

15 歴史科目における思考の評価

(HISTORY ASSESSMENTS OF THINKING)

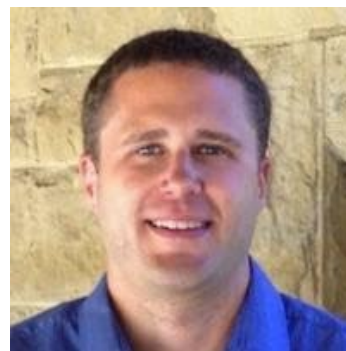
—認知的妥当性の探究 (An Investigation of Cognitive Validity)

Mark Smith, Joel Breakstone

広島大学大学院 小野創太
d200149@hiroshima-u.ac.jp

著者情報

Mark Smith : SHEG (Stanford History Education Group) の評価担当ディレクター。同僚の Joel Breakstone、Sam Wineburg とともに、SHEG の評価についてのウェブサイトである、「Beyond the Bubble」の開発を主導。2014年にスタンフォード大学教育学研究科で Ph.D.、アイオワ大学で中等社会科教育の M.A.T.、ノーザンアイオワ大学で歴史学、政治学の B.A. を取得。それ以前は、アメリカ各州で高校社会科教員に従事。研究関心は、K-12 の歴史科目の評価、特に妥当性と一般化可能性の問題に焦点を当てている。



Joel Breakstone : SHEG の責任者。スタンフォード大学教育学研究科で Ph.D. を取得。同じく「Beyond the Bubble」を開発。全米社会科評議会 (NCSS) よりラリー・メトカーフ模範論文賞を受賞。ブラウン大学で歴史学の B.A.、ダートマス大学でリベラル・スタディーズの M.A. を取得。大学卒業後は、高校歴史教員に従事。研究テーマは、「教師が評価データをどのように指導に役立てているか」。



<https://sheg.stanford.edu/about/people> (2021年2月20日閲覧)

重要用語

- multiple-choice questions : 多肢選択式問題
- Document-based questions : DBQ : 文書資料ベースの課題
- HATs : History Assessments of Thinking : 歴史科目における思考の評価
- cognitive validity : 認知的妥当性
- sourcing : 出所を明らかにすること
- think-aloud : 思考発話

議論したい論点

- ・読んでいて他の HATs への言及がほぼなく、この研究を賞賛するのみだったので、単純に研究成果が「本当か？」となった。この評価課題は他のものと比べて何が優れているのか？
- ・日本で活用する場合、どのようにしていけばよいのか。小テストみたいな位置づけでやっていく？
- ・出所を見ていなくても、歴史に向き合おうとする態度や主体性が見て取れたのだが、こうしたものを「出所を明らかにしていない」でバッサリ切ることは本当に妥当なのか。個別の認知的スキルを評価するのみでよいのか？

イントロダクション

○米国の歴史教師の評価ツールは多肢選択式問題とエッセイ課題が主流であり、ほとんど選択肢がない (Martin, Maldonado, Schneider, & Smith 2011)。

→より広い範囲の評価手段を必要としている。

- ・新たな評価の必要性は、米国の新たなコモン・コア・スタンダードでの要求を考慮すると、より顕著になる。

→一連の歴史的思考の範囲を広げるというもの。(例：複数の歴史的出来事の比較、出典情報の検討、議論の中での証拠の利用、歴史的論証の文章化)

- ・しかし、伝統的な教科書のナラティブを受動的に学習することが中心であることは変わっていない (Paxton, 1999; Ravitch & Finn, 1987; Wineburg, 2007)。

◎コモン・コアの野心的な目標を達成するためには、教師は離散した (discrete) 多肢選択式問題や、長くて退屈な文書資料ベースの課題 (Document-based question、以下 DBQ) 以上のものを必要とする。新しいタイプの歴史科目の評価を生み出す必要があり、新たなツールがどのように機能するかをより良く理解するための研究が求められる。

歴史的思考の評価 (Assessments of Historical Thinking)

○多肢選択式問題⇔DBQ → それぞれスペクトラムの反対側に存在する。

多肢選択式問題：研究者側は主に事実の想起と認識を求め (Reich, 2009; Wineburg, 2004)、生徒は非歴史的な受験戦略を用いる (Smith, 2011)。解答に至った生徒の思考が明らかにされないため、教師には限られた情報しか提供されない (Haney & Scott, 1987; Madaus, Russell, & Higgins, 2009)。

DBQ：複雑な課題。歴史的な問題を、一連の文書資料から答える。(問いの解釈→文書資料の読解→主張を裏付けるための文書資料の使用の吟味→エッセイ・ライティング)

→コモン・コアの目標と一致はするが、**限界**がある。

APを受講する生徒への調査

- ①歴史的な史資料として分析するのではなく、適切な引用や事実を探するために文書資料をあさってしまうことが多い (Young & Leinhardt, 1998)。
- ②場合によっては、文書資料を注意深く問わなくても、しっかりとした歴史的論証にまで高めることもできる。
- ③課題が複雑であることから、何を評価するのが分かりにくい。(例：歴史的論証を展開する能力なのか、原資料を素早く処理する能力なのか、など。)
- ④生徒が作成するエッセイが長いことから、教師が注意すべき特定のスキルや概念をすぐに見極めることが困難。

☞歴史教師は、生徒の思考に関してより良い情報を得ることができる対処しやすいタスクから利益を得ることができる。

○これを受けて、多肢選択式問題の単純さと DBQ の複雑さの間の空白を埋めるために、タスクの構築、パイロット、改訂を始めることにした。

・開発プロセスの一環として、歴史的思考の構成要素を明示的に定義した (cf. Ercikan & Seixas, 2011; Holt, 1990; Lévesque, 2008; Wineburg, 1991)。

→現在の計量心理学の実践の指針となる領域の概念が貧弱であるとの Stake (2007) の批判に応えるため。

・歴史的思考の構成要素を、「歴史的知識」、「証拠の評価」、「証拠・論証の利用」とした。この3つの柱を構成する具体的なサブプロセスも定義 (表 15.1 を参照)。

◎この定義をする作業が、中等教育課程の生徒が進むべきプロセスを生み出し、新たなタイプの評価の開発の指針となった。

○HATs (History Assessments of Thinking : 歴史科目における思考の評価) の開発

・歴史的思考のマッピングの後、歴史的思考の特定の側面に対応する短い課題のドラフトを作成した。各項目では、生徒は一次資料を分析し、推論の基礎となる洞察力を提供する短い回答を書くことを求められた (各課題は 10 分以内に終了させることができる)。

・例①：文書資料が作成された日付に注意を払っているかどうかを評価する課題 (「1621 年における初めての感謝祭：The First Thanksgiving, 1621」と題された 1932 年の絵画を取り上げたもの) (詳細は p.236)

問い：「1621 年のピルグリムとワンパノアグ・インディアンの関係を理解しようとしている歴史家にとって、この絵画は役に立つか？」

・HATs を試験的に実施した際、多くの生徒はこの資料の特性 (attribution) を完全に無視していた (生徒の反応は p.237)。

→3 世紀後に制作された絵画の限界を量るのではなく、多くの生徒は絵画の内容のみに焦点を当て、出所の情報を無視した。

・しかし、日付の重要性をしっかりと理解している生徒もいた (生徒の反応は p.237)。

・例②：生徒が公民権運動のナラティブの横糸（arc）をどれだけ理解しているかを測ることを目的とした課題（図 15.2）

問い：「**2つの手紙（日付が削除されている）のどちらが早く書かれたか？**」

手紙A：当時のファーストレディが NAACP（National Association for the Advancement of Colored People）の事務局長に宛てたもの。アフリカ系アメリカ人に対する残虐なリンチを止めるために国家レベルで介入することに大統領が消極的であったことに言及している。

手紙B：従来、全てが白人で構成されていた学校でのアフリカ系アメリカ人の子どもたちが直面していた問題について述べており、大統領が懸念していたことに言及している。

・人種隔離の問題がBで取り上げられ、Aが当時のその問題への対応？

→**間違い**。1950年代に最高裁が学校の人種隔離を命じるまでに、リンチは事実上根絶されていた。さらに、公民権運動が進むにつれ、連邦政府の介入が増えていた。

→生徒の理解の重大なギャップについての情報を教師に提供をしている。

→ここまでのパイロットの結果は満足できるものだが、各項目の**妥当性**をさらに研究した。

妥当性（Validity）

○認知的妥当性（cognitive validity）についての考察

・特定の評価が実際に測定するように設計されたものを引き出すのかどうかを検討。

・同時並行的および回顧的な**思考発話（think-aloud）**プロトコルは、認知的妥当性についての証拠を収集するための試みの中心。

研究方法（Methods）

・思考発話インタビューの項目は、専門家のレビューと大規模な教室でのパイロットを含む厳格な項目開発プロセスに既に持ちこたえてきたもののみを使用した。

参加者（Participants）

・中西部の都市部の公立学校区の高校3年生を戦略的にサンプリングした。

→年間のアドバンスド・プレースメント（AP）の米国史コースを修了し、国家試験（national exam）（AP試験？）で3以上（5段階評価）を獲得しており、生徒の出身州にある3つの公立大学のうち2つで単位を取得。

→生徒が**テスト対象の教材を学習する機会が与えられていたこと、リテラシーの要求が不必要な障壁とならないことを保証するもの。**

研究手続き (Procedures)

- ①研究者全員がインタビューを行う前の研修を受けた。
- ②インタビューのウォームアップを実施した。
- ③ウォームアップの項目を終えた後、生徒が1人ずつ HATs を完成させた。
→生徒に問題を解かせながら中断することなく同時並行的に思考発話をする。生徒が3秒以上沈黙した場合にのみ、研究者がリマインダーを差し挟む（「何を考えているのですか？」と問うことで思考発話を促す）。
- ④問題を完成させたのち、生徒に回顧的に思考発話をするよう求め、(もし可能であれば) 生徒の思考をさらに探るための一連の質問をスクリプトに沿って行った。

・次の分析 (Analysis) については省略します。どのように分析したか (録音から文字起こし→分析までの流れ) が常識的に書かれています。

結果 (Results)

- 分析の結果、HATs は歴史的思考の重要な側面を実際に活用していることが明らかになった。
 - ・例えば、感謝祭の HATs では、**出所を明らかにすること (sourcing) の重要な側面における生徒の習熟度を示す指標を一貫して示した** (p.241、スーザンの事例)。
 - ・スーザンの書面による回答は、彼女の思考の洗練度を反映していた。問題を解決するために使用された認知プロセスと、記述された解答によって提供された思考の証拠との間のこうした整合性は、生徒の歴史的思考を測定するために HATs を用いることの妥当性に関する論拠を補強するものであった。
 - 口頭と筆記の回答は、出所を明らかにする重要な側面に関与する能力についての推定を裏付けるものであることを明確に示していた。
- 予想通り、歴史的思考の他の側面にも取り組んでいた。
 - ・出所を明らかにすることに加え、しばしば感謝祭の文書資料が作成された文脈を考え、証拠を用いて回答を作成していた。
- 他の HATs でも同様のパターンが見られた (出所を明らかにするスキルを使用していた)。
- HATs は主な誤概念化 (misconceptions) も検出した (p.242、ジョナサンの事例)。
 - ・正しい回答と同様に、こうしたタイプの不正確な回答は、中等段階の生徒の間で歴史的思考を測定するために HATs を使用することについての妥当性の論拠を支持する証拠を提供していた。
 - ・HATs とは対照的な全米教育発達評価 (NAEP) での多肢選択式問題でも思考発話研究を行った (2010年に実施された12年生のアメリカ合衆国史の試験で出題された多肢選択式問題)。

→意図された高次の思考プロセスに触れることがほとんどないことが一貫して明らかになっている (p.243、ジュリアの事例)。

→多肢選択式では、単に単語を推測して正解にたどり着くことが出来る場合もあったが、HATsでは観察されなかった。

結論 (Conclusion)

・研究では、テストが授業の指導を形作ることが示されており、歴史的思考の有効なテストが無ければ、歴史的思考の指導が授業の実践の一部になるとは考えにくい (Frederiksen, 1984; Koretz, 2008; Madaus, West, Harmon, Lomax, & Viator, 1992)。

・残念なことに、現在生徒の学習を測定するために使用されているテストは、こうしたスキルを適切に測定していない可能性がある。

・HATsは、我々が21世紀の市民の間で開発しようとしている思考を引き出す評価のニーズを満たすために開発された。