

リーディングスキルテストの センター試験および「言語運用力・数理分析力」テストとの相関分析

石岡 恒憲¹, 菅原 真悟²

¹ 大学入試センター, ² 国立情報学研究所

tunenori@rd.dnc.ac.jp¹, sugawara@nii.ac.jp²

1 はじめに

リーディングスキルテスト (以下 **RST** と略す) は, 国立情報学研究所社会共有知研究センター (センター長: 新井紀子) らが考案した, 教科書や新聞, マニュアルや契約書などのドキュメントの意味および意図を, どれほど迅速かつ正確に読み取ることの能力を測定するために開発されたテストである[1]。出題文は主として検定済みの中学, 高校の教科書から採っている (国語と英語を除く) が, 一部, 辞書から採ったものや作問者が独自に作成したものも含んでいる。**RST** では読解力を以下の6つの独立した能力からなると定義している。

- 1) 係り受け解析: 語句の間にある「修飾する」「修飾される」関係の理解
- 2) 照応解決: 指示語や省略された主語が何を指しているかの理解
- 3) 同義文判定: 二つの文が同じ意味かどうかの判断
- 4) 推論: 論理や常識を使って文章を読み解けるか, 文章で書かれていない部分 (省略されている部分) を理解できているか
- 5) イメージ同定: 文章と図表が対応しているか
- 6) 具体例同定: 定義と具体例が対応しているか

これら6つの読解力は, 2つに大別することができる:

- 文の表層的な情報を読み取れる能力として「係り受け解析」「照応解決」「同義文判定」
- 文の意味を理解できる能力として「推論」「イメージ」「具体例」

開発リーダーの新井らによれば人工知能 (東ロボくん[2]) は前者 (「係り受け」「照応」「同義文判定」) を得意とし, 後者 (「推論」「イメージ同定」「具体例同定」) を苦手としているが, **RST** における中高校生の結果も全く同様だとしている[3,4]。

本試行調査では **RST** の比較的難しめの問題を選び, 都

内8つの国公立大学1年生に受験させることを行った。この被験者には当年のセンター試験 (本試と追試) を本番よりほんの少しだけ遅らせて受験してもらうので, これにより両テスト (センター試験と **RST**) の関係が明らかになることが期待された。

2節には本モニター試験で実施する試験形態 (試験時間, 解答時間, 問題数等) について述べる。3節にはセンター試験の総点, および教科科目ごとのスコアとの関係を述べる。4節にはセンター試験と緩やかな相関関係がある「言語運用力・数理分析力」テストと **RST** との相関関係について報告する。5節はまとめである。

2 RST 試験の外形的仕様

RST は主に中高生に対して通常の授業時間内に実施される。このため試験時間が正味30分程度になるように調整されている。本実験では難易度が高めの問題を中心に出题するため, 従来の **RST** より負荷のかかる試験になっている。その詳細は以下の通りである。

(1) 試験時間: 本試行調査での試験時間は解答時間だけで正味38分 (6分×5セッション+8分×1セッション) である。1セッションに何問あるかは被験者には教えない。またセッションの時間も教えない。

(2) 問題数: 6つのセッションに対して各12問とする。計72問を正味38分で解く。

(3) 問題の秘匿: 本 **RST** テストは繰り返し使用するために, 問題は秘匿される。試験で用いた問題冊子は厳密に計数したうえで, 全て回収される。残部についても適切に処理する。

被験者は都内の国公立大学の一年生であるが, その所属大学と人数は表1に示す通りである。被験者には謝金が支払われるが, 成績によって金額の多寡は生じない。

表 1：被験者の属性

大学	文理の別	人数 (計 352 名)
A 大学	文系と理系	35 名
B 大学	理系	51 名
C 大学	文系	70 名
D 大学	文系と理系	65 名
E 大学	理系	51 名
F 大学	文系と理系	18 名
G 大学	文系と理系	35 名
H 大学	理系	27 名

3 センター試験の教科科目スコアおよび総合点と RST スコアとの関係

RST スコアを都内の国公立大学別に箱ひげ図に示したのが、図 1 である。縦軸が RST6 領域の得点率を示す。箱ひげ図では箱がデータの四分位範囲（上位 25% から下位 25% まで）を示し、箱の中の線が中央値（50% 点）を示す。A, B, C の大学は入学偏差値の高い大学であり、これらの大学の RST スコアが総じて高いことがわかる。D 大学はこの中では比較的入学偏差値の高くない大学であり、RST スコアも低いことがわかる。

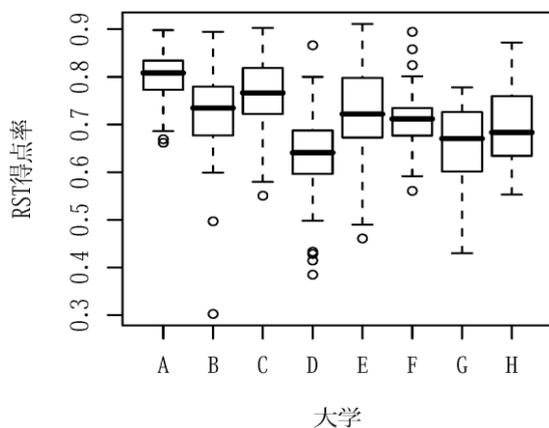


図 1：大学別の RST スコアの分布

紙面の都合で割愛するが「照応解決」と「係り受け解析」の表層的観点のみでみると、RST 得点率が平均で 8 割を超えているために大学間の差は目立たないものや、やはり差がある。それ以外（すなわち「同義文判定」「推論」「イメージ同定」「具体例同定」）の深層的な観点で見ると得点率に大学間の違いが大きく現れることが確認されている。

次に RST スコアとセンター試験スコアとのピアソン

相関係数を算出した。センター試験スコアは本試と追試の合計点を用いた。RST スコアはセンター試験総計スコアとの相関が ($r=0.50$) であり、各教科との相関は、国語 ($r=0.30$) ; 地歴 ($r=0.20$) ; 公民 ($r=0.25$) ; 数学 ($r=0.44$) ; 理科 ($r=0.39$) ; 英語リスニング ($r=0.26$) ; 英語 ($r=0.25$) となった。これより、RST は読解力を測るテストであるにもかかわらず、読解力と関係の深そうな国語や英語、英語リスニングとの相関が 0.30 以下と小さいことがわかる。最も相関の大きいのが（特定の教科学力ではない）センター試験総計スコアであることから、RST スコアがほぼ総合的学力を測定しているらしいことが示唆される。ただしその一方で、明確な相関を示しているわけではない。

なお RST の基本統計量としてクロンバックの α 係数を算出すると 0.69 となった。センター試験の同様の指標は科目によるが、おおむね 0.8 程度ある。試験の内的一貫性はセンター試験における各科目試験ほどには高くないことがわかる。ただ RST は試筆テストではなく CAT (Computer-adaptive Testing : コンピュータ適合型テスト) を想定して開発を進めてきており、CAT にすることで限られた時間での信頼性は向上するものと思われる。

4 「言語運用力」「数理分析力」テストスコアと RST スコアとの関係

大学入試センター研究開発部では (AO 入試や推薦入試などの) 志願者の多様化に伴い大学入学後の履修に必要な寄与的学力を担保する試験を開発してきた[5]。この試験は 2006 年と 2009 年の PISA における「評価の枠組み」を参考に「言語運用力」と「数理分析力」の 2 つのテストから構成されている。このテストスコアはセンター試験と緩やかな相関関係があることがわかっているため、本節では RST と「言語運用力」「数理分析力」テストとの関係について調査した。

今回、実施した「言語運用力」「数理分析力」テストは各 20 問、60 分である。このテストスコアと RST スコアとのピアソン相関を示したのが表 2 である。RST と「言語運用力」「数理分析力」とのピアソン相関は大きくはなくそれぞれ 0.43 と 0.42 であるが、両領域との相関はそれよりも大きく 0.50 である。RST も「言語運用力」テストのいずれも読解に関するテストであるにもかかわらず、相関が大きくないことは興味深い。

表 2: 「言語運用力」「数理分析力」テストと RST スコアとのピアソン相関

	言語運用力	数理分析力	両領域合計	RST
言語運用力	1.00	——	——	——
数理分析力	0.48	1.00	——	——
両領域合計	0.87	0.82	1.00	——
RST	0.43	0.42	0.50	1.00

センター試験, RST, 「言語運用力」「数理分析力」3つのテストの関係を相関係数という視点で模式図風に示したのが図 3 である。3つの試験がお互いに同じような相関係数で結ばれている。センター試験が高校学力を測定するという意味で真性な学力と呼ぶならば, RST も「言語運用力」「数理分析力」のいずれもこの学力と緩やかな相関を持つ学力と呼ぶことができよう。

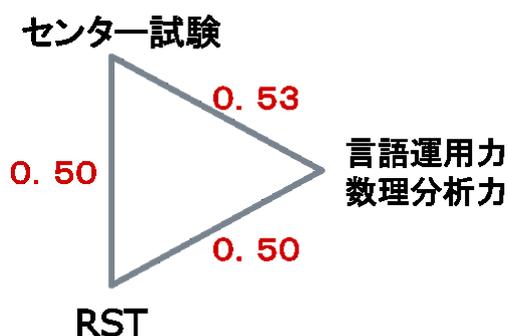


図 3: 3つのテストの相関

次に RST の 6 領域と, 「言語運用力」「数理分析力」「両領域合計」の 3 領域, 併せて 9 領域のそれぞれの相関を図示したのが図 4 である。上から「言語運用力」「数理分析力」「両領域合計」「照応解決 (ANA)」「係り受け解析 (DEP)」「推論 (INF)」「具体例同定 (INST)」「同義文判定 (PARA)」「イメージ同定 (REP)」を示す。対角線上にそれぞれのスコアの度数分布を, 対角線の左下に散布図と回帰曲線を, 対角線の右上に相関係数を示す (大きな値ほど大きく表示される)。

これを見るに, 「言語運用力」「数理分析力」「両領域合計」の 3 つはお互いに大きな相関があるのは当然だとしても,

- RST6 領域は, 「言語運用力」と強い相関はない
- 「数理分析力」も同様である
- 「両領域合計」との相関が最も強くなる

ことがわかる。RST の 6 つの領域での関係を見ると, 回帰曲線のグラフはどれもほぼ横に寝ていることから, 互いの相関の無さがわかる。また「係り受け解析 (DEP)」は被験者にとって易しく, モード (最頻値) が全問正解の J 字の形をした分布であることがわかる。

本稿では RST スコアを正解数で表したが, その後, RST の統計班より提供された項目反応理論に基づく領域ごとの能力値 (項目反応理論において想定・算出される受験者の能力水準を数値化した母数の推定値) を得てこれに置き換え, 同様の解析を行ったが, 結果に違いはほとんど生じなかった。

5 まとめ

RST が定義する読解力は 6 つの独立した能力から構成されるが, これらは互いに相関が大きくない。もっとも今回は都内 8 つの国公立大学の 1 年生という相対的に偏差値の高い人向けに難易度の高い問題を選択して実施したため, 一種の選抜効果により小さめの相関になったことは否定できないだろう。一方でこれらの合計がセンター試験スコアと相関を最も大きくもつことから, 漠然とした「いわゆる学力に関係する基礎的な能力」を測定するツールであることが確認された。このことは同じく大学入試センター研究開発部が考案した「言語運用力」「数理分析力」との緩やかな関係や, 大学別のスコアなどからも傍証されるといってよいだろう。

謝辞

本調査を実施するにあたり国立情報学研究所・社会共有知研究センターの新井紀子センター長・教授を始め, 統計グループの諸先生方, 関係各位に多くのご協力を賜りました。ここに記して心より謝意を表します。

参考文献

- [1] 国立情報学研究所ニュースリリース (2016 年 7 月 26 日) 文章を正確に読む力を科学的に測るテストを開発/産学連携で「読解力」向上を目指す研究を加速, ニュースリリース:
http://www.nii.ac.jp/userimg/press_20160726.pdf
 別紙資料 1: [press_20160726-1.pdf](#), 別紙資料 2: [press_20160726-2.pdf](#)

[2] ロボットは東大に入れるか。Todai Robot Project : <http://21robot.org/>

[3] Yahoo!ニュース (2016年11月14日) 『AI研究者が問う ロボットは文章を読めない では子どもたちは「読めて」いるのか?(湯浅誠)』

[4] Noriko H. Arai, Naoya Todo, Teiko Arai, Kyosuke Bunji, Shingo Sugawara, Miwa Inuzuka, Takuya Matsuzaki, Koken Ozaki (2017): Reading Skill

Test to Diagnose Basic Language Skills in Comparison to Machines, Proceedings of the 39th Annual Cognitive Science Society Meeting (CogSci 2017) 1556-1561.

[5] 大学入試センター研究開発部(2017)「大学での学修に必要な基本的能力の測定 最終報告書」, 平成23-27年度特別研究「新しい試験の開発に関する調査研究」報告書, 平成28年3月.

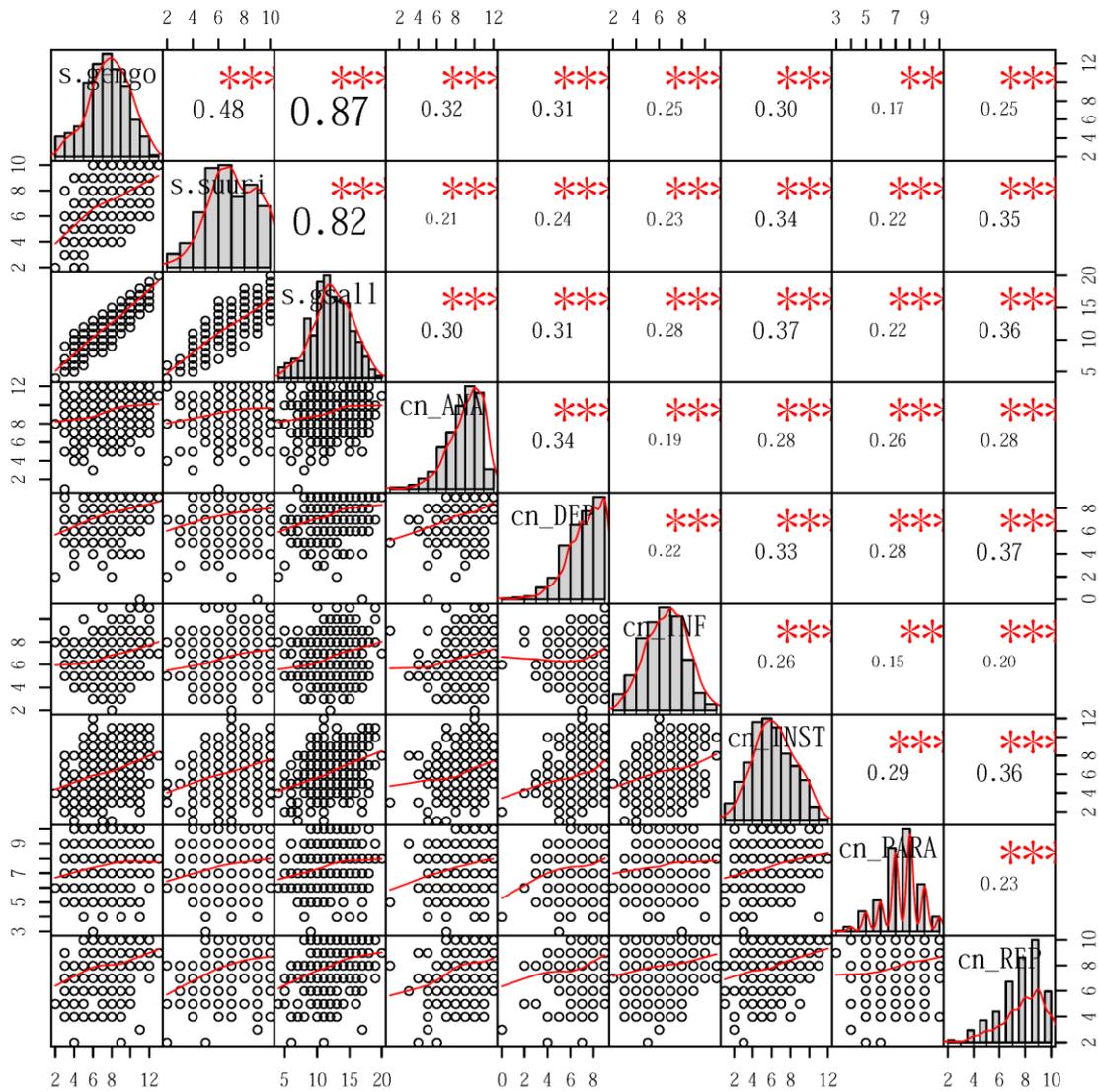


図4: 言語運用力・数理分析力・両領域(3領域) × RST(6領域)