

I・HEAP (New Directions in Assessing Historical Thinking)

## イントロダクション：歴史の評価の新しい形（後半）

担当：池尻良平（東京大学大学院情報学環）

ikejiri@iii.u-tokyo.ac.jp

著者（後半パートのため詳細は省略）

- ・ Peter Seixas
- ・ Kadriye Ercikan

### ■用語

- ・ historical thinking | 歴史的思考
- ・ historical consciousness | 歴史意識
- ・ validity evidence | 妥当性の証拠

### ■議題

- ①各自が興味のあるテーマや章は何かの共有と、担当する章の調整

-----  
※可読性を上げるため、本文中には書かれていませんが章タイトルを挿入しています。

### ■本書の構成（pp.7-12）

#### ●パート1：歴史教育のゴール：歴史の認知・学習モデル

- ・ 歴史教育のゴールや評価に向けた示唆に関する異なる表現について描写し、議論する。
- ・ なお、1970年代と1980年代の British Schools Council History Project (SCHP) は「二次的な歴史的思考概念」を提唱した重要なモデルだが、本書では章がない。

**1章**：Carlos Kölbl and Lisa Konrad | ドイツにおける歴史意識：コンセプト、実践、評価  
→ドイツの「歴史意識」を中心に、過去20年間の用語の系図を包括的に説明している。

**2章**：Abby Reisman | 学問的な歴史読解を評価することの困難さ  
→アメリカ合衆国における歴史の学問的な読解とライティングの研究やトレンドについて説明している。

**3章**：Carla van Boxtel, Maria Grever, and Stephan Klein | 歴史的思考の向上と評価のためのリソースとしての遺産：オランダからの振り返り  
→オランダの歴史カリキュラムの新しい要素、"heritage education"などを紹介している。さ

らに、オランダチームのアプローチの中に、二次概念（歴史的意義など）との交差を見ることがもできる。

#### 4章：Catherine Duquette | 評価を通して歴史意識と歴史的思考を関連づける

→評価をベースにして、カナダの歴史的思考プロジェクトと歴史意識のモデルを一緒に編み上げている。

### ●パート 2：歴史的思考の評価をデザインする際の論点

・歴史的思考の様々な側面の測定を狙った評価モデルのデザインや提供を行う。

#### 5章：Bruce VanSledright | 歴史の教室の学習に対する評価

→タスクデザインのアプローチを使った、教室レベルの評価を考察している。「重み付けの多肢選択」を提案している。

#### 6章：Andreas Körber and Johannes Meyer-Hamme | 歴史的思考、コンピテンシー、それらの測定：挑戦とアプローチ

→「歴史意識」を評価可能なコンピテンシーモデルと、大規模評価に適したタスクに変換している。(Carlos Kölbl and Lisa Konrad の歴史意識のモデルの一部と関連)

#### 7章：Peter Seixas, Lindsay Gibson, and Kadriye Ercikan | 歴史的思考の評価に向けたデザインプロセス：1時間テストのケース

→個々人のタスクではなく、複数の短い文書の抜粋を使った1時間のテストや、そのデザインを説明している。(Carlos Kölbl and Lisa Konrad の歴史意識のモデルの一部と関連)

#### 8章：Monika Waldis, Jan Hodel, Holger Thünemann, Meik Zülsdorf-Kersting, and Béatrice Ziegler | 生徒のナラティブなコンピテンシーを評価するためのマテリアルベースでオープンエンドなライティング課題

→複数の文書を用いて小さな質問に回答させた後、パネルディスカッション、ブログ、生徒新聞の中から1つのナラティブを構築させる課題や2種類のテストについて述べている。また、「質の特徴 ("quality features")」をスコア化している点が特徴である。(Carlos Kölbl and Lisa Konrad のモデルや、ナラティブなコンピテンシーのバリエーションに位置付けている)

コメンタリー：Josh Radinsky, Susan R. Goldman, and James W. Pellegrino | 歴史的思考：概念的・実践的なデザインガイダンスの研究と生徒のコンピテンシー評価の使用において

### ●パート 3：歴史的思考の大規模評価

・歴史的思考に関する3つの大規模評価（アメリカから2つ、スウェーデンから1つ）を紹介する。

・アメリカやドイツの認知モデルの特徴を、実際の大規模評価に落とし込んでいる点がパート3のポイントである。

#### 9章：Stephen Lazer | 歴史的な知識と推論の大規模評価：NAEP U.S. History Assessment

→1万人の異なる学年の大規模評価への挑戦について説明している。参照資料を用いるためのテクノロジーを使った評価や、シミュレーション、ゲームを使ったコラボレーションを含む評価まで言及している点が特徴。

**10章**：Lawrence G. Charap | アメリカ合衆国の歴史の AP コースと試験のリデザインにおける歴史的思考の評価

→多肢選択テストに対する批判を受けてリデザインし、記憶の再生の多肢選択問題ではなく、歴史的思考に従事させることを意図した多肢選択やライティング問題を作っている。

**11章**：Per Eliasson, Fredrik Alvéén, Cecilia Axelsson Yngvéus, and David Rosenlund | スウェーデンにおける大規模評価に反映された歴史意識と歴史的思考

→アメリカのモデルと比較して、より幅広いコンピテンシーと関連している歴史意識に焦点を当てた、スウェーデンの国としての歴史の評価を紹介している。また、この国の評価は行政と評価のスコアリングの両方の観点において、教師も巻き込んでいる。

コメンタリー：Susan M. Brookhart | 実践における歴史的思考の評価

#### ●パート4：スコアの解釈の妥当性

・歴史的思考の評価における妥当性の研究について議論している。

※池尻補足：「信頼性」「妥当性」、

**12章**：Pamela Kaliski, Kara Smith, and Kristen Huff | 歴史の評価の妥当性の証拠を構築することの重要性：見逃されやすいことや誤解されやすいことは何か

→妥当性の証拠を集めるためのアプローチとして、次元的な妥当性の証拠（評価の各項目が、構成概念の定義との一貫性の観点で相互に関連しているか）と、認知的な妥当性の証拠（意図した認知プロセスに従事しているか）の2つを紹介している。

**13章**：Kadriye Ercikan, Peter Seixas, Juliette Lyons-Thomas, and Lindsay Gibson | 歴史的思考の評価を有効にするための認知的な妥当性の証拠

→12章の議論をフォローする形で、認知的な妥当性の証拠を同定するための3つのステップを提案している。

**14章**：Gabriel A. Reich | 推定する？：多肢選択問題

→多肢選択問題における認知的な妥当性の証拠の重要性と、歴史的思考を評価する際の限界を論じている。

**15章**：Mark Smith and Joel Breakstone | 歴史の思考の評価：認知的妥当性の調査

→教室の中で教師によって用いられる特定の歴史的思考に向けた課題について論じている。シンクアラウドのプロトコルを分析することで、実際に歴史的思考を捉えられている点も特徴。

コメンタリー：Denis Shmeilt | 歴史的思考の評価の妥当性

2020/05/01

東京大学

池尻良平

・各パートの最後にあるコメントリーで、そのパートの要約や、各章の類似点や相違点の整理、批評をしている。

## ■評価の用語に関する池尻補足

出典:清水裕子 (2005) 測定における妥当性の理解のために -言語テストの基本概念として- . 立命館言語文化研究 16(4), 241-254.

### ・信頼性

「同じ対象(例えば受験者)に対して, ある観測(例えばあるテスト)をした場合, 同じような測定値が得られているかどうかを観察する指標」

→再テスト法 (いつでも同じ) や折半法 (誰でも同じ)、クロンバック・アルファ等の手法を使うのが一般的

### ・妥当性

「ある尺度(scale)やテストの, 特定の変数(例えば, 語彙力など)に関する測定道具としての適切さを示すものであり, 意図した特質や能力を測定できているかどうかを示す指標」

→因子分析や相関分析等の手法を使うのが一般的

→従来は表 1 が伝統的だったが、近年は表 5 の区分が重視されている。

表 1 伝統的観点からの妥当性の分類

基準関連妥当性	
併存的妥当性	そのテストが, 同時に測定されるある基準をどれだけ適切に推定しているか。当該テストとほぼ同時に収集された基準との相関により検証。
予測的妥当性	そのテストが, そのテスト実施以降の変化等をどれだけ適切に予測しているか。当該テストと, それより後の基準との相関により検証。
内容的妥当性	そのテストを構成している項目が, 全体を偏りなくどれだけ適切に代表しているか, また構成概念をどれだけよく反映しているか。その分野の専門家により判断, 検証。
構成概念的妥当性	そのテストが, 測定しようとする構成概念や潜在特性をどれだけ適切に反映しているか。因子分析や相関などの種々の検証方法がある。

表 5 Sources of evidence in the 1999 Standards

Sources of evidence based on.....	
Test content	項目の形式や採点方法などの変数とテスト内容の関連性について
Response processes	構成概念の性質とテストによる受験者の行動や回答プロセスの連関の重要性について
Internal structure	次元性 (Dimensionality) にも関わることで, テストの項目とその結果 [得点] が, 構成概念の構造と対応しているかどうかについて
Relations to other variables	テストと他の測定との相関関係に関して
Consequences of testing	テスト得点の解釈と使用が与える影響について