

学習を支援する評価

— キャロライン・V・ギップス著(鈴木秀幸訳) —

『新しい評価を求めて—テスト教育の終焉—』

堀川 有美

要 旨

本書は、Gipps, C. V. (1994) *Beyond Testing-Towards a theory of educational assessment* の全訳である。教育評価は、精神測定の分野から独立し、目的も方法もより多様化したものへと移行してきた。本書は、その移行を「パラダイムの転換」としてとらえ、教育評価が学習のための評価であることを強調しながら、新しいパラダイムの方向を示したものである。はじめに、学習モデルとの関連、測定中心の評価のもたらす弊害が述べられ、続いていくつかの評価方法について各々の意義と問題点などが英国の教育評価の事例を交えて解説されている。その中には、クライテリオン準拠評価、パフォーマンス評価、教師の評価と形成的評価など、日本語教育においても必須のテーマが含まれている。最後に教育評価の枠組みが提示され、信頼性と妥当性の概念の再定義が試みられている。

【キーワード】 教育評価、 学習を支援する評価、 アカウンタビリティ、 教師の評価、 形成的評価

1. はじめに

本稿では、2001年に論創社より出版されたキャロライン・V・ギップス著、鈴木秀幸訳の『新しい評価を求めて—テスト教育の終焉』(原書 Gipps, C. V. (1994) *Beyond Testing -Towards a theory of educational assessment*. The Falmer Press.)を紹介する。著者のギップスは、初等学校教師として勤務した後、研究職に転じ、パフォーマンス評価を組み込んだ全国学力調査、評価での公平性の問題、形成的評価などの研究を行ってきた人物で、1994年当時ロンドン大学教育学部教授であった。現在は英国の University of Wolverhampton の学長である。

訳者序文によると、本書は、「教育評価の新しいパラダイムの方向を解説したものとして、英国のみならず広く世界で教育評価の研究者に読まれており、多くの論文に引用されている」という。「測定から評価へ」という近年の教育評価の潮流に影響を与えた書物の1つであろう。本書は、10年以上前に出版されたものであるが、教育評価の理論および具体例を豊富な先行研究や英国の具体的事例に基づいて論じており、多くの言語教育関係者にとって現在でも役立つ内容であると考えられる。

本書の対象読者は、「なんらかの形で教育にかかわる人々や、評価に興味を持つ人々」(pp.2-3)であ

る。著者は、本書が「テストや評価についての基礎的知識を前提としてかかれている」(p.23)と述べているが、統計処理等の専門的知識が必要なわけではない。巻末には、訳者による用語の解説もついている。テストや評価の入門書程度の知識はあったほうが議論を理解しやすいかもしれないが、本書の大部分の内容は、教育に携わっていれば身近に感じられる問題を扱っていると考えられる。

本書の構成は、次のとおりである。

- 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
- 第4章 妥当性と信頼性
- 第5章 クライテリオン準拠評価
- 第6章 パフォーマンス評価
- 第7章 教師の評価と形成的評価
- 第8章 倫理と公正
- 第9章 教育評価の枠組み

本稿では、第1-3章、第4-7章、第8-9章の3つに分けて、各章の内容を紹介する。

2. 書誌情報

キャロライン・V・ギップス(著)、鈴木秀幸(訳)(2007)『新しい評価を求めて—テスト教育の終焉』論創社(原書 Gipps, C. V. (1994) *Beyond Testing – Towards a theory of educational assessment*. The Falmer Press.) 3,500 円(税抜) ISBN4-8460-0286-1

3. 各章の紹介

3.1 教育評価のパラダイム転換(第 1-3 章)

第 1 章から第 3 章では、本書のテーマである教育評価におけるパラダイム転換の経緯とその背景が説明され、本書全体の議論の導入がされている。

まず第 1 章では、教育評価が、精神測定学をもとにした測定中心の評価から、より多様で幅広い評価へと移行してきたことが述べられている。従来の精神測定学にもとづく評価方法では、個人の固有性を示すことが特徴であり、その固有性は変わらないものと考えられていた。しかし現実には、学習者は常に変化しているものであり、そのような限定的な見方には問題があると著者は考えている。また、測定では、信頼性や標準化などの技術的な問題が最優先されるため、採点だけでなくテストの実施や内容まで標準化されてしまうことも、問題点として指摘されている。

ただし、これらの批判は全ての教育評価に対して向けられているわけではない。著者の主張は、評価の目的を明確にし、その目的に適合した評価方法を選択すべきだということである。評価の目的として、「学習を支援する評価」と「アカウントビリティーのための評価(教育システムや学校の成果を知らせるための評価)」の 2 つが区別されており、後者のテストを前者の目的で使用することを問題視している。教室内テスト等、学習を支援する目的で行われる評価では、アカウントビリティー目的の評価ほど高い信頼性は重視されず、また、他の生徒との比較ではなく個人として捉えることが望ましいとされている。

第 2 章では、近年の学習モデルの変化について述べられている。学習モデルの変遷も、教育評価パラダイムの転換の重要な要因となっているためである。著者は、行動主義心理学に基づく「ブロック積み学習モデル」を批判し、それに代わる学習モデルとして認知心理学や構成主義に基づくモデルを説明している。このモデルでは、学習を知識の構築過程とみ

なし、学習者は受動的に知識を吸収するのではなく能動的に学習を進める主体として考えられている。そのため、推論したり思考したりする高次の学習技能とメタ認知の力が重視される。

このような学習モデルにおいて学習を支援するためには、どのような評価方法が求められるであろうか。本書では、ヴィゴツキーの学習モデルに基づく「足場組評価」や「動的評価」などの評価方法が紹介されている。これらは、評価者と子どもの対話に基づいて、必要な支援を与えながら子どもの学習の可能性を見ようとする評価方法である。指導と伝統的な試験との中間点に位置する方法であるという。著者が指摘しているように、教師の専門的スキルが必要になるなど実施上の問題はあっても、このような学習と評価に対する見方は、学習のための評価を考えるうえで大きな示唆を与えてくれる。2 章の結論部分では、「評価の新しいパラダイムへ移行するためには(略)生徒から支援や道具を引き上げるのではなく、反対のことが必要である」(p.41)と述べられている。

第 3 章では、テストが学習やカリキュラムに与える影響が説明されている。特に、ハイ・ステイクスなテスト(テストの結果が、生徒の将来の進路や学校の評価となり、社会全体の注目を浴びるようになるテスト)が、いかに学習指導やカリキュラムを歪めるかが述べられている。テストには否定的側面と肯定的側面があり、肯定的側面を高めるような評価の計画が必要であるとされている。テストが強力な力をもつことは明らかであるが、その力を適切な方向に向けるために評価方法を注意深く設計することが重要だということである。

3.2 新しいパラダイムにおける評価(第 4-7 章)

次に、新しいパラダイムにおける多様な評価方法が個別に検討されている。まず第 4 章で、続く 3 つの章の議論の前提として、妥当性と信頼性の概念が説明されている。妥当性(テストが意図したものを測定しているか)については、Messick(1989)の議論を中心に、妥当性は複数の側面からなる統合概念であること、また、テスト使用の社会的な結果も妥当性に含まれること等が説明されている。信頼性(テストが測定しようとしている技能や達成事項をどの程度正確に測定しているか)については、採点の信頼性に関する記述が多い。評価の一貫性を確保する

ための方法を英国では「モデレーション」と呼ぶとのことであるが、これには得点分布を調整するための統計手法もあれば、グループ・モデレーションと呼ばれる研修会や反省会なども含まれているという。モデレーションについての日本語での説明は、管見の限り本書以外に見当たらず、大変興味深い。

第5章から第7章では、新しい教育評価パラダイムにおいて注目される3つの評価方法について特徴や限界が論じられている。まず、第5章で、クライテリオン準拠評価(criterion-referenced assessment)が扱われている。クライテリオン準拠評価は、一定の基準(criterion)と比較して評価の結果を示す評価である。得点の分布に基づいて結果を示すノルム準拠評価(norm-referenced assessment)と区別される¹。学習を支援する評価では、学習者間の比較よりも各個人の成長に関心があるため、クライテリオン準拠評価のほうがふさわしい。ただし、クライテリオン準拠評価とノルム準拠評価は密接に関係しているという論も紹介されている。

また、「クライテリオン準拠」の解釈の変遷についても述べられている。厳密な形のクライテリオン準拠テストを実施するためには、評価基準を詳細に示す必要がある。そこで提唱されたのが評価領域やそれに基づく問題作成の方法を細かく規定する「ドメイン準拠テスト」であった。しかしこれは目標の過度の細分化・矮小化につながるとして批判され、より現実的で適切な方法として、スタンダード準拠テストが提唱されたという。この評価方法では、言語表現と具体的事例の組み合わせからなる「スタンダード」を用いて、教師の質的判断による評価が行われる。これらの記述から、クライテリオン準拠評価においては、評価基準の詳細さの度合いが、妥当性や信頼性に影響を与える重要な要因となることが読み取れる。

第6章では、パフォーマンス評価が取り上げられている。パフォーマンス評価とは、「生徒に取り組んで欲しいと願っている現実の学習活動をモデルにして評価しようとするものであり、択一式テストのようにこれらを分断して評価することではない」(p.135)。関連する用語として、「オーセンティック評価」があるが、これはオーセンティックなコンテキスト(つまり通常の学習活動の中)で行われるパフォーマンス評価であるとされている。パフォーマンス評価は、表面的妥当性、結果妥当性などが高い傾

向にあるが、伝統的な意味での信頼性は高くない。また、特定課題のパフォーマンスに基づいて評価を行うことから、領域全体でのパフォーマンスの推定を困難にし、一般化可能性に問題があることも指摘されている。それでもやはり、著者はパフォーマンス評価を学習に有効なものとしてとらえており、さらに、アカウントビリティ目的の評価としての実施も可能であるとしている。

第7章では、教師の評価と形成的評価について述べられている。教師による評価は、一定の期間にわたって複数の活動を対象として繰り返し行われるため、評価の妥当性も高いという。形成的評価とは、評価から得られた情報を指導・学習にフィードバックする機能の評価である。教育評価の分野においては、評価の機能別に診断的評価・形成的評価・総括的評価という3分類がされることが一般的だが、本書では、「診断的評価は本質的に形成的評価の一部であり、それ自体学習指導の技術の重要な部分である」(p.174)と指摘されている。

教師の評価の信頼性と妥当性については、第1章で述べられた評価の目的が関連している。つまり、教師の評価が形成的な目的で実施されるのであれば、妥当性が重要であり信頼性はそれほど重視されないが、教師の評価がアカウントビリティ目的などで使用される場合には、信頼性も重要となる。教師の評価において信頼性を高めるには、教師間の討議に基づくグループ・モデレーションが有効であるとされている。

3.3 これからの教育評価の視点(第8-9章)

最後に、第8章と第9章では、新しい教育評価パラダイムにおいて求められる視点と枠組みが示されている。第8章では、評価の倫理と公正の問題について述べられている。著者は、「評価が生徒の一生にたいへんな影響をあたえることを考えて、これからの評価についての研究やその使用を規定する枠組みのなかに、倫理的な問題をも含ませておくべきである」(p.23)と述べ、あえて1つの章をこの問題に充てている。まず、Messick(1989)による結果妥当性の概念を示して、テスト開発者はテストの構成概念のみならずテストの適切な使用方法やテストによる社会的帰結までも考慮しておかなければならないとしている。また、公平の問題として、テストが特定のグループに対して不利になる可能性について様々

な角度から検討している。さらに、テストの準備、つまりテストに向けた指導の倫理性についても先行研究を紹介しながら論じており、現場の教師にとっても興味深い内容であると思われる。

第9章では、本書全体の総括が行われるとともに、新しい教育評価の枠組みが提案されている。まず、第8章までの議論をもとに、教育評価の定義を10点にまとめている。その1つを例として挙げる。「教師自身による生徒の評価は、教育評価の中で重要な構成要素である。そのような評価は、生徒と十分に関わり合い、理解していることや勘違いを見つけることで、相互作用をもつ。それは、学習過程の支援や足場組となり、様々な学習活動でのパフォーマンスの評価を可能とする」(p.222)。

また、教育評価の理論の展開のために、信頼性の再定義を課題として挙げている。著者は、伝統的な信頼性という用語ではなく、評価の一貫性を基礎とする「評価の統一」という用語を用いるべきだとしている。妥当性についても、Messick(1989)による妥当性の包括的見方はあまりにも複合的であるため優先順位を示す必要があると述べ、新しい評価においては結果妥当性が特に重要であるとしている。

最後に、教育評価の質を判断するための新しい基準として、(1)カリキュラム適合性、(2)評価の統一性、(3)信任性、(4)社会的信用性、(5)コンテキストの記述、(6)公平、の6つが提示されている。

4. おわりに

日本語教育においても、教育における評価の重要性は当然認識されていると考えられる。しかし、現実に利用される評価方法は種類に限られており、評価の目的と方法が合っていない場合も多いのではないだろうか。その原因として、学習や教育に役立つ評価という視点で、評価の理論的側面や具体的な評価方法がまだ十分に検討されていないということが考えられる。

そのような状況において、本書から得られる示唆は大きいと思われる。特に、教育評価では従来の信頼性・妥当性とは異なる判断基準が必要となること、評価の研究では倫理や公正の側面も重視されるべきであること等は、現在の日本語教育においてあまり指摘されていない点であろう。

本書では、評価における教師の役割の重要性も強調されている。評価が教育の重要な一部であるとすれば、評価に対する正しい知識とスキルは、教師に求められる専門性の一つであると言える。学習と教育のための評価について知識を広げたいと考える方には、一読をお勧めしたい文献である。

謝辞 本書は、お茶の水女子大学大学院 2005 年度後期の「教育実践学演習」ゼミで講読した文献です。本書をご紹介、ご指導くださったお茶の水女子大学の富士原紀絵先生、および同ゼミの院生の皆様にお礼申し上げます。授業での議論は本書の理解に大いに役立ちました。尚、本稿は筆者の個人的見解に基づいてまとめたものであり、文責はすべて筆者にあります。

注

1. criterion-referenced assessment は目標基準準拠評価、norm-referenced assessment は集団基準準拠評価と訳されることが多いが、本書ではより原語に近い訳語としてクライテリア準拠評価を使うとされている。

参考文献

- Messick, S. (1989) Validity, In R. Linn (ed.), *Educational Measurement (3rd ed)*, American Council on Education, Washington, Macmillan.

ほりかわ ゆみ／東洋大学 文学部
yumih@hotmai.com