

CASEC と TOEIC の相関に関して (1)

上 田 恒 雄

はじめに

コンピュータを使って手軽に受験できる CASEC という英語能力判定試験の結果と認知度の高い TOEIC 試験の結果に相関があるかどうかを検証する。具体的には CASEC は試験スコアと共に、受験時点での TOEIC 予測点を提示するので、その予測スコアと、実際の TOEIC スコアをグローバル英語学科学生 120 名を対象に統計分析し、果たして相関があると言えるかどうかを見てみる。

CASEC とは

CASEC (Computerized Assessment System for English Communication) は、(財)日本英語検定協会が基礎開発し、現在、(株)教育測定研究所が開発・運営しているインターネット上で受験できる英語コミュニケーション能力判定テストである。会社、学校、自宅から、インターネットに接続すれば、時と場所を選ばずに受けられる。所用時間は TOEIC の 120 分に対し、30 分～40 分。オンラインなので、結果は試験終了後すぐに分かり、現在の実力や、今後の勉強の指針などが画面に表示される。

CASEC は、能力推定のアルゴリズムに IRT¹⁾ (項目応答理論) に基づいた CAT²⁾ (Computerized Adaptive Testing: コンピュータ適応型テストシステム) 方式を採用している。CAT のシステムは、試験中にコンピュータが常に受験者の能力を計算し、1 問ごとに受験者に最適な項目を提示しながらテストする。故に、少ない実施項目数で、より高精度の測定が可能。瞬時に計算し、能力に合った問題を出題するには、複雑な演算も必要とする

ので、コンピュータの発展で実用化された形式といえる。開発元の試験結果からは、通常の紙テストの 40% 程度の項目数で、同程度の測定精度が得られることが確認されている。なお、テストの信頼性実験として約 500 名の受験者に、CASEC と紙ベースのテスト (同じデータベース内から構成されるテストセット) を両方受験させ、何問で紙ベースのテストの信頼性と同等になるかを計測している。

CASEC と TOEIC の一般的相関関係

CASEC が認知される過程においては、現時点でのスタンダードといえる TOEIC と相関関係が現状では必要である。CASEC のスコアを TOEIC に換算できることで、採用する企業・学校側も TOEIC で蓄積されたデータと比較できるからである。TOEIC との相関関係についてだが、開発元によると CASEC と TOEIC のスコアには、両スコアを保有する受験者、約 2000 名のデータの分析結果、相関係数³⁾で 0.860 の関連が認められるとのこと。TOEIC とは目的や測定方式や取得時期の違い、スコアの申告値の信頼性なども差し引く必要があるが、0.860 の値は、高い関連性を示すと言えるかと判断している。

ここで CASEC と TOEIC の試験内容を概観しておく (2007 年度時点)。

CASEC :

特徴

1. 短時間 (平均は約 30～40 分) で、測定精度の高いテスト
2. テスト終了後、すぐにスコアが表示される
3. インターネット上で受験者の好きな時間に受

験できる	600	—	825
4. TOEIC のスコア目安・英検級の目安等がわかる	550	—	755
測定内容	500	—	675
	450	—	590

- Section 1 語彙の知識——4 択, 15 問, 配点: 250 点, 解答時間: 60 秒
- Section 2 表現の知識——4 択, 15 問, 配点: 250 点, 解答時間: 90 秒
- Section 3 リスニングでの大意把握——4 択, 15 問, 配点: 250 点, 解答時間: 60 秒
- Section 4 具体情報聞き取り——ディクテーション (書き取り), 10 問, 配点: 250 点, 解答時間: 120 秒

スコア (トータル1000点) と英語コミュニケーション能力に対するアドバイスを技能別に表示

TOEIC :

問題構成の特徴

- ・ 2 時間で200問990点マークシートのテスト
- ・ 問題は, 問題形式の説明を含めて英語
- ・ リスニングは45分で100問495点 Section I (Part I ~IV)
- ・ 写真描写, 応答, 会話, 説明文問題 (リスニングの設問文は, 一度しか流れません)
- ・ リーディングは75分で100問495点 Section II (Part V ~VII)
- ・ 文法・語彙誤, 文訂正, 読解問題

レベルスコア評価

- A 860~ Non-Native として十分なコミュニケーションができる
- B 730~ どんな状況でも適切なコミュニケーションができる素地を備えている
- C 470~ 日常生活のニーズを充足し, 限定された範囲内では業務上のコミュニケーションができる

CASEC≒TOEIC ?

・ CASEC スコアと TOEIC スコア			
750	—		990
700	—		950
650	—		895

600	—	825
550	—	755
500	—	675
450	—	590

(CASEC の Sec1~3 の750に対する TOEIC スコア)

上記の CASEC スコアと TOEIC のスコアの対称関係からも分かるように CASEC は TOEIC と強い相関があるということの特徴にしているようで, その辺りを少し眺めてみると, 過去のデータから下記の換算式を導きだしている。

TOEIC 換算点: $1.0911 \times \text{CASEC スコア} - 92.167$

CASEC の HP にも, TOEIC/CASEC 点数換算表と相関グラフが掲載されている。CASEC の点数はいずれもセクション1~3までの750点と TOEIC のスコアとの相関を取っている。理由としては TOEIC の方にセクション4のようなディクテーション形式の問題がない, そしてセクション4を入れない方がより強い相関を示したということである。ではなぜ CASEC にセクション4があるのかというと, ディクテーション形式の問題はインタビューテストと最も高い相関を示すために盛り込んであるということ, その分, TOEIC より広い範囲の能力を把握しようとしていることと言える。それらのデータからももちろん相関があることは分かる。ではどのくらいのばらつきがあるのかは CASEC から提供して頂いた scatter chart (Figure 1) を見ると分かる。なお, データは JIEM より掲載許可を頂いたものである (株式会社教育測定研究所 e-Testing 事業部 <http://www.jiem.co.jp/>)。

TOEIC との相関は CASEC の Section1~4 の合計点から計算してある。この散布図から, 受験者が TOEIC で取ると思われる得点の平均点を予想し, TOEIC スコアの目安として提示している。予想と実際のずれが100点以内におさまる人は全体の約74%程度である。このグラフはセクション1~3の750点でなく, CASEC トータルのスコアと TOEIC との相関を比較してある。それで

CASEC と TOEIC の相関に関して(1) (上田)

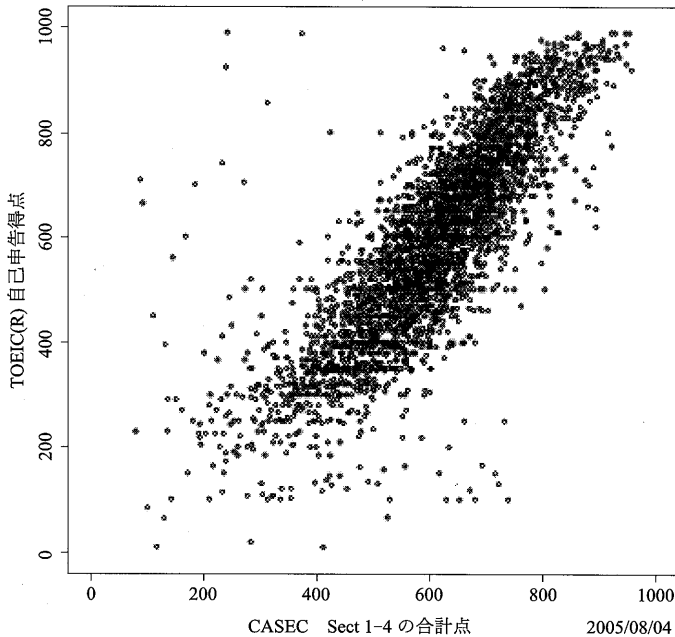


Figure 1 CASEC Section 4 までの合計点と TOEIC (R) 自己申告得点との関係
相関係数=0.83 (対象：2003/05/01～2004/10/31の受験者：3767人)

も相関係数が0.83である（係数が1.00で100%完璧な相関）。もちろんどちらもコミュニケーション能力を測るテストであるから正の相関があるのは当然であろう。

被験者の CASEC/TOEIC スコア

今回の検証・分析対象となる被験者（120名）の CASEC/TOEIC スコアを概観しておく。先ず第

1 回目の CASEC の統計値を Table 1 に示す。

続いて第 2 回目の CASEC の統計値を Table 2 に示す。

Figure 2 は第 1 回目のスコアと第 2 回目のスコアの得点分布をグラフにしたものである。

Table 3 は TOEIC の統計値を表したものである。

Figure 3 は実際の TOEIC の各セクション (Listening/Reading) の点数の分布表である。

Table 1 1st CASEC

	Sec 1	Sec 2	Sec 3	Sec 4	Sec 5
平均	124.0167	123.91667	132.45833	118.84167	499.2333
標準誤差	2.066641	2.3674847	2.6829135	2.0242921	7.271512
中央値 (メジアン)	124.5	121.5	134.5	118	503.5
最頻値 (モード)	128	124	137	144	492
標準偏差	22.63892	25.934495	29.389844	22.175009	79.65543
分散	512.5207	672.59804	863.76296	491.73102	6344.987
歪度	-0.2493	1.1939778	3.3064489	0.2169437	0.329423
歪度	-0.14311	0.1779422	-0.947223	-0.329127	-0.0691
範囲	111	168	207	120	464
最小	63	51	0	47	289
最大	174	219	207	167	753
合計	14882	14870	15895	14261	59908
標本数	120	120	120	120	120

Table 2 2nd CASEC

	Sec 1	Sec 2	Sec 3	Sec 4	Total
平均	129.167	124.642	136.375	123.067	513.25
標準誤差	2.09733	2.44344	2.4039	2.14994	7.2676
中央値 (メジアン)	127	124	135	124.5	508.5
最頻値 (モード)	127	125	153	132	472
標準偏差	22.9751	26.7666	26.3334	23.5514	79.612
分散	527.854	716.45	693.446	554.668	6338.1
尖度	0.11845	1.61364	0.45606	-0.2107	0.6338
歪度	0.14075	0.22543	-0.2823	-0.319	0.1253
範囲	117	177	140	107	449
最小	71	34	59	65	299
最大	188	211	199	172	748
合計	15500	14957	16365	14768	61590
標本数	120	120	120	120	120

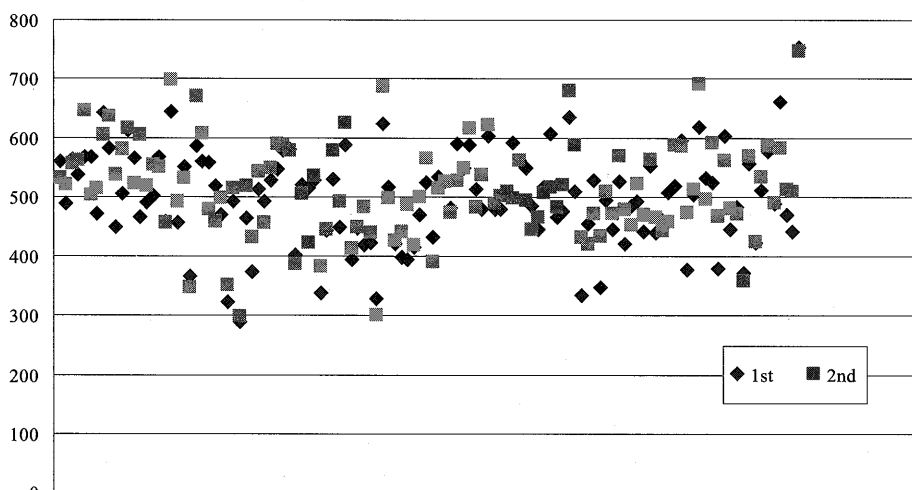


Figure 2 CASEC 得点分布

Table 3 TOEIC

	Listening Section	Reading Section	Total
平均	226.167	155.0833	381.25
標準誤差	4.5723	4.340216	8.076197
中央値 (メジアン)	220	155	375
最頻値 (モード)	220	110	440
標準偏差	50.087	47.54469	88.47031
分散	2508.71	2260.497	7826.996
尖度	3.22432	-0.25978	1.271392
歪度	1.12622	0.269117	0.697034
範囲	320	215	515
最小	130	60	200
最大	450	275	715
合計	27140	18610	45750
標本数	120	120	120

CASEC と TOEIC の相関に関して (1) (上田)

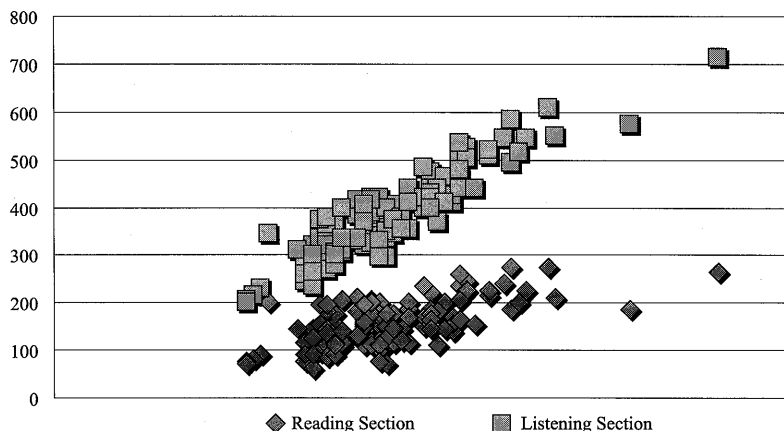


Figure 3 TOEIC セクション別得点分布

被験者全般に言えることは明らかにリーディングセクションのスコアがリスニングセクションのスコアを大きく下回っていることである。これがCASECスコアとTOEICスコアの相関を検証する際の結果に何らかの影響を与えることは十分考えられる。

検 証

被験者は平成19年度入学グローバル英語学科学生120名で、平成19年9月末に第1回目の

CASECを受験、その後同年11月末にTOEICを受験し、更に平成20年1月末に第2回目のCASECを受験した。各学生の2回のCASECとTOEICのスコア結果を基に両英語能力試験の間の相関を見るためにデータ分析を行った。分析に先立ち、まず2回のCASECのスコアから予測されるTOEICのスコアと、実際のTOEICのスコアがどのようなになっているかをFigure4に示した。

実際のTOEICスコアとCASECが予測したスコアの間にはかなり開きがあると言わざるを得な

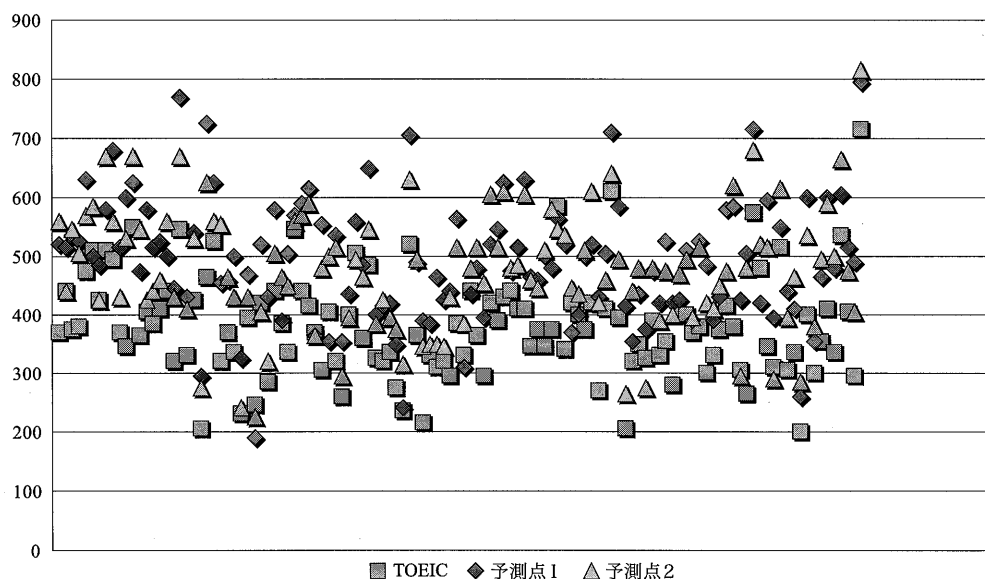


Figure 4 予測点

い。全体的に2つの予測スコアより実際のTOEICスコアの方が低い傾向が一目で見て取れる。特に2回目より1回目のCASECの予測点はかなり実際のTOEICのスコアと差があると言える。CASECによるTOEIC予測点は今回の被験者について全般的に言えば信頼性があまりないと言えるのかどうかを両試験の相関を具体的に統計分析することで検証してみる。

分 析

本研究では、TOEICのスコアとCASECのスコアとの相関をとり、CASECがTOEICと相関関係があるかどうかを総合的に検証するために4つの分析を行った。なお、相関係数とは、2つの変数の相関関係を示す指標である。相関係数は必ず-1~1の範囲に収まり、0.7以上あれば正の相関(片方の値が上がれば、もう片方の値が一定の割合で上がる関係)があると考えられる。さらに回帰分析におけるR2値とは、ある現象がその回帰式で表される確率というように考える。R2値は一般的には、0.5~0.8なら、回帰式成立の可能性があるようで、0.8以上ならばかなり成立の可能性が高いことを示す。(直線近似の場合は、R2=相関係数の2乗になる)。いずれの回帰分析の場合もCASECスコア(CASECからTOEICの予測

点)を説明変数とみなし、プロフィシエンシー・テストとしてはすでに標準化されているTOEICを目的変数として扱った。回帰分析には、重相関係数、決定係数、自由度補正済み決定係数、Y評価値の標準誤差を計算し、参考までに回帰式の有意性を検定するための分散分析、回帰係数の検定と信頼区間を計算した結果も載せてある。

まず分析1ではTOEICのスコアと1回目のCASECのスコアについて回帰分析を行った。結果をTable 4に示す。

相関関係は0.796とかなり高い相関である。R2値も0.632と回帰式成立の可能性を示している。決定係数は全データの変動のうち回帰式で説明できる割合を示しているが、この結果では63%のデータが回帰で説明できることを示している。標準誤差は実測値と予測値の差の標準誤差を示しているが、約54点くらいの変動がある。

次の分析2では、1回目のCASECのスコアから予想されるTOEICのスコアと実際のTOEICのスコアについて回帰分析を行った。結果をTable 5に示す。

相関関係は0.800となり、かなり高い相関を示した。R2値も0.640となり、第1回CASECスコアとTOEICスコアの相関同様に、回帰式成立の可能性はある。決定係数は全データの変動のうち

Table 4 第1回CASECとTOEICの相関

回帰統計	
重相関 R	0.795519
重決定 R2	0.632851
補正 R2	0.629739
標準誤差	53.83336
観測数	120

分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	1	589444.9	589444.9	203.395	1.94E-27
残差	118	341967.6	2898.03		
合計	119	931412.5			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-59.8493	31.317	-1.91108	0.05842	-121.865	2.166919	-121.865	2.166919
X 値 1	0.883553	0.061953	14.26166	1.94E-27	0.760869	1.006237	0.760869	1.006237

CASEC と TOEIC の相関に関して(1) (上田)

回帰式で説明できる割合を示しているが、この結果では64%のデータが回帰で説明できることを示している。標準誤差は実測値と予測値の差の標準誤差を示しているが、約53点くらいの変動がある。

分析3では同様に TOEIC のスコアと2回目の CASEC の総合点について回帰分析を行った。結果を Table 6 に示す。

相関関係は第1回 CASEC よりはやや低くなっ

たが、それでも0.776とかなり高い相関である。R2値も0.602と第1回 CASEC より下がってはいるが、回帰式成立の可能性は十分ある。決定係数は全データの変動のうち回帰式で説明できる割合を示しているが、この結果では60%のデータが回帰で説明できることを示している。標準誤差は実測値と予測値の差の標準誤差を示しているが、約56点くらいの変動がある。

最後に分析4では、2回目の CASEC のスコア

Table 5 第1回 CASEC 予測点と TOEIC の相関

回帰統計	
重相関 R	0.80008
重決定 R2	0.640129
補正 R2	0.637079
標準誤差	53.29711
観測数	120

分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	1	596223.8	596223.8	209.8949	5.92E-28
残差	118	335188.7	2840.582		
合計	119	931412.5			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	56.86718	22.91266	2.481911	0.014476	11.49387	102.2405	11.49387	102.2405
X 値 1	0.681477	0.047038	14.48775	5.92E-28	0.588328	0.774625	0.588328	0.774625

Table 6 第2回 CASEC と TOEIC の相関

回帰統計	
重相関 R	0.776036
重決定 R2	0.602232
補正 R2	0.598861
標準誤差	56.03314
観測数	120

分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	1	560926.4	560926.4	178.6553	2.24E-25
残差	118	370486.1	3139.712		
合計	119	931412.5			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-61.3682	33.50746	-1.83148	0.069552	-127.722	4.985673	-127.722	4.985673
X 値 1	0.862383	0.06452	13.3662	2.24E-25	0.734617	0.99015	0.734617	0.99015

Table 7 第2回 CASEC 予測点と TOEIC の相関

回帰統計	
重相関 R	0.753629
重決定 R2	0.567957
補正 R2	0.564296
標準誤差	58.39739
観測数	120

分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	1	529002.3	529002.3	155.121	3.03E-23
残差	118	402410.2	3410.256		
合計	119	931412.5			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	70.90113	25.48196	2.782405	0.006284	20.43991	121.3623	20.43991	121.3623
X 値 1	0.631004	0.050664	12.45476	3.03E-23	0.530676	0.731332	0.530676	0.731332

から予想される TOEIC のスコアと実際の TOEIC のスコアについて回帰分析を行った。結果を Table 7 に示す。

相関関係は0.754となり、今回の分析対象の中では一番低い数字ではあるがそれでも正の相関を示した。R2値も0.568と最低の数字になったが、回帰式成立の可能性は残している。決定係数は全データの変動のうち回帰式で説明できる割合を示しているが、この結果では56%のデータが回帰で説明できることを示している。標準誤差は実測値と予測値の差の標準誤差を示しているが、約58点くらいの変動がある。

今回の4つの分析結果すべてから、CASEC スコアが実際の TOEIC のスコアとかなり高い相関関係があることが明らかになった。4つの回帰分析結果の R2値もすべて回帰式成立の可能性があることが判明 (56%~64%) し、本研究の目的である両試験の相関は認められたと言えよう。なお、実際の TOEIC のスコアより53点~58点の範囲で CASEC の予測点は高い数字を出す可能性があることも記しておく。

おわりに

これまでの分析の結果から確かに CASEC と TOEIC スコアの間には数値上の相関は認められ

る。しかし本論の最初に記した今回の被験者の TOEIC スコアの特徴を考慮すると、CASEC のスコアから予測される TOEIC スコアは受験者のリーディング能力を十分に反映したものとなっていない可能性もある。この2つの試験の出題形式(問題構成の特徴を参照)を比較すると、CASEC の試験の内容は必ずしも TOEIC のリーディングセクションの出題形式に沿った問題内容にはなっていない。つまり TOEIC に設定されているリーディングプロパーの問題形式がないのである。TOEIC の場合はリスニングセクションとリーディングセクション各々50%配点されているため、スコア上どちらかの能力が劣る場合不利になることは否めない。そこで考えられることとして、例えばリーディングが多少不得意でもリスニングが得意な CASEC の受験者のスコアから予測される TOEIC の予測点が高くなる傾向があっても不思議ではない(分析結果の標準誤差を参照)。この点に関しては今後も更に検証していく。

注

1) IRT「項目応答理論 (Item Response Theory)」は、従来の素点を基礎とした評価(合計点/平均点/偏差値など)では実現不可能であった、受験者集団やテスト問題などに依存しない「絶対尺度の評価」を可能とした新しいテスト理論である。事前に行った予

CASEC と TOEIC の相関に関して (1) (上田)

備テストデータから算出した各問題の特性値（難易度、識別力など）をもとに、新規受験者の能力を絶対尺度上で推定しようとするものである。

- 2) CAT「コンピュータ適応型テストシステム (Computerized Adaptive Testing)」は、受験者のそれまでの問題に対する解答情報に基づいて、次に出題する問題を既存の問題群から選出していくようなテストである。

CASEC では、それまでの問題に対する正誤情報をもとに、次に問題を出す問題の難易度を調整している。この手法を用いることで、受験者にとって極端に難しい問題や、易しい問題の出題を避けることができる。自分の実力にふさわしい問題だけを解いていくため、従来の紙のテストと比較して少ない問題数で精度の高い測定が可能となっている。

- 3) 相関係数とは、2つの変数の直線的な関係の強さおよび向きを表す指標で、2つの変数の関係が完全に線形であれば、相関係数の絶対値が1になる。

一方の値が大きいほど他方の値も大きいという場合には相関係数の符号が正になり、一方の値が大きいほど他方の値が小さいという場合には符号が負に

なる。相関係数の値の解釈に関しては、心理学や教育学では一般に次のように解釈している（肥田野 1961）。

- 0.00～±0.20 ほとんど相関がない
0.20～0.40 (-0.20～-0.40) 低い相関がある
0.40～0.70 (-0.40～-0.70) かなり相関がある
0.70～1.00 (-0.70～-1.00) 高い相関がある

参考文献

- Bachman, Lyle F. 1990. *Fundamental Considerations in Language Testing*. Oxford University Press.
Bachman, Lyle F and A. S. Palmer. 1996. *Language Testing in Practice*. Oxford University Press.
Hatch, Evelyn and Anne Lazaraton. 1991. *The Research Manual*. Newbury House Publishers.
TOEIC 運営委員会 2000 『TOEIC 公式ガイド&問題集』(財)国際ビジネスコミュニケーション協会
林規生 2001 「英語能力測定における CAT の適応例と効果測定」『計測と制御』40(8), pp. 572-575.
肥田野直, 瀬谷正敏, 大川信明 1961 『心理教育統計学』培風館