
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

●はじめに

[評価におけるパラダイム転換]

従来の評価方法：実技、対話、筆記試験、標準テスト

↓

多様な評価方法の登場：教師評価、標準課題、コースワーク、ポートフォリオ

* 標準課題 (Standard Tasks) はイギリスの7歳児におこなわれるテスト

パフォーマンス評価をもとにしている

* コースワークはイギリスの中等教育終了資格試験のパフォーマンス評価部分

* ポートフォリオは、学習活動のなかで示した成果や事例、作品を価値として評価する

[評価方法におけるパラダイム転換]

ノルム準拠テスト (偏差値、順位で結果を出す、相対評価)

↓

クライテリオン準拠評価 (絶対評価、到達度評価)

形成的評価 (学習過程で得た評価を学習の改善や向上に用いるための評価)

パフォーマンス評価 (求める技能や能力を活用できるかを評価)

まとめると「精神測定学」を基礎とした測定中心型評価からの変化である

この変化した評価における課題とはなにか？

「信頼性」

- ・教育的に意味のある評価の内容の確保
- ・評価から導かれる決定の妥当性の確保

→しかし「高度の標準化されたテストの手続きはこうした新しい評価にとって

信頼性の確保のためにふさわしくない

* アカウンタビリティ

運営資金を提供している納税者に対する説明責任

●評価の目的適合性（何のための評価か）

[評価の信頼性と妥当性]

評価の目的

→専門情報をくみあげること

→しかし主流は、生徒のテストや試験の成績（学習を支援する以外の目的で開発されたテストによる評価

（アカウンタビリティに供する情報：学校の成果を知りたくて、その情報にアクセスしたい→テストが主流→信頼性が高い→表面的になりがち）

いま考えなくてはいけない評価

「望ましい学習指導や学習活動を阻害することなく、生徒や学習状況についての質的にすぐれた情報を提供し、アカウンタビリティの目的にも応用可能な評価の方法を考え出すことである」（p.5）

「求められている目的に適合し、かつ学習指導や学習活動の向上に役立つような評価方法を開発しなければならない」

→指導や学習過程を支援するための評価

どのような種類の学習を私たちはもとめているのか？

→学習活動と評価の関係

→学習のあるべき姿を明晰化して、評価の枠組みを構築する

●精神測定学

[精神測定学による評価の枠組み]

精神測定学の特徴

（1）個人の固有性の限定

→知能と知能テスト研究によって発達した

→知能を生来の特徴と同定する

→知能は測れるものである（＝個人と同一のもの）

→能力別クラス、グループ、学校へ選別

（2）個人の他者による格付け

→得点の解釈を全体の得点分布に対比して行う

→個人の成績が他の生徒の結果しだいでかわってくる

(3) 技術的なことに最優度がある

→標準化、信頼性、限定的な解釈など

(ex. テストや評価は全員同じようなやり方で実施され、採点や得点の解釈も同じように行わなくてはならない)

→このような形式のテストで評価できる題材や課題は限られている

精神測定学に対する批判

精神測定学による数式、数量化は客観的であるのか？

(1) テストは万人に本質的であるという普遍性の仮定がある

→テストが開発される前に、テストが評価しようとしている構成概念を規定することからはじめている

*「構成概念」=表面に現れた現象の背後にある技術や属性

→テストが属性を評価しようとしていることを確認し、それによって「妥当性」を確保しようとしている

(ex. 「読むこと」の例

「読むこと」の評価例 (評価すべきこと): 音読と黙読、正確さ、淀みなさ、文章の解釈、読書への関心

*標準テストでは、一部の技能しか評価しない。例えば、簡単な文章の解釈。

このことは個人の「読むこと」の能力を代表してはいない

(2) 単一次元性の仮定

→個人のもつ能力の属性1つだけを測定する

→不適合な問題や全体の得点や相関関係が低いものを排除する

→個人の能力をなす要因の一つだけを評価

→構成概念が意図的に単純化される

→問題が多く属性を測定しているとしても少数の問題で少数の属性を評価している

→構成概念を広く考えようとする得点の解釈には応用できない

精神測定学の方法に対する批判

(1) 個人の変わらない特性を測定しているという信念

→子供がどれくらい教育不可能か

(2) 信頼性のために、テストの実施や課題や内容を標準化していること

→結果、評価される課題の形式を制約

→妥当性の問題や教師自身の問題意識は後退してしまうか無視されてしまう

●教育測定

[精神測定学から教育測定へ]

グレイサーの1963年の論文から「教育測定」が展開される

クライテリオン準拠の教育測定

→精神測定学から教育評価の分離

→評価の結果をノルムを用いて解釈しない

→他人との比較で個人をとらえるのではないテストを作り出す

(例：絶対評価)

クライテリオン準拠の教育測定の問題点

→ノルム（偏差）がクライテリア（基準）の設定や解釈に用いられる

→評価のよりどころがクライテリアかスタンダードしかない

ウッドによる教育測定の定義

1. 個人の達成状況を他の者とではなく、彼自身の以前の達成状況と比較する
2. 知能をテストするのではなく、実行能力をテストする
3. 評価の手順はこれまでと異なり比較的自在な状況で実施され、そのため厳密に統制された条件下のデータではない
4. 型にはまったパフォーマンスではなく、ベストを尽くした場合に注目する
5. 標準テストの特徴である規則や規制が緩和された場合に効果的である
6. 個人をおとしめるのではなく、支援のための、その場しのぎの評価ではないことを示そうとする

ウッドの教育測定の特徴

- ・実行能力という言葉を使うことで、知能のような持って生まれた特徴ではなく、教育や訓練、その他の経験によって作り出されて能力を言いあらわそうとしている
- ・評価されるパフォーマンスや特性は異なった特徴をもっているため教育測定を精神測定学に基づいて考えるべきではない
- ・型にはまったパフォーマンスをいながらにして見るのではなく、大人が生徒を支援することを可能にする

実行能力とパフォーマンスの区別

パフォーマンス

→パフォーマンスは普通の状況下で何ができるのかを見ようとする

実行能力

- 実行能力はその人が理想的な状況下で何ができるのかを見ようとする
- 課題の達成度に影響する動機や好み、認知的な要因などに加えて、自分の持つ知識体系を動員して用いる能力

実行能力の評価

- 実行能力とパフォーマンスの区別は教育評価をかんがえるうえで重要
- ベストを尽くした場合のパフォーマンスこそ考慮すべきである
 - * 「理想的な」は「ベストを尽くした」と同義と考えるとわかりやすい

●教育評価

[教育評価に関する例]

グレイサー

- 評価は生徒や教師にとって、実行可能な改善方向を示すものでなければならない
- 評価はそれ自体で有効でなければならない、獲得した知識や技能を生徒が使えるかどうかに焦点をあてるべき

ラヴェン

- 認知的要因、感情的要因、意志的要因を分けるのではなく、価値ある目標を設定してそれに対するパフォーマンスを評価すべきである
- * メタ認知過程を重視する認知科学や学習理論の考え方から歓迎されている

ゴールドスタイン

- 生徒はテストを受けているときも変化している。相互の影響も考慮して評価を考えるべき

* オーセンティック評価

評価に用いる課題をパフォーマンスによって（オーセンティックな形式を用いて）、普通の授業で実施できるものとする

* パフォーマンス評価

テスト内容を、基準となる（＝オーセンティック）パフォーマンスで示される批判的な思考や知識の総合を求めるもの

（例：「択一式のテストのように間違っただけの解答の排除ではなく、解答をつくるための思考にある」）

パフォーマンス評価の課題

- 評価の統一に関して信用を得ることができるのか
- パフォーマンス評価を精神測定学の枠組みでかんがえるのではない
- 評価の統一の提示と学習過程に即した評価の必要性
 - *このことが評価の統一を困難にしている
 - (例) アカウンタビリティのための評価と教室での評価は根本的に違う
 - 標準テストをパフォーマンス的にする
 - 教室での評価をもっと標準化したりする
 - 全員に共通の属性を探り出そうとする
 - 複雑な現実から単一の要素を抽出して部分を評価する

スティギンズの2つのテストシステム

- ・「上から下降する」テストシステム
 - 標準化が先決であり形式はペーパーテストまたはパフォーマンス評価
 - 結果は通知しておしまい
 - ・「下から上昇する」テストシステム
 - 標準化されておらず、広い範囲の学習活動を対象とする
 - 指導方法を選択するために実施される
 - 生徒にとって指導と学習過程が適切であることを評価する
 - 評価は、狭い範囲で、多様な達成目標によって詳しく評価する
 - 実行能力、動機づけのために評価の統一は二の次
- *教師の役割は、実施者と結果を解釈する者
 - *生徒はこの評価を「学校そのものに対する態度を決定する情報」として用いる
 - *保護者は、進歩の確認、潜在的可能性を判断するために用いる

●精神測定学の遺産

- ・簡単に数量化できる技術や知識を重視
- ・個人的な活動での認知活動を教育上の進歩とみなす
- ・教育上の進歩を科学的な測定の対象と考える
- ・学習活動を序列化するためのテスト

●教育評価のこれからの課題

- ・評価の形式を議論するのではなく、評価が学習指導や学習活動に与える影響について理解を深める必要がある

- ・評価は学習指導や学習活動の外にあるのではなく、緊密な相互関係にあることろ理解する必要がある
- ・多様な評価の手段を用いることのできるシステムを作り出す

教育評価に関する困難な課題

→高い信頼性を重視する評価のあり方は考え直す

→複雑な課題を評価できる評価の一貫性と統一性を確保する方法の開発