

新 SAT テストについて

— 改訂の背景ならびに改訂点を中心として —

石岡 恒憲, 鈴木 規夫 (大学入試センター)

アメリカの4年制大学への進学希望者を対象に行われる全国共通試験の一つである SAT が 2005 年 3 月に改訂された。大学の授業内容により即した試験内容に変更され、エッセイ試験が追加された。また Analogies など学力の判定への寄与が少ないと判定された幾つかの出題形式が削除された。本稿では、今回の改訂の背景と、改訂に至った幾つかの研究成果について紹介する。また今後の方向について言及する。

1 はじめに

SAT とはアメリカの教育団体である College Board が主催・運営しているアメリカの4年制大学への進学希望者を対象に行われる全国共通試験である。大学での学習についていけるだけの論理的思考力 (critical thinking) があるかどうかを測定することを目的としている。

SAT は 1994 以降、問題解決力 (Reasoning) を測る SAT I と従来からあった Achievement Test (達成度判定テスト) を改称した SAT II (科目別テスト) の 2 つがあった。2004 年には数字の I と II は脱落して、それらのテストはそれぞれ SAT Reasoning Test と SAT Subject Tests となった。単に SAT と呼んだときは、SAT Reasoning Test のことを指す。SAT はいまや何の略語 (頭文字) でもない。

2005 年 3 月から新しい SAT が導入され、大学の授業内容により即した試験内容に変更された。

2 節では、新しい SAT の出題内容/スコアについて要約する。3 節では今回の改訂が行われた理由と、それを支持する研究成果について報告する。4 節には、今回新たに設けられたエッセイ試験において、採点基準および採点方法について、より詳しく紹介する。

2 新 SAT テストの概略

2.1 出題内容

テストには、(1) Reasoning Test と (2) Subject Tests の 2 種類がある。Reasoning のもともとの意味は「問題解決」であるが、「論理的思考力」として理解することができる。新しい SAT では Reasoning Test に Writing が加わり、それによって Subject Tests にあった Writing がなくなった。なお、SAT 公式サイトにある SAT Preparation Center では、各セクションの解説や練習問題を紹介している。

(1) Reasoning Test 出題内容

Reasoning Test は以下の 3 セクションから成る。また、スコアにはまったく影響しない Unscored Section (25 分) があり、3 セクション (Writing のエッセイを除く) のいずれかから出題される。試験時間は合計で 3 時間 45 分である。

Critical Reading (70 分)

従来の SAT の Verbal Section に当たり、語彙力と文の構造、理論の展開を理解する技量が測られる。大別すると、Sentence Completions (文章完成) と Passage-based Reading (読解) の 2 つから成る。Sentence Completions では、2 カ所の空欄を埋める穴埋め問題が、Passage-based Reading では、長文・短文の与えられたパッセージの内容や書き手の意見、さらに文章から推測される内容を問う問題が出される。

Verbal Section からの大きな変更点とし

ては、Analogies (類推)が削除され、新たに設けられた Passage-based Reading では、従来の Critical Reading で出題された long reading passages に short reading passages が加わっている。

Writing (60分)

新たに加わったセクションで、英文法の知識や文章構成力の技量が測られる。Multiple-Choice (多肢選択問題)と Short Essay (エッセイ)の2つから成る。Multiple-Choiceには、Identifying Sentence Errors, Improving Sentences, Improving Paragraphs の3種類があり、文章構造や文法、語法などの間違い探しや、単文・エッセイの主旨を読み取って答える多肢選択問題が出題される。Short Essay では、与えられたトピックについて自分の見解をまとめる。単語や文法の正しい用法はもちろん、明確でかつ論理的に仕上げることがポイントになる。試験時間は Multiple-Choice が 35 分、Short Essay が 25 分である。英語が母語でない受験者にとっては、難易度が高いといえる。

Math (70分)

Math は、数学的な能力を測るセクションである。問題は主に、Statistics (統計学)、Algebra (代数)、Geometry (幾何)などから出題される。出題は Multiple-choice (多肢選択問題)が中心だが、Student-Produced Response Question (通称、grid-in)という形式の問題が 10 問採用されている。grid-in の場合、計算して出した解答をマークシートに記入し、それぞれ回答欄(grid)に該当する数字の番号(および/と、を示す楕円)を塗りつぶさなければならない。塗りつぶしにはいくつかのルールがあるので、事前に知っておくことが必要である。なお、このセクションには、計算機の持ち込みが許されている。

従来の Math Section から Quantitative Comparisons (数量比較)が削除されたが、estimation (推計)や number sense (数感

覚)などの重要な技量は Multiple-choice で測定されるようになっていく。以前は含まれていなかった代数Ⅱ、具体的には指数成長(exponential growth)、絶対値、関数表記、負や分数の指数などが新たに含まれるようになった。

(2) Subject Tests 出題内容

Subject Tests には、英語(つまり国語)、歴史・社会学、数学、自然科学、語学の5分野がある。これらがさらに細かく分けられており、英米文学、アメリカ史、世界史、数学(レベル1、レベル2)、生物学、化学、物理、リーディング問題のみの語学(フランス語、ドイツ語、近代ヘブライ語、イタリア語、ラテン語、スペイン語)、リスニング問題を含む語学(中国語、フランス語、ドイツ語、日本語、韓国語、スペイン語)がある。

受験科目は、あらかじめ指定する大学もあれば、出願者が選べる大学もある。また、スコアを求めない大学もある。この Subject Tests のスコアは、学生にとって志望校へ自分の得意分野をアピールする機会となる。一方、大学にとっては入学審査の判定基準としてだけでなく、学生の入学後のコース振り分けや、進路相談の参考とするなどの利用価値もある。テスト時間はいずれも1時間で、形式は多肢選択式となっている。

2.2 スコア

Reasoning Test のスコアは、Writing、Critical Reading、Math の各セクションで、それぞれ200~800の間の10刻みで表される。Writing セクションでは、多肢選択問題を20~80、エッセイを2~12というサブスコアもレポートされる。このエッセイのサブスコアは、中にはあくまでも参考資料とするところもあり、各大学で扱いが異なる。

Subject Tests のスコアも、科目ごとに200~800の間で表される。語学には、サブスコアがレポートされるものもある。

スコアはオンライン(要登録)、電話、郵送

の3つの方法で知ることができる。通常、オンライン、電話は受験日から約2週間後、郵送は約3週間後に通知される。4校までは無料(受験料に含まれる)で、受験時に指定した希望大学へ College Board 本部から直接送付される。追加でスコアの送付を希望する場合は、1校当たり US\$ 9.0 かかる。

オンラインのスコアレポートは無料で24時間確認できるだけでなく、エッセイの成績通知、正答・誤答・無回答数、素点、スコア履歴などの確認ができ、利用価値が高い。

3 SAT改訂の理由

3.1 ブルーリボン委員会

College Board によれば、改訂の理由は、受験生が高校時代に学習してきたことをより反映させるためだという。Writing を含めることは、それが大学や将来において成功するための重要なスキルであると考えられることによる。Math に代数Ⅱが追加されるのは、高校生の70%が卒業前年の終りまでに代数Ⅱ(もしくはそれと同等科目)を履修していること、および進学希望者の97%が3 years of math (9-12年生の3年次の数学)を終えており、69%が4 years of math を終えていることによる。4年生大学の大半は、入学要件として3 years of math を要求する。Analogies が削除されたのはこれが高校のカリキュラムと直接的に結び付いていないことによる。

一方で今回の変更は、カリフォルニア大学による働きかけによるものと言われている。事実、カリフォルニア大学の研究グループは Analogies を削除し Writing を加えるべきであるという助言をしている。彼らの研究によれば、旧 SAT に Writing Subject Test を加えることで全体的な予測妥当性が .06-.08 上がるという。Analogies については、その出題形式が語彙の単純暗記を促していると指摘している。

College Board もカリフォルニア大学の助

言に注目してきたことを認めている。しかし College Board の主張によれば、改訂は(カリフォルニア大学を含む)全てのメンバーからのフィードバックを考慮しているのであり、他の多くの要因も変更にも寄与しているという。実際、今回の改訂の種は、1990年のブルーリボン委員会とその結果である“Beyond Prediction (予測を越えて)”という報告書にあり、1994年の改訂はこれに基づくものであるという。(1994年の改訂では、反意語 antonym の問題が削除され、より長文の読解が加えられた。数学では grid-in と呼ばれる自由解答形式の問題が加えられ、電卓の使用が認められた。)しかしながら、Writing Section を加えるべきであるという提言は、検討されたものの採択には至らず、今回はじめて導入されることになった。

Writing Section が当時、採択されなかった理由の1つは、技術的な処理能力の問題、すなわち百万以上もの受験生のエッセイを採点のためにプロの評価者へ転送することができなかったことによる。現在は Pearson Educational Management 社のインターネットに基づいた採点システム ePen を用い、これを可能としている。

3.2 フィールド試行テスト

2006年度入学に向けて2005年3月に新SATが導入されることは、2002年6月にアナウンスされていた。新SATの実施に向けて実に多くの事前研究が実施されている。これらの研究は、全て

www.collegeboard.com/research/から入手できるが、ここでは重要だと思われるものを以下の3つの段階に分けて紹介する。

(1) フィールド試行テスト以前の研究 信頼性確保の問題:

新SATの最も突出した変更は、Critical Reading の変更と Analogies 問題の削除である。これらを変更、削除しても測定精度の信頼性が保証されていなければならない。そ

ここで実際の SAT のデータを用いて検証が行われた。その結果、Analogies の問題を削除しても Verbal のテストの信頼性が維持できることが可能であることがわかった。その一方で、アイテム困難度の分布を修正する、具体的にはスコアスケールの分布の両裾において精度を高める必要のあることがわかった。言い替えれば、非常に易しい問題と、非常に難しい問題を他の Verbal タイプの問題(文章完成と読解)に増やす必要があることがわかった(Liu, Feigenbaum, and Cook, 2004)。

拡張した数学の問題:

新 SAT ではより進んだ内容(代数 II)が追加された。このことが被験者の成績に及ぼす影響を検証した結果、この内容の存在それ自体が成績に影響を与えることはほとんどなく、むしろ設問の困難度に依存することがわかった(Liu, Schuppan, and Walker, in press)。

疲労の影響:

新 SAT では試験時間が 45 分、増えることになる。College Board ではこの試験時間の延長が被験者の疲労をもたらさないこと、及び成績を低下させないことを保証したいと考えた。テストにおける疲労を 100 人の学生を用いて調べた結果、読解および数学の試験において被験者は 5 時間から 6 時間で疲労に耐えられなくなることがわかった。また複雑なタスクよりも単純なタスクの方がより多くの疲労をもたらすことも示された(Liu, Feigenbaum, Oh, and Burton, 2004)。

エッセイのタイプ:

Writing セッションにおけるエッセイ試験でどのようなタイプの質問文にするかを決定することは重要である。新しい SAT で提案された新しい質問文のタイプに対して、人種、母語、性差の違いで成績の違いが生じるかを調べた結果、特定のグループに不利であるとは認められなかった(Breland, Kubota, Nickerson, Trapani, and Walker, 2004)。

(2) フィールド試行テスト

2003 年 3 月に新しい SAT をより周知させる目的で、フィールド試行テストが実施された。実施側からみた場合の目的は、新しい SAT の内容、統計的性質、時間配分を評価し、そこで得られた SAT スコアが現在(当時)の SAT スコアと比較可能であるかを調べることである。679 の高校から 45,000 人以上の学生が公立/私立、都会/地方、全米の全ての地理的地域の別に渡って集められた。特に人種、民族の違いに基づく研究結果を保証するために、アフリカ系アメリカ人、およびスペイン系(ヒスパニック)やラテン系の学生が高い比率で集められた。結果は以下のようなものであった(Liu and Feigenbaum, 2003)。

- ・新しい Critical Reading と Math のセッションはテストの困難度に影響を与えない。すなわちそれぞれのアイテムの困難度は正常な範囲に納まった。
- ・新しい SAT の信頼度は現在(当時)のそれと極めて同じ程度である。
- ・新しい SAT スコアと現在(当時)のそれとの相関は、3 つの全てのセッションにおいて極めて高い(.95-.97)。すなわち新 SAT の Critical reading と Math は、現 SAT の Verbal と Math に等しいと見なすことができる。
- ・性や民族の違いによるスコアの差を悪くすることはなかった。

(3) フィールド試行テスト後

フィールド試行テストの後もさらに SAT を洗練させるために幾つかの研究がされた。そのうちの一つは、制限時間を様々に変えて、学生が答えることのできる問題数との関係の評価である(Allspach and Walker, 2004)。その結果を受けて、Writing スコアの信頼性を上げるために、Writing のセッションを 10 分多くすることが決まった。別の研究では、大半の学生はテストの最初にエッセイを選択する傾向があり、試験の最初にエッセイ試験をおいた方が、少しスコアがよくなることがわ

かった(Oh and Walker, 2003)。このために、新 SAT ではエッセイは常に最初に置かれることになった。

新 SAT の予測的妥当性:

新 SAT テストに Writing セクションを追加することの最終的な目的は、大学での成功(college success)を予測するための妥当性を向上させることにある。College Board との協力のもとで American Institutes for Research 社は、13 大学から約 1,200 名のデータを基に、新 SAT Writing スコアの予測妥当性を検証した。その結果、新 SAT Writing スコアの総計と大学 1 年次の成績との相関は.46 で、英語の文法(composition)との相関は.32 であった。この結果は前の SAT に比べ、大学 1 年次の成績の予測においては少し良くなり(.43 → .46)、また、英語の文法においては少し悪くなった(.35 → .32)。また、Writing セクションを追加することにより、前の SAT と高校の成績の妥当性を.01-.02 増加させる。この結果は Writing セクションの追加が SAT の予測的妥当性を増加させ、実施する価値のあることを示唆している(Norris, Oppler, Day, and Adams, 2004)。

4 エッセイについて

4.1 評定の観点

ショートエッセイは、批判力やアイデアを掘り下げる能力や思想に富んでいるか/説得力があるか/首尾一貫しているか、といった観点から評価測定するように設計されている。

エッセイでは、問題文が直接与えられる場合と、資料や短いパラグラフを提示しそれに対して意見を求めるものと 2 つのタイプがあるが、いずれにしても自らの経験や学習したこと、及び質問された問題への視点について様々な観点から記述することが要求される。

受験生は自分のライティング・スタイルに最も適した修辞方法で記述してよい。たとえば議論を展開してゆくスタイルでも良いし、比較・対比を用いても良い。他のテクニク

でも勿論よい。College Board の The Essay と題された Web には、エッセイ試験での問題例が紹介されている。

「問題：人は富や名声よりも個人的な成功をなしとげることにより動機付けされる。この問題についてあなたの視点を掘り下げたエッセイを計画し、記述しなさい。理由付けや読書、学習、経験、観察した例を示しながら自分の見解を支援しなさい。」

Assignment: Are people motivated to achieve by personal satisfaction rather than by money or fame? Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations.

スコアリング・ガイドに示されている評価の視点は以下の通りである。

- ・テーマに対して効果的かつ洞察力をもって意見の展開がされており、批判や例示等による補助的な情報提示があること。
- ・組織化されており、焦点が明確であり、アイディアの挿入がスムーズであること。
- ・言語の使用が効果的であり、語彙が多いこと。
- ・文脈の構成が意味を持ちながら多様であること。
- ・文法上の誤りがないこと

4.2 解答時間について

エッセイ・テストの解答時間はわずか 25 分である。(アメリカではエッセイ・テストにワード数の制限はない。) このため、College Board では、エッセイが十分に推敲されていることを期待していない。アメリカの大学の授業に on-demand writing というのがあり、このエッセイ・テストはそれに近いものとの認識である。

実際、スペリング・エラーはそれが異常に多く読み手の理解に支障がない限りスコアに

影響を与えない。句読法や文法のエラーがあった場合でさえも、最高点を得ることが可能である。同様に文字の巧拙も評価されないが、読み手が判読できる程度には丁寧に書く必要がある。コンピュータを使用すれば解決するような作文上の問題については、このような便宜が図られており、このことは College Board ガイダンスにも示されている。

4.3 採点の仕方

大卒以上の学位をもった最低3年の教育経験のある高校および大学の教員が、採点のための教育を受けた後、その資格を得て、はじめて採点することができる。時給は\$17.00から\$22.00で、APやSAT Subject Testでの採点経験が考慮される。訓練教育時間は1回のSAT試験につき約10時間である。採点者は自宅のIBM Compatible PC (Windows)を用いて、朝6時から深夜12時までの好きな時間に採点することができる。

エッセイは、2人の評価者によって、互いに独立に(互いの評定を知ることなしに)評定される。各々の評価者は1つのエッセイを1~6点(6点が最高)で、エッセイ全体の総合評価を行う。エッセイ評価の観点には、考えの複雑さ、議論の深め方、言葉の流暢さなどが用意されているが、これらの観点毎にポイントを与え積算するというやり方はしない。評価の観点を参考にして、全体的な評点を唯一つ決定する。もし2人の評点が1を越える場合、すなわち2点以上異なったときは、scoring leaderと呼ばれる第3者が、その違いを解決する。エッセイ得点は、Writing セクション全体の1/3を占める。

5 今後のSAT

College Boardは、当面2005年3月から実施された新テストについてのあらゆる様相と、どのような学生層にどのような影響があるかについて調査する。中でも重要なのは、新しいWriting セクションを200-800のスケー

ル上にどのように位置付けるかを新たな妥当性研究に基づいて決定すること、及び新しいテストの技術的な特徴がうまく機能しているかを観察することである。その研究を踏まえて、約10年後のSATが検討されることになるだろう。

参考文献

- College Board: Research,
www.collegeboard.com/research/
- College Board: SAT Preparation Center,
www.collegeboard.com/student/testing/sat/prep_one/prep_one.html
- College Board: SAT The Essay,
www.collegeboard.com/student/testing/sat/prep_one/essay/pracStart.html
- College Board: The New SAT 2005,
www.collegeboard.com/newsat/index.html
- Wikipedia: SAT,
<http://en.wikipedia.org/wiki/SAT>
- スペースアルク: SAT試験について,
<http://alc.co.jp/sabrd/test/sat/>
- Bennet, R.E. and Bejar, I.I. 1998.
 Validity and automated scoring: It's not only the scoring, *Educational Measurement: Issues and Practice*, 17(4): 9-17.