

## 35 Fifteen Ways to Improve Classroom Assessment

James Dean Brown and Jonathan Trace



## 1. 本発表の構成

- I 著者紹介 (2)
- II 論文の構成 (3)
- III 論文の要約 (4、5)
- IV 考察 (6)
- V 参考文献 (7)
- VI 追加 (8)

## 2. 著者紹介

著者：James Dean Brown

- ・ハワイ大学マノア校
- ・第二言語学習

著者：Jonathan Trace

- ・teaching at SFC in 2016
- ・英語、言語

## 3. 論文の構成

4.1	はじめに	目的
4.2	計画評価	自分の選択を分かる
		教室評価を学習に合わせる
		教室評価によって学習を促進する
4.3	アイテムを作る	最高のアイテムを作成するため
		十分なアイテムを書く
		もう一度アイテムを確認する
4.4	テストの編成	アイテムの整理
		スコアリングツールの作成
		テストの校正
4.5	テストの使用	テスト管理を計画する
		学習者にフィードバックを与える
		教師がフィードバックを用いて
4.6	テストの改善	アイテム分析
		依頼性のチェック
		妥当性のチェック
5	結論	結論

**【おねがい】** このレジュメは、東京学芸大学大学院国語教育専攻日本語教育 コースの授業「日本語教育方法論演習」(授業担当者：南浦涼 介)での大学院生で取り扱った、Hinkel, E. (Ed.) 2017. **Handbook of Research in Second Language Teaching and Learning.** の発表のレジュメです。教育的価値、資料的価値としてウェブでの掲載を行っておりますが、いわゆる「論文」ではありませんので、論文等への引用 や掲載は固くお断りいたします。また、分析対象の著作権は著 作者、資料文書の著作権は発表者に記しますので、無断転載は ご遠慮ください。質問については、東京学芸大学南浦研究室 <http://minamiura-lab.com/>までお願いいたします。

## 4. 論文の要約

### 4.1 はじめに

主な目的： 言語教師が実践的な立場から、学習者との教室評価を計画、設計、実行するのを役に立てることである。

### 4.2 計画評価

#### 4.2.1 自分の選択を分かる

- ・ 評価を学生の学習にかつ効果的なものになりたいと考える教師は、最初に自分の選択肢を知ることによって、評価活動を最初からカリキュラムに計画して、統合する必要がある。

- ・ 多くの語学教師は、他の分野の教師よりもはるかに多くの評価選択を持っていることを認識していません (Brookhart, 1999 : ブラウン 2013b)

- ・ **選択応答アイテム (複数選択、真偽選択とか)**

一> 彼らは主に言語の受動的知識 (理解、認識など) に限定されている。

一> 選択応答アイテムの作成が難しい傾向がある。

一> 正しい/間違いによって採点されるので、アイテムに複数の可能な答えがあることを避けるように作る必要がある。

- ・ **プロダクティブ・応答アイテム (実際に目標言語を用いて答えること)**

アイテム	内容 (例)
fill-in or cloze-type items	単語やフレーズを空白に入れる
short-answer items	1語か2語、もしくは1文から2文で解答する
performance items	プレゼンテーション、ロールプレイング

一> (Bachman&Palmer, 1996, 2010) パフォーマンスはより長い回答を必要し、実際の生活の中で最も必要性が高いと示唆した。

一> パフォーマンスの難しい所 :

- ・ 実運用と全く同様の環境でない限り、結果には不確さがあること
- ・ 大規模なクラスでは管理と採点が難しい
- ・ 慎重な計画設計しないと、確実に有効なスコアを出さない

- ・ **個人応答アイテム : 学習者を中心に**

(教師がタスクに幅広いガイドラインを設定するが、学習者に実際の内容を決定させる。)

一> 教師と学生とのインタビュー (Jang, Dunlop, Park, & van der Boom, 2015) 自己評価/ピア評価 (Butler&2010; Dippold, 2009, Matsuno, 2009; Ross, 2006) 。

一> 学習者の発達と、その発達に対する理解とかに焦点を当てる。

一> 他の形式の評価、特にパフォーマンスアイテムとペアになる傾向がある

一> (ReaDickins, 2006, Torrance&Pryor, 1998) 学習が教室での評価の目的や機能を認識していることを事前に確認する必要がある

一> 十分な言語能力を持っている

\* 明確な指示や目標なしに学習者に直接教室評価を提出すると、学習者はプロセスを誤解することにつながり、スコアの信頼性に影響を与え、教室評価を完全に損なう可能性がある。

#### 課題

一> 信頼性と妥当性の明確な基準を保つことである。

解決法： ルーブリックやチェックリストを使用することによって、明確な評価ガイドラインを確立できる。

・ 複数選択



一> 教師はどのような形で授業評価を行うのかを考えて、学習者の学習を促進するためにどの解釈するのかについて、多くの選択肢を考慮する必要がある。

#### 4.2.2 教室評価を学習に合わせる

・ 学習者を学習教材や学習経験としっかりと結びつけていること

(Hall, Webber, Varley, Young, & Dorman, 1997) コースには明確な学習者の学習成果 (SLOs) があれば、学習者の学習と合わせて教室評価事が簡単であるが、そうでない場合は、各評価に直接関連するコースシラバス、教科書などをレイアウトすることで、教室評価を学習者の学習に合わせることができる。

・ 行った教室評価を毎回慎重に調べ、教えているものと比較すべき。

・ テストを学習者の学習と合わせることで、(特にプログラムレベルと関連すること) 教室評価に役立つ。

一> 教師は、自分の教室評価がどのように教室で起こっているのかを反映するだけでなく、もっと幅広いに注意を払う必要がある。

#### 4.2.3 教室評価によって学習を促進する

(メリット：学習者に教室評価の目的と意識を高めることができる)

・ 教室評価は、言語知識とスキルを得るための不可欠な要素であるため、とても重要である。

・ 教室評価は、最後に向けられたものではなく、カリキュラム内で継続的なプロセス

と認識すべきである。

- ・ 教師は、どのようにテストされるのを教えるだけでなく、どのようにフィードバックを受けるのか、そしてそのフィードバックで何ができるのかについて、学習者に認識される必要がある。

#### 4.3 アイテムを作る

##### 4.3.1 最高のアイテムを作成するため

- ・ 「書き方」に関する文献を読む
- ・ アイテム規格を使用する

一> アイテム規格は、プロンプトの特定の属性（つまり、学生が何を与えるか、アイテムをどのように見えるか）と応答（アイテムの採点方法、パフォーマンスの記述方法、分類方法、構成方法を記述することによって、アイテムの各部分を定義する方法である。  
(Davidson&Lynch, 2002; Popham, 1981)

##### 4.3.2 十分なアイテムを書く

- ・ 実際に書く際には、時間を十分に配分することが重要である。

一> 良質なアイテムを書けるだけでなく、十分な数のアイテムを書ける時間もある。  
(より良く学習者へ評価できるし、いらぬアイテムを後で取り除くことができる)

- ・ どれくらいのアイテムがあれば十分と言えるか
  - (a) 評価されるか学習目標の数
  - (b) クラス内の学習者の数
  - (c) 試験に分ける時間
  - (d) 利用可能な資源
  - (e) 使用されたアイテムの種類

##### 4.3.3 もう一度アイテムを確認する

- ・ 良いアイテムを書くことだけでなく、アイテムの機能、フォーマット、外見が自分の意図通りかどうかを確認することも重要である。

一> (数日か週末か) もう一度校正すること

#### 4.4 テストの編成

##### 4.4.1 アイテムの整理

- ・ テストを作成するというのは、単にアイテムをまとめることではなく、論理的に編成すること も必要である。

一> 同じフォーマットのアイテムをグループ化する。(例：真/偽選択、複数選択)

一> 同じスキルのアイテムをグループ化する。(例：リスニング、文法)

- ・ アイテムとタスクをテストに編成する際に、アイテムの依存性も考慮すべきである。
- 一> 手がかりが現れて、他のアイテムも答えられる。
- 一> あるアイテムに答える能力が、他のアイテムを回答の手掛かりになれる。

- ・ アイテムの独立性も問題点
- テストの公平性と妥当性

#### 4.4.2 スコアリングツールの作成

- ・ 多くの場合、教師はルーブリックやチェックリストのようなスコアリングツールに依存する必要がある。
- このようなツールが持っていること と その使い方を知っていることはまったく違うことである。

- ・ テストを開始する前に、手順、記録装置、およびルーブリックを熟知したり、再学習したりすることは、依然として重要である。

#### 4.4.3 テストの校正

- ・ テストを完全に組み立てたら、全体的校正することが不可欠である。
- 学習者に提示されるすべてのアイテム、スペルミス、フォーマットミス、などの不慮の手がかりなどを慎重にチェックする必要がある。
- テストを床に置き、上に立って見下ろすことである。
- 自分がテストを受けることである。

### 4.5 テストの使用

#### 4.5.1 テスト管理を計画する

- ・ 教室評価の1つの重要な要素は、学習者が評価されていることを認識していることである (Clarke, 1998; Hill&McNamara, 2012; Rea-Dickins, 2001)。

- ・ 1つの重要な考慮事項は、部屋の雰囲気
- ① 学習者の不安に対して、教師 → 冷靜的
- ② できるだけ通常の教室活動のように

- ・ テスト前の授業では、学習者に質問があるかどうかを確認
- 学習指導書を配布とか不明瞭な領域をより明確に説明とか

- ・ 公平性を注意すべきである
- ① 学習者が持っているすべての質問に答える
- ② テストの間に余分な助けや余分な時間を与えないこと など...

- ・ テストの後、フィードバックを与え、テストの進め方について議論する時間をあげる
- 学習者が自分の経験を振り返られる（内省させる）だけでなく、教師にも貴重なフィードバックを提供できる。

#### 4.5.2 学習者にフィードバックを与える

- 学習者が自分の強みと弱み知らせたり、彼ら（そしてその教師）が自分たちの言語発達を強化するために何をできるかを明確する。

・ ほとんどの場合、いくつかの種類のコア（パーセンテージとか）と関連づける。  
しかし、理由と意味を学習者に説明する必要がある。

- ・ Hill and McNamara (2012) は、教師が3つのカテゴリーを使用できると示唆した。
  - (a) 確証的 eg: パーセンテージ (限界性がある)
  - (b) 肯定的フィードバック eg: よくできました
  - (c) 訂正フィードバック eg: 誤りの訂正

個々を心配って指導できる

- ・ フィードバックは必ずしもテスト後に起こるものに限らない
- ・ 評価後のフィードバックと組み合わせることで、その戦略を継続して使用するかどうかを判断するのに役立つ。

#### 4.5.3 教師がフィードバックを用いて

・ 教師は、意思決定、カリキュラムの調整、クラス内での学習促進するため、評価から得られたフィードバックを自分の目的に使用すべきである。

・ 教師は、評価に関するパフォーマンスに基づいて、学習者の学習目的の変化に合わせて、教授法を調整すべきである。

→ 例え: 診断テスト

・ パフォーマンスの平均値をクラスの全員に共有して、学習者が彼らのパフォーマンスを理解することに役立つ。

・ 大多数の学習者にとって問題があると思われるテストのアイテムを見直す。

・ コアを記録させることにより、学習者に縦断的な感覚を提供する。各学習者（および教師）は、必要なときにいつでも参照できるようにする。

#### 4.6 テストの改善

##### 4.6.1 アイテム分析

試験が実施され、その成績やフィードバックが学習者に与えられた ~~≠~~ 試験は終了

##### 教室評価に共通する戦略

・ ステップ1、プロセスに振り返ることである。おそらくテスト行った後、コアをつける直後に自分の考えを書き込むこと。（特に以下のa~gに関する）

- (a) どのようにうまく機能しているか
- (b) テストをどう改善するか
- (c) コアの解釈を使用してどう学習をより良く促進するか
- (d) 個々のアイテムがどれくらいうまく機能しているか
- (e) テストがどれくらい期待に合っているか
- (f) フィードバックがどれほど信頼できるか
- (g) 管理中に何の問題があったか

・ ステップ2、学習者に、特定アイテムの体験及びテストの前後で自分の期待や経験を反映させることである。

・ ステップ3、試験中または試験後に学習者がした質問のリストを保存することで、わからない所/混乱しやすい所を明らかにすることができる。

\* 次回にテストを使用するまで修正しないと、まったく同じミスを繰り返す可能性があります。

#### 4.6.2 信頼性（そのテストが信頼できるか、という指標のことである）のチェック

- ・ 測定結果に安定性と一貫性があるかどうかを確認すべき
- 一> 選択応答アイテム以外のほとんどのアイテム形式では、信頼性が学習者を個別に測定する必要があると思われるが、クラス全体で安定性（同一個人に同一の条件で同一のテストを行った場合、同一の結果が出るかどうか）と一貫性（同一個人が同じような質問に対して、同じような答えをするか）があると判断できる基準である。
- ・ 教師は教室で真実性と信頼性のバランスをより良くとるために、いくつかの方法をとることができる

1. 目的を目立つ（太字またはイタリック体とか）ことを確認し、明確に書かれ、同じ方法であるすべての学習者に伝える。
2. 事前に計画などをお互いに交換して、教師と学習者の間に自然言語の使用することが効果的な評価方法になれる（ロールプレイ）
3. より公式化されていない評価では、教師は学習者のパフォーマンスを追跡するための計画を確立する必要がある。
4. 、最良のルーブリックであっても、それを使用している人は潜在的な偏見を受ける可能性があるため、信頼性を向上させる方法として、複数の評価者を採用すること。

#### 4.6.3 妥当性（測ろうとしているものが、適切に測れているかどうか）のチェック

- ・ 授業評価では、教師が何を測っているかは、学習者が学んでいること関わっている
- 一> 学習者の学習に合わせて、教室評価を計画する必要がある。
- ・ しているテストを確認することも重要である。教師の評価は少なくとも学習者が学習状況と一致する必要がある。
- 一> 出題範囲から偏りなく出題されているかどうか

##### 妥当性チェック

1. アイテムの内容とフォーマットは、学習目標とどれくらい一致するか？
2. アイテムの内容とフォーマットは、授業でカバーしている教材とどの程度一致するか？
3. 学習者はこのテストのアイテムが学習目標と教材とマッチしていると思っているか？
4. コースの学習目標が学習者のニーズにあわせるかどうか？
5. テストのスコアは、どの程度で学習者がこのコースで何を学んでいることを示しているか？
6. 私の評価は、自分のコースで学習を促進するため、どれくらい効果があるか？
7. テストのスコアは、私の価値？学習者の価値？彼らの両親の価値？私の雇用主の価値？と一致？
8. 私の学習者のテストスコアに基づいて、その両親、私、私の雇用主、どう決定するか？

## 5. 結論

- ・ 健全な授業評価はスコアや成績よりもはるかに優れているというだけでなく、教師が学生の学習をサポートし、促進するために使用できるツールであることは明らかです
- ・ 評価は、学習目標、教材、教師の間で必要なつながりを明らかにし、学生の評価パフォーマンスだけでなく、学習と発達を最大限に高めることができると考えています。
- ・ Rea-Dickins (2026) が、これは学習者にテストの目的、使用方法、成果についての意識を高めることによって行われる。そのような意識を形成することは、学習者と交渉あるいは相談をとることができることを指摘していた。
- ・ 学習者を関与させることなく、評価リスクは無意味である

## 6 考察

① 私は自国で日本語と英語を外国語として勉強した時、学生たちはテストでいい成績を取るために一生懸命暗記して練習問題をやっていた。中間テストと期末テスト以外、授業中でもいつも単語や文法を聞き取りや小テストを行っていた。しかし、用いたテストはほぼ学校あるいは教育局が指定されたテストで、教師自分が作ったアイテムではない。こういう場合は、このテストは学生の現状を適切に測るか、この結果によって、より良く評価とかフィードバックを行うかどうか疑問だと思う。このような場合、学習者の現状や学習目的に合わせて、教師側が何について配慮する必要があるか？

② 言語教育の場合、評価を行う時、言語についての理解能力や話す能力等の評価をあわせる必要がある。その際、学習目的に応じて、紙テストだけではなく、プレゼンテーション、ロールプレイング、スピーチ等の評価等、様々なやり方の中からその場面における学習者の現状を的確に評価できる方法を選択することが大切だと思う。

## 7. 参考文献

- 大中逸雄・工藤一彦 「学習・教育に関する達成目標の評価方法」 (2014)  
ホームページ<<http://www.u-gakugei.ac.jp/~kishilab/validity-reliability.htm>>  
(2017. 7. 30 参照)
- 近藤安月子・小森和子 (2012) 『研究社日本語教育事典』 研究社

## 8. 追加 (ディスカッションのまとめ)

・テストの結果によって、学習者にフィードバックを行う。学習者が自分の強みと弱み知らせたり、自分たちの言語発達を強化するために何をできるかを明確する。教師の場合も、評価に関するパフォーマンスに基づいて、カリキュラムの調整、クラス内での学習促進する、学習者の学習目的の変化に合わせて、教授法を調整することができる。しかし、多くの教育現場では、テストを行う前に、教師が学習者に出題範囲などを教えて、あるいは学習者がテストの前に猛勉強をするというような現状が少なくない。テストの成績を高めるための暗記や詰め込み等の現状をどのように改善すれば良いのかをこれから意識してカリキュラムを修正して、テストアイテムを作るほうが良いと思う。

・テストの結果は教師が学生の現状をより良く分かって、評価をする不可欠な手段であるが、その結果だけで学習者を判断するのが不十分である。例えば、言語教育の場合は、学習者の文化的背景や今の到達度、更にテスト当日の体調等によって、テストの結果に影響が多少ある。これをテストのアイテムを作成する時や学生にフィードバックを行う時、配慮する必要がある。

・私は今ボランティアとして J S L 児童生徒を支援している。より良く支援活動を行うために、ただの日本語指導だけではなく、学生の現状 (学習面、生活面) を分かって、これからどう支援するのも明確する必要がある。現在の段階では、大体授業様子の記録や子どもとのやり取りの記録してフィードバックしているが、これから子どものテストを分析して、教師から評価とや教師とのやり取りも記録を含めてフィードバックする必要があるかをこれからも試して、考察したいと考える。