

I・HEAP (対象: Ercikan, K., & Seixas, P. (Eds.). (2015). *New directions in assessing historical thinking*. Routledge.)

担当: 空 健太 (岐阜工業高等専門学校)
sora@gifu-nct.ac.jp

歴史の評価における妥当性の証拠を構築する重要性 : 見逃されたり誤解されやすいことは何か

PART IV Validity of Score Interpretations,

Chapter 12: The Importance of Construct Validity Evidence in History Assessment: What Is Often Overlooked or Misunderstood? pp. 195-205

本章の著者 (*所属は2020現在)

Pamela Kaliski (所属: American Board of Internal Medicine)

Kara Smith (所属: College Board)

Kristen Huff (所属: Curriculum Associates)

■ 用語

- validity | 妥当性 • dimensionality | 次元性 • cognitive validity | 認知的妥当性
- think-aloud protocol (TAP) | 思考音読プロトコル (思考発話法)
- factor analysis | 因子分析 • correlation coefficient | 相関係数

概要と結論:

この章では、歴史の評価デザインにおける妥当性の問題に焦点を当てている。タイトルの副題になっている「見落とされたり誤解されることが多いこと」とは、次元性の証拠と認知的妥当性の証拠を指す。本章では、これら2つの妥当性の証拠とは何か、なぜ重要か、そしてこれらの証拠をどのように収集するかについて例を挙げながら説明されている。ハイスタークスの歴史の評価において、評価しようとする項目が評価問題によって適切に測定されているかどうかは極めて重要である。

■ 議論の提案

1. 因子分析を行った歴史の評価研究はどのようなものが存在するか?
2. TAPを行った歴史の評価研究はどのようなものが存在するか? (Cf. 益川ほか (2018))
3. 因子分析やTAPを用いて、どのような方向性の研究が今後期待されるか?

■ 導入部 [p.195]

- 本章の焦点: 歴史の評価デザインの文脈における妥当性の問題
 - 本章は、妥当性に関する最も包括的な議論を行っている Kane の見解を参考にしている。
- Kane (2013) の妥当性のフレームワーク
- 妥当性の証拠を形成する2ステップ (議論に基づくアプローチ)
 - (1) テストスコアの特定の解釈/使用に内在する主張を特定する
 - (2) 実証的な証拠や論理的な議論に基づいてその主張の評価を提示する

■ 本章の目的 (Purpose of the Chapter) [p.196]

- 本章の目的: 歴史の評価の構成が妥当性かどうか→次元性の証拠、認知的妥当性の証拠
 - 次元性の証拠
 - 歴史的思考、歴史的知識などの次元が評価されているという証拠
 - 認知的妥当性の証拠
 - 意図された認知過程 (=歴史的思考) を引き出しているという証拠
 - なぜこの2つか?
 - 歴史的思考の定義は依然として異なっている。次元性の分析は、項目が目的の構成要素に関連しているかどうかを検証し、構成要素の定義を洗練させる上で重要 →因子分析 (factor analysis)

→歴史的思考の構成要素に含まれる高次の性質が測定されているかどうかを判断するには、従来の方法に変わる方法が必要 →思考音読プロトコル (think-aloud protocols : TAPs)

■ 論点#1 : 歴史的思考の次元を評価する

(Issue #1: Assessing Dimensionality of Historical Thinking) [p.196-]

①歴史的思考の評価を行うために必要なこと

- ・評価が歴史的思考を測定できない理由の1つは、歴史的思考の次元が理解されていないことにある(歴史的思考の多義性、構成要素の複雑さ)。
- ・テスト開発者が評価をデザインする際、その歴史的思考に内在する主張を特定し、その立場をとることが重要。

②歴史的思考に関する前提

- ・合意されている点
 - ・生徒は歴史的な内容知識と認知の両方を理解する必要がある。
 - ・歴史的知識と歴史的思考は関連し相互に依存している。
- したがって、歴史的知識と歴史的思考の両方を評価する必要がある。
- ・本章では、Peck and Seixas (2008)の表現を採用。
- 修正された各次元

①歴史的意義 (historical significance) ②一次資料の使用 (using primary source evidence)
③連続性と変化 (continuity and change) ④原因と結果 (cause and consequence) ⑤歴史的パースペクティブ (historical perspective) ⑥倫理的次元 (ethical dimension)
+ ⑦事実に基づく知識 (factual knowledge)

□ 次元性の証拠を収集することの重要性 (The Importance of Gathering Evidence of Dimensionality)

①構成要素の妥当性の証拠を収集する重要性

- ・構成要素の妥当性は、テストが、測定しようとしているものを測定しているかどうか、意図しない構成要素を評価していないかどうかを確認するもの。そのためには複数の(出来るだけ多くの)証拠が必要。
- ・証拠を収集する2つの一般的な方法として、相関係数と因子分析。

□ 相関係数 (Correlation Coefficients)

①相関係数を調べる

- ・歴史的思考を測るテストは、他の歴史的思考を測るテストと高い相関性を持つはずである。
→異なる2つのテストのスコア間の相関係数(方向性や関係性の度合い)を示すことで、歴史的思考を測定しているかどうか、どの程度まで測定しているか、歴史の評価が意図しないスキルを測定しているかなどを判断できる。

□ 因子分析 (Factor Analysis)

①EFA と CFA

- ・探索的因子分析 (exploratory factor analysis: EFA)
→その項目が表す因子(=次元)の数とタイプを決定しようとする時に使用する
- ・確認(確証)的因子分析 (confirmatory factor analysis: CFA)
→歴史的思考に関する理論的な仮説がある場合、歴史的思考の構成要素が測定されているという証拠を得るために使用する

②因子分析のプロセス

- ・前述の7つの歴史的思考の評価を採用する場合、もしそのテストが7つの次元のそれぞれを十分にカバーしていたとしたら、7つの要因が現れることを期待することになる。

- ・例：「歴史的意義 (historical significance)」を測るために作成された項目
→全て同じ因子に負荷がかかると予想される。

ステップ1 因子が特定されたら、データの性質を把握し、モデルの仮定を確認するために、予備的な記述的統計解析を行う	ステップ2 因子モデルのパラメータを推定し、モデル適合性 (*) を評価する	ステップ3 パラメータの推定値を検討する。テスト開発者が、そのモデルが良好な「適合」を生み出し、パラメータ推定値が適切であると判断した場合、評価が歴史的思考スキルを測定しており、意図しないスキルを測定していないことを示す	ステップ4 適切な統計が特定されたら、分析結果を仮説の構造と比較し、類似性を評価する
--------------------------------------------------------------------	--------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------

(*) モデル適合性 (model fit) については、Hu & Bentler, 1999, and Marsh, Hau, & Wen (2004)

□ 次元性の証拠を収集することの一例 (An Example of Gathering Dimensionality Evidence)

① NAEP の妥当性研究パネルの作業

- ・ National Assessment of Educational Progress (NAEP)は、合衆国史に関する全国的な評価を実施し、国内の8年生の生徒の知識とスキルを測定している。
- ・ NAEP 妥当性研究パネル (NAEP Validity Studies: NVS)
→目的：関心のある構成要素を定義し、妥当性を測定するために様々なエビデンスを収集し、その評価が関心のある領域をカバーしているかどうかを実証的に評価する

② 構成要素の妥当性研究の限界と注意点

- ・ 因子分析法では、項目の回答のサンプルサイズを大きくする必要があり、テスト開発の状況では必ずしも利用できない場合がある。
 - ・ 相関研究と因子分析の両方とも、常にサンプルに依存しており、結果を解釈する際には留意する必要がある。
- スコアの妥当性が、あらゆるタイプの受験者に適用可能であることを保証するためには、異なるサンプルを使用して、同じ評価に対して多次元性の研究を実施することが最善。

■ 論点#2：歴史的思考の認知的妥当性の証拠を集める

(Issue #2: Gathering Cognitive Validity Evidence of Historical Thinking) [pp.200-]

① 認知的妥当性の証拠とは何か

- ・ 認知的妥当性の証拠＝項目が意図した認知プロセスを引き出している証拠 (テスト項目と対話する際、受験者が採用する問題解決や推論が、構成要素の定義に含まれているものと同じであること)
→受験者の認知プロセスの理解はテストスコアからの推論の妥当性を確保するために不可欠。
- ・ 評価デザインの初期段階での検討も重要 (Kane のステップ(1)) だが、生徒が意図した方法でタスクを実施しているという証拠が必要 (Kane のステップ(2)) 。
- ・ 次で紹介する方法は、テストスコアが生成される前に認知的妥当性の証拠を集める方法の一例。

② 認知的妥当性の証拠を収集する方法と意義

- ・ 一般的に質的な方法で証拠が収集される。
- ・ 推奨されるアプローチ
→生徒との TAP (テスト項目に回答する間の生徒の考えを言語化したレポートを集める方法)
具体的には「声に出して考える」＝生徒の口頭による報告
- ・ TAP は生徒が採用する認知プロセスや知識の構造を特定するために有用 (Ercikan et al., 2010; Leighton, 2009)
- ・ このような証拠を集める意義
→「テスト項目が体系的に誤解されている場合、(a) その評価は意図したものとは異なる内容の理解やプロセスを引き出すか、(b) スコアから導き出される推論が不正確であるか、またはその両方であることを意味する」 (Leighton, 2004, p.8)

□ TAP を用いて認知的妥当性の証拠を収集する方法 (How to Gather Cognitive Validity Evidence Using Think-Aloud Protocols)

① TAP の 2 つのタイプ

- ・ concurrent think-aloud protocols (同時 TAP) ・ retrospective think-aloud protocols (回顧 TAP)
- リサーチクエスションでどちらを実施するかを決定する。
- 例：「生徒がその項目と対話している間、どんなスキルが引き出されていたか？」 → 同時 TAP
- 例：「生徒はこの項目を難しいと感じたか、そうならばなぜか？」 → 回顧 TAP

② TAP を用いた認知的妥当性の研究の計画

ステップ 1：目的の明確化→TAP のタイプの決定
ステップ 2：実施項目の開発 留意点：簡単すぎるあるいは難しすぎる項目を避ける、利用可能な時間を考慮する
ステップ 3：生徒のリクルート 留意点：母集団を考慮する、背景情報データも収集する
ステップ 4：スクリプトの開発 留意点：目的の説明など。何かを評価されるわけではなく、声に出して考えることに正解も不正解もないことを説明する
ステップ 5：デモンストレーションと練習項目の開発 留意点：ファシリテーターによるデモンストレーションを実施し、少なくとも 2 つの項目で TAP の練習をさせる
ステップ 6：パイロット研究の実施
ステップ 7：生徒の口頭の報告の分析 (フレームワークに基づくコード化) 留意点：信頼性の確保ためコード化は 2 人以上で行う。
↓
認知的妥当性の証拠の収集、あるいは項目の改善

□ TAP に関する実践上の限界と注意 (Practical Limitations and Cautions About Think-Aloud Protocols)

- ・ TAP の実施にはコストが大きい (だからこのアプローチが取られないことが多い)。
 - ・ 歴史的思考の評価をこれまで以上に重視する現在の歴史評価デザインの新しい方向性にとっては、大きな価値がある。
 - ・ TAP の使用について次章で詳述
- HTAT (Historical Thinking Assessment Tool) と呼ばれるアセスメントの開発中の TAP の使用について詳しく説明

* 結論は省略。

* 参考引用文献は原文を参照