

2.5 学習評価

(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を確認するためであった。

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。言い換えるならば、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ない。

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。

評価の原則を教員が十分理解していない。

明確に設定されていない教授目標(一般目標、行動目標)を評価しようとしている。

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。

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评价が不十分である。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、純然たる形成的评价とはいえない。形成的评价とは教授・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。

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、特に行動目標の表現は、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。

行われたテストの質の管理は全くなされていない。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) 学生成績評価

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

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。

資料 2.15 教育評価の原則

ここでは教育を評価する場合の原則を整理している。

(1) だれを・なにを評価するか 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)	whom	what
(2) どの行動領域を評価するか 1. 認知領域 2. 精神運動領域 3. 情意領域	where	
(3) いつ評価するか 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	when	
(4) いかなる目的で評価するか 1. 形成的評価 2. 総括的評価	why	
(5) だれが評価するか 1. 教員 2. 学習者 3. 教育専門家 4. その他	who	
(6) いかに評価するか 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. 論文(レポート)	how	

資料 2.16 評価の持つべき性格
—よい測定のための条件—

成績評価では、たとえば、ある科目を数名の教員で担当するとき、同じシラバスで授業を展開し、評価は同じ基準で行う必要がある。そこでは以下のことが求められる。

妥当性

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。

信頼性

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。

客観性

同じ領域の専門家の間で、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。

効率性

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。

特異性

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。

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

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ぬ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。

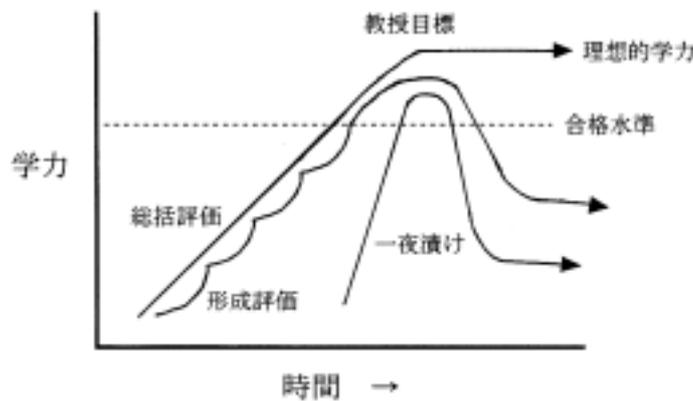
大学教育においてはもちろん総括的評価は重要で

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しろ少ない。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。

このような学生の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なであろう。

資料 2.17 評価方法による忘却曲線の違い



資料 2.18 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ことの助けとなり，学習の難点の原因をみ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

資料 2.19 評価方法と評価可能な行動

試験の方法と判定可能な目標

試験には様々な方法がある。またこれらは判定可能な目標とそうでない目標がある。

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, 合否判定などには不向き

形成評価にもちいるのに有意義

資料 2.20 評価方法と分類目標との関係

それぞれの評価方法が測れる目標の範囲を示している。

方 法	知識			技術 態度		測定範囲
	想起	解釈	問題 解決			
論述試験 (ペーパーテスト)	I	II	III			I
口頭試験	I	II	III			I
客観試験	III	III	I			III
Simulation test						
筆記型	I	II	III			I
模擬患者・模擬来談者	-	=	I	III	III	I
モデル				III		I
コンピューター	III	III	III			II
実地試験	-	=	I	III	II	I
観察記録	-	=	I	III	III	II
レポート	I	II	III		II	I

grading: - < = < I < II < III

狭い I → 広い III

(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)
100 点の内		
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。

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

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

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