

# 読解速度に基づく英文読解能力推定法の評価

小谷克則<sup>†\*</sup> 吉見毅彦<sup>‡†</sup> 九津見毅<sup>††</sup> 佐田いち子<sup>††</sup> 井佐原均<sup>†</sup>  
独立行政法人情報通信研究機構(NICT)<sup>†</sup> 龍谷大学<sup>‡</sup> シャープ株式会社<sup>††</sup>  
\*(kat@khn.nict.go.jp)

## 1 はじめに

本稿では、英文読解能力を推定することを目的とした読解速度テストの妥当性と信頼性を古典的テスト理論[2]の枠組みにおいて検証する。

第二言語による読解能力を評価する際、言語処理の効率が重要な判断材料となることから、読解速度など処理効率を示す指標に基づくテストは重要である。文献[1]、[6]は、第二言語教育において言語処理の効率を明らかにする手法として、コンピュータを用いて読解速度を計測する方法を示した。文献[1]では、紙の試験と異なりコンピュータを用いると読解単位が制限されるなどの問題を指摘するに留まっており、読解速度が言語テストの評価基準として有効かどうかについては触れられていない。文献[6]では、読解速度計測に必要なツールを提案するに留まり、読解速度の評価基準としての妥当性には触れられていない。

筆者らはこのような先行研究を踏まえ、言語テストの評価基準として読解速度が有効であるかどうかを検証する必要があると考えた。そして、読解速度を評価基準としたテストの有効性を、古典的テスト理論の枠組みにおいて検証することにした。

## 2 読解速度テストの有効性検証実験

古典的テスト理論では、有効な言語テストが充足すべき条件として、テストの目的が開発目標と合致していること（合目的性）や、運用面の問題が少ないことなどが挙げられている。合目的性に関する条件では、テストの信頼性や妥当性を扱う。テストの信頼性は、テストが同等の能力の学習者を等しく示せるかどうか（結果の安定性）を検証することによって確認できる。また、テストの妥当性は、開発したテストが、信頼性や妥当性が確認されている既存のテストと同様の結果を示せるかどうかを検証して確認できる。これらの条件を満たすテストが有効なテストであるとみなされる。

読解速度テストの有効性を信頼性と妥当性の観点から検証するために、TOEIC（Test of English for International Communications）のテキストを被験者が読む際の読解速度を計測する実験を行った。

読解速度計測実験の被験者は、参加報酬を支払うことを条件に募集した日本語母語話者102名である。読解用テキストとしてTOEIC準拠の問題集から84テキストを抜き出した。各被験者に14テキストずつランダムに割当てた。被験者には、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。また、テキストを全て読み終えた後に設問への解答を行うよう指示した。しかし、設問への解答に際してテキストを再度確認することは許した。読解時間計測ツールにおける読解単位は文とした。被験者は、ツールの操作に慣れるため、事前に操作練習を行った。

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、設問と解答の選択肢の部分は算出対象外とした。

実験の結果、1,232件の英文読解時間データが得られた。実験では設問解答時にテキストを再確認することを許可したため、収集されたデータには、設問解答時にテキストの再確認を行ったデータと行っていないデータの二種類が含まれる。読解速度テストの有効性を検証するに当たり、再確認を行ったデータは排除した。なぜなら、これらのデータは、設問を見る前の初回の読みと再確認のための読みといった異なる読解パターンが含まれるからである。

再確認のための読みを含むデータを排除した結果、138件のデータが得られた。このデータから、各テキストにつき4人以上のデータが存在する11テキストのデータ56件を分析対象とした。さらに、平均読解速度の三倍の速度（1分当たり750語）を閾値とみなし[4]、閾値より大きな値を含むデータ1件を取り除き、最

最終的に 55 件のデータを得た。各テキストに含まれる文の数と各テキストを読んだ被験者の数は表 1 に示すとおりである。

テキスト	文数	被験者数
T1	4	6
T2	6	4
T3	5	4
T4	5	4
T5	24	4
T6	6	4
T7	11	6
T8	11	5
T9	13	7
T10	4	7
T11	6	4

表 1：各テキストにおける文数と被験者数

### 3 有効性の検証

#### 3.1 基準の妥当性：基準関連妥当性

読解速度テストの基準の妥当性（基準関連妥当性）を検証する。一般に、テストに用いる基準が妥当であれば、テスト結果は正規分布するとされている[3]。したがって、読解速度がテスト基準として妥当であれば、テスト結果が正規分布すると予測される。そこで分析対象データの正規性をコルモゴロフ・スミルノフ検定により検証した。結果は、片側確率  $P$  (0.19) が有意水準  $\alpha$  (0.05) 以上であることより、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判断した。分布を示すヒストグラムを図 1 に示す。図 1 に示される平均読解速度とは、被験者が読んだ全テキストの読解速度の平均値である。

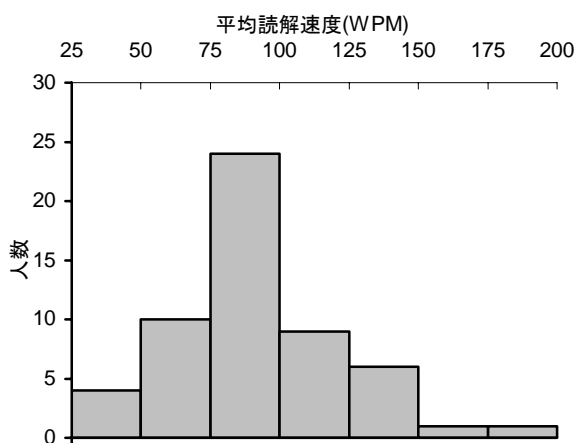


図 1：平均読解速度データヒストグラム

#### 3.2 結果の妥当性：目標関連妥当性

読解速度テストの妥当性を、学習者の読解能力を適切に反映しているかどうかを調べることによって検証する。具体的には、テストを構成する要素の妥当性（構成概念妥当性）である。

読解速度テストの構成概念妥当性の検証を、目標関連妥当性の観点から行った。目標関連妥当性とは、TOEIC など一般的に妥当と考えられているテストの結果との相関により示される妥当性である。

読解速度テストの目標関連妥当性を検証するために、各被験者の TOEIC スコアと読解速度データの相関を算出した。なお、被験者の TOEIC スコアは、被験者が本稿の実験以前に受験したときのスコアである。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。

読解速度テストでは、速度が高ければ読解能力が高いと推定されることから、TOEIC 読解スコアと読解速度データの間には正の相関関係があると予測される。ピアソンの積率相関係数の有意性検定を行った結果、統計的に有意な相関関係 ( $r=0.57$ ;  $p<0.01$ ) がみられ、この予測は正しいことが確認できた。被験者の平均読解速度と TOEIC 読解スコアの散布図を図 2 に示す。筆者らは、この結果から、速度データが読解能力を推定できることを確認した。

さらに、実験条件として設定した『一定の理解度』が適切に機能していることを確認した。今回の実験では、理解度に関する条件として、『TOEIC の設問に答えられる程度の理解』といった指示を与えた。この一定の理解度といった制限が適切に機能していなければ、読解速度から読解能力を推定することは困難である。しかし、読解速度データが TOEIC スコアと有意な相関関係にあることから、『一定の理解度』という条件を受験者に課すことによって、読解速度テストの結果だけから言語処理能力を推定できると考える。

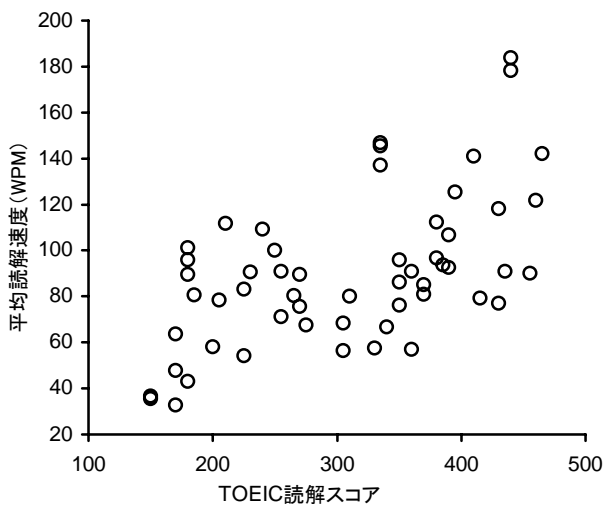


図2：平均読解速度と TOEIC 読解スコアの散布図

### 3.3 結果の妥当性：構成概念妥当性

読解速度テストの妥当性を，異集団法による構成概念妥当性の観点から検証を行った．異集団法による構成概念妥当性は，テスト結果から被験者群を能力別に弁別できるかどうかによって示される．

もし読解速度テストが妥当なテストであれば，被験者を TOEIC 読解スコアに基づいていくつかのグループに分類した場合，各グループ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，文献[5]に倣い，被験者を TOEIC 読解スコアに基づいて上位 25% (14 人)，中位 50% (27 人)，下位 25%(14 人)のグループに分類した．

各グループの平均 TOEIC スコアと平均読解速度は，表 2 に示すとおりである．これらのグループが TOEIC スコアと読解速度において，有意差があるかどうかを分散分析により検定した．その結果，全てのグループ間で統計的な有意差 ( $p < 0.01$ ) が確認できた．

クラス	TOEIC スコア	読解速度 (WPM)
上位	424.3	117.2
中位	311.1	88.9
下位	182.5	66.4

表 2：平均 TOEIC 読解スコアと平均読解速度

### 3.4 結果の信頼性

テストの信頼性は，安定したテスト結果が得られるかどうかを検証することによって確認できる[2]．安定したテスト結果が得られるかどうか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，同一被験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確認する内部一貫性法を用いる．

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・ $\alpha$  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・ $\alpha$  係数は値に関係なく用いることができる[5]．そこで，次の式(1)に従って，クロンバック・ $\alpha$  係数を算出した ( $k$ =文数， $S_Y$ =全文の速度の標準偏差， $S_j$ =各文の速度の標準偏差)．

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

55 件の分析対象データからテキスト毎に信頼性係数を算出した．各テキストにおける信頼性係数は，表 3 に示すように 0.45~0.97 であった．また，データ全体での信頼性係数の平均は 0.78 であった．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と考えられている[2]．したがって，読解速度テストは信頼性の観点から有効なテストとみなすことができる．

テキスト	信頼性係数 $\alpha$
T1	0.81
T2	0.87
T3	0.82
T4	0.75
T5	0.97
T6	0.72
T7	0.96
T8	0.88
T9	0.91
T10	0.46
T11	0.45

表 3：各テキストにおける信頼性係数

## 4 結論と今後の課題

本稿では、読解速度テストの有効性の検証を、速度データをTOEIC読解試験と比較することによって行った。読解速度テストの信頼性と妥当性の検証は、TOEICのテキストを被験者が読む際の読解速度を計測する実験を通じて得た結果に基づいて行った。テストの信頼性は、実験結果（読解速度データ）から算出されるクロンバック・ $\alpha$ 信頼性係数（ $\alpha=0.78$ ）により確認できた。また、テストの妥当性は、被験者のTOEICスコアと読解速度データを比較し、ピアソン積率相関係数（ $r=0.57$ ； $p<.01$ ）により確認できた。

今後、実際に言語テストとして、読解速度テストを単独で実行することによって検証をする必要がある。言語テストとして運用する際には、評価法を検討すべきである。また、評価の一環として、学習者の弱点検出方法も課題となる。さらに、言語テストとして活用する場合、読解速度テストを長期間にわたって用いることにより習熟度を評価できるかどうかも課題である。

## 参考文献

- [1]Alderson, J. C.: “Assessing Reading,” Cambridge University Press, Cambridge (2000).
- [2]Brown, J. D.: “Testing in Language Programs,” Prentice-Hall, Upper Saddle River, NJ. (1996).
- [3]Carver, R. P.: “Optimal rate of reading prose,” Reading Research Quarterly XVIII, pp. 56-88 (1982).
- [4]Ng, M. H., Hall, W., Maier, P., et al.: “Using effective reading speed to integrate adaptivity into web-based learning,” Lecture Notes In Computer Science, vol. 2347. Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Springer-Verlag, London, UK, pp. 428-431 (2002).
- [5]大友賢二: “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 (1996).
- [6]吉見毅彦, 小谷克則, 九津見毅, 佐田いち子, 井佐原均: “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, vol. 22, pp. 24-29 (2005).