

(城西人文研究第20巻第1号)

【研究ノート】

教育評価について

細部国明

1. 教育評価 (educational evaluation) の位置づけ

生徒に目標を与えて励ましたり、自分の接し方を反省して改めて計画を立て直したりする教育活動の中には、本人が意識するしないに拘らず、①教育目的(目標)、②教育内容、③教育方法、④教育評価という4領域が含まれている。教育評価とは、教育活動の成果を調べ、それに検討や反省を加える最後の段階である。しかし、その反省が次の教育目標や方法などにフィードバックされ、改善をもたらして、教育活動は更に続いていく。上記の①②③④の順は大まかな時間的流れに沿ったものであるが、実際の教育活動を観察すると、単元の途中であっても、反省して少しぐらいの修正が行なわれる場合もある。その流れは部分的には逆になることもあり、それらの関係は図1のように相互関連的なものであることが理解できるであろう。

2. 測定 (measurement) から評価 (evaluation) へ

人間の捉え方や選考方法が主観的過ぎるなら、もう少し客観的に測定

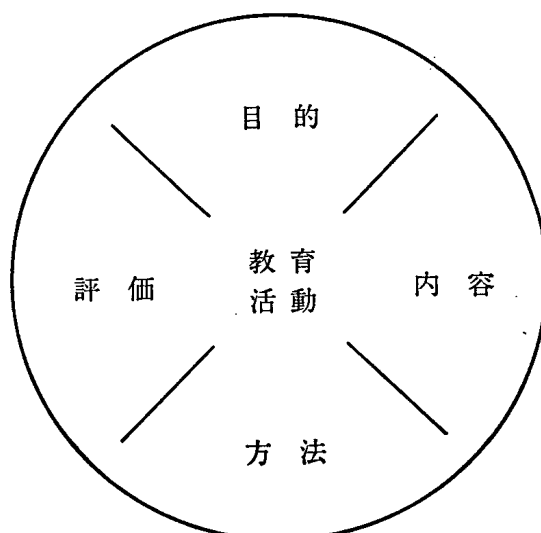


図1 教育活動の相互関連性

しようとする運動が起ってくるのは最もなことである。その芽生えをビネー・テスト開発（1905年）に見ることができる。ビネーの精神年齢の測定法を応用して、生徒の客観的把握に役立つ各種の標準学力検査の作成に力を注いだ Thorndike¹⁾は、「教育測定運動（educational measurement movement）の父」とまでいわれた。しかし、その後、生徒に関する情報である測定資料の量産は、教師側の指導の質的向上に必ずしも直結しないのではないかと、この反省が大きな形で現われたのが、アメリカの「8年研究（1933～1940）」²⁾である。測定に欠如していたもの、一言でいえば「指導」的観点、をも含んだ新しい視点は「教育評価」といわれた。

表 1 測定と評価

教 育 測 定	教 育 評 価
1. 〔横断的関係の明確化〕 個人の集団内位置が明らかになる	〔縦断的関係の重視〕 他との比較ではなく、個人の未来へ向けての指導と結びつけようとする
2. 〔量的産出〕 客観的な数量的資料を多く提出する	〔質的産出〕 ある資料が提出されることに如何なる価値（value）があるかを問題にする
3. 〔部分的把握〕 測定されるものが細分化していき、個人の際立った特性のみに注目しやすい	〔統合的把握〕 人間の全領域、即ち知情意・興味・態度適応面などを関連的に全人的に把握しようとする
4. 〔客観化できる表面層についての話題〕 測定し数量化できる資料だけが話題になりやすい	〔主観的な無意識層へ開かれた態度〕 数量化できにくい、各人の主観的な特性に対しても開かれた態度を持つ

表 1 には両者の違いのみを強調したきらいがあるが、評価的立場から測定への批判は理解できると思う。勿論、両者には共通点もあり教育実践の中で各々

- 1) 米の心理学者（1874～1949）で“try & error learning”の研究でも知られている。
- 2) USA の進歩主義連盟が30の高校と 300 の大学の協力を得て、教育課程の妥当性を追跡研究したものである。その成果は1942年に発刊され、教育評価の研究に大きな影響を与えた。

表 2 学校教育に関する伝統的見解と進歩的見解

項目	学 校 教 育 に 関 す る	
	伝 統 的 見 解	進 歩 的 見 解
教育測定の目的	測 定 評 価 配 置 (ク ラ ス 編 成)	診 断 ガ イ ダ ン ス
志 向 性	知 的 伝 統 的 保 守 的	人 間 的 実 験 的 進 歩 的
関 心	教 科 社 会 的 諸 要 求 思 考 達 成 学 習 成 果	児 童 ・ 生 徒 個 人 的 諸 要 求 感 覚 興 味 学 習 過 程
強 調 点	知 る 事 物 理 解 す る 事 物 何 事 か 成 し 遂 げ る 事 物 習 得 収 束 的 思 考 (convergent T.)	見 る 事 物 行 動 す る 事 物 何 も の か に 成 る 事 物 創 造 拡 散 的 思 考 (divergent T.)
目 標	教 師 が 設 定 す べ き も の 能 力 卓 越 性	児 童 ・ 生 徒 が 決 め る べ き も の 適 人 個 性
教 師 の 役 割	教 え る 事 物 方 向 づ け る 事 物 学 識 経 験 者 で あ る 事 物	意 欲 づ け る 事 物 促 進 す る 事 物 仲 間 で あ る 事 物
児 童 ・ 生 徒 の 役 割	従 う 事 物 学 習 す る 事 物 勉 強 す る 事 物	先 頭 に 立 つ 事 物 生 活 す る 事 物 参 加 す る 事 物
唱 導 者	ア リ ス ト テ レ ス コ メ ニ ウ ス, J. A. ヘ ル バ ル ト, J. F. バ ー ク リ ー, W. C.	ル ソ ー, J. J. ペ ス タ ロ ッ チ, H. フ レ ー ベ ル, F. デ ュ ー イ, J.

1) 引用文献 西山啓他監修 1978「目で見る教育心理学」ナカニシヤ (Ebel, R. L. 1972)

重要な役割を果たしている。しかし、教育評価に対する考え方は、その国の学校教育についての見解でもあり、文化環境によってその重心は異なる。エーベル¹⁾は、学校教育について表2のように比較している。

3. 評価の対象と評価の方法

(1) 評価の対象

測定の対象は主に生徒であったが、評価の対象は教育活動に関するすべてのものに及ぶ。一応つぎの5領域にまとめておこう。

- ① 学習効果の評価（知識科目，技能科目，教科外活動など）
- ② 個性の評価（身体・健康面，態度，興味，性格，職業適性など）
- ③ 教師の評価（適性，人格，学力，指導法など）
- ④ 学校の評価（校舎，校地，教具，図書施設，教育課程，教員管理など）
- ⑤ 家庭や地域の評価（家庭環境，地域の文化，人間関係，思考傾向など）

(2) 評価の方法

何によって評価するかという観点からみると、①観察による評価，②話し合いによる評価，③テストによる評価の3つに大別できる。

4. 評価の機能（または目的）

評価にはどのような働きがあるか、また評価はどのような目的に寄与するかという観点からまとめたものが、教育評価の機能（または目的）と言われるものである。教育評価の働きを幾つかに分けると、次のような働きがあることに気付くであろう。

(1) 指導機能

どのような状態にある生徒には、どんな指導目標が適しているか、またその

目標にどのような方法で導いていったらよいか、また指導結果をどう整理したら次の指導に役立つかなど、主に指導面に役立つ働きを指摘したものである。

(2) 学習機能

指導機能は主に教師側に重点があるのに対し、学習機能は生徒の学び取る力や自己学習の働きを強調したものである。生徒からみた場合、教師や仲間から適切な情報をフィードバックされることにより、目標意識が芽生え、やる気が出てきたりする。また、生徒が自分で実験したり、自分の学習活動を振り返ってみることが新たな学習方法の気づきや学習の促進につながることが多い。

(3) 管理機能

より良い教育活動に向けて教員組織や設備などに関する資料を集めて、教育を運営するための状況を改善したりする働きである。指導体制・通知票・学級間交流・入学者の選抜などの運営・管理などがこの働きを含んでいるが、これだけが強調されると評価は片手落ちになる。

(4) 調査・研究機能

ここでいう調査・研究機能とは日常の個々の教育活動の実践を直接対象にしたものではなく長期的・巨視的な観点からの働きを示している。例えば、教育制度のあり方を考えるために国際的比較の資料を作ったり、学力の全国的調査をしたり、教育評価の変遷の研究班を作ったりすること等がある。このように現実を一步離れたところからながて見ることは、創造的な教育活動を産み出すことにつながることもある。この機能が大きな使命として与えられているのが研究機関であるが、これだけが強調されると教育活動の実践的使命が薄らぐことになる。

5. 診断的評価・形成的評価・総括的評価

評価を時間的流れに沿って、初期的評価、中期的評価、後期的評価はどうあったらよいかを研究していくと、ねらいや方法に違いのある診断的評価 (diagnostic evaluation)、形成的評価 (formative E.)、総括的評価 (summative E.) の区別が可能になってくる。

(1) 診断的評価は、指導に入る前や指導初期に、既習の到達度や生徒の性格などを調べて、これからの学習に適したクラス編成を考えたり、授業目標を作るための資料を分析したりすることである。

(2) 次に学習・指導活動に入ったら、生徒の理解状況を常に確認しながら授業を進めることが重要となる。この段階で学力の形成に必要な評価が形成評価であり、それは大きく2つに分けられる。1つはテストなどを使用しない観察・口問口答・挙手などの評価活動である。生徒の表情や動作、クラスの雰囲気などを確認しながら、それらを指導活動にフィードバックさせて、時によっては授業の途中でも、学習内容をより望ましい方向に微調整する必要があるかも知れない。2つめは形成的テストの活用である。例えば、1時限ごとは無理だとしたら、3時限ごとに形成的テストを実施する。これは最後に行う総括的なものである必要はないので、学習を進めていく上にどうしても理解しておかねばならない基本的事項について5～10分程度のテストが一般的であろう。小さな単元が幾つか連続しているようだったら、単元の終りごとに重要事項確認のために、形成的評価の意味あいを強く持たせた簡単なテストを、教師がクラスの進度に合わせて自作することも有効であろう。評価が学習を促進し、学力を形成するというのが形成的評価のねらいである。

(3) 数単元後とか学期末とか教授目標が一通り終わった後で、どの程度到達されたかを調べて、次回の学習活動にフィードバックさせようとするのが総括的評価である。期末テストもこの類である。

これらの概略を整理したのが表3である。従来の日本は形成的評価より総括的評価に重点があり、本人の興味や意欲の評価より学力の評価に重点があった

のではないだろうか。

表 3 指導・学習活動の関連から見た教育評価のタイプ¹⁾

種類 基準	診断的評価	形成的評価	総括的評価
時期	・単元始め, 学年始め ・授業進行中	・学習活動進行中	単元末・学期末
目的	・学習の水準や前提条件などの診断や確認 ・適切なクラス編成や指導法を決定する	・教師と生徒間のフィードバックを盛んにし学習を促進する ・修得程度及び問題点を発見する	・目標達成度を把握し成績決定のための基礎資料を整理する ・指導計画の検討や改善を図る
方法・用具	・観察や面接 ・学力診断テスト ・各種心理診断検査など	・観察や口問口答 ・単元ごとに教師作成の形成的小テスト(5~10分)など ・教育機器の使用	・観察や面接 ・期末テスト(50分) ・標準学力テスト
用いられる 主な評価法	絶対評価 相対評価	絶対評価中心	相対評価>絶対評価

6. 絶対評価・相対評価・個人内評価

何を基準にして評価するかにより、3種類の評価法が区別される(表4参照)。

(1) 相対評価 (relative evaluation)

他人を基準にするもので、集団の中で個人がどの辺に位置するかを見る。ク

- 1) 参考文献：橋本重治，1976「新・教育評価法総説」(上・下) 金子書房
 Bloom, B. S. et al., 1971 (梶田叡一ら訳 教育評価法ハンドブック, 第一法規)
 西山啓他監修, 1978「目で見える教育心理学」ナカニシヤ出版
 岸田 元美ら, 1985「教育心理学」ナカニシヤ出版

ラスの得点分布，標準学力テストの規準（norm），正規分布が背景になっている偏差値や5段階評価など，客観的数量的記述に依存する評価法である。評価者の主観が混入する余地が少ないが，個人差や質的側面が軽視されやすくなる。相対評価の与える影響は，いつも比較されて育つ同胞のことを思い浮かべればその一端がわかるであろう。

表 4 評価基準とその影響¹⁾

評価の 名称		相 対 評 価	絶 対 評 価		個 人 的 評 価
			到 達 度 評 価	認 定 評 価	
評価の規準		規準集団の他者	外的客観的な規準や到達目標	評価者の内的直観的規準や目標	当人の属性（過去・他科目など）
受止め方の傾向	教師	優劣度の評価	外的基準の満足度の評価	内的基準の満足度の評価	進歩程度の評価
	学習者自身	優劣度の自己評価〔優越—劣等感〕	目標到達度の自己評価〔達成—未達成感〕	内的基準満足度の自己評価〔満足—不満足感〕	進歩度や努力度の自己評価〔進歩—停滞感，努力—怠惰感〕

(2) 絶対評価 (absolute evaluation)

他人と比較する相対評価のもつ弊害が大きくなってきた反省の中で，絶対評価が重視されてきた。この中を一先ず2つに分けておこう。1つは科目などの到達目標を基準にして，その到達状況を評価するもので，到達度評価（criterion referenced E.）とも呼ばれる。到達度評価を効果的にするには目標の難易度・段階性，下位目標間の関連性など，目標分析が重要になる。達成し易い学習の順序なども生徒皆同じとは限らない。各人の目標をどこに置くかを考えることが重要である。皆んなの目標が同じであれば相対評価に類似してく

1) 参考文献：梶田叡一他編 1982「評価と指導の研究課題」(シリーズ 評価と指導の一体化 I) 教育出版
村瀬 隆二編 1989「教育実践のための教育心理学」新曜社

る。到達度評価は個人の進歩・努力状況を反映させやすい。一般に絶対評価といえは到達度評価を意味する。

2つめは認定評価といわれるものである。到達度評価は科目など、いわば外的な客観的目標を基準にしているのに対し、認定評価は科目のみに限らず評価者の内的な主観的目標を基準にする、といえる。評価者によっては学習者が今までより向上した場合や努力点に大きな評価を与えるかも知れない。評価者の経験や専門的知識や学習者の意欲などを表現できる長所はあるが、評価者の主観から信頼性がおちてくる危険がある。教師作成テストの結果が予想に反して、クラス全体が低すぎたり、ばらつきが大きすぎたりした時は、客観的知識などの到達度評価だけに束縛されず、評価者の認定的事項を取り入れて、最終的評価を決めた方が良い場合がある。

(3) 個人内評価 (intra-individual evaluation)

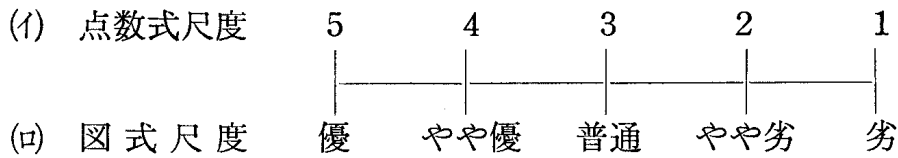
その生徒がもっている他の属性を基準にして評価する。他の属性としては、過去の成績、現時点の他教科、パーソナリティ特性などがある。それ故、相対評価で4の項目が、個人内評価では3になることもあり得る。全人的観点から評価するもので指導者の教育的価値判断が重要になる。学習者の個性を重視した建設的な評価が求められる。小学校の通知票で項目ごとに◎○△をつけるのも、この精神を活かしている。

7. 評定法 (rating method)

資料整理やテスト結果の表示の仕方などテスト全般の関連事項に若干ふれておく。対象や資料をある見地から分類したり段階づける方法の総称を評定法という。評定法には評定尺度法、順位法 (または品等法)、分類法などがある。

① 評定尺度法：最も一般的に使用されるもので、あらかじめ数段階の評定尺度を作り、対象をどこかの段階に当てはめていく。評定尺度には点数式尺度、図式尺度、記述尺度などがある。

<五段階評定の一例>



② 順位法（または品等法 method of rank order）：40名の成績を上位または下位から1，2…40と順位づけたりすることで，ここで取扱う数字は一定の方向をもった順位の意味しかない。これは Cattell が科学者の貢献度を品位づけし等級をつけるのに適用したため，品等法とも呼ばれる。

③ 分類法（classification method）：対象を定められたカテゴリー，例えば動物，植物，鉱物等に分類する。YG 性格検査や MMPI など多くの項目を，客観的・協調的などの幾つかの目録（項目）に分けて示す目録法（inventory）も，分類法の類である。

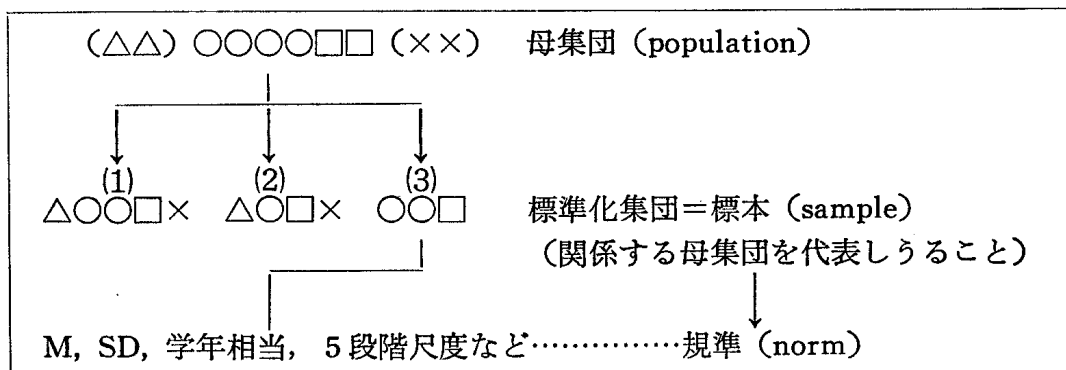
8. 標準化されたテスト

信頼に足るよきテストは，それなりの条件を備えていることが必要である。そのようなテストを標準化テスト（standardized test）という。

(1) 標準化過程がしっかりしていること

例えば，中学2年の英語の標準学力テストを作るとする（表5参照）。そのテストを受けると予想される中学2年生全部が母集団（population）となる。

表5 標準化過程



このテストを受けない会社員は母集団に入らない。次に母集団の中から、適正な標本 (sample) を抽出し、母集団を代表しうる標準化集団を作る。この標準化集団から出された平均や学年相当、標準偏差、百分率などが規準 (norm) とされる。各人の得点は、この規準をもとに解釈される。よいテストは標準化過程がしっかりしていることが必要だが、その他に信頼性や妥当性などを備えていることが必要である。

(2) 信頼性 (reliability) が高いこと

検査する場所や時間、また採点者が異なっても、検査結果に一貫性があることをいう。相関係数 (1 ~ -1) で表す。日時計よりは電子時計の信頼性が高い。信頼性の中味は大きく2つに分けて考えることができる。1つは測定用具側の信頼性の問題である。測定用具の変化性、テスト問題の意味の不明確さ、採点基準のあいまいさ等がある。2つめは主に評定者側の主観や個人差によって生ずる評定者間信頼性 (interrater reliability) の問題がある。

〔事例〕 スターチとエリオット (Starch, D. & Elliot, E. D) が、ある高校生の英語の答案を142人の教師に採点させたところ、最低が50点、最高が98点という極めて広い範囲の採点の分布を得ている。また幾何の答案では、115人の教師がなんと最低28点から最高92点までの採点をしたという¹⁾。

しかし、問題によっては、評定者の主観を生じさせやすいものもある。主観による不一致を除けるテストは、特に客観性 (objectivity) があるともいう。客観性条件は信頼性の一部と考えられる。

(3) 妥当性 (validity)²⁾ が高いこと

測定しようとするものを、どの程度測定しているかという合目的性のことを

1) 引用文献：氏原寛他編，昭和61年「臨床教育心理学」創元社，p. 203

2) アメリカ心理学会の「教育・心理テストとその手引の規範 Standards for Educational and Psychological Tests and Manuals」1974年版では、妥当性を、内容妥当性、基準関連妥当性、構成概念妥当性の3つに分類している。

テストの妥当性という。妥当性は次の3つの観点から調べられる。

① 内容妥当性：例えば「個々のテスト項目は教科内容に即しているか、範囲外のものや、逆に欠落した内容はないか」など、調べようとしている内容をテストがうまく引き出しているかどうかである。

② 基準関連妥当性：これは常に何か外的に関連する基準（変数）に対する妥当性を問題にするもので、幾つかに分類できる。例えば、外的関連基準が、時間的に未来に属する「将来良き教師になること」だとして、その教員選抜テスト結果が、それをある範囲で予測できたとすれば、この場合の基準関連妥当性は予測妥当性（predictive V.）という。また、選抜テストが「現在、子どもに愛情がある」という関連基準に対しても有効である場合は併存的妥当性（concurrent V.）を持つという。

③ 構成概念妥当性（construct V.）：これは抽象（abstraction）的な構成概念に対する妥当性を問題にしている。例えば、知能、不安、社会性、創造性、達成動機などは特殊な形や実体を備えていない、心理学的に構築されてきた構成概念（construction）である。そうした構成概念をテストが実際にどの程度測定し得ているかを構成概念妥当性とよぶ。この構成概念の妥当性を調べるためには、㉑理論的な方面からの妥当性と㉒実証（実験）的方面からの妥当性（統計的妥当性、因子的妥当性など）の両方からの接近が必要である。

例えば、ウェクスラーによると、知能検査の結果を因子分析で解明すると、知能に属すると考えられるものは全体の60%を越えることはまれで、いつも解明できない残在要素がある（因子的妥当性）。

9. 性格、適性の測定

テストで測定する対象は、一般に知能・学力・性格・適性の4分野に分けて考えることが多い。ここでは性格と適性に関するテストについて多少ふれておく。性格テストは次の3つに分類される。(1)質問紙法（questionnaire）……虚偽尺度（lie scale）をもつミネソタ多面人格目録（MMPI）、矢田部ギルフォ

ード性格検査 (Y-G) など。(2)作業検査法……Kreapelin test など。(3)投影法。適性検査 (aptitude test) の種類には, 一般的職業適性検査, 各種の職業適性検査などがあり, 本人の興味や特定の職業集団の特性などを調べる。ストロング職業興味検査が有名である。

10. 学力の評価

(1) 学力とは何か

各教科などいろいろな分野によって, また, 教育観によってその学力は同じとは言えないが, 広い意味で学力を図2のように四層構造でとらえておこう。学力の評価も観察・話し合い・テスト等を通して行なうことができるが, ここではテストによる場合について述べる。

(2) 標準学力テスト (standardized achievement test)

一般に標準学力テストでは次のような規準が算出されている。

① 教育年齢 (EA: educational age): 標準学力検査の得点規準に照して, ある生徒の学力の程度が何歳何か月相当か, 何学年相当かが算出できる (知能検査の MA に相当する)。

② 教育指数 (EQ: educational quotient): 次式で算出する (知能検査の IQ に相對する)。

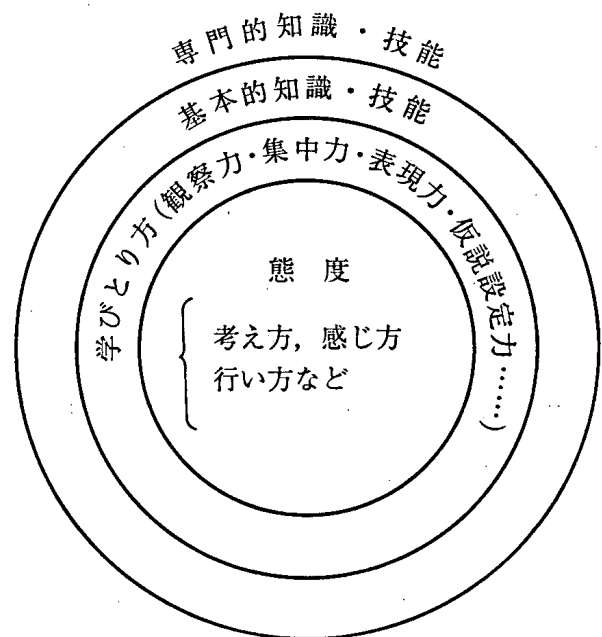


図2 学力の構造¹⁾

1) 参考文献: 広岡亮蔵 1971 特集 学力の伸ばし方「児童心理」金子書房 25
(広岡の学力の三層構造をもとにして, ここでは専門的知識・技能の層を加えて四層構造にした)

$$EQ = \frac{\text{教育年齢 (EA)}}{\text{生活年齢 (CA)}} \times 100$$

以下の③④は共に、ある生徒の学力が知能と比べてどのような状況にあるかを見る。

③ 成就指数 (AQ accomplishment quotient) : 学力と知能の関係を比で表す。

$$\begin{aligned} AQ &= \frac{\text{教育指数 (EQ)}}{\text{知能指数 (IQ)}} \times 100 = \left(\frac{EA/CA}{MA/CA} \times 100 \right) \\ &= \frac{\text{教育年齢 (EA)}}{\text{精神年齢 (MA)}} \times 100 \text{ に変形可能。} \end{aligned}$$

指数の代わりに偏差値を用いると

$$AQ = \frac{\text{学力偏差値 (ASS)}}{\text{知能偏差値 (ISS)}} \times 100$$

知能に比して学業成績がすぐれている場合を、オーバー・アチーバー (over achiever) といい、逆の場合をアンダー・アチーバー (under achiever ; 学業不振) という。

④ 成就値 (AS : achievement score) : 学力と知能の関係を差で表す¹⁾。

成就値 (AS) = 学力偏差値 - 知能偏差値

中学校などにおける学力の5段階評価の1・2・3・4・5に含まれる人数は、正規分布に関連があり、各々、7・24・38・24・7%という比率が伝統的に使用されることが多い。

1) 明確な根拠に基づくものではないが、AQ の場合は、100より±10以上の差をもって、即ち110以上をオーバー・アチーバー、90以下をアンダー・アチーバーという。AS の場合は±7以上の差。このうち精神遅滞を学業遅進 (slow learner) と呼んで、アンダー・アチーバーと区別することがある。また、アンダー・アチーバーを相対的学業不振と呼び、知能検査の得点とは関係なく学業成績だけに基づいて、学年相応の学力がかなり遅れている場合を、絶対的学業不振と呼ぶことがある。

一般に知能の高い者の成就値が負になり、又、創造性 (creativity) が高ければ、かなりの学力水準を維持するという報告もある。

(3) テストの形式¹⁾

通常の学力テストをその形態に従って分類すると、(A)口頭試問、(B)筆記試験に大別でき、後者は更に幾つかに分けられる。

(A) 口頭試問は、わが国では対話が苦手な国民性のためか、教育の現場で余り活用されていなかったが、教育における対話 [dialog(ue)] の役割を重視する欧米諸国では最終試験に口頭試問を含ませることが多い。口頭試問の特徴は学力ばかりではなく、人間性をも同時に見れることである。教師は必要に応じて口頭試問を行なえるだけの工夫と見識をもつことが望ましい。

(B) 筆記試験の種類には、専門家により厳密な手続きを経て作られた「標準化テスト」や、教師自らが適宜作成する「教師作成テスト」がある。筆記試験の形式を一先ず3つに分けてみよう。

① 論文体テスト (essay-type-test) : これは「…について説明せよ」「比較せよ」「論述せよ」など、ある主題を与えて、自由記述させるもので、伝統的な試験方法である。被検者の高度な構想力・思考力、個性を試すという高邁な目標とは裏腹に、問われている内容が余りにも不明確だったりすると現実にはうまく機能しなくなる。

② 客観テスト (objective test) : ややもすると論文体テストの欠点となりやすい信頼性の低さや採点の主観性を是正しようとして生まれたもので、教師が自由に作成するようになったのは1920年ごろからといわれる²⁾。客観テストは心理学的概念に従って再生法と再認法の二つに分けられる。

① 再生法 (recall method) : ㉑短答式テスト、1～2行の比較的短かい解答を求める記述問題などである。論文体テストより信頼性や採点の客観性が高まりやすい。短い和訳・英訳・数学問題などもこの範疇に入る。㉒完成法、いわゆる「穴埋め」で、選択肢が示されていない場合である。短答式テストの一種といえよう。

㉓ 再認法 (recognition M) : これはマークシート方式などの主要な出題形

1) 氏原 寛ほか「臨床教育心理学」創元社 S. 61.

2) 高嶋正士ほか「現代教育心理学」芸林書房 S. 50.

順序法（整序法，配列法）は，ある観点から項目を並べ換えるもので，余分な項目を加えると難易度は上るが，別の観点からの配列が可能になると採点が複雑になる。

真偽法（いわゆる○×法，または二者択一法）は，1つの設問に2つの選択肢しかない多肢選択法に相当し，理論的には50%の得点が，まぐれ当りによって失われる。最も好ましくないテスト形式であるとされる所以もそこにある。

再認によるいずれのテストでも，選択肢の数を増やし，まぐれ当りの得点圏を最小にすることも重要だが，問題そのものを充分検討することである。

③ 問題場面テスト（problem situation test）：数学の応用問題等のように新しい問題場面を示して解答させる形式である。記憶，思考力，発想力，構想力に限らず，図の活用などを含めた表現力，各自の個性を感じる創意工夫が見られることもある。総合的な問題解決力の相違により，幾通りもの解答の仕方が可能になる。

(4) 論文体テストと客観テストの比較

それぞれのテスト方法の特徴をわきまえ，効果的に使用することであるが，両テストの一般的な相異は表7の如くである。

表7 論文体テストと客観テストの長短の比較

	論文体テスト	客観式テスト
被験者側	問題の大きさや数	比較的小さい範囲からの多くの問題に答える
	とらえ得る学力の種類や複雑さ	体系的知識，比較力・総合力・価値観など
	連続使用による生徒の学習方向	断片的知識・理解など
採点者側	生徒の反応の型	断片的事実の習得に注意を向ける
	採点の主観性・客観性	複雑
	採点基準	主観的
ほか	ハロー効果（字の上手下手・採点者の先入観など）	比較的一定しない
	作問	影響あり
	容易	困難
	困難	容易

11. 評価者のおかし易い過誤

評価は一定の方向に歪み易いことがあるので、次のような歪みをおかさないよう注意する必要がある。

(a) ハロー効果（光背効果，後光効果：halo effect）：1，2の特に顕著な良い（あるいは悪い）特徴があると，その人物の他の面についても不当に高く（あるいは低く）評価しがちな傾向をいう。

(b) 中心化傾向（central tendency）：評定尺度の中心附近に評価が集まり易い傾向をいう。評定に自信が持てなかったり，極端な判断を避けようとする気持などから起る。

(c) 寛大効果（leniency effect）：評価者の抱く感情などによって左右されるもので，例えば性格判断などについて言えば，評価者が好ましいと思っている特性については評価は甘くなり，悪いと思っている特性については評価が辛くなる傾向をいう。これと関連するものに，自分のクラスの生徒は知らないクラスの生徒より過大評価され易いことがある。

(d) 論理的誤謬：「この性質はあの性質と結びついていた」という個人的経験による理論的な先入観を持っていることから生じる誤りをいう。

(e) 黙従傾向（acquiescence）：諾否法（yes or no method）形式では，一般に「はい」と答えやすい傾向が指摘されているが，そのような傾向をいう。特に，回答者が質問内容に確信が持てない場合などに生じやすい。

(f) 社会的望ましさ（social desirability）：質問紙法の性格検査への反応に影響する1つとして，文化や所属集団のもつ社会的望ましさの次元がある。

12. 今後の課題——自己評価能力の育成——

評価は誰がするかにより，他者評価，自己評価（self-evaluation），相互評価に分けられる。日本が西欧に追いつけ追いこせの時代は，評価の基準は西

欧、即ち、他者であった。この構図は学校という小宇宙でも同じであった。即ち、評価を受けるのは学習者であり、評価するのは教師であるという他者評価を強く意識させる構図があった。日本が良くも悪くも国際的に注目されるようになった現在、日本らしい生き方を日本自身で模索しなければならないような機運が生じてきた。時を同じくして、学校の中でも総じて各自が個性的に生きる自己評価能力の育成が重要度を増してきた。自己評価能力の育成には、それ相応の社会的・文化的背景が必要だったのである。価値観が自己評価を待つまでもなく社会的に固定されているところでは、自己評価能力は育ちにくい。現代の価値観の動揺が、他者評価社会から自己評価社会への過渡期ともなれば幸いである。今日では生涯学習の見地から「自己教育力」の育成が必要である。自己評価力は自己教育力の下位概念である。生物には自己治癒力が存在する如く、自己評価力も生来備っている。

他者評価は学習活動が受け身になりやすく、学習に対する責任が育ちにくい。それに対して、自己評価には学習者自身が学習を調整する人物であり、学習に対する責任者である、という自覚の発生を期待できる。また、一般に自己評価は他者評価より不安や緊張が少ないので、現代社会の中で起きている評価に伴う弊害を改善するために、自己評価の中に働らく原理や方法を活用できることが望ましい。自己評価能力の育成に関する今後の課題をあげる。

- ① 自己評価は全人格的領域にかかわり、長期的なフィードバックを必要とする。自己評価が必要となってくる領域は、知的・技術的側面に限らず、態度・行動・性格・習慣といった情意的側面にまで及ぼす必要がある。
- ② 評価活動の中の「指導へのフィードバック機能」、特に学習者自身へのフィードバックに関する研究が必要である。
- ③ 相互評価の訓練：自己評価能力が育たなければ、公正な相互評価は実施しにくい。諸種の活動を通じて相互に評価し合える機会を作って指導していくことが必要である。