

報告番号	※	第	号
------	---	---	---

## 主 論 文 の 要 旨

論文題目

センター試験・SAT・AP 試験「世界史」  
で測定される歴史的知識と歴史的思考力

氏 名

中 切 正 人

## 論 文 内 容 の 要 旨

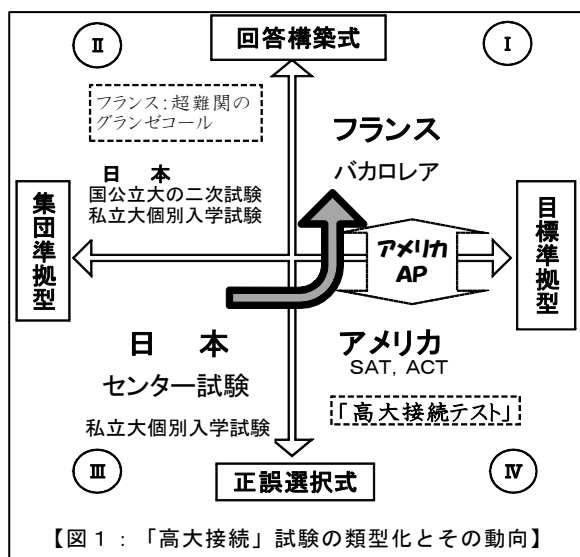
### 1. 本研究の目的およびその背景と分析対象・分析方法

本研究は、高大接続に関わるわが国の大学入試センター試験（以下センター試験と表記）「世界史」と、アメリカのSAT（Scholastic Assessment Test：以下SATと表記）科目別テスト「世界史」およびアドヴァンスト・プレイスメント・プログラム（Advanced Placement Program：以下APと表記）試験「世界史」で測定されている、「歴史的知識の質」と「歴史的思考力や技能」を比較し、日米の試験で測定されている学力の特色を明らかにすることを目的とする。そして、その分析結果を補強し、その背景的理解を深めるために、AP「世界史」コース（カリキュラム相当）と試験で育成かつ測定されている学力の全体像を示し、わが国の「世界史」教育で育成かつ測定されている学力と対比する。

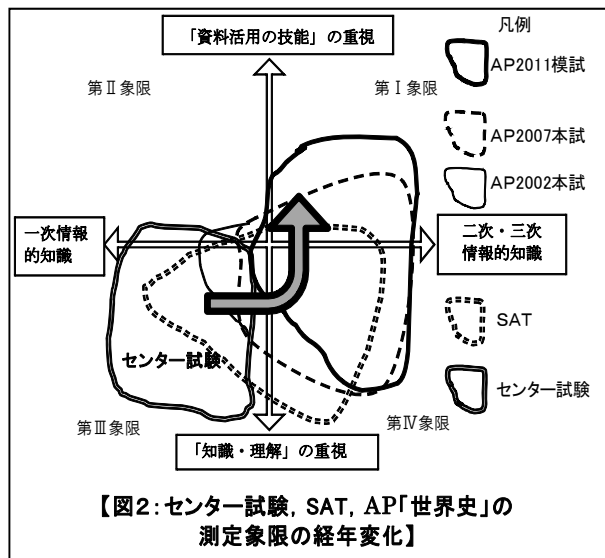
グローバル化する知識基盤社会（knowledge-based society）における知識は、周囲に共有されながら学際的な課題解決を志向して新たな知見を生み出す動的な広義の知識（OECD, 2008）や「新しい能力」（松下, 2010）であり、その中には「歴史的思考力」が含まれる（『高等学校学習指導要領解説地理歴史編』）。しかし、日本学術会議（2011）によると、わが国の後期中等「歴史教育」では、車の両輪として「歴史的知識の伝達」と「歴史的思考力の育成」が進められるべきところが前者に偏重し、大学入試では静態的な狭義の知識を問う傾向が強いことから、高校では教科書叙述の丸暗記が助長され、センター試験のマークシート方式では歴史的思考力を問うような出題は困難である。

大学入試とほぼ同義のセンター試験で測定される学力に関わる先行研究を分析すると、「多肢選択式試験で測定可能な歴史的知識の質および歴史的思考力や技能の分析<第一の課題>」が不十分であり、その分析の枠組みとして「歴史的思考力や技能の定義および構造が不明瞭<第二の課題>」な上に、「センター試験『世界史』の各設問で測定されている学力の明示<第三の課題>」がなされていない。よって、本研究の分析対象は、「第一の課題」と「第三の課題」を合わせた、「センター試験『世界史』を含む多肢選択式の大学入試で測定可能な歴史的知識の質および歴史的思考力や技能」である。ただし、わが国の多肢選択式試験では抽象度の高い歴史的知識、および、歴史的思考力や技能がほとんど測定されていないことから、海外の高大接続に関わる試験を類型化したところ【図1】、アメリカでCollege Boardにより多肢選択式の試験科目「世界史」がSAT科目別テストとAP試験で実施され

ているため、これらを分析対象に選定した。なお、【図1】の縦軸は試験の回答形式に着目して、回答があらかじめ提示されている「選択式」と提示さない「回答構築式」に大別し、日米のような試験の信頼性と公平性に重きを置く正誤選択式試験を行うグループと、当初のエリート育成機能を有する論述式試験を保持するフランスを対置させた。横軸は高大接続の二側面として「教育接続：学校教育の連続」と「選抜接続：入学志願者からの選抜」に着目し、前者には学習目標への到達度が大きな「スコア：段階」で絶対評価される「目標準拠 (criterion-referenced)」型試験が、後者には「1点刻み」の相対評価による統計で選抜を目的とする「集団準拠 (norm-referenced)」型試験が対応している。



【図1：「高大接続」試験の類型化とその動向】



【図2：センター試験, SAT, AP「世界史」の測定象限の経年変化】

分析対象の試験は、限定的に公開された SAT 科目別テスト「世界史」(2002年；2005年)と AP 試験「世界史」多肢選択部 (2002年；2007年)、さらに 2011年 AP「世界史」夏季講習会 (7/24～7/29：スタンフォード大学)において、2012年からの改訂 (選択肢が5個から4個に減少、誤答に対する減点措置の消滅、歴史的思考力の育成と測定の強化)の周知のために、異例の措置として配布された 2011年模擬試験 (完全版)、そして、以上の選定過程や公開年度を考慮したセンター試験 (2007年度；資料活用の多い 2009年度)である。また、分析に当たっては、試験で測定される「歴史的知識の質」と「歴史的思考力や技能」を分析軸として設定し、この二本の分析軸が構成するマトリクスの中に3つの試験の各設問を1つ1つコーディングすることによって、3つの試験が位置付けられる象限を検証した。そのマトリクスの【図2】の横軸「歴史的知識の質」では、歴史教育関係者の共通理解を参照して三段階に区分し、歴史的知識の基層部に個別的・事実に知識を一次情報的知識として位置付け、その上に一次情報的知識の解釈や因果関係の分析が施されて得られる二次情報的知識、さらに転移・応用性を高めた三次情報的知識を想定し、一次情報的知識と二次・三次情報的知識の境目は歴史的思考力の活用の差異が顕著であることからその境目を判別した。次に、縦軸「歴史的思考力や技能」では、「知識・理解」と並び「技能」は「見える学力」として客観式試験で測定され得ると見なし (梶田, 2008)、地理歴史科・公民科特有の「資料活用の技能」は歴史的思考力の一部として機能しつつ、資料を読み取る行為自体は選択肢に反映されていると考え、この技能の測定を判別した。

## 2. センター試験・SAT・AP試験「世界史」で測定されている学力の特色

3. 「<中略>もし人が奴隷の目をつぶすか、奴隷の骨を折れば、彼はその値の半額を払う。」

引用したハンムラビ法典は、バビロニア社会に関する下記のどれを例証するか。正解(C)：第Ⅰ象限

(A) すべての階級の経済的幸福に関する規定を設けた。 (B) 体罰依存から離れた。

(C) 社会的不公正を特徴とした。 (D) 王は神の力によって神聖であると見なされた。

上は【図2】のコーディングの事例で、2011年AP模擬試験「世界史」第3問であるが、これは資料の分析・解釈の過程を経て獲得された二次情動的知識が測定されているため第Ⅰ象限に位置づけられる。ただし、この法典は、当時としては報復の限度に合理的な規準が設けられ、奴隷は法文の範囲内で刑罰を受ける法の保護が確認されるという解釈に従えば(吉川, 1968)、選択肢(B)のように「体罰以外のお金」で処理され、かつ、奴隷といえども折られ損にならずに明文化された後始末が保障されたと考えることも可能で、わが国では教育関係者や予備校からクレームがつくことが予想される。

このような分析過程を経て作成された【図2】では、3つの試験の各設問を1つずつコーディングした設問数全体がそれぞれ一塊の面積を持つ図形として表され、センター試験では一次情動的知識しか測定されていないのに対し、SATでは3分の2の設問で高次情動的知識が測定され、AP試験では資料を活用して高次情動的知識を測定する第Ⅰ象限の比率が年度を追う毎に上昇し、【図2】中の矢印のように、「センター試験→SAT→AP2002→AP2007→AP2011 模擬試験」の順に、「第Ⅲ象限→第Ⅳ象限→第Ⅰ象限」へ向かってその測定の中心が移動していることが確認された。

ここに、【図1】中の矢印で日米の試験の移行動向「わが国＝SAT タイプの『達成度テスト』へ移行；アメリカ＝論述重視のAP試験の急速な拡張」を並べて比較考察してみると、両図に配置された各象限の試験が共通していることから、両図にはある程度の関連性が想定される。まず、両図において共に第Ⅲ象限に位置するセンター試験は、選抜を目的として一点刻みの点数が合否判定に機能する集団準拠型試験であることから信頼性が最優先される結果、試験形式は主観の入り込みにくい多肢選択式に統一され、さらに、設問では主観的解釈の入り込みやすい資料の活用が控えられると同時に、解釈が挿入されにくい一次情動的知識が測定されている。また、選抜の公平性を確保するために「初出」の試験問題が実施され、試験問題の「完全公開性」によって公正さが担保されている上に、測定される内容は学習指導要領に準拠した一次情動的知識の共通性の高い教科書が支えており、センター試験「世界史」は「試験問題の公開制度」と「設問の信頼性と正解の一義性」と「教科書の一次情動的知識の共通性」とが三位一体となって機能していると考えられる。これに対して、第Ⅳ象限のアメリカでは、SATとAP試験において、測定する学力が到達目標に達しているかどうか判定する目標準拠型試験が行われ、AP試験では設問の信頼性よりも妥当性の方が重視される論述部を含む試験結果がある一定の「段階：スコア」で評価され、多肢選択部では主観的な解釈が挿入されやすい高次の概念的知識が測定され、項目応答理論IRTに基づいて項目プールに保管されている設問が繰り返し使用されるため「非公開」となっている。その上、20世紀末まで歴史教育の全米基準が無かったアメリカでは、その抽象性の高さゆえに却って共通性が高い高次情動的知識が主に測定されている。つまり、SATやAP試験「世界史」では「試験問題の非公開制度」と「設問の妥当性と正解の多義性」と「教科書の高次情動的知識の共通性」とが三位一体となって機能していると考えられる。

この日米両国の対照性について、本研究の主題と離れる試験制度(公開・非公開)は脇に置くが、アメリカの「正解の多義性」を支える要因には目標準拠型試験に伴う「設問の妥当性」以外に何が存在し、そしてそれは「教科書の高次情動的知識の共通性」にどのように関わっているのか分析・考察

を進めることが必要である。その分析対象は SAT 受検者の 16 倍以上の受験者(188,417 名:2011 年)を集める AP 試験であり、分析に当たっては、多肢選択部と配点を折半している論述部で測定されている歴史的思考力に着目しながら、多肢選択部の位置づけを AP「世界史」コースと試験を合わせたプログラム全体の中で総合的に分析・考察した結果を以下に示す。第 1 に、AP「世界史」学習が記憶の訓練の場ではなく思考方法や問題解決の場であることが強調されるようになった流れに沿って、年度を重ねる毎に高次の情動的知識の測定が増加した。第 2 に、論述部では回答された三次情動的知識そのものではなく二次情動的知識を元に三次情動的知識を編集する思考過程の方が重視される評価基準に対応する形で、多肢選択部ではその土台となる二次情動的知識の測定が基本となっており、歴史的知識の質に関わる多肢選択部と論述部との相互補完関係が見られる。第 3 に、多肢選択部で「資料活用の技能」の測定が増加してきた背景には、2012 年から論述部において歴史学の基本「歴史的証拠に基づく歴史的論証」を中核に据える姿勢が明確に打ち出されたことが反映されている。そして、以上の観点が AP「世界史」教科書にも反映され、歴史的思考力を鍛える素材として二次・三次情動的知識が中心に叙述され、かつ、掲載された多彩な資料を活用した歴史的思考力の育成が図られながら、より高次の情動的知識を理解し、習得する素材として教科書が機能していると考えられるのである。

### 3. 本研究の結論と今後の課題

以上より、わが国のセンター試験「世界史」では歴史的思考力の一部をなす「資料活用の技能」の測定が少なく一次情動的歴史知識に偏った測定がなされているのに対し、アメリカでは高次の情動的歴史知識が測定され、特に AP 試験「世界史」においては多様で豊富な資料を用いた「資料活用の技能」の測定に伴って、ほぼ高次の情動的知識に偏重した測定がなされているという対照性が確認された。そこでは、センター試験側の「試験問題の公開制度」と「設問の信頼性と正解の一義性」と「教科書の一次情動的知識の共通性」の三位一体的機能に対し、AP 試験側の「試験問題の非公開制度」と「設問の妥当性と正解の多義性」と「教科書の高次情動的知識の共通性」の三位一体的機能が対照的特色として析出された。さらに、AP プログラム(コースと試験)「世界史」では、「歴史的知識の質」と「歴史的思考力や技能」が相互補完的關係で歴史教育の両輪として機能し合い、歴史理解を深めることに貢献している。しかし、AP 試験で測定されている高次情動的知識は、センター試験が測定する一次情動的知識に比べ、多様な歴史的思考力が機能することに対する裏腹の關係として、解釈や価値判断等が挿入されやすいことから、設問の良否が問われたり、正解の多義性や試験の信頼性の低さが指摘されたりする可能性を避けて通ることが難しいことが検証された。この点で、センター試験の設問のイデオロギー的中立性と、正解の一義性および試験の信頼性の高さは好対照をなしている。梶田叡一(2008)「新学習指導要領の実施に伴う移行措置をとらえるために」文溪堂『hito\*yume』1:10-15。

松下佳代(2010)「<新しい能力>概念と教育」松下佳代編(2010)『<新しい能力>は教育を変えるかー学力・リテラシー・コンピテンシーー』ミネルヴァ書房:1-42。

日本学会会議(心理学・教育学委員会・歴史学委員会・地域研究委員会合同)高校地理歴史科教育に関する分科会(2011)『提言「新しい高校地理・歴史教育の創造ーグローバル化に対応した時空間認識の育成ー」』(平成 23 年 8 月 3 日)

OECD, (2008). *Tertiary Education for the Knowledge Society: Volume1: Special features: Governance, funding, quality*, Paris: OECD.

吉川守(1968)「ハンムラビの時代」岸本通夫他『世界の歴史 2 古代オリエント』河出書房新社:81-118。