

# 第7章

## 教師の評価と形成的評価

### 本章の構成

- ① 教師の評価
- ② 形成的評価
- ③ 教師の評価の問題点
- ④ 評価におけるフィードバック
- ⑤ 教師の評価の信頼性と妥当性
- ⑥ モデレーションを通じた教師の評価の信頼性の強化
- ⑦ ナショナル・カリキュラム Key Stage 1 での教師の評価のモデレーション
- ⑧ 結論（まとめ）



## ① 教師の評価 (pp. 171-173)

### 教師はクラス内で様々な場面で様々な形の評価を行う

- 生徒から得る情報は断片的であるが、そのような情報を積み重ねることによって、生徒の学習をよりよく理解することができる
- そのような積み重なった評価は、内容や構成概念の点で高い妥当性を有しており、一般化が可能である。(例、様々な課題についての評価を重ねることにより、計算能力や読みの能力を見ることができると)
- 教師が形成的な評価を実行すると、生徒たちの学習に貢献する。(教師→教育の修正・調整／生徒→学習を見直す)

### 外部の試験

- 限られた範囲の情報をもたらすだけ

### 教師の評価は、構成主義の学習モデルの上に成り立っている

- 構造主義：子供が自身の前提知識をどのように使って新しい知識を自分のものにしていくか。このモデルの枠組みでは、学習とは理解を伴うプロセスである。
- 心理学の見地：学習課題が生徒の理解の程度に適合していることを必要としている→生徒の今のレベルと同じくらいの課題をさせるか、少し高いレベルの課題をさせるか。

形成的評価を通して生徒が何を知り、どのように学習したかを理解することはその後の指導設計に有益である。それゆえ形成的評価は連続的に行われる学習指導の一部となっている。生徒自身も、教師からのフィードバックに向き合うことによって自身の学習を改善するのに何が必要か知る必要があるだろう。

## ② 形成的評価 (pp. 173-176)

### サドラー (1989) による定義

- 学習をより促進するために、どの程度生徒の実行能力の形成や向上に役立つかに関わるもの。
- 2つのフィードバック  
教師へ：レディネス、診断、補習などについての決定材料となる  
生徒へ：学習の得意な部分と弱点を発見し必要に応じて強化・補強する (診断的評価)

### 教師の形成的評価 (フィードバック) が生徒の学習を改善させる条件 (⇒生徒のメタ認知を促進)

- 学習の達成目標について知っていること

- 目的を達成するために必要な行動・テクニックを導き出せること
- 次に求められるパフォーマンスと実際のパフォーマンスを比較できること
- それら2つのパフォーマンスのギャップを解消するために適切な行動を導き出せること

上記の条件を満たさない例：グレード使用（よくできました：7/10）→サドラー的には形成評価ではない。

### 形成的評価の過程で生徒の学習を改善するための不可欠な要素

- ◆ 生徒を「自己制御できるレベル」に引き上げること（→自律を促す？）  
ただし、そうなった生徒にも引き続き教師からのフィードバックは必要である（一流スポーツ選手とコーチの関係に似ている）
- ◆ 教師は、生徒と自己制御モデルに従って共に目標を設定するべきであり、生徒自身でも目標を設定できるようになるべきである。
- ◆ ブランジャー（私信）：「パフォーマンスの質を判断する方法を、批判的に再考するといった反省や熟考を伴わなければ、自分の学習をうまく利用できない（p. 175）」（メタ認知ができなければ認知スキルを使うことができない？）
- ◆ 具体的で明確で、挑戦的な目標がパフォーマンスに対して最も効果的である
- ◆ 反対に、「自分のベストを尽くせ式」の目標は、目標のない場合と比べて対して効果的ではない

### トーランス（1993）：指導⇔学習のサイクルにおける形成的評価の位置付け・機能

- 対話による評価の機会を設け、ある課題の成果物を作り上げる過程の評価を考慮すべき
- 形成的評価は、構成主義と相性が良い（発達の最近接領域）

### ③ 教師の評価の問題点（pp. 176-180）

学校内評価が総括的な目的に用いられる限り、教師は生徒によって援助者ではいられず判定者になってしまう。**形成的評価（援助者）と総括的評価（判定者）**は対立する。

#### サドラーによるこれら2つの評価の対立についての記述

こまめに行われる形成的評価は、最終的にまとめられるようになると形成的な機能を失う  
→そのような状況下では、ある評価が最終評価に関係ないとき、生徒は、「やる価値がない」とってしまう。  
(p. 176)

#### ハーレン（1992）による、総括的評価における「要約すること」と「確認すること」の区別

**要約**…生徒の達成状況を見るために一定期間累積された情報で、個々のプロフィールとして保存される。  
→形成的

確認…特定の時期におけるパフォーマンスを記録するために行われるテストや課題→総括的評価

### 「質問すること」(→形成的評価の一種?) (p. 177)

- 役割: 「子供の解釈や概念を理解し、かつそのような概念や疑問を持つことを促す (p. 177)」だが同時に、**教師の質問は社会的な機能も持っている。**
- 教師の質問は答えが 1 つしかない場合が多いため、生徒は質問について考えるのではなく、教師の求めている答えは何かについて注意を注ぐようになる。(エドワーズ&マーサー, 1989; ポラード, 1985)
- **教師は、異なった種類の質問の役割を理解し、クラス内での学習活動についての生徒の受け止め方を理解する必要があるだろう。**→年齢や性別、人種などによって異なる。
- サドラーによると、形成的評価の枠組みでのフィードバックは、ほとんどすべての科目や状況を含む。算数の問題のような多くの場合○か×かの問題であっても当てはまる。
- 教師の評価基準は、生徒にとって透明であるべき。
- そのため、教師が生徒に具体的な評価の特徴を説明することも有効である。(ルーブリックなら可能?)
- 「よくやった」という発言は、生徒からするとそれ以外の成果は価値がないと見なされてしまい、さらには、生徒が教師の評価基準から意識をそらせてしまう。

### 生徒が形成的評価に対応できるようになる年齢について、PRAISE チーム (ブロードフット等, 1988) によって発表された、「達成事項の記録」に関する調査報告

- ✓ 中等学校の生徒は、自己評価に困難を感じている。その原因は、評価基準の曖昧さにある。明確でないと、達成事項についてノルム準拠的な見方をする傾向がある。そうなると、生徒が学習集団の中での自分の位置付けを気にしてしまい、そのことが自己評価に影響してしまう。
- ✓ ジェンダーや人種によっても自己評価に対する姿勢に違いがみられる。  
男子: 反論したり、周りの視線を気にしたりする / 女子: 男子よりもっと教師と議論する
- ✓ 自己評価と教師の評価が、生徒の学習を促進するようになるには、教師の技術向上が不可欠  
→評価基準について生徒に理解させ、生徒が自身の長所と短所を自覚するように促す努力が必要。

### ④ 評価におけるフィードバック (pp. 180-189)

フィードバックとは…

- 自尊感情を通じて生徒の学習を促進する
- 指導の過程で重要なのは、フィードバック、訂正、強化である。(ブルーム, 1976)

- 教師のフィードバックは生徒の学習活動への参加と理解であり、学習を成り立たせるための重要な変数である。（その他の変数：課題の提示、系統化、教師の期待等）（ベネット，1982）
- 生徒の学習過程に教師がフィードバックを与えることは、その生徒に次にどのような指導をするか、そのレベルの課題を設定するのかといった判断の助けとなる。（ギブス，1990；上のサドラー，1989）

### 誤りを訂正するタイプのフィードバックの重要性についての引用（p. 181）

（フィードバックは）正しい解答が何であるかを確認し、生徒が学習内容をどの程度理解しているかを示し、誤っているところを見つけ訂正する。または、生徒自身に訂正する機会を与える。この訂正機能がフィードバックの最も重要な働きであろう（カールハビー，1977）

**クルックス（1988）**・・・クラス内での評価が生徒に与える影響に関する研究→生徒の学習レベルに合った教材と適度なフィードバックは学習を向上させる。

**有効でないフィードバック**・・・点数を示したり正しい解答を確認したりするだけのフィードバックはその後のパフォーマンスに有益な影響を与えない。

**有効なフィードバック**・・・事実に基づく詳しいフィードバック、概念的な支援または使用した解決方法についてのフィードバック

### そのほかの教師のフィードバックに関する先行研究

|                          |  |
|--------------------------|--|
| <b>キューリックら（1990）</b>     | フィードバックとその改善策の実施が、とくに優秀でない生徒に対して有効であることを示す研究結果を得た。   |
| <b>クルックス（1998?）</b>      | 教師は、等級づけではなく学習に関連したフィードバックを使用すべきである。   |
| <b>ベネット&amp;ケル（1989）</b> | 教師の評価は「正しいか、誤りかを判断の中心としていて、注意や×か、コメントか、あるいはそれに準ずるものに陥りがちである」   |
| <b>ポラード（1990）</b>        | 学習過程において教師は「思考の支援者」である<br>→「意味のある適切な助言をし、子供が元々持っている経験を基礎として認知構造や技能の発展を図ること」<br>「このようにして子供が意味を見出し、発達の最近接領域を通過していくことを支援するのである。彼らの思考はこのようにして新しい経験の中で再構成されていく」 |
| <b>ベネット（1982）</b>        | 誉めることによる診断的評価の重要性。誉めることが効果的であるには、考慮されている内容に関連し、本質的で、信用できるものである必要がある。   |
| <b>ドウエックら（1978）</b>      | 学習活動に関係のないフィードバック（素行や容量の良さ）がフィードバックそのものの効果を損なう。フィードバックは生徒に対して彼らの学習上の力と価値について伝えることになる。  |

まとめると、フィードバックは生徒にどうすれば理解や概念化、そして学習を改善できるかを示すものでなければならない。

## 自尊感情とフィードバック

クーパースミス（1967）による自尊感情の定義

各人が普段から持っている自分についての価値判断—それは、肯定的なものであれ否定的なものであれ、各人が自分の能力、重要性、成功、価値についてどの程度信じているかを示すものである。

学習と自尊感情は密接に関わっており、**自尊感情が高まる**要因は他者（両親、教師、仲間）からの有意義なフィードバックである。特に学業においては、教師の評価が1番影響する。子供は学校という社会の中で他者との交流の中で自尊感情を形成していく。（クロッカー & チーズマン, 1988a, 1988b)

### そのほかの自尊感情とフィードバックに関する先行研究

|                  |   |
|------------------|---|
| シャンセン<br>(1985)  | 子供の自己概念を高めることが学校の1つの大きな使命である。   |
| チャップマン<br>(1988) | 子供は、7歳～8歳に自分の能力を過小評価する傾向にある。特に、学業での失敗の積み重ねがその評価をさらに下げる要因となる。その結果、自尊感情も下がる。  |
| ガーニイ<br>(1988)   | 教師のフィードバックには、学習活動に関するコメントが含まれており、自己概念や自尊感情への影響が大きい  |
| ドゥエックら<br>(1978) | 教師によるフィードバックが、女子生徒には失敗の原因を能力に結びつけさせ、男子生徒には、努力不足や真剣さの不足に失敗の原因があると結論づけさせる傾向がある。一方でクラスメイトからのフィードバックにはそのような異なる効果はない。  |
| ドゥエックら<br>(1986) | 自尊感情が高い生徒はチャレンジングな課題に対して努力して時間をかけて取り組もうとする。 <sup>1</sup> そうでない生徒は難しい課題に直面すると努力をせずにあきらめてしまう。 <sup>2</sup> ➡「学習された無力感」<br>性別による差異：女子生徒は男子生徒に比べて学習された無力感を抱きやすい<br>(p. 186)<br><sup>1</sup> 適合型モチベーション：挑戦を求めて、達成目標の確立や継続に向かう<br><sup>2</sup> 非適合型モチベーション：目標達成への失敗体験 |

## フィードバックと自尊感情

### ピントリッチ&ブルーマンフェルド（1985）

学年による差異：低学年の生徒の方が、自分の能力、努力、行動について肯定的に考えている傾向があるため、学習活動について称賛されると自尊感情が高まる。そのため、彼らのレベルを考慮しながら、彼らが「称賛的フィードバック」を受けやすくなるような課題を提供することが重要

### バン・オーデンフーベン&シエロ（1985）

教師による評価的なフィードバックが、生徒の達成水準についての教師の予測が持つ影響（日本語理解不能）→教師の達成基準に関する予測が与える、教師の評価的フィードバックへの影響??

教師が、「①達成水準が高いと思う生徒」と「②達成水準が低いと思う生徒」に対して与えるフィードバックに違いが見られる。①（＝実力を十分に発揮していない生徒）には、「もっとできるはずだ」といった個人的な励ましを送り、そのことが生徒の自信につながる可能性がある。②には、教師は否定的な非言語的フィードバックと肯定的な励ましの言葉を送っている。

重要なのは、評価的フィードバックの内容ではなく、生徒にとってそれが何を意味するか。

### クロッカー & チーズマン（1988a）

5～7歳までの子供の学校活動において、次の相関を発見した。

順位に関する・・・子供自身の推定—仲間の推定—教師の推定

### ティザード（1988）

7歳の児童の推定は正確ではない。8歳までの子供、特に男子は自分の能力やパフォーマンスを過大評価する傾向がある。

ジェンダーや人種によって異なるフィードバック例：

- ・ 白人の女子生徒：叱られることが最も少ない。誉められることも少ない。  
→教師の目には止まらない
- ・ 黒人の生徒：ふざけ回る傾向があり、叱責されることが多い。
- ・ 女子生徒：読みと数学で自身を過小評価する傾向

フィードバックは、教室のあらゆる場面で行われており、生徒の成功や失敗に関する認知や自尊感情を決定づける重要な要因であるが、教師の生徒の成功に関する予測や生徒の年齢やジェンダー、人種などによって影響を受けている。

## ⑤ 教師の評価の信頼性と妥当性

**妥当性 (pp. 81-82)** : テストによる測定が、どの程度正確であるか。正確さを保証するには、特に構成妥当性が重要で、評価する構成概念の定義を意識することが必要。

例：読み {大きな声を出して読む、読みながら内容を理解する、読みの正確さ、読むことを楽しむ、etc.}

**信頼性 (p. 94)** : 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

サドラー (1989) によると、評価によって信頼性と妥当性の位置付けが変わる

- 総括的評価において、信頼性は妥当性を考える上での前提条件 (妥当性を成り立たせるためには信頼性を保証する必要がある)
- 形成的評価において、妥当性は信頼性のための十分条件 (妥当性が成り立てば信頼性も成り立つ)

フィードバックを与えた結果として学習が向上するのであれば、それは形成的評価が形成的な妥当性を担保していることを示す。そのために必要なこと

→ (フィードバックを伴う) 形成的評価の対話に生徒を参加させるようにする。そうしながら、生徒は自分の学習を客観的に振り返ることができる。

ここで、教師が考えなければならない課題→**評価の目的の問題**

評価が (外部試験のような) アカウンタビリティのために用いられるのであれば、(評価の統一を保証するために) 信頼性を必要とする。その一方で、評価が形成的な目的で用いられる場合、信頼性はさほど重要ではなく、むしろ妥当性 (内容、構成、概念、結果妥当性) の方が重要となる。

では、形成的評価がアカウンタビリティに部分的に関わってしまう場合、どのように信頼性を保証すれば良いのか? →**モデレーションを行う**

## ⑥ モデレーションを通じた教師の評価の信頼性の強化

ある学年の生徒全員が取り組んだ同じ課題を異なる教師が評価する場合、評価基準が統一されているのだろうか?

評価の一貫性を高めるためのポイント

- ◆ 明確なガイドライン

- ◆ 評価の過程でのグループ・モデレーション（のちに説明）
- ◆ 明確な評価基準と評価計画の共有
- ◆ 外部からのモデレーション→学校へのフィードバック

モデレーションの役割：教師としての技術の向上（教授法関連の研究会に似ている？）

グループ・モデレーション：生徒の成果物（サンプル）が教師らによって討議され、評価基準に関する共通理解、特に評価過程・評価結果の統一を図る。

コンセンサス・モデレーション：上記の議論を学校よりもっと高い次元（地域や州，国レベル）に引き上げることも可能。そうすること自体が，社会における評価の信頼性を高めるための監査機能の役割を果たしてくれる。（リン，1992）

## ⑦ ナショナル・カリキュラム Key Stage 1 での教師の評価のモデレーション

グループ・モデレーションが最も有効に働いた例

### 【手続き】

- ① ナショナル・カリキュラムの2つの結果について教師のグループが議論する。教師の評価とテストの本来の結果について不一致があるかどうかを精査する。
- ② 特定の学校に不一致がある場合は，成果物のサンプルや当該の学校の状況の情報を頼りに検討する。
- ③ それぞれの学校の生徒集団についての最終的なレベルの分布を決定する。
- ④ 教師が成果物を持ち寄り，外部テストの結果に，教師たちのレベルを適合させる，研究活動の場であり、この活動が教師の技能向上に貢献する。

TGAT モデレーション活動に対する学校試験評価委員会の否定的な反応

- 教師の負担が大きい
- SATs（英国のナショナル・カリキュラムの枠組みで7歳児に対して実施される全国テスト）のデータが達成目標に反映されていない。
- クライテリアン準拠評価（到達度）システムでのモデレーションは、生徒の集団ではなく個人に焦点を当てるべきだ。

信頼性に関する懸念→達成事項の記述について教師の理解が十分でなかった。

- そのため，モデレーショングループのメンバーは，議論を通して教師が評価基準をよりよく理解できるように支援をいくつかの学校で行った。（1992年まで）

- 実際に支援を行ったモデレーションメンバーによると、訪問した学校の教師が行う評価には妥当性について問題があることが明らかになった。「話すこと」や「聞くこと」について評価するとき、その技能ではなく、生徒の意欲や参加態度を評価していた。（→教師がその技能を評価するための訓練を受けていなかった）
- その一方で、ある技能の評価に関する構成概念にも評価者によって違いがあり、統一を求めることは現実的ではない（フィラー、1993）  
例）書くことの評価の構成概念→A：想像的なこと／B：構成や構造
- だがそれでも、グループ・モデレーションの過程で各教師が構成概念について考え、何を重視すべきかを議論することは実りのある活動であるといえる。

## 結論

### 「教師の評価」の有用性

- ✓ 学習指導や生徒の学習向上が期待でき、妥当性が高い。
- ✓ 様々な学習活動を用いて多角的に広い範囲の技能や過程をカバーすることができる。

### ただし注意が必要

- ✓ 教師の評価が何らかの形で教室の外部に報告される場合、評価の統一性を保証することが必要となる。

### そのためのグループ・モデレーション

- ✓ 教員間で学習の達成事項などについて討議することで教師の専門的な技術を向上させる可能性を秘めている。
- ✓ 作者の提案：「評価の過程と評価に用いる課題に焦点を当てることによって評価の統一を図ることは、評価結果の統一にも寄与する」