

テスト・データ分析とコンピュータ利用の一考察

実践女子大学

英語教育学

石川 祥一

東京経済大学

言語テスト論

中村 優治

ishikawa@univ.jissen.ac.jp

nkyj@tku.ac.jp

発表者を含めて7名は研究プロジェクト(日本学術振興会科学研究費補助金研究・基盤研究(B)(1)「グローバル化社会における発信型英語教育に対応する大学英语能力テストの開発」課題番号 15320074 研究代表:石川祥一,研究分担者:中村 優治,小林ひろみ,佐野富士子,シュナイダー・デニス,岡田敦子,松本佳穂子)により昨年度より取り組んできた。本発表はこれまでの研究の成果を基にしている。

コンピュータは言語教育及び言語テストの分野において非常に重要な役割を果たしてきている。テストの作成,テストの実施,テストの採点,item bankの構築,テスト・データの分析,テスト分析発表方法,言語資料の分析など言語教育の分野でコンピュータの貢献なしに語ることはできない。

本発表ではこれらのうち次の2点を取り上げ論じることとする。

- (1) テスト・データのコンピュータによる分析(実践報告を基にした展開)。
- (2) テスト開発に伴うコンピュータを用いた item bankの構築(理論的手続きを中心にした展開)。

1. テスト・データのコンピュータによる分析(実践報告を基に)

大学英语教育学会(JACET)では過去15年以上に渡ってListening Comprehension Testを用いて,日本人大学生の基礎聴解能力を中心とした英語能力を測定してきた。春季に実施されたForm Aと秋季に実施されたForm B(両テストは古典的テスト理論によって大まかに等化されたものとかんがえられる)の,これまでのデータをコンピュータを活用することによって,IRT手法のラッシュモデルを用いて分析した結果,大枠で以下のようなことが明らかになってきている。

1. Form Aに関して

- (1) 受験者に関しては第1回から第12回までは大きい変動があるものの,第13回以降は着実に減少の傾向にあるといえよう。
- (2) 受験者の得点平均値は全体的に下降線をたどっている傾向が見られ,また標準偏差も小さくなっていることから受験者の英語学力が確実に低下しているとも考えることもできる。同時に別の可能性として,低学力の受験者が集中的に受験者しているとも見ることができる。
- (3) 受験者の能力と項目の難易度の相対的位置関係の変化を見ると,第1回では明らかに40項目では測定できないほどの優れた能力を持った受験者が多くいたのに対し,後半になるにしたがって(特に第27回では)受験者能力と項目難易度はほぼ対応しており,テストとしては理想的である反面,受験者のレベルの低下は否定できないものといえよう。
- (4) このテストはテスト項目40問(4パート)により構成されているが,いずれの実施においても項目の難易度の相対的位置についての変化は見られなかった。

11. Form B に関して

- (1) Form B のテストは実施回数が Form A ほど多くないため、説得的な資料としては不十分ではあるものの、全体的に受験者数の減少に関しては Form A と類似している。
- (2) Form B の使用者の中には少なからず、このテストを Form A と併用して授業効果、学習効果の測定器具として考え、受験者の英語学力の伸長度を測定して英語教育改善の一助として利用してきたことも報告されている。この点からも Form B の十分な役割を果たしたといえよう。
- (3) Form A と Form B のデータの比較により、春と秋のテストをほぼ同一の学生が受験していると推測されることから、すべての実施回において秋のテスト成績が春を上回っていることから、学生の英語学力（リスニング力を通して見た場合）は大学のリスニング関連の授業効果により伸びているということも可能であろう。

以上、JACET リスニングテスト・データをコンピュータを用いて、IRT 手法(ラッシュプログラム)により分析した結果、多くの科学的、客観的な資料が得られ、学生の英語学力の変化を中心に、英語教育全体についてデータを基に検討する下地が整い始めたといえよう。

2. テスト開発に伴うコンピュータを用いた item bank の構築(理論的手続きを中心に)

ここでは、ラッシュプログラムを用いた場合の item bank の構築の一般的手続き、item bank の利点、item bank の限界について簡単にまとめることにする。

- (1) item bank の構築の第一段階は、データの第一回分析後、満点および 0 点の受験者の削除を行う。第 2 段階は全員正解および全員不正解の項目を削除する。第 3 段階は削除後の受験者、項目で再度分析をする。第 4 段階では、サンプリング、テストの長さなどが受験者とテスト項目とまったく依存しないように、logit score (通常 - 3 から 3 の範囲) に換算する。第 5 段階においては、ラッシュモデルを用いた場合は各項目の infit/outfit の欄に注目し model に fit しているかどうかを検討する。その際の許容範囲は Multiple choice test のような dichotomous data の場合 0.7 から 1.3 とするのが一般的である。この範囲を外れるものは再検討を加え、削除なども含めて最終項目の決定をしなければならない。以上のように Rasch calibration を終えて決定された項目は難易度に応じて item bank に store されることになる。
- (2) item bank の利点
項目の難易度順に並べることができるので、テスト作成が容易であると、同時にバラエティーに富んだ難易度の問題を作成することができる。また、受験者の能力をより正確に測定できる問題選出が可能である。さらにテスト結果の分析比較が容易に行えるので伸長度の比較も簡単で信頼性もたかくなる。IRT(Rasch model)で一度 calibration された項目は同じデータバンクに他の item を追加していくことが可能であるため、利用価値は大きいと考えられる。
- (3) item bank の限界
item bank は永久的なものではないのでその質、内容を一定に維持する努力が必要であるし、またそのことがテストの妥当性の検証とつながると思われる。

コンピュータの進歩により、item bank を用いた言語テストが英語教育に新しい側面から多大な貢献をするとおもわれる。コンピュータは言語教育及び言語テストの分野においてもこれからさらに重要な役割を果たしていくと考えられる。