

英文読解速度テストの有効性の検証

Verification of Effectiveness of English Reading Speed Test

小谷 克則[†] 吉見毅彦^{††} 九津見毅^{††} 佐田いち子^{††} 井佐原均[†]

Katsunori Kotani Takehiko Yoshimi Takeshi Kutsumi Ichiko Sata Hitoshi Isahara

1. まえがき

近年、コンピュータ支援言語教育 (CALL) システムの発展を背景に、読解速度を読解テストの評価指標として加えることが提案されてきた (Alderson 2000)。Alderson (2000) によると、読解速度を向上させるといった言語処理能力を育成させることが第二言語の習得に大きな影響を与える。したがって、読解速度といった観点から学習者を評価することは、非常に有益であると考えられる。

しかし、これまで、読解速度が言語テストの評価基準として有効かどうかは十分に確認されていない。このため、本稿では、読解速度テストの有効性を検証する。検証は、古典的テスト理論 (Brown 1996) に基づいて行う。まず、2節において、第二言語教育での読解速度の重要性を指摘し、さらに読解速度を評価基準とする読解速度テストの効果について議論する。3節において、言語テストの有効性を検証する際に用いられる古典的テスト理論について説明する。そして、4節において、読解速度テストの有効性評価実験の結果を報告する。5節において、本稿のまとめと今後の課題を述べる。

2. 読解能力

2.1 理解度と読解速度

読解能力といっても様々な能力が含まれるが、理解度と読解速度によって示される能力があることは門田他 (2001) も指摘するように古くから知られている (Tinker 1939; Davies 1975/1989)。このような能力の区別は、読解過程の相互作用モデル (Rumelhart 1977 など) にも反映されている。相互作用モデルは、文字認識から単語認識や統語解析といった過程を経て意味構築をボトム・アップに行う言語処理過程とテキストに表現されている情報を背景知識やスキーマを利用して情報の統合や予測などをトップ・ダウンに行う情報処理過程から成る。このように読解能力を二分した場合、読解速度は情報処理過程に関与する能力と比べ、言語処理能力においてより重要となると考えられている (Grabe & Stoller 2002)¹。このように読解能力の一部として読解速度が果たす役割は大きいと考えられる。

理解度と速度の関連性は、言語処理過程の自動化 (無意識化) の有無によって説明できる (阿部他 1999)。処理過程が自動化されることによって、処理速度が向上する。速度が向上することにより、短期記憶を通じた処理範囲が向上することなどから、理解度の向上が得られると考えられている。特に、第二言語の場合、処理の自動化を促す目的からも読解速度の向上が重要であると考えられている (Rasinski 1989; Day 1998; Koda 1996/2005; Segalowitz et al. 1991)。

これらの研究が示すように、読解速度が第二言語の読解において言語処理過程の自動化といった重要な役割を果たしていると考えられる。そのため、読解テストに従来の理解度に速度を加えることによって、読解能力を速度の観点から評価することは読解教育において有益であると筆者らは考える。

2.2 読解速度テストの効果

英語母語話者による平均英文読解速度は、200~300WPM (Words per Minute) である (Carver 1982) のに対して、日本語を母語とする英語学習者の場合、平均英文読解速度は 100WPM 程度とされ (高梨他 1987)、英語学習者の読解速度は英語母語話者と比較するとかなり遅い。英語学習者の読解速度を上げるためには、読解速度を意識した読みを学習者に行わせる必要がある。これを実現する一つの訓練方法として、読解速度テストを繰り返し受けさせることが考えられる。実際に、Tan & Nicholson (1997) によると、学習者に読解速度を意識した読みを行わせるためにチャンキングの指導を行い、その後の訓練を通じて、読解速度が上がる事が確認されている。

さらに、言語処理過程の自動化による読解速度の向上が読解能力以外の言語能力、例えば、リスニング能力 (Elly & Mangubhai 1998) や英作文能力 (Mason & Krashen 1997) に良い効果を及ぼすことが確認されている。このように読解速度の向上による第二言語能力全般への波及効果が期待できることから、読解速度テストは第二言語教育にとって有益である。

3. 古典的テスト理論における検証法

言語テストが有効であるかどうかを古典的テスト理論に基づいて検証する手法が提案されている (Brown 1996)。Brown (1996) はテストの有効性を統計的に検定する項目としてテストの信頼性と妥当性を挙げている。特に、信頼性の低いテストは、どれだけ妥当性が高くても有効なテストとして認めることが困難とされる。

3.1 テストの信頼性

テストの信頼性は、同一の受験者に同一のテストを繰り返し行うと、毎回ほぼ同じ安定した結果が得られれば得られるほど高くなる。このテストの信頼性は有効なテストの主たる条件であると考えられている。ある受験者の能力に変化がないにもかかわらず、あるテストを繰り返し受験し、その結果が大きく異なる場合、テストの信頼性は著しく低い。一方、テスト結果の差が少ない場合、そのテストの信頼性は高い。

読解速度テストを言語テストとして利用する場合も、学習者の能力やテキストの難易度に大きな変化がない限り、テスト結果が安定していることが望まれる。

3.2 テストの妥当性

テストの妥当性は、学習者の能力の評価がより適正であればあるほど高くなる。テストの妥当性の検証は、基準関連妥当性、目標関連妥当性、構成概念妥当性を確認することによって行える。

[†] 情報通信研究機構^{*} 龍谷大学^{††} シャープ株式会社¹ 情報処理過程の成功には速度よりも情報の質、つまり正確さが重要であると考えられる。

基準関連妥当性は、集団準拠テストの場合に検証すべき妥当性である²。本研究では読解速度テストを集団準拠テストとみなす。集団準拠テストでは、テスト結果が正規分布することが望ましい。したがって、読解速度テストの基準関連妥当性は、テスト結果が正規分布するかどうかによって検証する。

次に、目標関連妥当性は、既存のテストでその妥当性が確認されているものがある場合、そのテスト結果との相関によって示される。読解速度テストの目標関連妥当性を調べるための比較対象テストは、読解速度テストとしての妥当性が確立されていなければならない。

本研究以外の読解速度テストとしては、眼球運動の追跡によるテストなどがある。しかし、筆者らの知る限り、眼球運動の追跡による読解速度テストは第二言語の読解速度テストとして妥当であるかどうかは確認されていない。そこで、妥当性が既に確認されている TOEIC (Test of English for International Communication) のリーディングセクションを代替テストとして用いる。TOEIC は理解度テストであり、読解速度テストではない。このため、TOEIC が測定対象とする英語運用能力と本研究の読解速度テストが測定対象とする運用能力とは完全には一致しない。しかし、両者の間には共通する部分もあると考えられる。なお、Kitao (1995) や門田他 (2001) などによると、理解度と読解速度の間に、相関係数は 1.00 に近い値ではないが、正の相関が見られるという。また、第二言語の読解において、理解度が向上するにつれて読解速度も向上することは多くの先行研究により示されている (Smith 1988; Nutall 1996; Day 1998; Bell 2001; Kitao 1995)。これらのことから、TOEIC スコアを比較対象として利用することにした。

最後に、構成概念妥当性は、テスト結果から受験者を能力別に区別できるかどうかという弁別性によって示される。読解速度テストに弁別性があるならば、読解速度テストの結果から受験者をクラス分けすると、クラス毎の平均読解速度間に統計的に有意な差が確認できる。また、読解速度テストと TOEIC が測定対象とする能力に習熟度といった観点から関連があるならば、読解速度テストの結果によるクラス分けが TOEIC スコアにも反映されると予測される。

4. 読解速度テストの有効性評価実験

4.1 実験概要

読解速度テストの信頼性と妥当性を検証するために、実験を行なった。前節で概観した古典的テスト理論に基づいて次の検証項目を設けた。

- (1) 信頼性：読解速度テストの結果は安定しているか (4.2 節)
- (2) 基準関連妥当性：読解速度テストの結果は正規分布を示すか (4.3 節)
- (3) 目標関連妥当性：読解速度テストの結果は参加者の TOEIC スコアと相関関係にあるか (4.4 節)
- (4) 構成概念妥当性：読解速度テストの結果は参加者を能力に応じたクラスに弁別できるか (4.5 節)

実験への参加者の募集では、事前に TOEIC を受験していることを条件とし、参加報酬を支払うことを明示した。この募集により日本語母語話者 107 名が参加した。参加者の TOEIC スコアにおける読解部分の平均点は 310.4 点、

標準偏差は 99.2 であった。また、最低点は 105 点、最高点は 470 点であった³。

実験用テキストとして TOEIC 準拠の問題集から 84 テキストを抜粋した。84 テキストを 12 通りのテキストセットに分けた。各参加者をいずれかのテキストセットに振分けた。参加者には、TOEIC の設問に解答できる程度にテキストの内容を理解するように指示した。制限時間を設けずに、参加者が自己ペースでテキストを読み進めるように指示した。読解速度の計測にはコンピュータ上で指定した読解範囲における読解時間を計測するツール (吉見他 2005) を用いた⁴。指定読解範囲は文とした。つまり、参加者は TOEIC テキストを文毎に自己ペースで読み進め、設問の解答を行なった。参加者は計測ツールの操作に慣れるために事前練習を行なった。

得られた読解時間データと各テキストに含まれる単語数 (カンマやピリオドを除く) とから一分間当たりの読解量を読解速度として算出した。また、読解速度を算出する際、TOEIC の設問と解答の選択肢の部分は算出対象外とした。各参加者について、割り当てられた全テキストの平均読解速度を求めた。以下、このデータを読解速度データと呼ぶ。107 人の参加者のうち 4 人については、データに不備が確認されたため、対象外とし 103 人分の読解速度データについて分析を行う。

4.2 実験結果：信頼性

テスト結果の安定性を検証する方法として、再テスト法や内部一貫法などがある。再テスト法では、同一参加者に同一の試験を課して同一の評価を与えることができるかどうかを検証する。しかし、読解速度テストの場合、以前のテストの記憶や学習効果があるため再テスト法は望ましくない。このため、一つのテストを折半して安定性を検証する内部一貫性法を用いる⁵。

テキストセット毎に次の式 (1) に従って、読解速度データに対するクロンバック α 信頼性係数を算出した⁶。

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k S_j^2}{S_Y^2} \right) \quad (1)$$

但し、 k は各テキストセットにおけるテキストの数、 S_j は各テキストセットにおける全テキストの速度の標準偏差、 S_Y は各テキストセットにおける各テキストの速度の標準偏差である。信頼性係数が一般に 0.7 以上あればテストは有効である (Brown 1996)。それぞれのテキストセットにおける信頼性係数は表 1 に示すように、0.82~0.98 であった。また、全テキストセットでの平均信頼性係数は 0.91 であった (標準偏差は 0.05)。したがって、本実験による読解速度テストの結果は、安定しているといえる。

³ 参加者の TOEIC 受験時期は、2002 年が 12 人、2003 年が 26 人、2004 年が 59 人、2005 年が 10 人であった。実験は 2005 年 3 月に行った。

⁴ 提示範囲を文単位に設定することによって、文毎の読解速度を計測することが可能となる。この計測法では自由に読み返しを行なえることから、文単位での読解による負担は小さいと思われる。しかし、負担の縮小の程度に関しては今後の課題である。

⁵ 言語テストに対して再テスト法を用いることの問題として学習効果が指摘されている (Alderson 1991)。また、言語的情報を排除したテキストの読解であっても、学習効果により約 3% 程度の速度上昇が確認されている (Wilkins 1996)。

⁶ 内部一貫法による信頼性係数の算出には、一般に、KR-20 やクロンバック α 係数が用いられる。KR-20 が正解・不正解の二値にしか対応できないのに対して、クロンバック α 係数は値に関係なく用いることができる (大友 1996)。

² 目標基準準拠は授業において指導された項目の習得の程度を測定することから、得点分散が期待されないため、基準関連妥当性は無視できると考えられている (Brown 1996)。

テキストセット	信頼性係数 α
S1	0.86
S2	0.91
S3	0.91
S4	0.86
S5	0.93
S6	0.95
S7	0.97
S8	0.87
S9	0.91
S10	0.82
S11	0.90
S12	0.98

表1：各テキストセットにおける信頼係数

4.3 実験結果：基準関連妥当性

一般に、テストに用いる評価指標が妥当であれば、テスト結果は正規分布するとされている (Carver 1982)。したがって、読解速度が読解能力を測る評価指標として妥当であれば、読解速度データが正規分布すると予測される。そこで読解速度データの正規性をコルモゴロフ・スミルノフ (K-S) 検定により検証した。

参加者 103 人の平均読解速度は 69.6WPM、標準偏差は 24.0 であった。K-S 検定の結果、分布は正規分布と異なるといえないことから正規性ありとみなされた (K 統計値は 0.9, 片側確率 p 値は 0.2, 有意水準 α は 0.05)。これにより、正規性が確認できた。分布を示すヒストグラムを図 1 に示す。

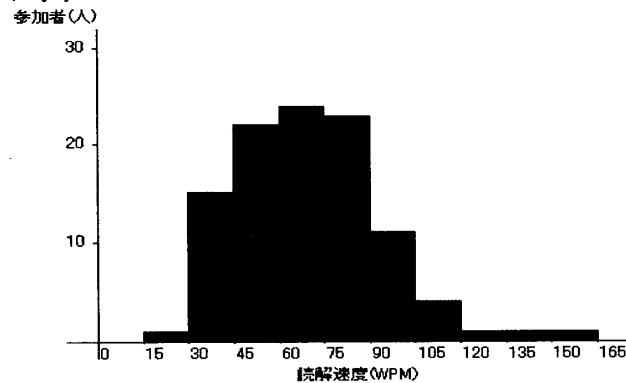


図1：読解速度データヒストグラム

4.4 実験結果：目標関連妥当性

読解速度テストの目標関連妥当性を検証するために、読解速度データと各参加者の TOEIC スコアの相関を算出した⁷。TOEIC スコアは、聴解テストのスコアと読解テストのスコアに分けられるが、本稿では読解テストの妥当性を検証することが目的であるため、読解テストのスコア (以下、TOEIC 読解スコアと呼ぶ) との比較を行った。

読解速度データが参加者の読解能力を反映しているのであれば、読解速度データと TOEIC 読解スコアの間には正の相関があると予測される。

⁷読解速度の比較は、TOEIC スコアと行うのではなく、今回の実験で読解速度と同時に得られた理解度スコアと行うことも考えられる。しかし、今回の実験は、読解能力の評価指標として読解速度と読解効率 (読解速度と理解度の積) のどちらが適切であるかを検証するために予備実験として行ったものである。読解効率の有効性検証では、読解効率を TOEIC スコアと比較するため、今回の実験でも TOEIC スコアを比較対象とした。

TOEIC 読解スコアの平均は 309.8 点、標準偏差は 99.7 であった。散布図を図 2 に示す。スピアマンの順位相関係数の有意性検定を行なった結果、相関係数は 0.5 であり、有意であった ($F(1,101)=36.4, p<0.01$)。説明率 (相関係数の二乗) は 25% であり、読解速度データと TOEIC 読解スコアの間には中程度の相関があるといえる。この結果は Kitao (1995) や門田他 (2001) の報告によっても支持される。

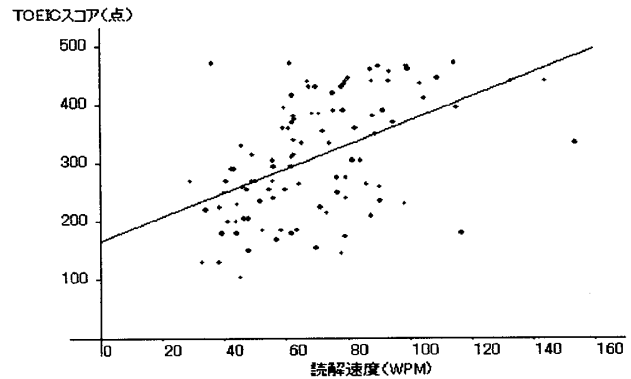


図2：TOEIC スコアと読解速度の散布図

4.5 実験結果：構成概念妥当性

読解速度テストにより参加者を読解能力別に弁別できるかどうかを検証する。参加者平均読解速度データに基づいて、参加者を 3 クラスに分類した。読解速度が参加者の読解能力を測る評価指標として妥当であれば、各クラスの平均読解速度の間に統計的有意差があると予測される。そこで、大友 (1996) に倣い、参加者を読解速度データに基づいて、上位 25% に当たる 26 人を「上位クラス」、下位 25% に当たる 26 人を「下位クラス」、その他を「中位クラス」に分類した。表 2 は、各クラスの読解速度と TOEIC 読解スコアの平均値と標準偏差を示す。

クラス	読解速度		TOEIC 読解スコア	
	平均値 (WPM)	標準偏差	平均値 (点)	標準偏差
上位	101.7	18.2	374.2	92.0
中位	67.2	9.0	315.2	91.1
下位	42.3	5.4	234.6	75.1

表2：読解速度と TOEIC 読解スコア (クラス毎)

読解速度データに対して各クラスの中央値に有意な差があるかどうかをクラスカル・ウォリス検定を用いて検定した結果、クラス分けの効果は有意であった ($H(2, N=103)=86.3, p<0.01$)。シェッフェの方法を用いた多重比較によれば、全てのクラスの間には有意差 ($p<0.01$) があった。従って、読解速度データに基づいて参加者を弁別できることが確認できた。

TOEIC 読解スコアに対してもクラスカル・ウォリス検定を行なった結果、クラス分けの効果は有意であった ($H(2, N=103)=24.5, p<0.01$)。シェッフェの方法を用いた多重比較によれば、全てのクラスの間には有意差 ($p<0.01$) があった。

従って、読解速度データに基づいて読解速度の観点からだけでなく TOEIC 読解スコアの観点からも参加者を弁別できることが確認できた。

5. まとめ

本稿では、読解速度テストの必要性を指摘し、読解速度テストが英語運用能力の評価指標として有効かどうかを古典的テスト理論に基づいて検証した。古典的テスト理論 (Brown 1996) において検証すべきとされる信頼性と妥当性 (基準関連妥当性, 目標関連妥当性, 構成概念妥当性) を検証した結果、読解速度テストは有効であることが確認できた。

今後、データの詳細な分析を行い、テストの有効性をさらに確認する必要がある。さらに、本研究の手法により収集した読解速度データから、テキスト特徴や学習者の習熟度に応じた最適英文読解速度を予測できるモデルを構築する。このモデルは読解速度テストの評価値として利用するなどしてCALLシステムの開発に役立つと考えられる。

参考文献

- 阿部純一, 桃内佳男, 金子康郎, 李光五 1999 人間の言語情報処理: 言語理解の認知科学. サイエンス社, 東京.
- Alderson, J. C. 1991 Giving students a sporting chance: assessment by counting and judging. In J. C. Alderson & B. North (Eds.) *Language Testing in the 1990s: The Communicative Legacy*. Macmillan, London.
- Alderson, J. C. 2000 *Assessing Reading*. Cambridge University Press, Cambridge.
- Bell, T. 2001 Extensive reading: speed and comprehension. *The Reading Matrix* 1(1).
- Brown, J. D. 1996 *Testing in Language Programs*. Prentice-Hall, Upper Saddle River, NJ.
- Carver, R. P. 1982 Optimal rate of reading prose. *Reading Research Quarterly* XVIII, 56-88.
- Davies, A. 1975 Two tests of speeded reading. In R.L. Jones & B. Spolsky (Eds.) *Testing Language Proficiency*. Center for Applied Linguistics, Washington, DC.
- Davies, A. 1989 Testing reading speed through text retrieval. In C.N. Candlin & T. F. McNamara (Eds.) *Language Learning and Community: Festschrift in Honour of Terry Quinn*. National Centre for English Language Teaching and Research, Sydney.
- Day, R. R. & Bamford, J. 1998 *Extensive Reading in the Second Language Classroom*. Cambridge University Press, Cambridge.
- Elly, W. B. & Mangubhai, F. 1983 The impact of reading on second language learning. *Reading Research Quarterly* XIX, 53-67.
- Grabe, W. & Stoller, F. L. 2002 *Teaching and Researching Reading*. Pearson Education, London.
- 門田修平, 野呂忠司 (編著) 2001 英語リーディングの認知メカニズム. くろしお出版, 東京.
- Kitao, K. 1995 Japanese college students' English ability. MS.
- Koda, K. 1996 L2 word recognition research: a critical review. In J. Devine, P. Carrell, & D. E. Eskey (Eds.) *Research in Reading in a Second Language*. TESOL, Washington, DC.
- Koda, K. 2005 *Insights into Second Language Reading: A Cross-Linguistic Approach*. Cambridge University Press, Cambridge.
- Mason, B. & Krashen, S. D. 1997 Extensive reading in English as a foreign language. *System* 25(1), 91-102.
- Nutall, C. 1996 *Teaching Reading Skills in a Foreign Language*. Heinemann, Oxford.
- 大友賢二 1996 項目応答理論入門—言語テスト・データの新しい分析法. 大修館書店, 東京.
- Rasinski, T. V. 1989 Fluency for everyone: Incorporating fluency instruction in the classroom. *The Reading Teacher* 42(9), 690-693.
- Rumelhart, D.E. 1977 Toward an interactive model of reading. In S. Dornic (Ed.) *Attention and Performance VI*, Lawrence Erlbaum, Hillsdale, NJ.
- Segalowitz, N., Poulsen, C., & Komoda, M. 1991 Lower level components of reading skill in higher level bilinguals: implications for reading instruction. *Reading in Two Languages: AILA Review*, vol. 8, Free University Press, Amsterdam.
- Smith, F. 1988 *Understanding Reading*. Lawrence Erlbaum Associate, Mahwah, NJ.
- 高梨庸雄, 高橋正夫 1997 英語リーディング指導の基礎. 研究社出版, 東京.
- Tan, A. & Nicholson, T. 1997 Flashcards revisited: Training poor readers to read words faster improves their comprehension text. *Journal of Educational Psychology* 89, 276-288.
- Tinker, M. A. 1939 Speed versus comprehension in reading as affected by level of difficulty. *Journal of Educational Psychology* 30, 154-161.
- Wilkins, A. J., Jeanes, R. J., Pumfrey, P. D., & Laskier, M. 1996 Rate of reading test: its reliability, and its validity in the assessment of the effects of colored overlays. *Ophthalm Physiology* 16, 491-497.
- 吉見毅彦, 小谷克則, 九津見毅, 佐田いち子, 井佐原均. 2005 英語学習者の読解能力推定のための読解時間測定法. 教育システム情報学会誌 22, 24-29.