

教育評価を教育する —教育統計学基礎における実践—

太 田 伸 幸

1 はじめに—教育評価とは

教育評価とは、教材の選択、授業の進め方、自己学習指針、進路選択、入学者選抜など、教育においてなされる様々な判断や決定のために、生徒や教師や諸々の教育環境に関する情報を収集して利用する活動を表す。梶田(2008)が、『教育活動の中で、どのような学びがなされたのか、どのような育ちが実現したのかの確かめであり、その結果の教育的な活用である』と述べているように、教育評価とは、単なる学習結果の確認のみならず、その効果的な利用まで含意されている概念である。

そのため、教育評価の訳語として、“evaluation”と“assessment”の2つが用いられることが多い。

“evaluation”とは、教育実践の目標に照らして達成度を価値判断する行為を表し、“assessment”とは、多角的な視点から、多様な評価方法によって評価資料を収集することを表す。

“evaluation”と“assessment”は、どちらも評価のための情報を集めるという点では共通している。しかし、その評価情報を基にしてなされる意志決定は多岐にわたっており、“evaluation”で表されるような学習者を対象とした価値判断は教育評価の一側面に過ぎない。教育評価は単なる「学力による序列づけ」でなく、「教育のための評価」なのである。

これまで教育評価を受けてきた学生にとって、“evaluation”的な側面—良い評価を得る—toに注目するのは想像に難くない。しかし、将来、教育評価を行う側に立つ身として“assessment”的な側面も考慮できるよう、教育する必要があると考えられる。

2 現代教育学部「教育統計学基礎」

筆者が担当する教育統計学基礎は、1年春学期に開講される学部共通科目である。カリキュラム上は選択科目であるが、選択履修するのは幼児教育学科の学生のみで、児童教育学科の学生に対しては指定科目とし、原則全員が履修することとなっている。講義は両学科混成で学籍番号順に2クラスに分割し、1クラスあたりの人数は50人程度となる。また、シラバスに掲載した講義の趣旨は「多様な評価方法の概要を理解し、目的に応じた評価方法の選択および評価情報の解釈に関

する基礎的な知識を習得する。教育場面において用いられる評価方法を概観し、教育評価に関する理解を深める。特に評価方法の多様性や評価の目標と手段の関係について事例を紹介しながら解説し、教育目標に対応した評価方法の選択や評価結果の解釈のあり方に対する態度を形成することを目的とする」である。

学部としては文系学部とされているが、入試科目で数学・理科を選択可能にしたこともあり、理系から進学してくる学生も少なくない。したがって学生の高校数学の学習履歴は様々で、数Ⅰしか履修していない学生から数Ⅲ・数Ⅲまで履修している学生まで混在している。そのため、数学的な内容を中心に統計学の講義を展開すると理解度のばらつきが大きくなる恐れがあることより、教育評価の内容・方法に重点をおいたカリキュラムを構成した(表1参照)。講義のカリキュラムはおおよそ以下の4つのパートに区分される。

- 1) 評価の意義と種類について解説(第1回～第5回)
- 2) テスト作成の方法と実施(第6回～第8回)
- 3) テスト結果の整理の方法(第9回～第12回)
- 4) それまでの内容を振り返りながら学校における評価の活用・記録についての解説(第13回・第14回)

このうち、第6回、第7回、第11回、第12回の4回でノートPCを用いた実習を行った。また、本講義では第8回に中間試験、第15回に最終試験を実施し、この試験問題は学生が作成した問題を基に構成することと、学生が作成した問題はWebにて公開することを、シラバスおよび初回講義で指示した。

表1 教育統計学基礎のカリキュラム

第1回	オリエンテーション/教育評価とは
第2回	到達度評価と相対評価
第3回	教育目標と教育評価の関係
第4回	学力評価の様々な方法(1)学力評価の方法の分類, 学力テスト
第5回	学力評価の様々な方法(2)ポートフォリオ, ルーブリック他
第6回	テスト作成の方法(1)問題の種類と特徴
第7回	テスト作成の方法(2)問題作成実習
第8回	テストの実施と採点(中間試験)
第9回	テスト結果の整理の方法
第10回	テスト結果の解釈の方法
第11回	テスト結果の分析(1)
第12回	テスト結果の分析(2)
第13回	指導に生かす評価のあり方
第14回	指導要録と通知表
第15回	最終試験

3 Blackboardの利用

3.1 Blackboardとは

本学では、学術情報センターにおいて教育用CMSとしてBlackboard（図1、図2参照）を導入し、講義における利用を呼びかけている。このBlackboardの主な機能として以下のものがある。

- ・シラバス提示
- ・教材提示
- ・講義ノート作成・提示
- ・課題提示（出題、回収、成績管理）
- ・アセスメント（問題作成、自動採点、成績管理）
- ・アンケート実施（質問作成、集計）
- ・コミュニケーションツール（ディスカッション、メール、チャット）
- ・学生管理データベース（成績管理、アクセス履歴）

これらの機能のうち、いくつかの機能（Webメール、掲示板、教材配布、課題提出）はTora-netにも実装されている。しかし、アセスメント機能や成績管理機能など、講義運営上有用な機能がBlackboardには提供されている。ログインもTora-netと共通のID、パスワードを使用して行うため、専用のアカウントを準備する必要はない。

教員は用意された機能のうち、自身の講義に必要とするものを組み合わせて活用することになる。教材提示やアセスメント実施などでは、Blackboardに教材作成支援機能が備わっているため、比較的容易にこれらの教材を作成することが可能となっている。1つのコース内に資料提示、問題演習、コミュニケーションツールが統合されており、学生は特定のサイトにアクセスすることで、講義時間内の学習支援だけでなく、講義時間外での自主学習においても必要な情報交換を行うことができる。

また、サイトの管理の面からみると、個人ごとのトラッキング情報や、テスト、課題の提出状況や採点結果をBlackboard上で管理・公開することができる。これらの情報は選択的に学生に公開することができるため、個別に成績を通知したり、各自で自身の学習履歴を確認したりといった利用も可能である。特に課題に関しては、eメールによる提出では提出者の確認やファイルの整理などに手間がかかっていたが、Blackboard内で一括して管理ができ、ファイル形式は限定されるがブラウザ上でファイルの内容の確認も可能である。電子ファイルでのレポート提出を求める場合、成績管理まで含めてかなりコストを低減させることが可能となる。



図1 Blackboard「教育統計学基礎」トップページ

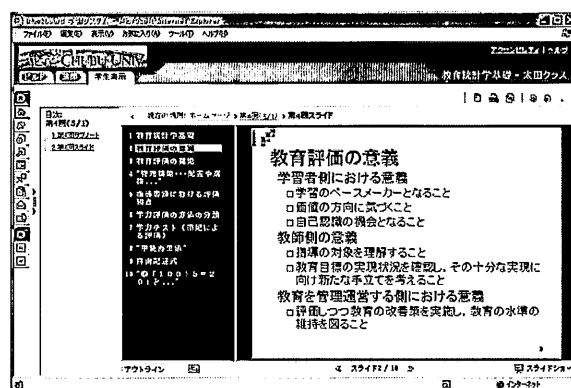


図2 PowerPointファイル表示画面

3.2 利用開始時期

春学期開始時にコンテンツ作成が間に合わなかったため、Blackboardを利用する回に間に合うように受講学生のユーザ登録を行い、5月下旬より公開した。全員がBlackboardの利用経験のない学生であったため、練習用のコンテンツを作成し、第6回の講義時に、本講義で利用を予定しているツールのチュートリアルを行った。

3.3 講義の進め方

プレゼンテーションソフトとしてMicrosoft PowerPoint2007（以下PPT）を使用した。1單元あたり6～14枚の提示用スライドを作成した。提示用スライドはWeb用のプレゼンテーションファイルに変換し、Blackboard利用開始時点で、それまでの回のファイルと講義用のサブノートのPDFファイルをBlackboardにアップロードした。以後、各回の授業後に該当回のファイルをアップロードした。講義はほぼ全てスライドを使用して進行するため、学生が復習用教材として利用することをねらいとした。

また、座学の合間に学習内容に関連した課題を実施し、学生どうしの意見交換や以後の学習事項の教材として利用した。

3.4 ディスカッション機能の利用

講義で主に活用する機能は教材提示とディスカッションであった。特にディスカッションは実際に問題投稿を行うため、講義内のチュートリアルでディスカッションへの投稿を経験させ、自分の入力した内容がBlackboardに掲載されること、および自分の投稿内容に対して他の学生から返信されることを理解させた。

実際の問題投稿は第7回の講義時間内に行った(図3参照)。各自試験範囲より問題作成および投稿を行った。途中より他の学生の投稿した問題に解答を投稿するよう指示を出した(図4参照)。なお、中間試験では講義時に問題作成を行ったが、最終試験は講義時間外での問題作成を指示した。

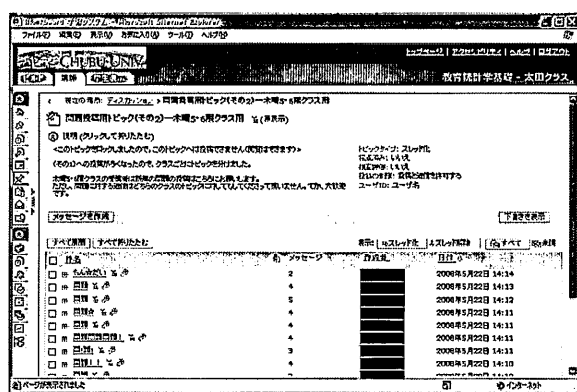


図3 ディスカッション投稿一覧画面

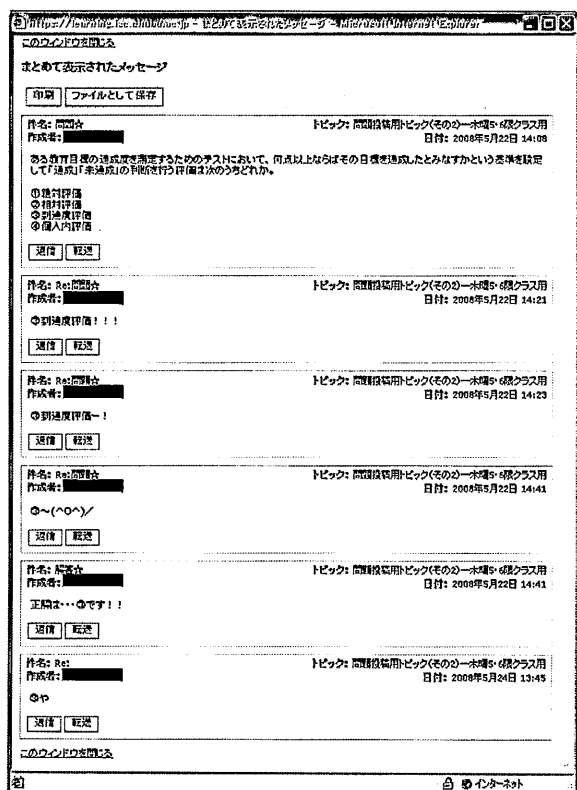


図4 ディスカッションに投稿された問題と解答

3.5 ディスカッションの利用状況

ディスカッショントピックは中間試験用と最終試験用に分けて作成した。クラスごとのトピックや共通投稿用のトピックも作成したため、中間試験用のトピックは3つ、最終試験用のトピックは4つとなった。投稿されたメッセージは、中間試験用トピックに対しては816件、最終試験用トピックに対しては1664件であった。このうち中間試験用トピックに対しては206件、最終試験用トピックに対しては201件が問題の投稿であった。残りの投稿は、それぞれの問題に対する他の学生からの解答、出題学生によるフィードバック、担当教員からの注意メッセージ(解答方法の補足、解答の間違いの指摘等)であった。

問題投稿数は平均4.51問であったが、標準偏差が3.61と大きく、これは投稿数のばらつきが多かったことを示す。事実、1問も投稿しない学生から23問投稿する学生まで存在していた。また、中間試験前と比較して最終試験前の投稿数が2倍以上となっているが、これはBlackboardの利用状況も最終評価に含めることを意識したことと、最終試験が評価の50%を占めるため、試験対策としてBlackboardを利用したことによるものと考えられる。

4 評価結果と教育評価に対する考え方の関連

4.1 最終評価結果

履修登録者は2クラス併せて92名であったが、最終試験を受験したのは87名であった。この87名の評価分布は表2の通りとなった。R評価は合計得点が57点以上の学生を対象とし、57点未満の学生はすべてE評価とした。評価の高い学生ほど、問題投稿数が増える傾向がうかがえる。

表2 最終評価および問題投稿数

評価	人数	平均問題投稿数	標準偏差
S	2	6.84	5.37
A	17	5.72	3.91
B	29	3.04	2.22
C	25	1.43	1.18
R	4		
E	10※		

※欠席者5名を含まない

4.2 授業評価アンケートの実施

最終試験終了後、授業評価アンケートを実施した。調査用紙は講義で使用した講義ツールに関する項目(28項目、5件法)、講義への取り組みに関する項目(10項目、5件法)、講義内容・講義ツールに関する自由記述(3項目)、そして、講義受講前と受講後の「教育評価」に対する考え方(自由記述、各1項目)により構成されていた。

表3 「教育評価」に関するイメージの記述に関する主題分析結果（受講前）

分類名	文書数	信頼度	代表語句	類似度	記述例
付け方	2	100	付け方/付け/成績	1.000	成績の付け方なんて考えたこともなかった
自分	3	75	自分/基準/興味/判断/成績	0.956	評価は常に自分が基準であると思っていた
通知表	3	100	通知表/表/通知/言葉/まあ感覚	0.920	まあ感覚で通知表や課題点などつけるだけでした
何	3	60	何/not 理解/考え方/基準/考え	0.920	複雑だと思っていた。何をするか分からなかった
授業	4	50	授業/態度/生活/点数/テスト	0.869	評価するのは点数や得点率などだけであるものだと 思っていた
教師	9	89	教師/主観/独断/好み/生徒	0.884	評価は教師個人の考えで全てつけていると思っていた
教育	9	89	教育/教育評価/成績/関心/テスト	0.872	教育評価というものを全く知らなかった
成績	14	38	成績/相対評価/相対/方法/疑問	0.884	評価の仕方について具体的に考えたことがなかった
イメージ	25	7	イメージ/数字/主観的/意味/抽象	0.836	簡単にできると思っていた

表4 「教育評価」に関するイメージの記述に関する主題分析結果（受講後）

分類名	文書数	信頼度	代表語句	類似度	記述例
知識	3	44	知識/基礎知識/基礎/身/受講	0.883	基礎知識を覚えることができた
付け方	3	75	付け方/付け/点数/担任/先生方	0.932	生徒の成績の付け方を知り、担任を持ってくださった先生方のすごさを知った
奥	4	75	奥/通知表/表/通知/こと	0.858	テストや通知表や成績の出し方など、いろいろと奥が深いと思いました。
自分	5	83	自分/判断/多角的/多角/何	0.910	評価は自分で見て判断するだけでなく、多角的に見ることも必要だと知った
用語	5	60	用語/パソコン/基準/編集/必要性	0.827	ちゃんとした基準が設定されているのだと知った
仕方	6	66	仕方/評価一つ/一つ/大分/子供	0.919	評価の仕方を学んで、テストを作る大変さを知った
テスト	7	44	テスト/苦勞/教師/ペーパー/ペーパーテスト	0.855	パフォーマンスとか、ペーパーテスト、たくさんの種類があったのについて分かって良かった
数字	16	6	数字/規則/説明/前/関数	0.879	いろいろな規則等によって決められている
評価	30	73	評価/種類/教育評価/教育/方法	0.823	具体的に評価の種類や目的等が明確になった

調査用紙には学籍番号記入欄を設けてあり、学生に対して、試験得点を含む評価情報と対応させるために学籍番号が必要であることを説明し、対応させた後は学籍番号情報を削除し個人特定はしないことを約束した上で記入を求めた。不安が残る学生には未記入を認めたため、最終試験受験者87名中6名に学籍番号の記入がなかった。したがって、以降の分析は、最終成績と対応させることが可能であった81名を対象とする。

4.3 「教育評価」に関する自由記述の分析

授業評価アンケートの内容のうち、「教育評価」に対する受講前・受講後の考えに関する自由記述を分析対象とした。また、自由記述内容を比較する属性として最終評価を用いた。ただし、SおよびR評価者は少ないため、区分はS/A、B、C、R/Eの4区分とした。

分析にはジャストシステム社製TRUSTIA /R.2を使用した。まず、記述のあった文章を全て入力した上で、同意語（「教師」と「先生」等）とストップワード（「ワード」等）を設定した。続いて受講前、受講後のそれぞれの記述について主題分析を行った。

まず、受講前の教育評価に対する考えについての自由記述の分析（表3）では、72の記述が9つに分類された。最も頻度が高い「イメージ」は、「評価」に対する漠然としたイメージの記述であった。他にも「付け方」「何」「成績」など、教育評価に関する知識の乏しさや、「自分」「教師」など、評価が単純な基準で行われるという考え方のように、評価に対するとらえ方の具体性が認められない記述が大半を占めていた。た

だし、唯一、他の記述と比較して具体的に評価の内容を記述していた「授業」に関しては、最終評価がB以上の学生の記述のみで構成されていた。

対して、受講後の記述の分析（表4）では、79の記述が9つに分類され、記述内容にも大きな変化がみられた。最も頻度が高い「評価」は評価の目的や種類に関する知識の構築を表す分類であり、漠然としたイメージが中心だった講義前の記述とは大きく異なる。また、「教師の主観」や「テスト」だけで成績が付けられているといった考えから、「付け方」「自分」「仕方」などの記述にみられるように、評価の多面性にも気付くことができたようである。さらに、分類ごとの成績評価の分布も偏りが認められなかった。これは、意識としては促すことはできているが知識として定着しているとは言い難いにとらえるべきであろう。

5 おわりに

本講義の実施により、教育評価に対する考え方に広がりを持たせることができた。しかし、評価活動への参加は部分的であるため、さらに多面的な評価活動を展開できるような活動を取り入れていくことが今後の課題としてあげられるであろう。

引用文献

梶田敦一 2008 教育評価入門 協同出版

〔 准教授 現代教育学部 児童教育学科 〕
〔 教育実習センター 〕