

## 13 章

歴史的思考の評価を検証するための認知的有効性の証拠

2020/11/20

I・HEAP

担当：渡邊竜平（広島大学大学院教育学研究科）

### ◎著者情報

**Peter Seixas**



- ・ ブリティッシュコロンビア大学名誉教授（2016 年退職）
- ・ バンクーバーで高等学校社会科を 15 年教えたのち、カリフォルニア大学で博士号取得。  
UBC 歴史意識研究センター長に就任。
- ・ 研究テーマは歴史教育と歴史意識，社会科教育研究，集合的記憶，歴史意識，カリキュラム研究など。

<https://edcp.educ.ubc.ca/faculty-spotlight-peter-seixas/>

**Lindsay Gibson**



- ・ UBC のカリキュラム・教育学専攻の助教授
- ・ オカナガン公立学校で 12 年，中等学校の歴史と社会科を教えたのち，UBC でカリキュラムと教育学の博士号取得。  
Historical Thinking Project の執行委員会のメンバー
- ・ 研究テーマはカリキュラム設計，歴史的思考，歴史意識，評価，教師教育，ディフィカルトパスト，教育と学習など。

<https://edcp.educ.ubc.ca/faculty-staff/lindsay-gibson/>

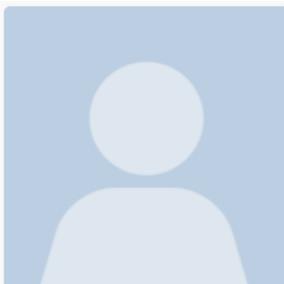
**Kadriye Ercikan**



- ・ 教育カウンセリング心理学科，特別教育学科教授
- ・ 教育評価と教育研究の方法に焦点を当てて研究。  
学校の評価，政策決定，各国の教育システムの有効性の比較，教育に関する研究を行うための情報を提供
- ・ 研究テーマは国際的評価，妥当性と公平性，教育測定など。

<https://ore.educ.ubc.ca/2014/08/29/kadriye-ercikan/>

**Juliette Lyons-Thomas**



- ・ 測定・評価・研究方法論を専門
- ・ UBC の博士号を 2014 年に取得
- ・ 思考発話法を用いた歴史的思考評価の博士論文を執筆。
- ・ Ercikan や Seixas の指導を受けた

<http://independent.academia.edu/JulietteLyonstomas>

<https://ecps.educ.ubc.ca/juliette-lyons-thomas-final-phd-defence/>

## ◎議題の提案

- ・思考発話法の妥当性の有無や限界

→生徒は自分の意図を全て話せるわけではないし、無意識で行っている思考もある？

思考を言語化できない生徒は歴史的思考をしていないとみなされ、

無意識で歴史的に思考しているかもしれない生徒は評価されない

喋りながら考えることが得意でない生徒…

コードの妥当性…

- ・この論文の結果が日本の歴史的思考評価研究に示唆するものとその限界

→この論文でうまくいったから日本にも導入しよう！でもいいの？

追試の必要性の有無、国の文脈が与える影響など

## ◎用語

- ・TAPs (Think aloud protocols) →思考発話法

- ・MC (multiple choice) →多肢選択問題

- ・CR (constructed response)

→構成的応答(生徒に知識, スキル, 批判的思考力を実世界の基準に基づいたパフォーマンス課題に適用するように求める評価項目。答えが一つ以上のことが多い。)

## ◎本文の内容

### 問題関心

歴史的思考などの複雑なものに関する評価

→ある評価タスクがどの目標や能力を測定しているのかなどの判断が難しい

→どのタスクが測定されているかを明確にする実証的証拠の必要性

### 【課題設定】

歴史的思考の評価の妥当性調査における認知的証拠について、生徒の言語化が生徒の歴史的思考の証拠を提供する程度を検討し、議論する

### 研究方法

- ・思考発話法

生徒が問題に取り組んでいる際に思考していることや行動の理由などを言語化させるもの。生徒の回答データや思考プロセスの可視化に資する

### 評価ツール

- ・歴史的思考の概念(歴史的意義 証拠 継続性と変化 原因と結果 歴史的視点 倫理的次元)

→証拠中心評価設計 (ECD) に基づいて開発

ECD について

- ① **認知学習モデル**…評価の対象となる知識, 能力, 高次思考の要素から構成。タスクモデル(どの評価タスクに重点を置くべきか)と証拠モデル(認知学習モデルで特定された能

力に関連してタスクへの回答をどのように解釈すべきか)の両方を決定。

②**タスクモデル**…生徒の反応を引き起こすことができるタスクの重要な特徴と、学習進歩における生徒のレベルの証拠を提供できる成果を記述したもの。①から導出。

③**証拠モデル**…認知学習・タスクモデルと密接に関連。タスクについて、生徒の行動をどのように観察するかが、学習の進歩のレベルとどのように関連しているかを記述。

→特徴：**評価したい能力（歴史的思考の概念）**と**タスク**，**評価の一貫性**

1時間のテストから適切な測定を行うため、6つの概念のうち3つに注目

#### ①**証拠**

・歴史的議論のための史料を発見し、選択し、文脈づけし、解釈すること

#### ②**歴史的視点**

・「過去は外国」であり、人々の生活や行動を形作ったさまざまな社会的、文化的、知的、さらには感情的な文脈があることを理解すること

・史料が作られた時の状況や世界観を考慮して読むこと

#### ③**倫理的次元**

・過去の様々な状況において、我々が現在どのように行為者を判断しているのかを理解する

・過去の犯罪や犠牲が今日どのような結果をもたらしているのかを理解する

・それらの結果に関連してわれわれが今日どのような義務を負っているのかを理解する

**題材**…第一次世界大戦中のカナダにおけるウクライナ人の抑留（11 学年）

・必要なすべての事実的知識は、6つのポイントからなる「背景情報」のページと

ウクライナの抑留に関連した史料からの5つの抜粋で提示

（事実的知識に依存しない歴史的思考測定のため）

#### **1~3の問い**

・生徒に各文書に示された見解の解釈、彼ら自身または歴史家が史料文書をどのように使用するか、文書に示された見解を比較し、それらが異なる理由を説明することを求めた

#### **最後の2つの問い**

・ウクライナの強制収容について倫理的な判断を下す前に、生徒たちが読んださまざまな文書を考慮することを求めた

#### **思考発話法の手順**

・問いを声に出して読ませ、問いを読んでいるときに考えていることを答えさせる

・問いの解釈や、何をしているのかを答えさせる

・何をどのように尋ねる問いか、問いに答えるために何をしなければならなかったか、どうやって答えを出したのかなど、問いに答えながら考えていることを全て答えさせる

## サンプル

- ・ 11 学年, 35 名(男性 11 人, 女性 24 人), 12 人が家庭で中国語, 13 人が英語
- ・ 17 人が社会科の成績 A, 12 人が B, 2 人が C+, 2 人が C
- ・ 10 人は充実したアカデミックプログラム, 残りの 25 人は普通高校

## 今回使用したタスク…全部で 11

MC…タスク 1, 2, 3, 6, 7, 9 CR…タスク 4, 5, 8, 10, 11

タスク 1~9, 証拠と視点の能力を要求 タスク 10, 11, 倫理的次元の能力を要求

## 生徒の言語化のコーディング

コードセット 1 : タスクの理解

- ①生徒が質問を明確に理解している度合いで, さまざまな理解度でスコア 0~2
- ②タスクの中で生徒が理解できなかった語彙があるかどうか, Yes か No

コードセット 2: 生徒の言語化における歴史的思考

(1) 証拠と歴史的視点のコード

- ①出典: 著者のアイデンティティ, 経験, 日付, 文書の性質に関する生徒のコメント
- ②視点: 史料またはその著者の視点に関する生徒のコメント
- ③目的: 著者の目的に関する生徒のコメント;
- ④比較: 他の文書や文章との確証や対比

(2) 倫理的次元のコード

- ①公正: 生徒は倫理または公正の原則を述べる(必ずしも歴史的思考の証拠ではない)
- ②距離: 文書の作成時から現在までの時間的距離に関する生徒のコメント
- ③集合的: 生徒は, 集合的責任の考慮に基づいて, 歴史的不正に対する賠償金(またはその他の手段)を課すことに賛成または反対する議論を組み立てる
- ④子孫: 生徒は, それぞれの現在の子孫への利益と損失の考慮に基づいて, 歴史的不正に対する補償(またはその他の手段)を課すことに賛成または反対の議論を構築する

## コーダー間の合意

- ・ コードセット 1…少数を除いて 80~100%  
→少数 (タスク 2: 68%, タスク 8: 51%)  
歴史における一次史料の使用法の知識不足, 問いが著者視点か生徒視点かで混乱
- ・ コードセット 2…ほぼ 60%から 70%, 一部 100%, 30%~40%

## コードセット 2 の体系的分析 (全三段階で構成)

- 第一段階…思考発話の記録からコードを抽出  
タスク 1~9, 証拠と視点の能力を要求

- ①出典：6%～89%の生徒に存在
- ②視点：43%～91%
- ③目的：2%～17%
- ④比較：100%と74%，ある課題のみ20%
- ・文書を事実として読む生徒：14%～44%，痕跡として認識している生徒31%～71%
  - タスク10, 11, 倫理的次元の能力を要求
  - ①公正：歴史的文脈や時間的距離へ言及した場合歴史的思考としてカウント(37%, 49%)
  - ②距離：54%と6%
  - ③集合的：37%
  - ④子孫：46%
- 第二段階…言語化された歴史的思考の証拠をコード間結合
  - 各タスクで、一つでもコードが含まれている生徒の割合を出す
  - 例) タスク1で4つのコードのうち一つでも含んでいた思考発話の割合は○%
    - 32%(タスク7の場合)から100%(タスク4の場合)
    - タスク7(32%)…ウクライナ系カナダ人新聞編集者六人が署名した書簡からの短い抜粋「新聞編集者たちは、自分たちが描写している状況の責任は誰にあると考えたのだろうか?」に答えた四つの選択肢の一つを選択
      - 必ずしも歴史的思考必要なし、抜粋されたものを理解することが必要
    - タスク4(100%)…「ウルリッヒは、ウクライナの囚人たちを、善良で法を守る住民だと描写している。ある文では、ウルリッヒがなぜウクライナ人をモリス神父とこれほど違ったように表現するのかを描写している。」という二つの別々の文書で提示された、アメリカ政府の役人と宗教指導者との間の視点の違いについて説明
      - 生徒は2つの文書の視点を比較することを生徒に明示的に要求されていた
      - 全ての生徒は言語化の過程で視点を比較
  - CR…タスク4, 5, 8, 10, 11
    - 4と11は歴史的思考の証拠を示す生徒の割合が最も高く、5, 8, 10はタスク1, 2, 3, 9などのMC項目のいくつよりも低かった
    - CR項目に関する歴史的思考の一貫した強い証拠はなかった
- 第三段階…言語化の証拠とパフォーマンスの対応
  - もし言語化が歴史的思考の証拠を示すならば、言語化で歴史的思考を示した生徒は、それらのタスクに対するペーパーテストでより高いスコアを持つことが期待される
  - 結果：歴史的思考の証拠を言語化に含めた学生の方がスコアは高い
    - 高スコアと低スコアの学生の間でアルファ=0.05レベルで統計的に有意
    - 36のコードのうち11で統計的に有意な差、25は統計的に有意でない
    - (35というサンプル数の少なさが影響)
  - タスク4の比較とタスク10の公正は、歴史的思考の証拠を提供する生徒とそうでない

生徒との差がない

→タスク 4：文書の視点を比較することを生徒に明示的に要求したから生じた

歴史的思考の証拠と歴史的思考のスコアとの間には何の関係も成立しない

タスク 10：公正の一般原則を述べる→文書に言及しながらであれば歴史的思考が  
欠如していることにはならない

MC タスクよりも CR タスクに対する生徒の言語化から歴史的思考のより強い証拠

→5つの CR タスクのすべてにおいて、少なくとも1つのコードが生徒の書面に基づく  
応答に基づくスコアと統計学的に有意な関連を持っていたが、MC タスクは1つのみ

### 歴史的思考の評価を検証するための推測

思考発話法のデータ…評価課題が歴史的思考に生徒を従事させる認知的証拠を提供

第一段階→各タスクがどのようなタイプの歴史的証拠を解明するかを決定

タスクによって取得される構成要素を理解するために必要な段階

第二段階→どのタスクが一貫した生徒の歴史的思考を必要とするかに関する情報を提供

強力で一貫した歴史的思考条件を持つタスクを評価に含めることができる

ように、タスクを修正または選択するための評価設計段階で有用

第三段階→生徒の言語化における歴史的思考の証拠と歴史的思考スコアの間関係を

調べることは、タスク全体のコードの大部分に対して一貫したパターンを示した  
分析の3段階は、タスクが何を測定しているかに関する補足的な情報を提供

### 思考発話法の限界

①労働集約的、サンプルサイズに限りあり→推論の強度に制限

強い体系的関係があっても統計的有意性は得られない可能性、

中程度または弱い関連は観察されない可能性

②生徒の言語化と能力の証拠との間には一対一の関係はない

生徒の意欲と能力、メタ認知能力、必要な言語化の種類にタスクがどの程度適するか…

③歴史的思考の証拠を含む生徒の割合が最も高いタスクは、歴史的思考を測定する

ための最良のタスクとは見なされない（タスク 4 の問題）

④高度なタスクは、そのタスクをこなせない歴史的思考のレベルの生徒の歴史的思考

レベルについて、何の情報も与えてくれない

### 【結論】

思考発話法とパフォーマンスを結びつけることは、歴史的思考の評価に必要な妥当性の  
証拠を提供する。

そのような証拠がなければ、歴史的思考のいかなる評価も、その評価が本当に測定してい  
るものについての主張を支持することは困難になる。