【原著】　　大学入試研究ジャーナル第 30 号，36-43，2020.

リーディングスキルテスト，センター試験，「言語運用力・数理分析力」テストの相関および因子分析

石岡 恒憲（大学入試センター），菅原 真悟（国立情報学研究所）

国立情報学研究所社会共有知研究センターが考案したリーディングスキルテストが，大学入試センターが作成するセンター試験，および「言語運用力・数理分析力」テストとどのような統計的な関連があるかを観察するために，相関分析および因子分析を行った。その結果，RST の各スコアはどのようなセンター試験の教科科目とも相関が大きくはなく，RST スコアの合計がセンター試験総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「言語運用力・数理分析力」テストとの緩やかな関係や，大学別のスコア，あるいはセンター試験科目とを合わせた因子分析の結果などからも傍証される。

キーワード：読解力，相関係数，信頼性係数，因子負荷量

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ができるかの能力を測定するために開発されたテストである（国立情報学研究所ニュースリリース , 2016）。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）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 (Arai, et al, 2017)。

本実験では 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）を本番とほぼ同じ時刻に受験してもらう。これにより被験者はセンター試験問題を事前に知ることはない。使用した RST テスト問題も全て非公開である。このため被験者は両テスト（センター試験と RST）冊子をともに初めて見ることになり，これにより両テストの関係を明らかにできると期待される。本稿の目的は，巷間大きな話題となっている RST のテストスコアが，比較的正当な学力と見なされているセンター試験スコアとどのような関係にあるのか，また大学入学のための基礎診断のために作られた「言語運用力・数理分析力」テストのスコアとどのような関係にあるのかを，比較的易しい統計解析である相関分析や因子分析を使って明らかにすることにある。

2 節には本実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 節にはセンター試験と RST スコアを込みにした因子分析の結果を示す。6 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題するため，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。

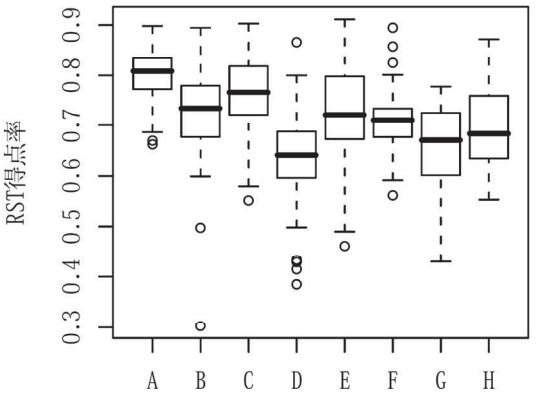
被験者は都内の国公立大学の一年生であるが，その所属大学と人数は表 1 に示す通りである。被験者には謝金が支払われるが，成績によって金額の多寡は生じない。

表 1：被験者の属性

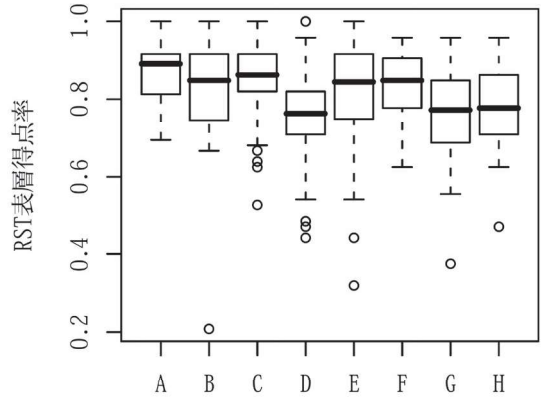
大学	文理の別	人数(計 352 名)
A 大学	文系と理系	35 名
B 大学	理系	51 名
C 大学	文系	70 名
D 大学	文系と理系	65 名
E 大学	理系	51 名
F 大学	文系と理系	18 名
G 大学	文系と理系	35 名
H 大学	理系	27 名

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

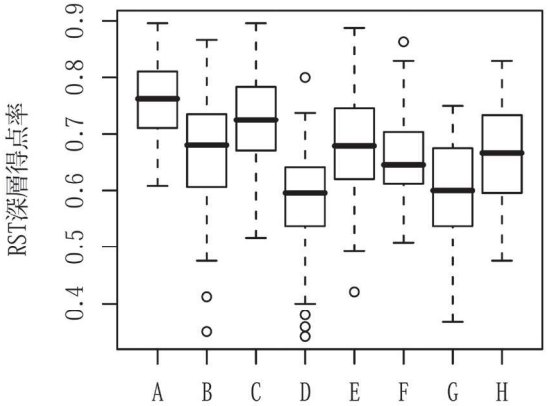
RST スコアを都内の国公立大学別に箱ひげ図に示したのが，図 1(a) である。縦軸が RST6 領域の得点率を示す。箱ひげ図では箱がデータの四分位範囲（上位 25%から下位 25%まで）を示し，箱の中の線が中央値（50%点）を示す。A,B,C の大学は入学偏差値の高い大学であり，これらの大学の RST スコアが総じて高いことがわかる。D 大学はこの中では比較的入学偏差値の高くない大学であり，RST スコアも低いことがわかる。



(a) 全部の観点



(b) 表層的観点（「照応解決」と「係り受け」）



(c) 深層的観点（「同義文判定」「推論」「イメージ同定」「具体例同定」）

図 1：大学別の RST スコアの分布

図 1(b)「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のやはり差がある。図 1(c) 深層的な観点（「同義文判定」「推論」「イメージ同定」「具体例同定」）でみ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る。「同義文判定」は表層的観点からのみでは判定できないこ

とも多く（国立情報学研究所，2013），ここでは深層的な観点のカテゴリーに入れる。

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もない。

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 程度以上ある。これより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。ちなみに，この試験について，よく知られた真の得点の 90％信頼区間の公式，

$$\mu + \rho(X_i - \mu) - 1.65\sigma\sqrt{1-\rho} < T_i < \mu + \rho(X_i - \mu) + 1.65\sigma\sqrt{1-\rho}$$

ただし，X_i は個人 i の RST の得点，μ は平均値，σ は標準偏差，T_i は同じ個人の真の得点とし，ここに例えば，

$$X_i = 60,\ \rho = 0.69,\ \mu = 50,\ \sigma = 10$$

という値を代入してみると，

$$47.7 < T_i < 66.1$$

ということになる。これはある人があるとき，偏差値 60 相当の得点をとったとしても，実際には平均（偏差値 50）以上の「能力」かどうかさえ確実には言えないことになる（あくまでも，α 係数を ρ と等しいとしての話ではあるが）。これより RST は，少なくとも大学教育レベルにおける個人差の評価には用いられないということはいってよいだろう。もちろん，同じ大学から 50 人程度の学生をランダムに選んで，RST の平均値を求めれば，そこからその大学の偏差値をかなりよく予測することはできるだろうし，集団間の差異を判別する指標としては有効だろう。もっとも RST は紙筆テストではなく 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紙筆テストよりも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 スコアとの相関分析

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基礎的学力を担保する試験を開発してきた（大学入試センター研究開発部，2017）。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ので，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して示したのが図 2 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。

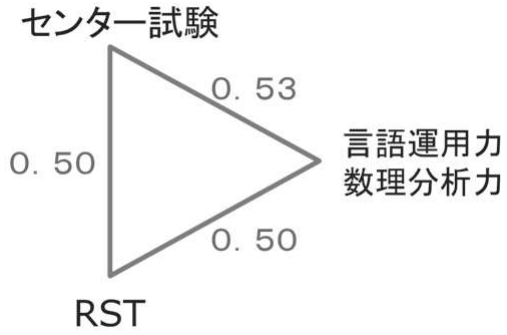


図 2：3 つのテストの相関

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 3 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）。

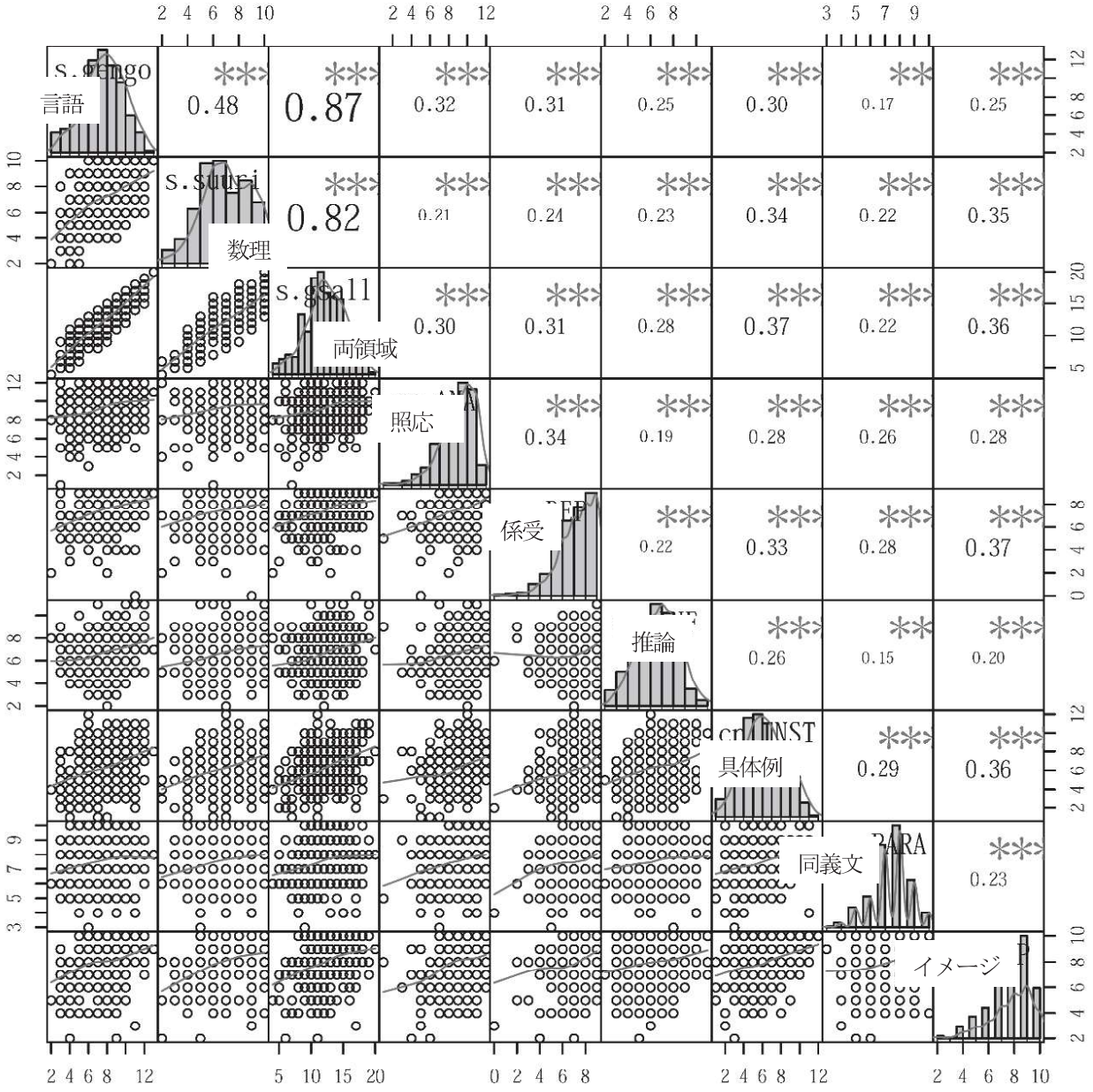


図 3：言語運用力・数理分析力・両領域（3 領域）× RST（6 領域）のヒストグラム（対角線），散布図（左下３角），相関係数（右上３角）

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」とも同様に強い相関はない
- 「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 因子分析(センター試験 7 科目+ RST6 領域)

因子分析は観測変数に影響を与えている共通因子を抽出する統計的手法である。テストデータ分析においては得点という観測変数が,潜在的な学力の発露として考えるのが自然であるため(主成分分析ではなく)因子分析が多用される。本稿でも慣例にしたがい,RST(6 領域)とセンター試験(7 科目)とを合わせて,これらの得点に関わる因子を抽出する因子分析を試みた。それぞれ単独の因子分析を試みるのではなく両テストの科目を合わせて因子分析を行うのは,そのことにより RST テストのセンター試験科目との関係がわかるようになると期待されるからである。

本稿では現在最も標準的とされる,因子の抽出法に最尤法を,因子軸の回転法にプロマックス法を用いる因子分析をおこなった。因子分析で因子をいくつ抽出すればよいかを判断するために,横軸に因子を,縦軸に固有値の大きさをプロットしたもの(いわゆるスクリープロット)を図 4 に示す。因子数決定の基準には従来からも,固有値が 1 以上の因子を採用するガットマン基準や,スクリープロットにおいて推移がなだらかになる前まで選ぶスクリー基準や,累積寄与率がある閾値(たとえば 50 ～ 60%)以上になる因子までを採用する方法などがある。これら基準を勘案すれば因子の数は 2 ～ 4 を抽出すればよいように思われる。

因子独自性と因子負荷量を伝統的な表示方法に従い,表 3 に示す。A ～ J までがセンター試験科目であり,α～ζまでが RST 領域である。これより第 1 因子は,RST 能力にセンター試験の数学を加えた「論理読解力」のような因子と判断できる。第 2 因子は,センター試験能力を主としつつも数学が逆に作用していることから「ある程度の勤勉力=勤勉に勉強のできる能力」を示す因子であろう。第 3 因子はセンター試験の英語力と RST の具体的同定が高いことから「語学力」を,第 4 因子はセンター試験の理科系科目と RST の同義文判定が高いことから「理科的な能力」を示しているように思われる。

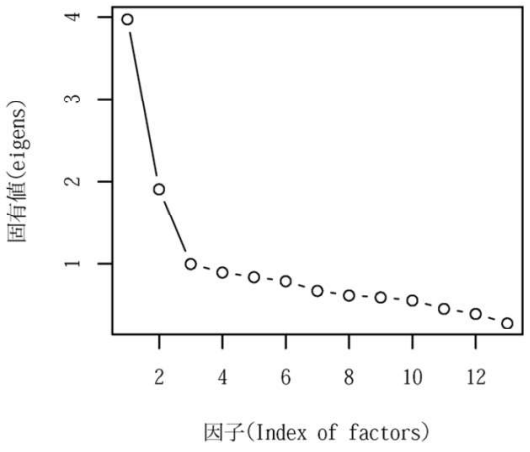


図 4:スクリープロット(センター試験 7 科目+ RST6 領域)

因子間相関行列を表 4 に示す。これより,第 1 因子(論理読解力)と第 4 因子(理科的な能力)との相関が比較的強いこと,また第 2 因子(ある程度の勤勉力)と第 3 因子(語学力)との相関も同程度に強いことが確認できる。これは普段我々の感じる印象とかなり符合する。

表 3 の情報を可視化するために,各変数の因子共通性(= 1 −因子独自性)を図 5 に示す。因子共通性は共通因子によってどの程度まで説明できているかを示す指標であるが,これよりセンター試験科目の方が RST よりも因子共通性が多い(独自の変数に係る影響が少ない)ことがわかる。逆に言えば,RST6 領域は独自性がより大きいことを示している。図 6 には 4 つの因子における各変量の因子負荷量を図示する。

また,付録には読者が再分析できるように,13 の変数間の相関行列と,各変数の平均値と標準偏差を示す。

表 3：センター試験 7 科目＋ RST6 領域による因子分析

因子独自性

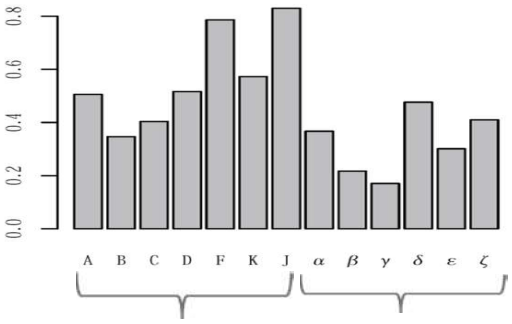
A:国語	B:地歴	C:公民	D:数学	F:理科	K:リス	J:英語
0.494	0.654	0.596	0.483	0.214	0.427	0.170
α::照応	β:係受	γ: 推論	δ:具体	ε:同義	ζ:イメージ	
0.633	0.783	0.829	0.524	0.699	0.590	

因子負荷量

	第 1 因子	第 2 因子	第 3 因子	第 4 因子
A:国語		0.606	0.150	
B:地歴	-0.104	0.514		0.191
C:公民		0.652		
D:数学	0.538	-0.373	0.172	0.287
F:理科		0.212		0.830
K:リス		0.184	0.642	
J:英語		0.150	0.824	
α:照応	0.454	0.347		
β:係受	0.443	0.108		
γ:推論	0.381			
δ:具体例	0.655			
ε:同義文	0.483		-0.116	0.134
ζ:イメージ	0.716			

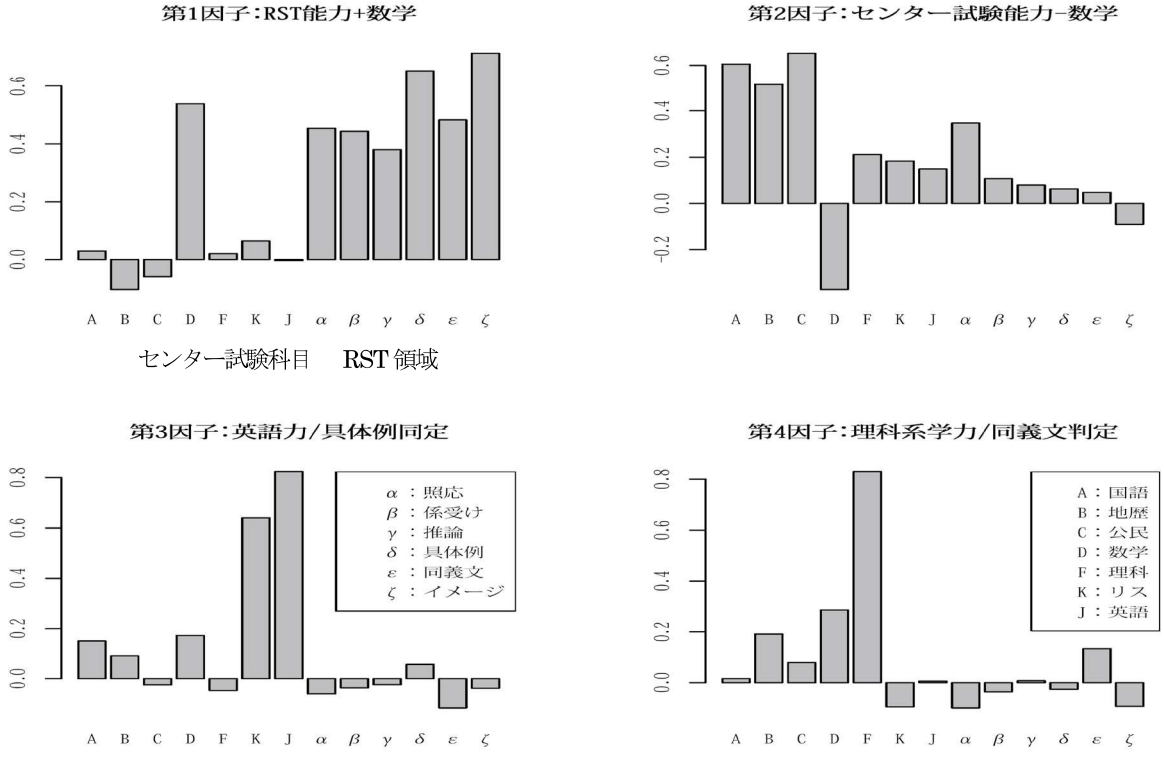
表 4：因子間相関行列

	第 1 因子	第 2 因子	第 3 因子	第 4 因子
第 1 因子	1.000			
第 2 因子	0.252	1.000		
第 3 因子	0.300	0.483	1.000	
第 4 因子	0.520	0.173	0.372	1.000



センター試験科目　　RST 領域

図 5：因子共通性（1 −因子独自性）



センター試験科目　　RST 領域

図 6：4 つの因子における各変量の因子負荷量

6 考察とまとめ

信頼性係数が ρ_0 であるようなテストを，同質な項目を加えることによって n 倍に引き伸ばしたときの信頼性係数を与える式として，よく知られた Spearman-Brown の公式がある。

$$\rho = \frac{n\rho_0}{1 + (n-1)\rho_0}$$

これは，相互相関が ρ_0 の等質な 6 つのテストの総点の信頼性を与える公式と解釈できるから，この左辺に（RST の α 係数である）0.69 を代入し，$n = 6$ として解くと，$\rho_0 \fallingdotseq 0.27$ を得る。これを図 3 における RST の 6 つの下位テスト間の相関と比較すると，ほぼ同じであることがわかる（実際に 15 個の相関係数の平均値は，ちょうど約 0.27 となる）。これがもう少し低ければ，α 係数の低い理由も，異質な下位テストを（無理に）加算したからということになるだろうが，この結果はむしろ 6 つの下位テストは，区別のつかない程度に「漠然とした」共通の成分を反映しているということがいえる。つまり RST が定義する 6 つの（下位）尺度間の相関は高くないが，この結果はこれら 6 つの尺度の信頼性の低さに起因するものであろう。また因子分析における第 1 因子がセンター試験の数学と RST 能力で代表されることから，RST は数学で高得点をとるための処理能力の一端を多く反映していることがうかがえる。

今回の受検者は都内 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層に偏っており，それに対応して難易度の高い問題を選択して実施したにもかかわらず低めの相関となった。したがって，彼らを対象とした場合には，RST は漠然とした「いわゆる学力に関係する基礎的な能力」を測定するツールとしてのみ機能するということがいえよう。更に踏み込んでいえば，読解力よりもむしろ数学（や理科）などの処理能力の一端を反映しているようである。もっともセンター試験は読解力そのものを測定する試験ではないので，RST が読解力（reading skill）を測定するテストであるという証拠は少なくとも本実験からは得られていない。ただ RST が「いわゆる学力に関係する基礎的な能力」を測定することは，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。なお RST の主催元である「一般社団法人 教育のための科学研究所」の Web ページ https://www.s4e.jp/about-s4e には RST の例題が示されている。必要に応じて参照されたい。

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。

また 2 名の査読者には本当に多くの貴重で有益なコメントをいただきました。特に 6 節の RST の信頼性係数に関する考察や 3 節の α 係数のもつ具体的な意味についての記述は査読者からのコメントに基づくものです。ここに記して心より謝意を表します。

参考文献

Arai, H. N., Todo, N., Arai,T., Bunji, K., Sugawara, S., Inuzuka, M., Matsuzaki, T., and Ozaki, K. (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.

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.

国立情報学研究所ニュースリリース (2016). 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 2016 年 7 月 26 日：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

国立情報学研究所 (2013). 問われるのは意味を理解する力。暗記だけでは解けない社会科科目 , NII Today, No.60, 8−9.