evaluation: Classic Works of Ralph W. Tyler*, Kluwer Academic Publishers, 1989, pp. 53-54.

大学入試制度の持つ中等教育カリキュラムへの規制力を的確に把握できていなかった点に見いだされる。そこで、委員会は、公私立、規模の大小、及びアメリカの各地域の代表という点に留意しつつ、30の進歩主義(progressivism)にたつ実験校を選択するとともに、約300校のカレッジとユニバーシティに在来の入学試験を免除して、これらの実験校の卒業生を受け入れることを約束させるという画期的な活動計画を立てた。

しかしながら、大学入試を免除されたということは、大学入学に対する資格が不問にされたということではない。実験的なカリキュラムのもとで育成されてきた生徒が、入試を免除されて、はたして実際に大学教育の場において成長、発達するのかわ、通常のコースから入学してきた生徒との比較研究で明らかにすることも要請された。大学入試という既成の評価方法を否定することで出発した研究は、ここであらためて教育評価という行為に対する本質的な究明と、実験的な教育実践に対応する多種多様な評価方法・技法の創工夫という課題に直面した。そこで、委員会は、当時オハイオ州立大学で注目すべきテスト研究に従事していたタイラーを招聘して、彼を中心として評価委員会を組織する(1942年)。

「8年研究」の輝かしい成果は、タイラーが主導した教育評価の理論と方法の構築である。「教育評価(evaluation)」とは、「測定(measurement)」や「試験(examination)」とは違って、生徒の序列づけや組分けのために行うのではなく、カリキュラムとそこに内包されている仮説を吟味、改良するため、さらには生徒たちにも学習に有効な情報を与えるために行われるものである。また、そのためにはカリキュラムの実現に向かう生徒の行動を多面的重層的にとらえる評価方法・技法(質問法、インタビュー法など)の開発も必要となってくる。この「8年研究」で確立された理論と方法は、後にカリキュラム研究における「タイラー原理(The Tyler Rationale)」として定式化されるとともに、グループたちの教育目標の「分類学」の成立にも影響を与えた。

「タイラー原理」とは、I〈目標の選択〉「学校はどのような教育目的を達成するように努めるべきか」、II〈学習経験の選択〉「どのような教育的経験を用意すれば、これらの目的は達成できるか」、III〈学習経験の組織〉「これらの教育的経験は、どのようにすれば効果的に組織できるか」、IV〈結果の評価〉「こ

コラム1-3 行動目標の二つの潮流

同じく「行動目標」と呼称されても、タイラーとブルームの系譜とは明らかに異なる点に注意しておきたい。タイラーが指摘するように、後者が想定する学習論はかつての「メジヤメント」運動と同質であって、教育目標を微細に分解して、それを訓練主義的な刺激と反応からなる授業で詰め込むものであり、プログラム学習はその典型である。それに対して、タイラーとブルームは高度な教育目標を設定して、それに向かつて教材や授業の質をより豊かに開発していかうとする点で異なる。しかしながら、この両者は後に見るような行動目標批判のなかでは区別されず混同されてしまった。このことが、「目標に準拠した評価」の発展を鈍らせた原因のひとつと考えられよう。

参考文献：Tyler, R. W., "The Father of Behavioral Objectives Criticizes them: An Interview with Ralf Tyler," *Phi Delta Kappan*, September (1973).

これらの目的が達成されているかどうか、どのように判定できるか」という四つの問いを提案したもので、カリキュラム-授業-評価を一貫したものととして把握することを求めたものである。この「タイラー原理」によって、「エバリエーション」とは、単に評価方法のみにかかわるのではなく、カリキュラムや授業の改善のために機能するものであることがより明確にされた。

しかしながら、このタイラーによる「エバリエーション」の提唱は、本国アメリカにおいても、また日本においても、すぐに教育評価研究に受容されることはなかった¹⁰⁾。そこには両国にそれぞれの特殊な事情があったとはいえず、共通に指摘できることは、「エバリエーション」が否定したはずの「メジヤメント」運動とそれを支えた「相対評価」が、その後の両国の教育世界においてはむしろ主流を形成することになったことである(コラム1-3参照)。したがって、「エバリエーション」の復権には、「相対評価」へのさらなる挑戦が必要とされた。次に、この間の事情を日本の戦後教育史に探ってみよう。

- 1) 「教育評価」を歴史的な概念として使用するときは、カギ括弧をつけて「教育評価」と表記し、一般的な意味で使用するときはそのままで記述する。
- 2) Starch, D., *Educational Measurements*, Macmillan, 1916.
- 3) 「骨相学」の意味については、小牧佳代子「骨相学の教育論的展開」佐藤達哉編「心理

学史の新しいかたち」誠信書房、2005年を、ターソンについては、ポール・ティビス・チャップマン著、菅田洋一郎・玉村公二監訳「知能検査の開発と選別システムの功罪」晃洋書房、1995年(原著1988年)を参照のこと。

4) ゴールトンについては、岡本春一「フランク・ゴールトンの研究」ナカニシヤ出版、1987年を参照。

5) ダーウソン著、八杉竜一訳「種の起源」岩波文庫、1963年(原著1859年)、第2,3,4章参照。進化論と人類との関係については、ハクスレー(Huxley, T. H., 1825-1895)が、「進化論の人間社会への適用については、スペンサーが大きな役割を果たした。ダーウソンのこの点への重要な言及は、ようやく1871年の『人類の起源』においてなされた。

6) 進化論とアメリカの保守思想との奇妙な結合を提示したR. ホフスタッター著、後藤昭次訳「アメリカの社会進化思想」研究社、1973年(原著1944年)参照。

7) ゴールトン著、甘粕石介訳「天才と遺伝」上・下、岩波文庫、1935年(原著1869年)。

8) ビネーとターソンの相違については、滝沢武久「知能指数」中公新書、1971年に詳しい。また、佐藤達哉「知能指数」講談社現代新書、1997年も参照のこと。ターソンの主張は、Terman, L. M., *The Measurement of Intelligence*, The Riverside Press, 1916とTerman, L. M., "The Use of Intelligence Tests in the Grading of School Children," *Journal of Educational Research*, Vol. 1, No. 1 (1920)を参照のこと。

9) 「メジヤメント」運動批判については、ステイヤー・ダン・ジュール著、鈴木善次ほか訳「人間の測りまちがい」河出書房新社、1989年(原著1981年)に詳しい。

10) バグリーの主張は、Bagley, W. C., *Determinism in Education*, Warwick & York Inc., 1925を、デューイの主張は、Dewey, J., "Mediocrity and Individuality," と "Individuality, Equality and Superiority," *New Republic* 33(1922)、さらにDewey, J., "Progressive Education and the Science of Education," *Progressive Education* 5(July 1928)を参照のこと。なお、この時期のデューイ思想の特徴については、古屋憲太「IQ論争期におけるジョン・デューイの「個性」(Individuality)概念の展開」『教育学研究』第68巻第4号(2001年12月)に詳しい。

11) バグリーが開発した評価問題が、榑橋健治「アメリカ社会科学習評価研究の史的展開」風間書房、2002年の第2章で紹介されている。

12) タイラーの初期論文については、Madaus, G. F. and Stufflebeam, D. (eds.), *Educational Evaluation: Classic Works of Ralph W. Tyler*, Kluwer Academic Publishers, 1989を参照のこと。

13) Tyler, R. W., "The Relation between Recall and Higher Mental Processes," in *Education as Cultivation of the Higher Mental Processes*, Judd, C. H., Breslich, E. R., McCallister, J. M. and Tyler, R. W. (eds.), The Macmillan Co., 1936, p. 17.

14) Madaus, G. F. and Stufflebeam, D. (eds.), op. cit., p. 29.

15) Tyler, R. W., *Basic Principles of Curriculum and Instruction*, University of Chicago Press, 1949. 邦訳は、ラルフ W. タイラー著、金子孫市監訳「現代カリキュラム研究の基礎」日本教育経営協会、1978年(原著1949年)。

16) 米国の大学入試は、1980年代に登場した「8年研究」をはじめとするさまざまな改革案の中から、教育測定派が主導するSAT(Scholastic Aptitude Test, 邦訳は「大学進学適性テスト」)を採用するようになる。このSATは専らIQテストであって、大学入試の妥当性に欠けると批判されるようになったのは、半世紀余り経った1990年代になってからである。1983年のレポート「危機に立つ国家」以降、「教育的スタンダード」の設定という、「目標に準拠した評価」への転換がSAT批判を押し進め、その結果1994年には「AI」が「Aptitude」から「Assessment」に置換され、SATが「Scholastic Assessment

「test」となったことは象徴的な出来事であった。まさしく、大学入試が「教育評価」の一環として把握し直され始めたのである。ニコラス・レマン著、久野温穂訳『ビッグ・テスト』早川書房、2001年(原著1999年)と荒井克弘・橋本昭彦編『高校と大学の接続——入試選抜から教育接続へ』玉川大学出版部、2005年を参照。

第2章

「教育評価」の混迷と再生

第1節 「エバリュエーション」と「教育評価」のあいだ

第二次世界大戦直後の日本において、アメリカの進歩主義の考え方とともに、「エバリュエーション」も紹介され、「教育評価」と翻訳されることになった。そして、この時期に『教育評価』の題名を持つ著作が多く出版されることになった。しかしながら、この「翻案」事情は、それほど単純なものではなかった(図2-1参照)。

そのひとつには、「エバリュエーション」概念の発信源となったアメリカでは、その創始者であるタイラーの理論と、マジヤメント派のある人々(修正測定派)がその理論を併呑・修正した理論が、同じく「エバリュエーション」を呼称していたことがあった。事実、この時期に日本で紹介された文献には、この二つの系譜があった²⁾。そのため、「エバリュエーション」と「マジヤメント」の区別が判然としない中で、「教育評価」が語られるようになる。

そして、この事態を助長した日本側の条件として、その概念の紹介を担った主体が、おもに「測定(マジヤメント)」概念の研究を戦前に行っていた教育心理学者とその門下生にあたる人たちであったことである。ただし、それらの教育心理学者も一様ではなく、その立場の相違が翻訳に際しても反映されることになった。また、「エバリュエーション」概念が導入されようとした日本の土壌は、明治以降からの「考査」概念が支配的であった。これらの事情も、その紹介に際して独特の色彩を与えることになった。以下、詳しくみていこう。

(1) 「エバリュエーション」の紹介

ここに波多野完治(1905-2001)の興味深い思い出話がある³⁾。おそらくは第二次大戦直後の頃であろうが、波多野によれば、日本では「評価」ということばは、授業の結果が子どもにどのよう⁴⁾に現れたかを「ネブミ」することの意味で