

Flucher, G. (2010). *Practical Language Testing*. London: Hodder Education.

Chapter 3

8. Dependability

- 教室内テストがどれだけ信頼できるものかを示す指標が **dependability** である。
- **dependability** は、測定目標を基準とした信頼性の相関として定義され、生徒が目標の学力を達成したか否かについての教師の「判断 (decision)」と、それぞれのレベルの「基準点 (cut score)」が問題になる (Ch.8 で解説)。
- 最もよく用いられる **dependability** の算出法が、**threshold loss agreement approaches** である。これは同じ測定を 2 回行う。そして教師による、学習者が **master** か **non-master** かの判断が一貫するかどうかを見る。
- 算出にあたって、まず一致係数 (**agreement coefficient: P_0**) を計算する。これはテストが複数回実施された際 (もしくは複数人によって判断された際) の一致率を示す。公式は以下の通り。

$$P_0 = \frac{A + D}{N}$$

N = 総数, A, D = 実施されたテスト (評価者) 間で判断が一致した数 ※要説明

- しかし一致係数は、判断の一致が偶然によるものである場合を考慮していない。よって、次にこの「偶然の一致率 (**Pchance**)」を計算する。

$$Pchance = \frac{(A+B) \times (A+C) + (C+D) \times (B+D)}{N^2}$$

- 一致係数と偶然の一致率を利用することで、偶然によらない一致率であるカッパ係数を出すことができる。

$$k = \frac{P_0 - Pchance}{1 - Pchance}$$

カッパ係数の判定基準

値	判定
.00 - .20	slight (低い)
.21 - .40	some (やや低い)
.41 - .60	moderate (中程度)
.61 - .80	substantial (かなり高い)
.81 - .99	very high (非常に高い)

■ カッパ係数は、データが名義尺度または順序尺度の際に適応されるノンパラメトリック検定で、評価者間信頼性・評価者内信頼性のいずれに対しても用いることができる。

■ 分類された2群の一貫性を調べる手法として、squared-error loss agreement がある。これは分類 (e.g., mastery and non-mastery) の一貫性だけでなく、分類された群それぞれの程度 (i.e., the degree of mastery or non-mastery) を考慮に入れる。

■ 程度を考慮するため、この手法を用いるスコアには幅があることが求められる。
→ 多くの項目を含み、正解・不正解でスコアを算出できるテストのみに利用できる。

■ squared-error loss agreement の代表例として、ファイ ラムダ ($\Phi \lambda$) がある。

$$\Phi \lambda = 1 - \frac{1}{K-1} \left(\frac{\bar{X}_p(1-\bar{X}_p) - S_p^2}{(\bar{X}_p - \lambda)^2 + S_p^2} \right)$$

K = 項目の数, \bar{X}_p = 平均正解率 (mean of the proportion scores), S_p = 正解率標準偏差, λ = 基準点

※ 基準点: 分類を分け目となるスコア (e.g., これ以上なら master, 未満なら non-master)。詳しくは8章を参照。

→ これを計算することで、設定した基準点の dependability を推定することができる。

■ 基準点の dependability を向上させるためには、基準点を操作するよりも、テスト自体の質 (e.g., 測定対象の構成概念をしかるべき方法で測れているか) に目を向けるべきである。

■ 2章で述べた標準テストでの標準誤差に当たるものとして、目標準拠テストでは信頼区間 (CI: confidence interval) を算出する。これによって、スコアがどれだけ上下し得るかを推定することができる。

$$CI = \sqrt{\frac{\bar{X}_p(1-\bar{X}_p) - \bar{S}_p^2}{K-1}}$$

ここで、分母に N を用いる点が他の統計手法との差異としてあげられる。

$$SD = \sqrt{\frac{\sum(X - \bar{X})^2}{N}}$$

■ 基準点の信頼区間内にスコアが収まった受験者に対しては、他の情報も加味して分類を決定する必要がある。

3.9 Some thoughts on theory

■ 教師が生徒を必ずしも評価する必要がないテストがあってもいいのではないか。

→ 教師は時に、生徒の知識や技能以外の要素から影響を受ける。特に人々が協働して学習する環境では、人との関わり (contact) が重要であり、学習は社会的な様相を呈する。

→ このような学習は、教師や学校が変われば、別のものになるもの。

→ よって、様々な生徒がそれぞれの学習環境の枠を越えて比較される high-stakes テストでは、外部テスト (external tests) を利用すべきである。この中の判断を下すプロセスを教師は担うべきである。

■ 外部テストは教師を守ることにもつながる。

→ high stakes テストを教師のみが担うと、個人的な偏見を挟んでいると非難される可能性がある。

■ テストが、「より良い人生を歩む人々を選定する」ために使われないならば、high stakes なテストなどは存在しない。この性格のため、テストには常に経済的価値がつきまどってきた。

→ だからこそ、教室の形成的評価は別ものとして扱われるべきである。形成的評価の目的は学習を促進したり、学習の目標や発達段階への意識を高めたりすることであって、high stakes な判断を下すためのものではない。

■ 学習のための評価 (assessment for learning) と DA (ダイナミックアセスメント) の理論的基盤はそれぞれ異なる。

・学習のための評価は、特定の大規模テストにおける調査で得られた成果を、学習環境に応用しようとするものである。確率的、経験的なスタンスを有し、教室で何が機能するかということへ、実践的なアプローチをとる。

・対して DA は、文化社会的理論を基盤にもつ。DA で用いられる多くのテクニックは、学習のための評価で使われるものほど画期的なものではなく、多くの教師が何らかのかたちで実践しているものとそう変わらないように見える。

■ 結果を一般化できないという点が、DA の最大の欠点である。

・実質的な理論は、教育的介入の効果を要因 (e.g., 個人差要因、環境的要因) ごとに分析し、予測可能、測定可能にしようとする。されにこれらの要因は、予測された効果に影響するものとしてリスト化される。

→ このような過程を経てのみ、理論が現象を正確に説明できるか否かを判断できる。

・だが、対象を全体論的に評価することを試みる DA には、この性格が本質的に欠如している。つまり、DA は、実質的な理論的正当性を主張してはいる一方で、その必要条件を放棄しているという根本的な矛盾を抱えている。

■ 教室内の形成的評価もまだ確立の途上にあるが、実用的な面で、教師や学習者が求めることは、言語習得やコミュニケーション能力向上につながるテクニックである。

→ この実用的な観点では、教室形成的評価は、教師や学習者の目標達成に資する発展を遂げてきたと言える。

☆ ダイナミックアセスメントとは

■ ヴィゴツキーの発達の最近接領域の考えに基づいて、学習者同士や学習者と教師とが関わり合うことで、生徒がどのような変化をとげたのかに焦点を当てる評価。

■ ダイナミックアセスメントの主な特質には以下の3点があげられる。

(1) 学習者と教師との相互作用

- ・教師は、学習者についての観察や推測に応じて、彼らの学習過程を明らかにしたり、変化を促したり方法を用いることで積極的な評価者としての役割を果たす。
- ・つまり、教師はペーパーテストなどによって生徒を静的に評価するのではなく、学習者のその時々状態に応答することで彼らを変化させ、その様子进行评估するのである。

(2) メタ認知への着目

- ・教師と学習者が動的、積極的に関わり合うことで、学習者がどのように問題解決に取り組んでいるか、自分の学習成果を評価しているか、学習計画を監督しているかなどのメタ認知的側面をより重視した評価が可能になる。

(3) 介入によってもたらされる情報

- ・教師の介入を評価過程の中に取り入れることで、学習者が介入に対しどのような反応をし得るのか、介入によってどのような変化を遂げ得るかという情報を得ることができる。これは、静的なテストでは決して得られないダイナミックアセスメント特有の情報である。

■ ダイナミックアセスメントの問題点

- ・ダイナミックアセスメントは、「できない」から「できる」へ学習者がどのように変化、移行できるか否かを問題とする。しかしながら、他の学習者や教師との関わりがこの移行の中で果たせる役割には限界があり、多くの場合、移行の成果は学習者の発達の最近接領域に依存する。つまり、彼らの発達の最近接領域を前進させる努力なしには、ダイナミックアセスメントは十分な教育的示唆を与えられないのである。

- ・また、発達の最近接領域に基づくことで、学習可能性を見出すというよりも、独力では達成できないが指導や協同によって学習者をいかに前進させられるかという指導法に主眼を置かざるを得なくなる。その具体的な指導法なしに、学習者の潜在的な可能性を相互作用の中で模索してゆくしかないという点もダイナミックアセスメントの限界点である。

参考文献

Lidz, C. S. and Elliott, J. G. (2000). Introduction. In Lidz, C. S. and Elliot, J. G. (Eds). *Dynamic assessment: Prevailing models and applications*. Greenwich, CT: Elsevier-JAL.