

### 1) 教育評価とは

#### (1) はじめに

近年、大学生の学力の低下、およびあるレベルの学力をもつようにして卒業させることの必要性から、厳格な成績評価が強調されている。北海道大学では、同じ科目なのに担当教官によって成績にバラツキがあるという学生の指摘を受けて、1997年に成績評価について教員・学生の両方に全学的アンケート調査を行った。成績評価は、単に試験成績のみをあらわすことではなく、教育の総体を表すというものであり、教育を総合的に理解している必要があるということ認識するためであった。

大学審議会答申(1998年)や大学評価機関の評価内容(2000年)では「厳格な成績評価」があげられている。ここでも単に成績評価を厳しくすることよりは、入学してきた学生にどれだけの学力をつけさせ卒業させるかという教育の質を問うこと、教員の教育の質、責任を評価なことに受け取った方がよい。卒業時に求められる学力を正しく評価し、そのような学力をつけたと云えるようになることである。

#### (2) 評価の意味

成績評価は教育評価の一部である。教育評価とは「試験をすること」と一般には考えられがちであり、しかも試験は60点以上で合格と簡単に処理されがちである。しかし教育評価は教育のなかでもっとも難しいものの1つである。

評価とは教授学習の過程における「成果」の判断である。すなわち評価とは、教育活動が効果的に行われたかどうかの判断を行う作業である。言い換えるなら、評価とは単に試験によって学習者を測定することではなく、測定の方法や測定結果についての価値判断を行い、その判断に基づいて、何らかの意志決定を行う一連の作業である。しかも得られた意志決定は、次の教育活動へフィードバックされるという継続的なものでなければならぬ。

現在の大学教育において「評価」がとくに問題とされる点に次のようなことが挙げられる。

- ① 評価の原則を教員が十分理解していない。
- ② 明確に設定されていない教授目標(一般目標、到達目標)を評価しようとしている。
- ③ 知識の領域、とくに想起または解釈レベルのみを評価し、問題解決能力、技能、態度の評価が行われていない。
- ④ 総括的评价が重視され、フィードバックを目的とする形成的評価が不十分である。ただ一回の進級試験のみで最終的結論を出すことは妥当でない。また中間テストを行っても最終成績の一部に加えられる場合は、純然たる形成的評価とはいえない。形成的評価とは教授・学習へフィードバックすることを目的とする評価で、学生がどこに強く、どこに弱点をもっているかを詳細に知るための評価である。それは教授法の改善にも関与し、教員の反省にもつながる。
- ⑤ 評価は、一般には学生が学習の結果として目標のどの程度まで到達したかを試験し、学業成績を求めることにある。したがって、目標、特に到達目標の表現は、試験の評価基準をあらわすことになる。評価方法には論述試験、客観試験、口頭試験、実地試験、レポートなどがあるが、それぞれの試験の有する長所、欠点および測定可能な範囲を教員がよく知らないことが多い。
- ⑥ 行われたテストの質の管理は全くなされてない。出題者は自分の出した問題が、適切であったかどうか反省する必要がある。

一方、評価は学生の学業成績を求めることのみではない。教員の教授法も評価される(学生による授業評価)。また、目標の立て方、学習方略、評価の方法もまた評価も対象となる。以上のように評価は総合的なものであるが、以下では主に学業成績評価について述べる。

#### (3) 学業成績評価

教育評価では、学生(学習者)のみが評価の対象ではない。教員やカリキュラム、授業も評価の対象になる。また評価される目標も特定されなければならない。また、試験にも様々な方法がある。教員は評価の原則や成績評価の基準について理解している必要がある。

#### (4) 評価の基準化

各科目の成績評価は基準化される必要がある。最近、成績の標記に米国で一般に採用されているGPAの制

度を採用することも推奨されている。この制度は、学業成績の表記として国際基準へ発展させることができ、平成 10 年 10 月大学審議会答申でも触れられている。

GPA (grade point average 平均成績係数) は、各科目の成績を以下のように 0 から 4 まで分布させる。

試験の成績	科目成績	GP (grade point 成績評点)
90 ~ 100	A	4
80 ~ 89	B	3
70 ~ 79	C	2
60 ~ 69	D	1
60 以下	E	0

試験の成績は、種々の評価法による総点であり、合格水準、配点は標準化されたものである。また、履修科目の総体的評価すなわち総合評価の標記は以下のように単位あたりの GP で表わされる。

$$GPA = (A \text{ 科目 } GP \times A \text{ 科目単位数} + B \text{ 科目 } GP \times B \text{ 科目単位数} + \dots) \div \text{登録科目総単位数}$$

成績標記は、各学期の GPA として一定期間 (学年, 学期, 総修学年度) にわたって表記され、様々な場面で利用される。たとえば、1 に満たなければ退学勧告、3 以上であれば奨学金、さらに優秀であれば受賞、交換留学派遣などである。また単位互換での成績表現、他大学へ転学するための指標にも利用される。

#### 参考文献

Downie, N. M. (1967) Fundamentals of measurement: Techniques and practices., New York, Oxford University press

## 2) 教育評価の原則

(1) だれを・なにを評価するか (whom, what)

①学習者 (学生) ②教育者 (教員) ③カリキュラム (目標, 方略, 評価)

(2) どの行動領域を (where)

①認知領域 ②精神運動領域 ③情意領域

(3) いつ評価するか (when)

①学習前 (プレテスト) ②学習中 (中間テスト) ③学習終了後 (ポストテスト) ④フォローアップ・テスト

(4) いかなる目的で評価するか (why)

①形成的評価 ②総括的评价

(5) だれが評価するか (who)

①教員 ②学習者 ③教育専門家 ④その他

(6) いかに評価するか (how)

①論述試験 ②口頭試験 ③客観試験 ④実地試験 ⑤観察試験 ⑥レポート

## 3) 評価の持つべき性格—よい測定のための条件—

(1) 妥当性 Validity

用いられる評価方法が、測定しようとする学習者の行動特性を本当に測定しているかどうかの正確さの程度をいう。

重さをはかるのに ”物差し” を使ったのでは妥当性がないのと同様である。

テストの妥当性をみる一方法として問題項目分析がある。試験問題の 1 問ごとに難易度 (正答率) と識別力を見る方法である。

(2) 信頼性 Reliability

評価の信頼性とは、結果の安全性を意味しており、同じ集団に同じテストを何回行っても同じ結果が得られる性質、いいかえれば再現性のことである。

重さをはかる ”はかり” が検定済みであるかどうかということである。

(3) 客観性 Objectivity

同じ領域の専門家の中で、問題の正解に関して意見が一致している度合をいう。誰がはかっても一定の長

さが計測されること。

(4) 効率性 Effectiveness

評価の容易さ、経済的にも時間的にも実用的であることをいう。ただしいかに効率が良くても、妥当性や信頼性が損なわれるようでは良い評価とはいえない。

(5) 特異性 Speciality

なぜそう解答されたかが分かるような評価の性格で、形成的評価では大切なことである。

Guillbert, J. J.: Educational Handbook. pp311-312, WHO, 1976

#### 4) 形成的評価と総括的評価

プロダクトの評価ではその目的から次の2種類が大別される。

(1) 形成的評価 (Formative evaluation)

科目の目標を習得しているか否か、もし習得していないなら、それを習得するのに何をしなければならないかを判定するための評価をいう。すなわち、学習形成過程の改善を目的とする評価である。したがって、科目終了時の試験とは異なり成績評価（最終判定の資料）とはしないのが原則である。その結果は、学習者が学習を矯正するため、また教師が教授方法を矯正し、学習指導の指針を得るためのフィードバック資料となる。

(2) 総括的評価 (Summative evaluation)

達成された学習成果の程度を総括的に把握する評価で、学科やコースを終了した時期に行われる。従来わが国の大学教育で行われてきた試験の大部分はこれに当たる。

大学教育においてはもちろん総括的評価は重要であるが、近年、形成的評価の教育における価値がきわめて重要視されるようになってきた。しかし、まだ日本の大学教育の現状では真の形成評価はむずかしい。たとえば普段の自習時間はきわめて少なく、成績に結びつかない試験は勉強しないという傾向がある。中間試験をくりかえし、これも成績に入れることも行わざるを得ない。

総括的評価のみで学科の教授目標に到達したか否かを判定しようとする場合には、これに合格したものの中

表 形成的評価と総括的評価の特徴

	形成的評価	総括的評価
1. 機能	以後の学習の手引となる、学習者一人一人の強点と弱点の詳細なデータを提供すること	学習者が満足すべき能力レベルに全体的に到達したかどうかを決定すること
2. 時期	各ユニットの途中と終了時	進級、卒業、免許取得についての重大な決定を行う時
3. 範囲	1ユニット又はユニットの一部分の内容	全カリキュラムまたは大きなカリキュラムブロック
4. 内容とサンプリング	ユニットに含まれる新しい内容要素または技能の一連の練習問題、内容要素の正確な認識からそれらの正しい適用までの、理解の各レベルにおける質問を含む練習問題群	上記の全領域から抽出した問題
5. 採点	内容ごとの誤り、理解や技能のレベルごとの誤りのパターンを指摘する多角的な採点	単一な全体的採点
6. 評点の報告	勧告のみ。学習者の正式記録には絶対に含めるべきではない。	学習者の永久記録の正式な一部として記録する。
7. 学習者へのフィードバック	何を学んだか、また何を学ばねばならないかを知るための助けとなり、学習の難点の原因を見つけるのに役立つように、学習者一人一人に、その正解と誤りを、直ちに詳細にフィードバックする	合否
8. 教員へのフィードバック	少なからぬ学習者たちが間違えたことを同定できる。詳細な報告を直ちに行う。	全体としての、また各学科や技能ごとの合否学習者数

医学振興財団、医学教育ワークショップ資料

に、たとえば「一夜漬け」で学習したものもいるかもしれない。

このような学生その後の忘却のスピードはきわめて速い可能性が大きい。これに対して各授業の区切りごとに形成的評価からのフィードバックを受けて、効果的な学習をし、前述の学生と同じレベルの合格点を総括的評価でとった学生の忘却曲線はおそらくなだらかであろう。

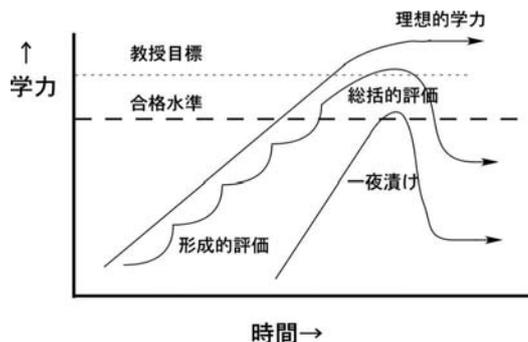


表 評価方法による忘却曲線の違い

### 5) 評価方法と評価可能な行動

それぞれの評価方法には、利点と欠点がある。

◎**論述試験**

知識（想起）、理解（解釈）、問題解決能力

◎**口頭試験**

知識（想起）、理解（解釈）、問題解決能力

◎**客観試験**（国家試験、統一試験などにみられる様式）

主として知識の想起 および単純な解釈、問題解決能力

◎**シミュレーションテスト**

問題解決能力

◎**実地試験**

問題解決能力、技能、（態度：試験のときだけよく見せることに注意）

◎**観察記録法**

態度・習慣 および技能の評価によい。

チェックリスト (check list)：多くの項目をあげると公正な評価となる。

評定尺度 (rating scale)：1～4点法、1～5点法がある。

逸話記録：特記すべきことを記録

◎**論文または研究課題（レポート）**

丸写しがあり、可否判定などには不向き

形成評価にもちいるのに有意義

◎**ポートフォリオ**

振り返りによる自分の学習履歴の確認、今後の学習計画の資料

学習活動の全体を評価

表 評価方法と分類目標との関係

方 法	知識			態度・習慣	技術	測定範囲
	想起	解釈	問題解決			
論述試験（ペーパーテスト）	3	4	5			狭
口頭試験	3	4	5			狭
客観試験	5	5	3			広
Simulation test						
筆記型	3	4	5			狭
模擬患者・模擬来談者モデル	1	2	3	5	5	狭
コンピューター	5	5	5			中
実地試験	1	2	3	4	5	狭
観察記録	1	2	3	5	5	中
レポート	3	4	5	4		狭
ポートフォリオ	3	4	5	5	4	広

grading: 1～5

## 6) レポートの評価基準の例— ポートランド州立大学の評価ガイドラインより —

### (1) 「フレッシュマンセミナー」のジャーナル（授業報告）

#### 評価Aの条件

①積極的参加について： 定期的なきちんと出席して記載する。週3回あるいはそれ以上の出席がこレベルに該当する。記載内容は、その筆者が指定参考書を読んできたこと、ピア・メンター指導の授業で活動したこと、知的に前向きな姿勢でエッセイや宿題に取り組んだことを示すものであること。記載の長さは場合によるが、自分自身で考えたことを反映し表現するために最低1ページが必要であろう。

②積極的姿勢について： 記載内容は、筆者が自ら進んで仮説をたて、疑問を持っていることを示すものでなければならない。例えば、自分が読んできた本の中の重要な文や授業の中で得られた知識などから始めても良いが、さらに進んでそれらを総合して記載しなければならない。それ以外に、授業で使う資料や他の授業の資料、または著者自身の実際の生活と結びつけるのも良い。

③使った時間： 記載内容は、筆者がきちんと前の記載内容を読み直し、コメントし、考え直し、筋道を見いだしたことを示すものでなければならない。時間の経過とともに、この「ジャーナル」は、この授業の中心にあって常に訴えつづけた一連の疑問、問題点、争点を明らかにすることになる。

これに対し、最低ランクのB評価は、例えば(1)については、「平均して週2.5回しか出席せず、授業への積極的参加が短く、断片的な場合」、また(3)については、「以前の記載内容をほとんど読み直したことがなく、理解が深まり進歩した形跡が認められない場合」と定義されている。

### (2) 大学院レベルのセミナーのレポート

評価A：セミナーにおけるすべての要素が網羅されていること。文章は明快で、文法的に正しく、「言葉の遊び」になっていないこと。最低2件の文献を引用してセミナーで話されたポイントに対してコメントし、それを発展させていること。

評価B：レポートに含まれている要素と文章の質はAと同等である。ただし、議論は知的ではあるが、十分な思考が反映されておらず、セミナーでのべられたことの繰り返しのみに過ぎない。2件の文献を読んではいるが、それについての議論がどちらかといえば表面的。

評価C：要素の一つが欠けている。文章はそこそこ明快であるが、雑であったり、仕上げが十分でない。文献は読んでいるが、セミナーの話題に知的に関連づけられていない。

評価D：要素のいくつかが欠けている。文体は不注意で、タイプミスあるいはスペルミスがある。読んだ文献がセミナーで述べられた問題と関連づけられていない。

評価E：このレベルは無し。

評価F：きわめて表面的で、レポートの文体および内容についてのガイドラインを無視している。文献は読まず、一見して不注意なレポート

## 7) 発表の観察評価

年月日 \_\_\_\_\_

授業名など \_\_\_\_\_

発表タイトル \_\_\_\_\_

発表者、発表グループなど \_\_\_\_\_

	満足度				
	高い		低い		
1. 発表テーマの目標が明確に示されたか？	5	4	3	2	1
2. 発表内容は量的に定められた発表時間に適切であったか？	5	4	3	2	1
3. 発表の全体構成は適切であったか？	5	4	3	2	1
4. 内容に関して十分な知識と理解を示したか？	5	4	3	2	1
5. 理路整然と発表されたか？	5	4	3	2	1
6. 発表者が、その内容や発表に熱意をもっているように見えたか？	5	4	3	2	1
7. 発表の話し方・態度は聴衆にとって適切であったか？	5	4	3	2	1
8. 発表の補助手段は適切に使用されていたか？	5	4	3	2	1
9. 聴衆の関心が高められたか？	5	4	3	2	1

平均点 (小数点以下3位まで) \_\_\_\_\_

\* その他の意見

評価者 \_\_\_\_\_

注：この尺度評価では評価者の満足度で評点します。主観による評価ですが、評価項目は、客観性のある観察可能な内容です。評価者の見識で評価することになります。なお、複数名で評価することが勧められます。

## 8) ポートフォリオ

ポートフォリオ (portfolio) は「書類かばん」「紙ばさみ」などを意味し、建築家、写真家、デザイナー、イラストレーターなどが自分の技術や技能、成し遂げた仕事やその成長の軌跡を、顧客や雇い主に見せるために持ち歩いた作品集のことをさす。教育の場では、「学生の努力・向上・学力を示す成果を集めたもの」(Paulsonら, 1991)「反省的な語り (reflective narrative) をともなう成果物を集めたもの」(Porter & Cleland, 1995)などと定義されている。目的や制作期間の長さなどによって、多種多様なポートフォリオが使われている。

### 事例：クラス・ポートフォリオ

(カーネギーメロン大学 2003 年春学期に開講された、“Rapid Design through virtual and physical prototyping”のシラバスから抜粋・改変した)

#### (1) 含めるもの

この授業を受講して身につけた知識やスキル示すために下記のものを含める。

- ①書いたもの (written documents)
- ②スケッチ
- ③CAD を用いて描いた図面
- ④作品の写真
- ⑤自分で制作したもの

#### (2) ポートフォリオを作成する意義

- ①各自がこの授業の中で何を学んだのかを理解し評価するために役立つ
- ②他者にこの授業で学んだことを示すために役立つ

#### (3) 内容

“Rapid design”に関して、下記の項目 (目標) の到達状況を記述する。

- ①デザインの過程——概念, 実行
- ②コミュニケーション—文書, 発表, 議論とポスター
- ③問題解決
- ④チームワーク
- ⑤個人及びチームの評価
- ⑥ Rapid prototyping の技法についての知識

#### (4) 留意点

最初は宿題をノートやレポート用紙に書き、ファイルに保存するだけでかまわない。しかし、宿題は自分が学んだことではなく、行なったことを示している。自分が学んだことをどのように記録し文書化したらよいかをよく考えること。一つの方法はエッセーを書くこと。もう一つの方法は自分の考えがどのように変化し進化したのかを示すこと。また、自分の最終課題 (プロジェクト) のアイデアがどのようにまとまったのか、プロトタイプをつくりあげる過程で学んだこと、友人から学んだことをまとめることもよいかもしれない。

#### (5) 評価

以下の各側面について5段階尺度を用いて評定する。

- ①全体の構成
- ②授業目標の到達度
- ③創造性

- ④自己評価の根拠
- ⑤表現力

参考文献

Paulson, F. L. Paulson, P. R., & Meyer, C.A. (1991) What makes a portfolio a portfolio? Educational Leadership, 48, 60-63.

Porter, C. & Cleland, J. (1995) The portfolio as a learning strategy, Portsmouth:Boynton/Cook Publishers.

### 9) ルーブリックの活用

ルーブリックとは、評価する項目とその達成度をレベル別に仕分けした基準が記載された評価指標表のことである。採点がしやすい客観試験ではなく、レポートなどの採点で重宝する。ルーブリックを作成して見ると、評価の妥当性の善し悪しが見えてくるし、信頼性・客観性が上がる。詳細なルーブリックを作成し、TA と採点作業を分担することで、効率性を上げることが出来る。

表 レポートのためのルーブリックの例

	誤字・脱字	文字数	独自性・新規性	論理性・構成力
1点	4ヶ所以上みられる	3500-6500に収まらない	引用元の紹介にとどまっている	論理性にかけ、思いつきの構成である
2点	3ヶ所以内である	4000-6000に収まらない	批判・感想にとどまっている	部分的に飛躍や解りにくい構成がある
3点	見当たらない	4500-5500に収まっている	自分の意見・発想が述べられている	飛躍なく順序立てられている

### 10) 学生間ピアレビュー

ピアレビューは、学生の学習成果を学生同士で評価する方法である。教員は、教えることでさらに自分の理解が深まることを知っているが、学生間ピアレビューはこれと似た効果を生む。他の学生の評価をすることで、学生は自分自身の提出物について客観的に評価・反省することになるはずである。

しかし、学生にそもそも評価をさせてよいのか、その能力があるのかという問題がある。そのため、学生間ピアレビューを行う場合、まず学生にルーブリックを徹底的に周知させることが原則となる。