

# 1910—30年代におけるアメリカ社会科の新しい展開

— 科学的決定論の導入による学習成果測定 —

平田嘉三・棚橋健治

(1985年9月10日受理)

The New Development of Social Studies in the United States  
in the 1910's—1930's  
— The Measurement based on the Scientific Determinism  
on the Outcomes of Social Sciences Teaching —

Kazou Hirata and Kenji Tanahashi

The purpose of this article is to make an inquiry into the history of the educational evaluation in social studies from the viewpoint of the rise in objectivity, which came into existence in the Report of the Commission on the Social Studies of the American Historical Association.

The studies on the objectification of the educational evaluation started in the early part of twentieth century. At that time, also in the field of teaching social sciences, the studies on the measurement of the outcomes of learning are made. Externally, the measurement movement in the teaching of social sciences is concerned with the improvement of the form of the tests in order to remove the vagueness inherent in the traditional examination. But the measurement movement changed not only mere the technique of constructing the test. The measurement in the teaching of social sciences aimed at the removal of empirical and speculative thinking from the educational research on social studies. The measurement in the teaching of social sciences was one aspect of the introduction of Scientific Determinism into the educational research on social studies.

The scientific nature which prescribes the substance of the measurement in the teaching of social sciences has four aspects. That is to (a) specify the object of measurement by dividing the goal of teaching social sciences, (b) raise the content validity by analyzing the textbooks in history and the social sciences, (c) adopt the type of test question which obtain the same result at any time by anyone, (d) rationally interpret the result of measurement by making the criterion based on statistical managements.

The measurement in the teaching of social sciences has two distinctive features. First, it regards the person as the gathering of elements, and measures the person elementary like the physical phenomena. It is the peculiar conception, opposite to wholism. Second, it is based on the theory that the children are the center of interest, so the criterion is made by the competition in the certain group of children who learn the social sciences. These have something to do with the social conditions in the early part of twentieth century in the United States. That is, the coming into existence and development of the capitalism and the request for the rationality.

## はじめに

今日わが国の社会科教育学における評価論は、評価における客観性向上をめぐる展開されている論と、教育全体の目的や価値観と評価との関係をめぐる

て展開されている論とに大別できる。今日の評価論が二つの要因に規定されている理由は、評価論の歴史自体がこの2点をめぐって発展してきたからである。評価論のいわば系統発生において現われた二つの問題が今日の評価論の論点を規定しているのである。した

がって、今日多様な展開をみせているこれら2点に関する研究の一層の進展のためには、評価論史における各要因の原点に立ち戻って検討することも必要であろう。本稿では、二つの要因のうち、評価論史における客観性向上という要因について考察してみる。

評価における客観性向上をめぐる論の発展により、評価論史に一つの転機をもたらしたのは、20世紀初頭にアメリカでおこった「教育測定運動」(Educational measurement movement)である。教育測定運動については、心理学や教育学の立場から多くの研究がなされてきて、大きな成果があげられている<sup>1)</sup>。教育測定運動が活発になった1910～30年代には、アメリカの社会科学においても学習成果測定の試みが盛んに行われていた。従来の研究の多くでは、教科における教育測定運動については、心理学の進歩によって開発された知能検査法を教科の学習成果測定に適用したのとして説明されているだけである。社会科学という一教科に限定して、そこにおける測定の論理と実践や、さらには社会科学教育研究の発展において測定論の担った役割などに焦点をあてた論はみられない。

本稿は、1910～80年代のアメリカにおける社会科学の学習成果測定論を考察することを目的とし、それを、教育測定運動の本質を規定している科学的性格という観点から論ずる。当時の社会科学における測定運動を、合理性追求という時代の要請を基盤とするものにとらえ、社会科学の学習成果測定の合理性を、社会科学教育研究の科学化という視点から考察しようとするものである。

社会科学の学習成果測定が具備している科学的性格は、(a)測定対象の特定、(b)テスト内容の妥当化、(c)テスト形態の妥当化、(d)測定結果の解釈と利用の合理化にみられる。これらは、教育測定運動以前のいわゆる伝統的な試験にはみられないことである。以下、まず教育測定運動以前の試験について略述し、さらに教育測定運動の背景として、合理性追求という時代の要請を社会および教育一般について略述してから、科学的性格の四つのアプローチにより、1910～30年代のアメリカの社会科学学習成果測定論を検討してみる。

## II 教育測定運動以前の成績考査

教育活動における広い意味での評価という行為は、教育測定運動以前からなされていた。たとえば、中世のヨーロッパにおいて成立した大学においても一種の資格試験が行われていた。しかし、中世の大学に端を発する論述試験は、直接、現代の教育評価の起源となったとは考えられていない。このころの論述試験

は、形態としてはきわめて概括的なものであった。社会科学の教科においても「フランス革命について論じなさい」というような出題のされ方が一般的であった。

伝統的な論述試験のもつ問題点については、次のように述べられている。「伝統的試験について研究したところ、同一の子どもの解答用紙を多くの教師が互いに話し合わずに採点したところ、50%以上の項目についての採点結果に相違が出たことがしばしばであった。また、1人の教師が同じ解答用紙を別の機会に、たとえば数日後に採点したところ、非常に異なったものとなった。同じ目的で別々に作った試験でも、及第とされる子どもの割合は年によって大きく変わり、その難易度には大きな差が出るだろう。…(中略)…伝統的な試験は、次のような事実を無視している。すなわち、ある学年の子どもの中には、その学年で学ぶべきとされている内容を学ぶだけの準備ができていない者もいる一方で、すでにより上の学年の内容も習得している者もいるという事実である。」<sup>2)</sup>このように、伝統的な試験では、採点に対する信頼性が低いと同時に、子どもの存在が考慮されていなかった。すなわち、伝統的な試験は学年なり学期なりの範囲で行われた授業内容の理論が中心であり、子どもを興味を中心とする理論にもとづくものではない<sup>3)</sup>。したがって、授業内容の習得状況の判定も、教師の一存で作られた基準により、教師ひとりの判断でなされるのである。

現代の教育評価は、このような伝統的な論述試験もっていた非科学性への批判にもとづく科学化運動の産物として成立したのである。

## III 学習成果測定論生起の基盤

通常、歴史家の間では19世紀末期から20世紀初頭にかけては、アメリカ史において特別な意義が認められ、いわゆる「世紀転換期」(The turn of the century)と呼ばれている。アメリカの社会は、世紀転換期に端を発する一大変化をみた<sup>4)</sup>。この変化は、無数の条件が絡みあって複雑な現象を呈しているが、一つの局面として、価値観の変化、すなわち、合理性、科学性志向が顕著になったことが挙げられる。この合理性、科学性志向が、社会、教育一般、そして社会科学という一教科の研究に大きな影響を与える要因となっていた。

### 1. 社会の変化と科学的管理論の生起

19世紀後半以降、アメリカ資本主義経済は急速に成熟し、工業生産額が農業生産額を追い越すとともに、工業生産額が世界第1位を達成するところまで工業化が進んだ。このような産業構造の変化は、企業の内的

環境の変化をもたらし、それらに対応するものとして新たな経営管理が求められることになった<sup>5)</sup>。この新たな経営管理は、経験的に得られた管理法を一部の者の間だけで独占・伝習する旧来の「伝習的管理」(Rule-of-thumb management)に対する概念として「科学的管理」(Scientific management)と呼ばれた。科学的管理のめざしたものは、企業経営の合理化と組織の能率化である。その方法の眼目は、正確な科学的知識と方法を用いて、目分量を排して法則の発見に努めることである。具体的には、標準作業時間、工具の標準化、課業観念と工具の能率などが科学的に実験、調査されている<sup>6)</sup>。

このような経済活動における科学的管理は、行政面からも支持され基礎づくりがすすめられた。そのような合理化、科学化志向の行政を推進した者のひとりがフーヴァー(H. Hoover)である。彼は、経済改造のための施策の眼目を、企業の能率化を中心とする国民的効率の増進に置き、その実現の基礎として厳密な科学的調査の必要性を強調した。彼自身、商務省を率いて各種情報の収集と配布に努め、企業合理化のデータ提供に多大な貢献をなした<sup>7)</sup>。

さらに、政治面における合理化、科学化志向は、20世紀初頭にアメリカ社会の諸矛盾に対処しておこった改革運動、いわゆる革新主義運動の中にもみられる。たとえば、専門家の科学的合理性を社会編成に生かし、科学的管理論を応用して企業経営の原則を市政に導入する要求や、その発展形態として1920年代に登場した計画的、合理的な都市社会の管理などが挙げられよう<sup>8)</sup>。

## 2. 教育科学論の発展

科学的管理論にみられる合理性・能率性追求の姿勢は、この時期の教育実践および教育研究にも反映されている。

### (1) 教育実践・研究における合理性

教育への科学的管理論の適用が最も直接的な形で現われたのは、ゲーリー・システム(Gary System)などにみられる学校の効率的運用計画である。ゲーリー・システムの目的は、義務教育制度の普及とU.S. スチール工場の進出によって学齢人口が急増したために発生した学校不足を解消することであった。これは、学校の施設や教員をむだなく活用することによって、同一施設で従来の2倍の生徒に、従来以上の学習時間を確保するというものである<sup>9)</sup>。

ゲーリー・システムほど直接的な形ではないが、科学的管理論、特にその中の最善の作業方法および標準作業時間に関する研究の影響を受けているとみられる

のが、1911年のN. E. A. (The National Education Association)「時間経済委員会」(Committee on the Economy of Time in Education)の設立である。時間経済委員会は、教育における時間の経済化を図るために、むだを取り除き、最低必要な内容を決定することをめざしたのである。たとえば、社会科に関しては、小学校の地理および歴史における最低必要内容の決定、アメリカ史の現行内容における欠陥の指摘などが行われた。ゲーリー・システムが科学的管理論における施設利用の合理性という発想を学校運営に反映させたのに対し、時間経済委員会の研究は、科学的管理論における作業プログラムの合理性という発想をカリキュラム決定にも反映させたといえよう。

教育への科学的管理論の適用は、このような直接的・実践的な形ばかりでなされていたわけではない。カリキュラム改造の研究手法への科学的手段の導入という、より基礎的な次元でも科学的管理論の発想はとり込まれたのである。

カリキュラム改造の研究は、すでに19世紀後半からおこっていたが、それが科学的研究となったのは、時間経済委員会における研究からである。時間経済委員会は、真に有用な内容の選択およびその効果的な学年配当を検討するために、代表的なカリキュラムで取り上げられている内容とその時間配当の調査、ならびに不要な内容を排除して能率化を図っている実験の調査を行った。このような教育調査は、カリキュラム改造等の教育研究に合理的な基礎を与え、研究の科学化の重要な要素となった。

このように、カリキュラムの科学的研究は、概括的・一般的な主題のもとに経験的・思弁的な考察を加えるという従来の研究方法と異なり、広範で総合的な教育現象の中から特定の局面を切りとり、具体的な調査を行って、その結果を解釈するという研究方法をとっている。これは、当時急速に発達してきた自然科学が自己の研究方法を科学化するためにたどった変遷を参考として、得られたものである。また、社会改造の合理的な基礎を求めて行われたシカゴ、ニューヨーク、ピッツバーグ等の社会調査の実施も、教育調査の実施に刺激を与えたといえよう。

科学的手段を導入して教育問題に対処しようとするこのような研究がめざしたことは、(a)客観的、(b)公平、(c)数学的に正確、(d)有能な観測者による実証を条件とする、ということである。カリキュラムの科学的研究においては、従来の経験的・思弁的思考が排除され、信頼できる統計的・実験的研究の裏づけが求められるようになったのである。このような教育の科学的研究の動きは、「科学的決定論」(Scientific

determinism ) と呼ばれた<sup>10)</sup>。社会科のカリキュラム研究においても、1910年代になると従来の先験的方法から、十分な調査を踏まえた方法へと変わっていった。

## (2) 教育測定論の生起

カリキュラム改造の科学研究は、裏付けを得るために学習成果に関する情報収集を欠くことができない。したがって、情報収集の手段である測定についての研究は避けられなかった。社会科においても、カリキュラム作成者は、測定によって「さまざまなレベルの能力が混在する学校において、どの程度まで社会諸科学の題材が適合するかを決定したり、われわれの目的の実現にとって不要な題材を除去したりすることができる<sup>11)</sup>」とした。

教育測定を成立させる基本的な論理は、次のように述べられている。(a)存在するものはすべて何らかの量として存在する、(b)量として存在するものは何でも測定することができる、(c)教育における測定は物理学における測定とおおむね同じである、(d)科学的教育の発達にとって測定は不可欠である、(e)生徒の自発的能力や可能性が測定不可能な範囲にあるなら、その生徒について知ることはできない、(f)どのような教育目標もそれが把握不可能な範囲にあるなら、価値はない、(g)教授の方法や材料の価値は、それらの効果が測定されてはじめてわかる、(h)教育や教育者がテストによって機械化されてしまうことはない、(i)テストが画一化を招いて子どもや授業の活気を奪うことはない<sup>12)</sup>。

ここに、カリキュラムの科学研究の基礎として、教育現象における数量的測定とそのデータの統計的処理が可能であり、かつ必要であるということが浮き彫りにされている。

## IV 社会科における学習成果測定論

教育測定運動は、合理性・効率性の追求という時代の要請を背景とし、教育問題の解決に科学的態度で臨むという科学的決定論の考えにもとづいてすすめられた。社会科教育研究の科学化を図るうえで測定運動が重要な役割を担ったのは、「測定」自体が旧来の「判定」とは異なる科学的性格をもっているからである。そこで、ここでは学習成果測定の本質を規定している教育測定の科学的性格を、社会科で考察したい。そのために、まず、社会科における測定運動の展開を概観しておく。

### 1. 社会科における測定運動の展開

社会科における学習成果測定の歴史を見ると、1915

年に最初の転機が見出される。この年、歴史テストの領域で、それまで支配的であった伝統的な論述テストに代わるものとして、客観テストがベル(J. C. Bell)とマッカラム(D. F. McClum)の手によって初めて作られた。これよりさき、読み、書き、綴り、算数の領域では、教育科学運動で要請されていた学習成果測定に使える客観的な用具の研究・開発がすでに進んでいた。それらに刺激を受けて、歴史テストでも学習成果が客観的に測れる用具が求められていたのである。ベルとマッカラムが1915年に作った「合衆国史のテスト」<sup>13)</sup>は、時を経ずして絶版となり、転機を示す指標としては微弱なものではあった。しかし、これを境として、歴史科およびその他の社会科的教科における客観テストの作成が活発に行われるようになったのであり、意義深い第一歩であったといえる<sup>14)</sup>。

ベルとマッカラムのテストに引き続いて、多数の客観テストが作成され、それらはニュー・タイプ・テスト(New-type test)と総称された。ニュー・タイプ・テストはそれ以前より主観的なテストにとって代わる傾向が強かった。その理由は、論述テストと比較してニュー・タイプ・テストが現場の教師にとって、二つの長所をもっているからである。すなわち、第1に同じ時間でより多くの標本が抽出できるということ、第2に、採点がより容易であり、しかも採点がより信頼できるということである<sup>15)</sup>。

ニュー・タイプ・テストの開発は、当初、歴史的重要な事項の年代や名称等の記憶に関するものが大部分であったが、しだいにその対象範囲を拡大し、1920年代後半に入ると、社会的態度や技能に関するものも現われた<sup>16)</sup>。それらの研究の多くは、測定対象の各領域ごとに孤立的になされており、全領域を視野におさめた体系的な研究は少なかった。このような状況下において、体系的なテスト研究を行ったのがアメリカ歴史学会(American Historical Association, 以下A.H.A.と略す)である。本稿では、このA.H.A.のテストと測定の研究を検討する。

A.H.A.は1884年に創立され、歴史研究の促進とともに、社会科のカリキュラムの研究を活発に進めた全国的な学会である。A.H.A.のカリキュラム研究は委員会によって行われ、「7人委員会報告」(1899年)、「8人委員会報告」(1909年)、「5人委員会報告」(1911年)などを次々に公にした。これらは、当時の社会科のカリキュラムの構成に多大な影響を与えた。しかし、これらのカリキュラム研究は、十分な調査を踏まえて科学的方法によって行われたというものではない。むしろ、先験的方法による委員の討議を概括するという性格のものであった<sup>17)</sup>。一方、先述のよう

にカリキュラム研究の科学化という運動が進む中で、A.H.A.の研究も科学化の方向をたどっていった。そのような状況の下でまとめられたのが、1934年から1941年にかけて次々に公にされた「社会科委員会報告」である。これは16の報告書から成り、その内容は、社会諸科学の性格、学校における社会諸科学の位置、社会諸科学教育の方法、教科としての社会科、社会科のカリキュラム構成、社会科的教科の各々、教育の社会的基礎、教師教育などである。その中の一つとして、テストと測定についての報告がなされた。

カリキュラムの科学的研究は、正確なデータ収集とその合理的な解釈を基盤とするため、そこでは、新たなカリキュラムの教授効果を示すデータ収集の用具の重要性が強調されることになる。1916年以来、活発に行われていたカリキュラム開発は、多彩なコース、科目配置、内容を呈していたが、それらの有益な比較材料は得られていなかった。このような状況下において、社会科委員会の研究は、「望むべき結果の獲得における、さまざまなカリキュラムや授業手順の相対的な効果を決定できるような用具を作り出すこと」<sup>18)</sup>を目的として行われたのである。以下、この研究を先述の科学的性格の四つのアプローチにより検討してみる。

## 2. 測定対象の特定

伝統的な試験にはみられなかった測定の科学的性格の第1は、測定対象を特定しているということである。すなわち、何を測定するのか、何が測定されているのかを特定することが、科学的な教育測定における第1の要件となる。伝統的な試験においては、解答や採点に、査定内容とは無関係の要素が非常に多く入った。たとえば、「歴史の答案の評点は、歴史の理解のみならず、いかに採点者の意にかなうように述べるかという表現力にも影響される」<sup>19)</sup>ことが多く、その試験によって何が測られるのかの特定は考慮されていなかった。

社会科における測定は、「社会科学教育の望むべき成果であると、委員会の構成員全員が受け入れることができる目標を述べることに努める」<sup>20)</sup>という活動から始められた。まず、生得的なものや習得的なものとの混同を排し、測定対象を習得的なもの、すなわち、学習成果に限定するのである。ここに、形態のうえで類似し、参考とされた心理学における知能検査と、社会科における測定の根本的な相違がある。

学習成果測定をすすめるには、測定対象を社会科学教育の望むべき成果に限定したうえで、さらに、社会科学教育の望むべき成果自体の特定がなされる。こ

では、測定対象を可能な限り正確に限定することがめざされ、学習成果が分析的に検討されて、次のようなリストが作られている。

### 〔理解の領域〕

「社会が機能するための重要な制度の理解。ここには、原理と理想が含まれる。

これは、社会科の機能として最もひろく受け入れられていることであり、しかも、最も正確である。理解と情報は同義ではなく、テストは理解を求める。この目標を効果的にテストするためには、いくつかに細分することが必要である。

政治制度 — 地方、国家、国際：原理と理想

経済制度 — 地方、一般：原理と理想

社会制度 — 国内、倫理的、宗教的：原理と理想

教育制度 — 原理と理想

美的制度 — 原理と理想

娯楽制度 — 原理と理想」<sup>21)</sup>

### 〔技能の領域〕

「社会についての情報源を用いる際の技能。

歴史科ならびに他の社会科の課程を履習する子どもは、現代社会に関係するさまざまな情報源をいかに見つけたり、知的に活用したりするかを知ることが期待される。

そのような情報源には、次のようなものが含まれる。

一般に知られたゴシップ — 口頭、印刷

理由をあげた討論 — 口頭、印刷

社会的活動 — 実物あるいは写真

偉業を示すもの — 活用中あるいは記念物」<sup>22)</sup>

### 〔視点、興味、態度の領域〕

「視点、興味、態度 — 視点、興味、態度は、これらの教科の教育成果として、ふつう受け入れられている。…(中略)…視点、興味、態度の中には次のようなものが含まれる。

〈現代のでき事に対する展望〉 — 子どもの周辺に生起するでき事をじっくりと見るための基礎を子どもに与えるということは、ほとんどすべての社会科でなすことである。

〈歴史的関心〉 — 常に、社会的でき事とその生起した時とを関連づけて考えること、としてふつう示される。

〈地理的関心〉 — 常に、社会的でき事とその生起した場所とを関連づけて考えること。

〈社会一般の善への関心〉 — この目標は愛国心と呼んだ方がより一般的である。これには、集団の福祉における自分自身の福祉を認識することの訓練や、共通の福祉を望ましい目的として尊重することの訓練が

含まれる。

〈民族的、宗教的、国家的、社会的な寛容性〉— 前項に深く関連し、道徳的不正に至るような生まれつきの偏見を特に扱う。他の民族、宗教、党派、社会集団を理解することは、それらに対するより公正な判断へ至ると信じる。しかも、それは個人にも社会にも結果的に利益をもたらすと信じる。

〈余暇の時間の興味〉— 社会や過去、現在について知ることは、同質の話題や一群の話題をできるだけ見つけようとする情熱や興味を子どもの中におこしうると考えられている。それらの話題を追求することで、その子の余暇を占める満足や喜びが与えられる。余暇の増大が一般化してくると、この目標はさらに重要になる。<sup>23)</sup>

このように、測定対象をどこまでも分析的に扱ってゆくという測定運動の性格は、人間を諸要素の集合体とみる見方を前提として成立している。つまり、単に伝統的な試験が概括的であるのに対して、学習成果測定は細分化されているというだけではない。その基盤となる人間のわかり方自体に相違がある。学習成果測定は、人間を要素的に測定し、その機械的総和をもってその人間の説明をする、という考え方を提示しているといえよう。このような考え方に立つからこそ、測定対象の特定が可能であり、かつ必要になるのである。

### 3. テスト内容の妥当化

伝統的な試験にはみられなかった測定の科学的性格の第2は、測定法を科学化しているということである。測定対象の特定は、測定法の科学化と相即不離の関係にある。測定対象を特定することによって、はじめてそれに対する測定法も科学化できると同時に、測定法の科学化が測定対象のより厳密な特定を可能にするのである。

社会科における学習成果測定の方法の科学性は、二つの局面から成り立っている。その第1は、テスト内容の妥当性の追求である。これは、測定するように意図していることと、実際に測定していることとの一致度の問題である。社会科の学習成果測定では、テスト内容の妥当化のためのテスト項目の選択は特に重要な課題であった。それは、社会科的諸教科の基礎となる社会諸科学の領域で、いわゆるミニマム・エッセンシャルズ (Minimum essentials) が確定できていないという事情による。

「理解」の領域におけるテスト項目の選択は、社会科的教科の教科書分析を中心として行われている<sup>24)</sup>。プレッシー (L. C. Pressey) 作成の歴史語彙テストの

場合は、次のような手順でテスト項目が選択されている。まず、第4学年から大学までで用いられている歴史および公民の23の教科書を分析し、そこに出てくる用語を抽出する。この際、各用語の登場頻度は問わない。次に、それらの教科書の中から、高等学校で特にひろく使われている歴史教科書6冊を選んで分析する。ここでは、各用語が何回ずつ使われているかを調べる。そして、歴史や公民の教科書を同様に分析して出版されている資料のデータをこれに加える<sup>25)</sup>。さらに、新聞の一面や社説、現在のでき事を扱った多くの論文や雑誌に出てくる用語についても同様に頻度を調べ、教科書分析によって得られた資料を補う。このようにして、1444の用語が選択され、各々の頻度が示される。

次に、列举された1444の用語を絞るために、高等学校の歴史担当教師64人と大学の歴史担当教授5人によって、これらの用語の等級づけがなされる。「必須」、「補助的」、「重要でない」の3等級である。

さらに、1444の用語は「社会学的価値」(Sociological value)によって等級づけがなされる。社会学的価値とは、歴史の授業以外の場でのこれらの用語の使われ方に関わるものである。この等級づけは、大学の歴史担当教授、社会事業家、歴史を教えていた高校長、公立学校の教育調査専門家、テスト作成の専門家などによって行われる。

最後に、これら三つの基準、すなわち、使用頻度、重要性、社会学的価値によって各々作成されたリストをつき合わせ、必須のものとして三つのリストのすべてに挙げられているもののみを抽出し、抽出された415語をテスト項目として選定している<sup>27)</sup>。

歴史の語彙として選定された用語の一例を挙げると次の通りである。

- A. 政府に関する用語 … 大使、民主主義、法案、認可、当局、帝政、布告、任命、ほか142語
- B. 政治学的用語 … 選挙運動、過半数、秘密投票、候補者、少数党、選挙、ほか30語
- C. 経済学的用語 … 業務、製造、従業員、インフレーション、商業、商品、雇用者、ほか30語
- D. 社会学的用語 … 貴族、田園、移住民、教育、農民、都市住民、ほか30語
- E. 法律上の用語 … 専断、国民性、詐欺、非合法、陪審、ほか37語
- F. 軍事的用語 … 同盟国、海軍、募兵、侵入、交戦国、ほか48語
- G. 地理学的用語 … 境界、地区、探検、農業、大陸、地域、ほか14語
- H. 宗教的用语 … 聖職、宗派、カトリック教、教

皇、宣教師、ほか8語

I. 年表や記録に関する用語 … 古代、歴史、文明化、世紀、ほか14語<sup>28)</sup>

このように、プレッシャー作成の歴史語彙テストにおける項目選定は、教科書等での使用頻度、教師の考える重要度、歴史の授業以外での価値、という三つの基準を設け、各々の基準にのっとってなされた調査結果にもとづいてなされているのである。

#### 4. テスト形態の妥当化

社会科の学習成果測定の方法における科学性の第2の局面は、テスト形態の妥当性の追求に現われている。学習成果測定の研究がなされたこの時期には、心理学者によって開発されたニュー・タイプ・テクニックの、教育測定に対する効果への信頼が絶大なものであった。さらに、この新たなテクニックを用いて、語彙、文法、読解に関する知識の進歩を測定した近代外国語の教科書専門家達の研究が認められた。これらのことから、社会科委員会は、この新しいテスト形態が社会科における測定でも高い妥当性を得られるだろうと考えた<sup>29)</sup>。

「理解」の領域におけるテスト形態の選定を、プレッシャーのテストの場合で見えてみよう。ここでは、テスト形態選定の基準として、第1に信頼性が高いこと、さらにそれに従属する基準として処理が簡単なこと、が設けられている。信頼性とは、「ある形のテストによる評点と、他の形のテストによる評点との相関関係がよく保たれていること。すなわち、1人の子どもに、同一の題材について二つまたはそれ以上の異なった形でテストしても、同じ評点になる傾向があること」<sup>30)</sup>である。歴史語彙テストでは、このような条件を満たすものとして、いわゆるニュー・タイプ・テストと呼ばれる客観テストが採用されている。客観テストの型には、正誤判定法、多肢選択法、組合法、単語解答法、完成法などがあるが、プレッシャーのテストでは多肢選択法が用いられている。テスト問題例を数例だけ掲げておく。

A-40 羊毛の衣類作成のすべての業務に関係している語句はどれか。(a) 毛糸店 (b) 羊毛産業 (c) 畜羊家 (d) 牧畜場

C-66 ある会社が他社から仕事を奪うための価格引き下げの実践、にあてはまるのは次のどの言葉か。(a) 謀反 (b) 移住 (c) 産出 (d) 競争

C-48 世界で最も強い国々は、しばしば何と呼ばれるか。(a) 帝国 (b) 国際連盟 (c) 列強 (d) 民主国家

B-78 従うべきものはどれか。(a) 法律 (b) 提案 (c) 報告 (d) 儀式<sup>31)</sup>  
(A-40などは、同テストの問題番号)

#### 5. 測定結果の解釈と利用の合理化

伝統的な試験にはみられなかった測定の科学的性格の第3は、測定結果の解釈と利用における合理性の追求である。

「理解」の領域における測定結果の解釈と利用について、プレッシャーのテストの場合でみてみよう。このテストでは、テスト結果を三つの形で処理している。

第1の処理の仕方は、子どもごとの正解項目数の度数分布表の作成である。ここでは、まず子どもひとりひとりの正解の総数を2点刻みに分けて記録し、中位数を出す。実験した第4, 6, 8, 10, 12学年各々についてこの操作を行い、一覧表にする。ここにその一部(80項目分)の表を掲げる(表1)。この作業によって、次の点が明らかになる。第1に、同一の教育レベルにあると分類されている子ども達の学習の達成に、実際にはどの程度の差があるのかが明らかになる。第2に、ある子どもの得点を他学年の中位数と比較することによって、その子どもが現在、実際にどの学年レベルにあるのかが明らかになる。第3に、隣接する学年の中位数を比較し、その差を求めることによって、1年間での子どもの変化、すなわち、どの学年と学年の間で著しく伸び、どこであまり伸びないかが明らかになる。これは、子どもの発達特性を示唆するとともに、どの学年の授業を改善することが必要かも示唆する<sup>32)</sup>。

テスト結果の第2の処理の仕方は、用語ごとの正解率とその度数分布表の作成である。まず、正解とされる用語ごとに、各学年ごとの正解者の数を数えて、百分率を出す。この作業によって、各用語の各学年ごとの習得状況とその差が明らかになる。さらに、正解率を5%ずつに区切り、さきに出した個々の用語の正解率をそこにあてはめ、何項目ずつになるかを一覧表にする。この作業によって、各学年における語彙のひろがりが見え明らかになる。また、授業を受けた子どものうち、何%の者がその用語を正解したら、その用語の学習が成立したといえるのかの基準を設けておけば、この第2の処理によって、どの学年で、どの用語の学習が成立したかの判別がつくことになる<sup>33)</sup>。

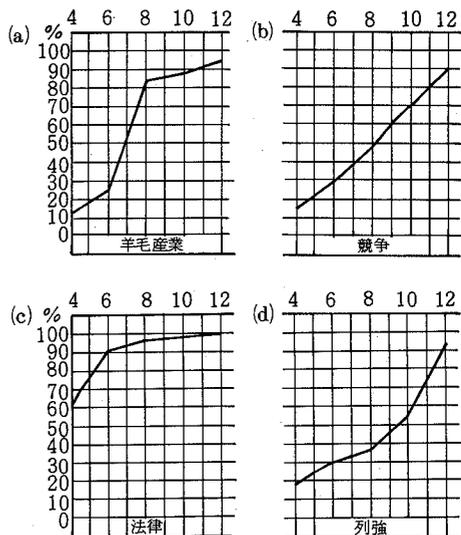
テスト結果の第3の処理の仕方は、用語ごとの正解率の学年進行による変遷のグラフ化である。ここでは、各用語の学習曲線が描かれ、そのタイプを分類することにより、各用語がどの学年からよく習得され、どのような速度でその習得状況が変化するかが明らか

〔表1 正解項目数の度数分布〕

項目数	学年				
	4	6	8	10	12
80.....	...	...	...	...	...
78-79.....	...	...	...	5	37
76-77.....	...	...	5	16	91
74-75.....	...	...	5	18	123
72-73.....	...	...	15	42	156
70-71.....	...	...	27	58	123
68-69.....	...	...	35	79	89
66-67.....	...	...	66	64	98
64-65.....	...	1	93	57	75
62-63.....	...	...	93	82	50
60-61.....	...	2	128	71	51
58-59.....	1	1	132	69	30
56-57.....	...	5	141	55	25
54-55.....	...	3	130	48	15
52-53.....	...	5	137	46	15
50-51.....	...	8	115	50	5
48-49.....	1	7	106	52	8
46-47.....	...	16	119	27	7
44-45.....	2	10	131	21	6
42-43.....	...	18	120	12	2
40-41.....	1	20	111	11	...
38-39.....	...	19	116	10	2
36-37.....	1	22	105	11	2
34-35.....	...	18	94	12	...
32-33.....	2	30	77	3	3
30-31.....	5	44	96	1	1
28-29.....	4	27	72	2	1
26-27.....	5	39	47	...	...
24-25.....	8	41	47	1	1
22-23.....	13	31	21	...	...
20-21.....	10	28	13	1	2
18-19.....	22	12	9	1	...
16-17.....	32	14	5	1	...
14-15.....	22	16	...	...	...
12-13.....	18	9	1	...	...
10-11.....	13	2	2	...	...
8-9.....	16	3	...	...	...
6-7.....	11	4	...	...	...
4-5.....	11	1	...	...	...
2-3.....	6	1	...	...	...
0-1.....	8	...	...	...	...
合計	212	457	2,416	926	1,018
中位数	16.0	30.0	48.4	60.8	70.4

になる。プレッシーのテストの場合、学習曲線は10タイプに分類されている。代表的なものとしては、(a)第6学年から第8学年の間で急速に上昇しているが、それ以外ではほとんど上昇していないもの、(b)第4学年では低い値だが、順調に伸びて第12学年では高い値を示しているもの、(c)早くから高い値を示し、学年進行にともなう上昇が小さいもの、(d)低い学年ではなかな

〔表2 正解率の学年進行による変遷〕



か上昇しないが、高学年になって急速に上昇するもの、などがある(表2)<sup>34)</sup>。

以上のような3通りの処理によって統計的に解釈された測定結果は、各学年の教育内容ならびに教科書執筆に利用すべき用語の決定のための基礎として利用される。このような処理によって、第4学年から第12学年での必須用語の学習の軌跡をたどるとともに、語彙の学習の一般的傾向をみることができるので、使用教科書と各学校・各学年での語彙の習得との関係や、教科書で既知と仮定されている用語や新聞などを読むのに必要とされる用語と、実際に子どもが知っている用語との関係、が得られるからである。プレッシーは、このデータをもとに、各学年ごとに90%以上の子どもが正解した用語の一覧表をつくり、教科書執筆者が準拠すべき使用用語基準作りを行っている<sup>35)</sup>。そして、実際、この測定結果を基準として、いくつかの教科書について、そこに登場する用語のうち、何%の用語が子どもに理解されているといえるかを論評している。

このように、従来の伝統的な試験におけるあいまいな査定基準に対し、学習成果測定においては、子どもの一定の学習集団の学習成果を査定基準として統計的に設定し、その集団内の相対的位置によって合理的に学習成果の査定をしている。そして、それがカリキュラム研究に合理的基礎を与えるデータ提供となっているのである。

## V おわりに

社会科における学習成果測定は、伝統的な試験のもつあいまいさへの批判を直接的な契機とするテスト形

態の改善という形をとっているが、そこで変革されているものは、単にテスト作成の技術だけではない。学習成果測定論は、社会科学教育研究における経験的・思弁的思考に代わる科学的決定論の導入の一側面である。20世紀初頭には、社会科学以外の教科においても学習成果測定は盛んであったが、特に、教科としての性格自体が未確定であった社会科学では、カリキュラム研究の合理的根拠の必要から、学習成果測定は重要な課題であった。

社会科学学習成果測定は、先述の四つのアプローチにより、従来の成績考査にはない科学性を追求している。すなわち、社会科学教育の目標の限定・細分化による測定対象の特定、社会科的教科の教科書の実態調査等によるテスト内容の妥当化、時や採点者の相違に関わりなく同一の結果が得られるテスト形態の採用、統計的处理による査定基準の設定にもとづく測定結果の合理的解釈である。

このような学習成果測定論には、特徴が2点ある。一つは、人間を諸要素の集合体とみなし、分析的にとらえようとしているということである。人間の知的活動を物理現象と同様に要素的に測定し、機械的に数量化してプロフィールをとらえられるものとして扱っているのであり、ドイツ流の全体論・観念論とは異なるアメリカ独自の発想がみられる。いま一つは、査定が教師の独断に委ねられたものから、子どもの存在を中心に据えたものになっているということである。一定の学習集団内の子ども相互の競争により、すなわち子ども達自身により、査定基準が設定されている。これは、資本主義社会が生成・発展しつつあったという当時の社会状況とも密接な関係があった。資本主義社会自体が競争を原理とする社会であり、競争社会の要請に応えるべく教育をより合理的・能率的にすすめるためには、伝習や独断に頼らない査定が求められたのである。したがって、測定では標準化が第1の目的となり、個々の子どもの実態は、その標準と個人の特性のずれでつかもうとしている。

以上のように、1910～30年代にアメリカの社会科学教育において盛んに行われた学習成果測定は、資本主義競争社会における合理性追求の精神にもとづき、統計的論理性を柱とするアメリカ流の科学的決定論を理論的根拠として成立・展開したのである。

なお、本稿では紙幅の都合上、「理解」の領域での測定を用いての説明が中心となった。「理解」以外の領域においても基本的原理は踏襲され、要素分析、数量的測定、統計的处理が導入されている。それらは、最近わが国で関心を集めているいわゆる情意的側面の評価に示唆する点が多いが、それらの検討は別の機会

に譲る。

## 註

- 1) 近代的な評価の成立期といわれる20世紀初頭の測定・評価の傾向は、いくつかの時期に分けてとらえられる。たとえば、スケータス(Scates)は、1897年から1906年までを抱卵期、1907年から1916年までを第2期、1917年から1926年までを第3期、1927年から1936年までを第4期としている。抱卵期は、ライス(J. M. Rice)の提議で始まり、ヒネー(A. Binet)の知能尺度の出版で最高頂となった。第2期は、ソーンダイク(E. L. Thorndike)の研究、最初の標準化測定尺度とアチーブメント・テストの出版などによる、テストにおける客観性追求の時期である。第3期は、教育測定の急速な拡大の時期である。第4期は、知能や学業成績の測定のためのみならず、人格の査定のためにも新しい改善されたテストや技能が求められた時期である。また、ライトストーン(J. W. Wrightstone)は、1900年から10年単位で時期を区切り、各々を次のように特徴づけている。最初の10年間は、近代的なテストのおこりの時期、第2の10年間は、アチーブメント・テストが教育者や大衆への浸透を図った時期、第3の10年間はアチーブメント・テストが受け入れられて急速に浸透した時期、第4の10年間は、測定対象の拡大とともに評価の概念が登場した時期、とされている。(J. W. Wrightstone, *Evaluation in Modern Education*, New York, American Book Co., 1956. pp. 4-7)
- 2) J.S. Orleans, *Measurement in Education*, New York, Thomas Nelson and Sons, 1937. p. 41.
- 3) Ibid. pp. 38-39.
- 4) 世紀転換期に端を発するアメリカ社会の変化は、海外への膨脹発展の開始、典型的資本主義社会の成立、革新政治の発展などがある。(アメリカ学会訳『原典アメリカ史 第5巻』岩波書店, 1957)
- 5) 経営管理への関心が呼び起こされた要因としては、次の諸点が挙げられている。(a)トラスト化の進行による経済の独占化傾向により、競争の焦点が技術革新による生産手段の合理化から、工場管理技術の発展へと移行した、(b)言語・習慣の異なる各国からの移民を主とする不熟練労働力を能率的に利用する必要があった、(c)労働組合による8時間労働制の要求の強化にともない、短縮された労働時間で従来の生産量を確保することが求められた、(d)機械化がすでに相当高度に進展しており、これを効率的に利用するための作業計画、統制、組織が求められた、

- (e)細分化された仕事を統合・調整し、混乱と無秩序の状態を回避する必要があった、(f)量的に増加したとともに、質的にも複雑化した管理業務への対応が必要であった、(g)賃金問題の合理的解決が求められた。(向井武文『科学的管理の基本問題』森山書店、1970、pp. 4～12)
- 6) F. W. Taylor, 上野陽一訳『科学的管理法』産業能率大学出版部、1969。
- 7) アメリカ学会訳、前掲書。  
関西アメリカ史研究会編著『アメリカの歴史、下』柳原書店、1982。  
青木怜子「フーヴァー政策再評価の限界と発展性」『アメリカ研究』12号、1978。
- 8) 志邨晃佑「アメリカ革新主義運動の解釈」広島史学研究会『史学研究五十周年記念論叢 世界編』福武書店、1980。  
平田美和子「1920年代のアメリカ都市政治 — 市政改革を中心に — 」『アメリカ研究』12号、1978。
- 9) 世界教育史研究会編『世界教育史大系17 アメリカ教育史 I』講談社、1975、pp. 232～236。
- 10) H.N. Rivlin, ed., *Encyclopedia of Modern Education*, The Philosophical Library of New York City, 1943. p. 714.
- 11) P.V. Sangren, “The Present Status of Measurement in the Social Sciences” *The Historical Outlook*, Vol. XXI. No. 2, February, 1930, p. 280.
- 12) W.A. McCall, *How to Measure in Education*, New York, The Macmillan Company, 1922. pp. 3-18.
- 13) J.C. Bell, D.F. McCollum, *Test in United States History*, 1915, Out of print.
- 14) R.M. Tryon, “Standard and New Type Tests in the Social Studies” *The Historical Outlook*, Vol. XVIII, No. 4, April, 1927, p. 172。  
客観テストの出版は、1920年代を中心に非常に多い。1930年代に入ると新たなテストの開発はやや下火となり、既成のテストの改訂版の使用や、一般的なアチーブメント・テストの一部として組み込まれた社会科テストの使用が中心となった。(W. S. Monroe, ed., *Encyclopedia of educational research, revised edition*, The Macmillan Company, New York, 1952, p. 1234.)
- 15) J.W. Wrightstone, “Recent Trends in Social-Studies Tests” *Social Education*, Vol. I, 1937. p. 246.
- 16) 社会的態度に関するテスト研究としては、たとえば、L.L. Thurstone, E.J. Chave, *The Measurement of Attitude*, Chicago, Univ. of Chicago Press, 1929, や R. Likert, “A Technique for the Measurement” *Archives of Psychology*, No. 140, New York, Columbia Univ. Press, 1932 などが挙げられる。また、技能に関するテスト研究としては、たとえば、図表からの情報抽出や参考情報源の選択についてのテストを作成した J. L. Rogers らの Iowa Every-Pupil Test of Basic Study Skills が代表として挙げられる。
- 17) National Society for the Study of Education, *26th Yearbook, Part I, The Foundations and Technique of Curriculum Construction*, 1927. pp. 63-64.
- 18) T.L. Kelley, A.C. Krey, *Report of the Commission on the Social Studies, Part IV, Tests and Measurements in the Social Sciences*, 1934, p. 121.
- 19) Rivlin, *ibid.* p. 279.
- 20) Kelley, Krey, *ibid.* p. 125.
- 21) *Ibid.* p. 153.
- 22) *Ibid.* p. 234.
- 23) *Ibid.* p. 340.
- 24) 学習成果測定における内容の妥当化のためには、プレッシャーが用いた教科書分析法以外に、次のような方策も用いられていた。(a)地方レベルあるいは州レベルでのコース・オブ・スタディーの分析、(b)むつかしいとされる内容の項目、話題の配列、ドリルの量と配分についての調査、(c)内容の社会的価値と利用の調査、(d)教師自作試験の問いや、よく知られた機関によって作成された試験の問い、の分析。(Orleans, *ibid.* p. 44.)
- 25) 社会科委員会では、社会科学で用いられる重要な用語を質的に分類し、固有名詞で表わされるもの(人物、場所、でき事、具体的な物体など)と、一般名詞で表わされるもの(関係を示すもの)とに整理している。さらに後者は、単に時と場所の関係を示すだけのものから、物質界の関係、人間の関係、短文では表わせないような複雑な社会関係までさまざまなものが含まれると分析している。テスト項目選定のための用語抽出の対象とされているのは、関係を示す一般名詞、特に人間の関係を示すものである。これによって、先に示した測定対象の特定が一層すすめられているのである。(Kelley, Krey, *ibid.* pp. 8-16, 476-477.)
- 26) この種の語彙調査としては、A.S. Barr, “The Vocabulary of American History” *The Journal of Educational Research*, XX, 1929, や A.W.

- Stephenson, "The Special Vocabulary of Civics"  
*The Journal of Educational Research*, XVIII,  
1928.
- 27) Kelley, Krey, *ibid.* pp. 155–157.  
28) *Ibid.* pp. 186–188.  
29) *Ibid.* pp. 106–113, 121.
- 30) *Ibid.* p. 157.  
31) *Ibid.* pp. 189–204.  
32) *Ibid.* pp. 164–166, 205–207.  
33) *Ibid.* pp. 166–169, 208–218.  
34) *Ibid.* pp. 169–174, 218.  
35) *Ibid.* pp. 175–179.