

---

## 新しい評価を求めて テスト教育の終焉

キャロライン・V・ギップス著  
鈴木秀幸訳  
論創社 2001年

---

(第1章 評価のパラダイムの転換)

(第2章 評価と学習の関係)

## 第3章 テストのもたらす弊害

フレダーリクセン 1984年

「学習内容に与えるテストの影響力が非常に大きい」

「択一式の最低習得度テストの義務的使用が、このような方法では容易に測定できない能力の指導を妨げたのではないか」

### ●測定主導の学習指導

ポファム 1987年

「測定主導の学習指導が公教育の質的改善のために最も効率的な方法である」

※測定主導の学習指導：テストの実施を梃子として学力の水準を高めようという考え方

「ハイ・ステイクスなテストはカリキュラムを誘導する強力な力を持っている」

※ハイ・ステイクス：その結果が社会的な関心を呼ぶテストや評価（例：大学入学共通テスト、大学入学試験）

「測定主導の学習指導が基礎的技能での学生の優劣の差を明らかに縮めた」

→測定主導の学習指導は公教育の質を改善する効率的な方法になる可能性がある

→ポファムが提示する測定主導の学習指導が成功する条件(pp.45-46)

アイラジアン 1988年

「ハイ・ステイクスの概念は絶対的ではなく、相対的であり、テストが実施される状況によって異なってくる」

→社会的重要性和標準レベルを分けて問題にすべき（例：高校卒業テスト）

### ●テストがカリキュラムや学習指導に与える影響

マダウスの調査から導かれた7つの原理(pp.47-51)

・原理1…影響するか否かは認知の問題→テストがハイ・ステイクスだと信じてしまうから

- ・原理 3・・・テストの結果により重要な決定がなされる場合、教師はテストに向けた学習指導をする
- ・原理 7・・・ハイ・ステイクスなテストにより、カリキュラムの統制機関が作問・監督部門に移る→公的・統一的テストの実施が学校の外部組織になる。教師による都合の良い結果の活用を招く。 等々

→結論「テストの持つ社会的重要性を低減するように働きかける必要がある」

「テストの結果は一つの情報にすぎず・・・他の指標と並行して用いるべきことを説得すべき」

### クルックスの調査 1988 年

教室での生徒に関する評価に焦点を当てた調査

対象とした評価活動：正式な教師作成テスト、カリキュラムに付随するテスト、口頭試問、その他の様々な学習活動の評価、動機や態度の変化についての評価、学習技能に関する評価など)

→評価のあり方が生徒の取り組む学習のあり方に影響する→学校教育の早い段階から深い学習を促す強力な根拠がある

→単に生徒を等級付けるのではなく、生徒の学習を助ける評価の役割をもっと強調すべき

\*クルックスは生徒の達成状況を標準分布させて等級付けするような評価の役割に批判的。等級付けによって行われる生徒間の社会的な比較や競争がほとんどの生徒に望ましくない結果をもたらすことを調査から明示（例：内発的な動機の低下、不安感の高まり、生得的な能力を強調することで生徒の努力を損なうこと等）

\*最も効果的なフィードバックの形は、求められた課題をどこまでできたかの進歩に生徒の注意を向けさせるもの。生徒の自己効力感を高め、努力を重視させ、仲間との比較にとらわれないようにしていくことが重要

\*異なる能力の生徒が集まっているクラスで、全員に達成基準を実施する場合の問題を解決するには、個別指導を取り入れたり、グループ内での協同学習を取り入れて、生徒の得意な部分と不得意な部分の相殺を図ること

### 最も重要な教訓としてのクルックスの結論

→教育者が最も大切だとみなしている技能や知識、態度について、評価でもそれ相応の重点を置くべき

→教室での評価は様々なやり方で生徒に影響する。例えば、何が学習上重要かについての生徒の判断を導くものであり、生徒のモチベーションや実行能力の自己認識に影響し、・・・永続的な学習戦略や技能の発達に影響する

→クルックスの調査は、テストが否定的な側面と同時に、肯定的な側面を持つことを明示

### ●評価がモチベーションに与える影響

**自己制御学習** 内発的な動機づけを促し、それが生徒の学習者としての自律を促進

**等級別評価** 他の生徒の学習結果と比較されることなく、評価と結びついた短い学習単位は、2年間を対象とした公的試験よりも多くの生徒に対してモチベーションを高める

**GCSE 試験**モチベーションを高めるためにコースワークの評価を導入。但し、生徒がコースワークの課題に熱心になりすぎて、教室での学習や宿題に障害となることが問題

→評価の形式と、モチベーションの種類を区別すべき

→生徒の自己評価がどのようにしてモチベーションを高めるかの判断は容易ではない(納得でき、適用可能なことや、肯定的、建設的に応用されることなどの条件が必要)

→ハイ・ステイクスな外部テストが生徒のモチベーションに与える影響として唯一明らかなことは、全ての生徒のモチベーションが高まるわけではないこと

### ●テストに向けた学習指導

テストに向けた学習指導、試験に向けた準備は、それなりに有効な指導方法だが、その結果が一般化可能性を備えているのかの証明が必要

テストに用いられる問題について教えられているかではなく、テストで測定しようとしている技能や知識を教えられているかどうかであり、生徒が構成概念を教えられているか、を知ることが必要

### ●テストの点数の汚染

※**汚染**:測定しようとする構成概念と何の関係もなくテストの点数が上がること

#### ハラダイナ 三つの汚染源 1991年

1. 学校や職員が生徒をテストに備えさせるために取る方法
2. テストの実施方法や実施条件
3. 学校や教師を越えた外部要因

#### レイク・ウォベゴン効果

ジョン・カネルが1987年に、アメリカ50州すべてがテストを受けた全学年で全国平均を上回るというテスト結果を公表。ウェスト・バージニア州において全生徒の基礎的技能の結果が全国平均を上回ったが、「どの州も平均より高いことになんの説得力もない」と結論づけた。教育省5つの原因を列挙(p.66)。リンらが1990年に、点数は上がっていても実際の達成状況は向上していない可能性を指摘した。

→レイク・ウォベゴン効果の原因如何にかかわらず、この研究と種々の批判から、テストに向けた学習指導やそれが学習内容や学習指導の方法にもたらす影響へ注目させることになった

#### コレッツらの分析 1991年

都市部の学校で数年にわたって用いられてきたハイ・ステイクスなテストの結果と、同じ学区で数年用いられてこなかったテストの結果を比較した。

→いつも実施しているテストで良い結果をとっても、用意をしていなかった別のテストで良い結果を残すことにはならない、ことが判明した

## ●教師はなぜテストに向けた指導をするのか、またその方法は スミス 1991年

アメリカの二つの小学校における外部テストの役割についての研究

15ヶ月にわたる学校生活のすべての側面を捉えた詳細なデータに基づいた分析結果は、**初等学校全国評価**に関する私たちの検討結果（以下の★）とも一致する

※**初等学校全国評価**: イギリスのナショナルカリキュラムに関して7歳段階での全国テストと教師の評価の実施が、学校のカリキュラムや教師の評価の在り方に与えた影響についての調査。1991年と92年。

→教師の感情（不安や恥ずかしい思い、尊敬を失うこと、テストの公表や使用による疎外感）が教師のアイデンティティに組み込まれ、それにより教師の学習指導についての定義をも形作ることを指摘。テストが教師に与える影響をいくつかのカテゴリーに分類。

1. テスト結果の公表が複雑な感情を教師に引き起こし、そのような感情を抱かずに済むように必要な手を打とうと教師に決心させる
2. テストの妥当性に対する疑いと、一方でテストの点を上げる必要性により、教師は矛盾や疎外感を感じている
3. テストが子供の感情に与える影響を考え、教師は心配したり自責の念を持ったりする
4. テストに要する時間が学習指導の時間を削減、テストされる内容に指導の焦点を合わせることでカリキュラムの内容を狭めることになる（カリキュラムの矮小化）

★「英国のナショナルカリキュラムの評価について、7歳児を担当する教師に見出された調査結果＝私たちの調査結果」になる点

- ・公表された結果に対する怒りと自責の念、将来そのようなことがないようにとの意志
- ・テストの内容の妥当性についての疑問がわかっているだけに矛盾や疎外感を感じている
- ・テストが子供に与える影響についての危惧の念

→スミスの正当な二つの発言

1. ノルム準拠であろうとクライテリオン準拠であろうとテストの形式が教師や学習指導へ影響を与えるわけではない。テストの点数を政治的または社会的にどう使うかにある。
2. 教師が何を教えるか、どのように教えるかに関して、ハイ・ステイクスな限られた範囲を調べるテストに完全に従属するようになると、テストされない題材を教えたり、その他の指導方法を実行する力を失ってしまう。

## ●学校システムへの影響

アイラジアン 1988年

「現存する異なった文化を考えた場合には、国の命ずるハイ・ステイクスなテストの実施はそれほど積極的な支持を得られないこともある」

テストの実施が学校教育の目的や構造、その過程の再検討を目指す手段とする場合、間違っている点として三つを指摘 (p.75)。

→テストの実施は個人の弱点を発見するだけになりがちであり、システムについての診断なしに、現在の目的や構造、過程のどこを変更すべきかについての指針を得られない

→(コルベットとウィルソン、アイラジアンによる、**最低修得度テスト**の導入を関連させた議論に対し)ただ特定の方式の評価を実施するだけでは教師の実際の活動を変えることにはならない。パフォーマンス評価の導入が、教師の学習指導の実践に大きく影響するためには、情報不足であり、中途半端に教えられている技能や指導過程についての説明や訓練を伴わなければならない

※**最低修得度テスト**：アメリカで1960年代から行われたテスト。基礎的な能力や技能を身に付けさせるため、一定の点数以上を取らなければ次の学習に進めないようにして反復学習をさせた。

#### シェパード命名、「テストを梃子にした改革」1992年

ハイ・ステイクスな最低修得度テスト導入の結果、アメリカで基礎的な技能のテストの点数が上昇したが、高次の思考力や問題解決の技能を犠牲にしたことを受け、1990年代初めにオーセンティック評価やパフォーマンス評価の導入を求める声が出た。

→ハイ・ステイクスなテストの影響について、シェパードの発見

1. テストの結果が政治的な圧力やマスメディアによってハイ・ステイクスなものになると、得点のインフレーションを生じるため生徒の達成状況について誤った印象を与える
2. ハイ・ステイクスなテストはカリキュラムの内容を狭める。テストされる内容はテストされない内容を排除しながら教えられる
3. ハイ・ステイクスなテストは、基礎的な技能に関するものであろうと、学習指導の方向を誤らせる。
4. テストが強化するドリルの繰り返しによる学習指導は、時代遅れの学習理論に基づく。学習の改善につながるのではなく、生徒の思考力や問題解決技能の発達の機会を実際に奪うことになる
5. テストの点をあげる圧力のため、学習指導の難しい生徒は学校システムから排除される
6. 外部から命令されるテストの専制は教師の専門的知識や地位を低めることになる

#### 1992年、レズニックとレズニックが指針として導き出したアカウンタビリティーのための評価とその影響についての3つの原理

- ・評価するものを手に入れることになる
- ・評価しないものは手に入らない
- ・教師に指導して欲しいものを評価すること

## シェパードによる、学習指導の方向を変えるためにパフォーマンス評価を導入する場合に考慮しなければならない要因

- ・パフォーマンスを求める課題は矮小化されないものであること（パフォーマンスが他の似た課題に対しても適用できるものであること）
- ・高次の思考や技能に向けた学習指導を振興するための、カリキュラムや学習指導に関する教師の研修

→テストそのものは私たちが求めている変化をもたらすことはできない

## シェパードの指摘

「公平性の視点から見ると、問題は試験が公正に測定しているかどうかではなく、試験が教育的機会にどのような影響を与えているかであり、つまり妥当性の問題である。テストの使用によって重大な結果が生じるほど、テストの妥当性とテストの使用の公正の問題に関心を払うことが大切である」

## 第3章のまとめとして

特にハイ・ステイクスなテストが、学習指導や学習に対する影響力を持つことは明らかである。どうすればこの強力な手段を利用して、評価が最も広い意味で高次の技能や過程を発達させるのを促進できるのか。

→教育評価パラダイムの中で評価をより広く概念化することが、その一部を担うことは確かだが、教育の本質とも言うべき評価の種類を論じる前に、信頼性と妥当性という技術的な問題を取り上げなければならない