

リーディングスキルテスト (RST)

—その教育測定学的性質と教育実践に対する示唆—

話題提供: 新井紀子 # (国立情報学研究所)

話題提供: 尾崎幸謙 # (筑波大学)

話題提供: 新井庭子 # (東京大学)

指定討論: 野口裕之 (名古屋大学)

話題提供: 菅原真悟 # (法政大学)

話題提供: 犬塚美輪 (東京学芸大学)

話題提供: 分寺杏介 # (東京大学)

企画・司会・話題提供: 登藤直弥 (大学改革支援・学位授与機構)

キーワード: 読解力, 項目応答理論

企画主旨

登藤直弥

本自主企画シンポジウムで取り上げるリーディングスキルテスト (Reading Skill Tests, RST) においては、いわゆる国語の文章読解問題のように、「文意が正しく読み取れるか」などといった大まかな能力が測定されているわけではない。RSTではより細かな認知能力、具体的には、「文の中における係り受け構造が正しく理解できるか」だとか「文間での照応関係が正しく認識できるか」などといった、より基礎的な読解能力が測定されるように設計されている。

このRSTに関しては、一昨年度よりその開発を始め、現在では、様々な学校・企業・自治体の協力を得て、データの収集や分析を進めているところである。その結果、以下に記載されているように、RST、さらには、「読解」という行為に関する様々な知見が蓄積されてきた。

そこで、本自主企画シンポジウムでは、RSTの開発を始めた経緯、RSTの仕様、RSTの教育測定学的性質、そして、RSTを用いて行われたいくつかの研究の結果について、以下に示す通り、計6名の先生方から話題提供をしていただく。

基盤となる読解力を測る

新井紀子

21世紀の知識基盤社会においては、人生を通じて必要となる知識やスキルが確立されたカリキュラムの枠組みだけでは十分には獲得できず、人が生涯にわたり学び・学習の活動を続けていく「生涯学習」が必要となる。その形態が、アナログかデジタルか、直接か遠隔かに関わらず、学習コンテンツの多くが書記言語の「読解」を前提として構成されている。そのため、知識基盤社会においては、書記言語の運用能力の如何が、労働市場での価値を大きく左右する。一方、学習者間で看過できない読解力の差が生じていることは動かしがたい事実である。読解力が知識基盤社会を生き抜くための核心的スキルであるならば、すべての学習者がそれを一定程度身に着ける教育プログラムがあることが、格差の少ない民主的な社会の形成の上で強く望まれる。

我々は、近年の深い言語処理の研究成果と対比

させつつ、従来の読解の認知プロセスモデルを精緻化した上で、項目応答理論と認知診断モデルを用いた適応型テストを用いることにより、従来の学習到達度テストや読解力調査では測定することのできなかった読解に関わる認知プロセスにかかわる能力（認知能力）を高い精度で診断することを目的とするRSTの研究開発に着手した。その成果について発表を行う。

RSTの各問題タイプの目的と作問方法

菅原真悟

RSTでは6つの問題タイプを設定し、それぞれ異なる能力を測ることを目的としている。各問題タイプの概要は、以下の通りである。

- (1)「係り受け認識 (DEP: Dependency Analysis)」: 係り受け関係を正しく認識できているかを測る。
- (2)「照応解決 (ANA: Anaphora Resolution)」: 指示照応やゼロ照応を理解できているかを測る。
- (3)「同義文判定 (PARA: Paraphrasing)」: 2つの文を比較し、それらが同義かあるか否かを理解できるかを測る。
- (4)「推論 (INF: Logical Inference)」: 提示された文から論理推論することで、新しい知識を獲得できるかを測る。
- (5)「イメージ同定 (REP: Representation)」: 提示された文を読んで、それがどのようなことを表しているかイメージできるかを測る。
- (6)「具体例認識 (INST: Instantiation)」: 概念または用語の定義を読み、それがどのような状況に当てはまるか、具体的に認識できるかを測る。

各問題は、主に中学校及び高等学校で用いられている検定済教科書に載っている文を基に作問している。なお、RSTで教科書等に掲載された文を使用するにあたっては、各著作権者と協議のうえ、著作権許諾契約を結び実施している。

本発表では、各問題タイプが、どのような目的で作問されているかを、具体例をあげて説明する。

RSTの信頼性と妥当性および質問紙調査項目との関係

尾崎幸謙・登藤直弥

本発表では、1) RSTの信頼性と妥当性、2) 質問

紙調査項目との相関について報告する。

RST の各項目が適切に読解にかかわる認知能力を測定していることを調べるために、まず項目特性図を描き、その形状によって適切な項目であるか否かを判断した。通常の項目特性図は、合計得点のグループごとの各選択肢の選択割合を示したものである。しかし、現在の RST は、受験者に対してランダムに項目提示を行っているため、受験者によって解いた問題が異なる。そのため、受験者に対する評価は項目応答理論を用いて行った。項目応答理論における受験者特性値 θ をすべての受験者について求め、これを合計得点の代わりとして項目特性図を描いた。

その結果、いくつかの問題については、問題文や選択肢の修正の必要があることが分かった。なお、RST では修正の必要の有無の判断・修正・再データ収集のループを回して実施し、適切な項目内容となるためのフローを作っている。

また、RST の信頼性と妥当性については、項目特性図によってスクリーニングされた適切な項目のみを使って問題タイプ毎に検討を行った。具体的には、信頼性であれば、たとえば ω 係数 (McDonald, 1999) を用いて検討を行い、妥当性であれば、受験者特性値 θ 間の相関係数や「各受験者特性値や正答率と年齢との関係」などを用いて検討を行った。その結果、現状では、全ての問題タイプにおいて、十分な信頼性と妥当性が示唆される結果が得られている。なお、問題タイプ毎に、項目応答理論を適用する際の前提が満たされているかどうかについても確認を行ったが、この点についても、適用するにあたっての前提は満たされているであろうことが示唆されている。

さらに、信頼性と妥当性が確認された項目群を用いて、質問紙調査項目との相関を求めた。なお、これらの分析についても問題タイプ毎に行っている。この質問紙調査項目には、読書習慣、読書の好き嫌い、得意な科目などについて尋ねるものが含まれている。したがって、読解力という漠然とした構成概念ではなく、読解力を構成するより細かな認知能力について、読書習慣等に関する質問紙調査項目との相関を調べることができた。

なぜ読むのが難しいのか：眼球運動の特徴から

犬塚美輪

RST の実施結果からは、教科書という基本的な学習リソースにおいて文理解が困難な学習者が少なくないことが示されたと言える。こうした困難の要因の一つは、教科書の文章の構造の複雑さにあると推測できる。一方で、学習者が十分な読解スキルを有していないことも要因であろう。読解スキル育成のためには、優れた（あるいは困難を持つ）読み手の特徴に基づいた介入案の検討が必要であるが、青年期の読み手の読解プロセスにつ

いては基礎的データが十分ではない。

そこで、本発表では、大学生を対象に、原文と構造を明確化した修正文における読解プロセスを検討する。予備的な研究からは、原文と修正文で読解時間は大きく異なるものの、修正文では低成績者の理解を底上げすることが示唆された。本発表では、眼球運動のデータを用いて、原文と修正文で学習者の読解プロセスがどのように異なるか、そこからどのような教育的提言が可能かを論じる。

小・中ギャップにつながる教科書テキストの特徴

新井庭子・分寺杏介

教科書の文章は正しく読めることが前提とされてきたが、RST の実施結果からは、学習者が正しく教科書を読めていない実態が示された。本発表では、学習者の読みを困難にしている教科書の具体的な特徴について検討を行った結果を報告する。

近年、教育現場でいわゆる小・中ギャップ、すなわち、小学校から中学校への進学において、新たな環境での学習や生活に対応できず不登校等に繋がる状況が問題視されている（中央教育審議会初等中等教育分科会, 2012）。小・中ギャップは学習者の学習面と生活面から検討が進められており、学習面では、担任制の相違や授業形態の違いなどの指導法の面からの研究はあるが（中央教育審議会初等中等教育分科会, 2012 ; 伊藤, 2014）、学習者が抱える困難がどのようなものか、例えば教科書を読んで理解し知識を獲得できているか等を具体的に分析した研究は少ない。

そこで、小学校と中学校の理科教科書を題材として、そこにどのような量的・質的なギャップがあるかを分析した。また、読みを困難にするテキストのパラメータを予測し、小・中教科書テキストの間にそのパラメータで表現できるギャップがあることを示した上で、RST の結果を用いてそのギャップが実際に人にとっての困難につながるかどうかを確認した。

「人にとってのテキストの難しさ」を扱う関連研究で伝統的に採用されてきたパラメータのうち、単語親密度・係り受けの数・距離の平均・ツリー構造の深さを採用し、特に係り受けの複雑さについて小・中間で明確な差があることを確認した。その他にも、これまでの研究では考慮されていなかった新規パラメータとして定義表現の数・分類表現の数について、小・中間で大きなギャップがあることを示した。