

Computer Based Testing TOEFL の概要とその諸問題

研究開発部進学適性研究部門 椎 名 久美子

研究開発部特別試験研究部門 内 田 照 久

1. はじめに

TOEFL (The Test of English as Foreign Language) は、英語を母国語としない者の英語能力を測定するテストであり、Educational Testing Service (アメリカ合衆国の非営利団体) によって開発・実施されている。

1998年7月から、アジアとアフリカを除いた地域で、CBT (Computer Based Testing) 版 TOEFL が導入され、日本でも2000年秋から CBT 版 TOEFL がはじまる。今後4年をめどに、アジアとアフリカでも、従来の一斉実施によるペーパー・ペンシル形式の TOEFL は廃止される予定と聞く。

本稿では、1999年4月に ETS を訪問した際に得た情報を基に、CBT 版の概要と問題点を整理してみたい。(椎名は平成10年度文部省長期在外研究、内田は平成10年度文部省海外研究開発動向調査等に係る調査研究の一環として ETS を訪問)。

2. CBT 版と従来版 TOEFL の違い

形式上の大きな違いは、CBT 版 TOEFL がコンピュータ端末を用いて実施されることである。従来版は、聴解 (Listening Comprehension)、文法問題 (Structure)、読解 (Reading Comprehension) の3つのセクションから構成されているが、CBT 版では、Essay が加わって4つのセクションから構成されている。Essay 以外の3つのセクションは、マウスによる多肢選択式であるが、Essay セクションは自由記述式であり、画面に直接タイプして答えるか、手書きで答えるかは、受験者が選択できる。

CBT 版の聴解と読解のセクションでは、実際の会話や大学の講義に近い場面を再現するために、写真や図などの画像を多用した画面構成になっている点が従来版と異なっている。

また、従来型では、志望する大学から要請がある場合のみ TWE Test (Test of Written English) を受験すれば良かったが、大学側からの強い要

望によって受験者全員に Essay が課されることになった。

従来版も CBT 版も、素点ではなく尺度点が受験者に通知されるが、現在は CBT 版と従来版の両方が実施されているため、一目でどちらの版の得点か判別できるように、得点範囲の重複を避けて、従来版が301~667点、CBT 版が0~300点に尺度化されている。利用大学側の要望により、両者の得点を比較可能な換算表が作られている。ただし、前述のように、従来版は文字だけのテストであるのに対して、CBT 版には画像による情報も含まれているため、厳密な意味では両者の得点を同じ尺度で比べることは出来ない、というのが、ETS の研究者の見方である。そのため、等化表 (equating table) ではなく、換算表 (concordance table) という名称で呼ばれている。

3. CBT 用チュートリアルの開発とコンピュータ親和度の調査

CBT 版の導入にあたっては、受験者がコンピュータに関する知識や経験の有無にかかわらずコンピュータ上でテストをスムーズに受けることが出来るように、マウスを用いた解答のしかたを説明するパート (チュートリアル) が開発された。ベースになったのは、既に CBT 化されていた GRE (The Graduate Record Examinations, 大

学院入学テスト) のチュートリアルであるが、TOEFL 受験者が英語を母国語としない者であることから、言葉 (英語) での説明は最小限に抑えられ、図やアニメーションを多用したものとなっている。また、解答方法を練習してうまく出来なかったら、繰り返し説明が受けられるように設計されている。

コンピュータ親和度 (コンピュータの使用頻度や熟練度、Computer familiarity) が得点に影響するかどうか、についても調査が行われ (Kirsch et al., 1998)、チュートリアルを受けた後では、コンピュータ親和度の違いは得点に影響を及ぼさないことが示されている (Taylor et al., 1998)。

4. CBT 化に伴う問題

4.1 膨大な数の項目プールとセキュリティ

従来版でも CBT 版でも、事前テスト (Pre-test) によって難易度のパラメタが既に推定されている項目 (問題) を貯めておいて出題するという点は同じである。

従来版では、難易度がある程度同じになるような問題冊子が何種類も用意され、冊子間の等化が行われて尺度点が算出される。それに対して CBT 版の聴解と文法問題のセクションでは、Adaptive な出題が行われており (注)、受験者の正誤の反応に基づいて、次に

出題される項目がリアルタイムに項目プールの中から選ばれる。すなわち、受験者が正答した場合には、より難しい項目が次に出题され、誤答した場合には、より易しい項目が次に出题される、というように、各受験者の能力値に応じた難易度の項目が出题されることになる。

Adaptive型では受験者ごとに違う項目を出題するため、従来版よりも多くの数の項目が必要となる。また、同一の受験者が複数回受験した際に同じ項目が重複して出題されると、難易度のパラメータが変わってしまう恐れがあるし、たくさんの受験者の目に触れれば、その項目がこれから受験する者に漏れてしまう恐れもあることから、項目を長い間プールに貯めておくことは好ましくない。従来版では、年間17種類の出題冊子を作成するために、1年あたり約3,700題の項目が作成されてきたが、Adaptive型の出題には、それをはるかに上回る数の項目が必要だと言われている。ETSでは、常勤の項目作成者に加えて、外部の項目作成者にも作題を依頼し、項目作成と事前テストに多大な労力を費やしている。

また、TOEFLが大学入学に非常に重みを持つテストであることから、非公開であるはずの項目を盗み出そうとする職業まで存在する。ETSでは、たとえ一部の項目を盗んでも、それに

よって得られる成果が盗む労力に見合わないようにするためのセキュリティ対策を取っている。従来版では、(1)事前テストとして1度、採点に使う問題として1度出題したら、以後その項目は2度と使わない、(2)事前テストと実際の出題との間隔をランダムにする、などの対策が重ねられてきた。CBT版では、前述のように、従来版をはるかに上回る数の項目を随時作成しているが、これは、受験者が一度見た項目が再び出される頻度を抑えるためのセキュリティ対策としての意味も持っている。

4.2 受験者の減少傾向

CBT版の場合、受験者は、試験終了直後に取り得る得点の範囲を知ることが出来る(ただしEssayセクションはその場では採点されない)。受験者にとっては、コンピュータ上で実施するからこそ得られる恩恵だが、ETSにとっては、受験者の減少という皮肉な結果となっていると言う。従来版では、試験を受けてから郵送で受験者に得点が通知されるまでに、約1か月かかっていたため、数か月連続で複数回受験する者が多かったが、CBT版では、ある程度の得点範囲をその場で知ることが出来るため、一人当たりの受験回数は従来型よりも減る傾向があると言う。逆に、CBT化のためにはテ

ストセンターに器材を新たに設置する必要があることから、受験者一人当たりの実施費用は従来版よりも高額になっており、コストという面から見ると厳しい部分があるようだ。

5. その他の問題

CBT版のEssayセクションは自由記述式であり、訓練を受けた採点者によって評価される。1つのEssayを2人の採点者が6段階評価し、2段階以上の差が生じた場合には3人目の採点者を加えている。自由記述式の作文はGREやGMAT(The Graduate Management Admission Test)でも出題されていて、GMATのEssay testでは、E-raterと呼ばれる自動採点装置と1人の人間によって採点されるが、TOEFLは英語を母国語としない受験生が書いた英文ということで、機械による自動採点はまだ実用段階ではない。

CBT版のEssayを採点するための支所は米国内に3か所あり、TOEFLとGREのEssay答案が毎日送られて来る。採点者はその支所に通勤して採点を行うが、現在、採点者の自宅に直接Essayを送って、自宅で採点できるようなシステムを作成中である。また、その支所では、採点者の訓練も、コンピュータを用いて行われている。

Essayセクションの信頼性係数は、GREでは2題で0.7程度、GMATは

5題で0.8程度であるが、TOEFLでは信頼性係数が0.5~0.6程度と低く、Essayの題材によっても信頼性が大きく異なる。GREやGMATは主に英語を母国語としている大学生が受験するテストであるが、TOEFLのEssayセクションは外国人の書く英語であり、安定した結果が得られにくいと考えられている。

CBT版TOEFLは、本来4つのセクションから成っているが、総得点を算出する際にはそれらを3つのブロックに再構成して行う。すなわち、総得点は、聴解セクションの得点、読解セクションの得点、文法問題セクションとEssayセクションの得点が半分ずつの重みを持つように合成した3つのブロックの得点に、それぞれ3分の1ずつの重みを持たせて合計したものとなる。文法問題とEssayセクションの得点を合成する理由は、Essayセクションの信頼性の低さを補うためと言われている。

6. おわりに

本稿ではCBT版TOEFLについて紹介したが、ETSでは、21世紀に向けて、より統合された英語能力を測定するためのTOEFL 2000を計画中である。食欲過ぎるように思えるほど、テスト開発に多大な資源を費やすETSの動向は、今後も常に視野に入れてお

く必要があるだろう。

ただし、CBT版TOEFLをコストという点から見ると、調査の時点(1999年4月)では、一人当たりの実施にかかる費用が受験料を上回っており、開発・実施に費やされる多大な資源に見合う結果が得られているとは言えない。今後、アジアやアフリカ地域で全面的にCBT化するためには、テストセンターの建設や維持に莫大な費用がかかると予想される。実施面から見ても、大量のテスト項目を作りつつ、セキュリティを強化することが要求されている。CBTが必ずしもバラ色ではないことを、ETSは既に実証してしまっているのも事実のようだ。将来的に我が国のテストにおいてCBT化の検討がなされるような場合には、このCBT版TOEFLの功罪を厳に見据えた議論が必要となろう。

参考資料

Computer-Based TOEFL Score User Guide, 1998-99 Edition, Educational Testing Service.

TOEFL Test and score manual, 1997 Edition, Educational Testing Service.

Kirsch, I., Jamieson, J., Taylor, C. and Eignor, D., Computer Familiarity Among TOEFL Examinees (TOEFL Research Report 59), Princeton, NJ: Educational Testing Service, 1998.

Taylor, C., Jamieson, J., Eignor, D. and Kirsch, I., The Relationship Between Computer Familiarity and Performance on Computer-based TOEFL Test Tasks (TOEFL Research Report 61), Princeton, NJ: Educational Testing Service, 1998.

(注)

読解のセクションは、1つの文章(250～300語)と数問の項目から1題が構成されており、各項目が独立ではないのでAdaptiveな出題になっていない。