

デジタルライゼーション時代に 求められる人材育成

21世紀はデジタルライゼーションと呼ばれる技術の大変革期であり、人間の労働に大きな衝撃がもたらされる。あらゆるコトが機械で可読・処理可能な形式でデータ化され、もはや情報を伝達するだけの仕事や定型的な仕事は機械で代替される。このような時代に、われわれ人間はどのような能力を身につけるべきだろうか。

2011年に開始したAIプロジェクト「ロボットは東大に入れるか」の研究の結果、AIは言葉の意味を理解し状況を判断することが苦手である、という結論に達した。逆にいえば、文脈理解や状況判断が要求される問題解決型の仕事がある限り、人間の役割は無くならない。このような仕事では、適切なコミュニケーションをとれる人間であれば、機械以上の能力を発揮することができる。しかし、筆者らが行ったリーディングスキルテストからは、言葉や文脈がもつ意味を理解しながら読む能力が身につけておらず、AIと同じような表層的な読み方しかできない者が少なからずいることがわかった。

これらの結果を踏まえれば、今後の国語教育を、読む能力を養う論理的活動へシフトさせるという案は一考に値するだろう。

新井紀子

Noriko Arai

国立情報学研究所
情報社会相関研究系教授

尾崎幸謙

Koken Ozaki

筑波大学大学院
ビジネス科学研究科准教授

第I部 ロボットと人の違い

1. デジタイゼーションという技術革新

デジタイゼーションとは何か

21世紀に入り、われわれは、デジタイゼーションと呼ばれる技術の大変革に直面している。デジタイゼーションとは、従来は記号化されずに流通していたあらゆる情報をコンピューターが可読・処理可能な形式で流通させることを指す。例えば、アナログの文書をスキャンしてパソコンに読み込んでも、機械がそれをテキストとして読めない限り、それはデジタイゼーションとはいえない。近年関心を集めているのは、さまざまなモノにセンサーを付け、それによってあらゆるコト(イベント・状況・状態やそれらの変化)を、コンピューターが解析可能な情報にすることを目指すIoT(Internet of Things)である。昨今、話題となっている人工知能(AI)技術も、「人と同じように思考する機械」を目指す技術というよりも、モノについてのセンサー情報から、機械にコトを「理解」させるための技術とみなされるべきである。言い換えるならば、デジタイゼーションとは、これまで人と人との間で進んでいた情報の処理を、機械に代替させるために不可欠となるインフラ整備といえる。

機械に代替される労働とは

デジタイゼーションは人の労働に大きな衝撃をもたらす。まず、情報を伝達するというだけの仕事は真っ先に機械に代替されると考えてよい。例えば、銀行や役所における帳票システムについて考えてみよう。帳票に人間が手で記入することを廃止し、最初からデジタルで入力すれば、手で記入されたことを転記するという単純な情報伝達の作業を担う労働は不要になる。また、「定型的」な処理も機械に代替されることになる。ここでいう「定型的」とは機械にとっての、という意味である。それは、入出力の型が数学的に決まっており、入力がある条件を満たすときにどのような処理をするかを数学的に記述できれば、機械による代替が可能だ。例えば、「お薬手帳」がデジタル化されたとして、必要以上の投薬がなされたり、飲み合わせに注意が必

要な薬が出ているときに、アラートを表示したりするのは典型的な定型的処理である。

近年、人間の判断を統計的に模倣する「機械学習」という方法論が開発されたことにより、機械ができる業務の幅も広がった。送電線の点検作業やX線写真での結核の有無の診断のように、入力として得られる情報が定型的で、出力すべき判断がYES/NOの2値であるような業務は相当数ある。さらに、その延長線に位置づけられるものとして、入力を2値より多い複数のタイプに分類するものもある。写真に写っている物体が何であるかを当てるような判断や、手書き文字や音声などの文字や言葉に該当するかを判別するようなタスクは、その代表的な例である。このような「分類判断」と呼ばれる課題は、入力データとそれに対する正解・不正解に関するデータが十分大量に与えられれば、機械はしばしば人間が行うより高い精度を発揮する。例えば、ゴッホの作品の真贋を見極めたり、めったにないタイプの病気を診断したりする場面で、人間を上回る精度を達成する¹。

AIが分類判断をする際、重要となるのが「ランキング」と呼ばれる技術である。コールセンターへの問い合わせに対して、FAQ表に照らし合わせてもっとも適切な答えを返す業務の機械化を例にとりて考えてみよう。本来ならFAQ表にある質問のうち、顧客の質問がどの質問に一番近いかの意味上の距離を測らなければならない。その距離は機械で算出するために数学的に定義できるものである必要があるが、数学では現実世界の「意味」を扱う方法論が確立されておらず、「意味上の近さ」を直接的に測ることはできない。そのため、次善策として、顧客の発する文章(言葉の列)と質問との表層的な近さを測ることになる。多くの場合は文字列や単語列をベクトルとして表現した上で、FAQ表の質問のうちのどれに近いかを過去のデータから学習させ、距離(近さ)の順に出力する。そのランキング1位のものが正解である確率が99%に達すれば、100かかってくる電話のうち99までを正しくさばくと期待できる。だが、それが85%にとどまる場合は、顧客の満足度が得られない恐れがある。そこで考え出された次善の策が、

1 現在、AI技術で探究されているのは、どのような入力のタイプ(画像・音声・自然言語・数値データ等)に対して、どの程度の多様性のあるデータが、どの程度の規模あれば、人間が犯すのと同じ程度の誤りで抑えることができるのか、という点だといえよう。

「分類判断支援ツール」としてのAIの導入である。ランキング上位の複数の選択肢を画面上でみながら、コールセンターの職員が問い合わせに対応することにより、1000のFAQをそらんじる必要や、自らFAQを検索する手間を軽減しようというのである。上位5位までに正解が含まれる確率が99%であれば、支援システムとして極めて優秀であり、コールセンター業務を30%効率化することができるといわれている。実際、こうしたタイプの技術は、すでにさまざまな場面で導入されている。

人間社会から労働は無くならない

ここで注意すべきことは、「顧客の相談を聞き、問題解決を図る」という点で、人間にとっては似ていると感じられるコールセンターと便利屋は、機械にとっては全く別のタイプの業務である、ということである。前者は、意味理解を要求しない分類問題を解いているにすぎないが、後者は真に状況判断と意味理解を要求する問題解決業務である。今後は機械の導入により前者は価格ゼロに向けて最適化される一方、後者のように意味理解に柔軟性が求められる仕事は引き続き人間が行うことの付加価値が高まっていくことだろう。

また、経営判断や介護、保育、あるいはベンチャーの起業や地方創生コンサルティングの仕事のように、どこからどこまでがその判断に必要な情報で、何が判断の結果なのか、またそれが正解なのかどうかを判定したい業務もある。さらに、類似のデータが十分でない場合には機械が判断を誤ることは避けられない。人間ならば過去に類似のデータがなくても、それ以外の情報(文脈等)から当然に正しく判断できるような場合でも、統計学的手法に基づく機械では柔軟に対応できないことがままある。

こうした仕事が残る限り、人間の社会から労働が無くなることはない。機械に代替されない労働を担うことができるスキルをもつ人の価値は一層高まり、他方、そのようなスキルを持たない人は、低賃金の労働にしがみつくと以外にはない。希少な労働とそうでない労働の間に格差が生じやすく、また後者は長時間労働に結びつきやすい点も懸念される。

ここに、デジタルライゼーション時代の教育についての具体的な検討が急がれる理由がある。上記でみたように、デジタルライゼーション前と後では労働市場で価値をもつスキルが大きく変容する。労働者が現代の労働市場に参入するには、デジタルライゼーション時代にふさわしいスキルを教育を通じて身に着けることが必須となる。

2. 「ロボットは東大に入れるか」

人材育成のあり方を認識するための東大口ロ

2017年3月、2020(平成32)年度から始まる新学習指導要領が発表された。まさにデジタルライゼーション時代を生きる子どもたちをいかに育てるかを規定するものである。しかし、その議論が始まった2011年ごろは、まだ、デジタルライゼーションの中でAIとともに生きる2020年代の明確なイメージは存在していなかった。AIが社会に導入されることは不可避だとして、どこまでホワイトカラーの仕事が代替し、何が人の仕事として残るのか、についての具体的なイメージは共有されていなかったのである。

そんな時代を背景としてAIプロジェクト「ロボットは東大に入れるか」(通称:東ロボ)は2011年に始まった。ホワイトカラーになることを目指すために(と考えられる)大学受験に向かう50万人の高校3年生と、現在の、そして近未来のAIの性能を比較しようという前代未聞の試みである。1990年代の第5世代コンピューターの手ひどい失敗がトラウマ(心的外傷)となった結果、当時の日本には大型のAIプロジェクトは皆無の状況だった。そのこともあって、日本では、どのようなタイプの粒度のデータがどの程度集積すれば、高精度の判断を機械にさせられるかの肌感覚が欠落しており、Google Carの出現や、IBMのWatsonのクイズ番組での勝利に右往左往する状態だった。そこで、産業界と大学の研究機関が連携して大学入試というベンチマークに取り組むことで、近未来AIの可能性と限界を明確にした上で、日本としてどのようなビジネスに取り組むべきか、またどのような人材育成をすべきかを正しく認識するために、東ロボは開始されたのである。

5年間の研究の結果、東ロボは「現在の理論とそれに基づく近未来のデータと技術では、相手と意思疎通をし、状況を的確に判断し、人と協力しながら問題解決を図るようなAIを生み出すことはできない」という結論に達した。深層学習と呼ばれる一連の技術は、人間のように少ない事例から一般化をすることはできないし、0、1や画像として表すことができないような、「意味」「技術」といった抽象概念は扱えない。冒頭でも述べたとおり、それを扱うための数学の枠組みがそもそも存在しないからである。なかには、「未来のことはわからないではないか」と反論する人もいるが、はっきりしていることがある。仮に、「意味を理解するAI」が生まれるなら、それは人工知能やハードウェアの世界で革命が起きるのではない。それらを支える理

表1 2016年度センター模試における東口ボの成績
【成績概要:2016年度進研模試 総合学力マーク模試・6月】

	国語 (200)	数学ⅠA (100)	数学ⅡB (100)	英語(筆) (200)	英語(リ) (50)	物理 (100)	日本史B (100)	世界史B (100)	5教科合計 (950)
全国平均点	96.8	54.4	46.5	92.9	26.3	45.8	47.3	44.8	437.8
東口ボ得点	96	70	59	95	14	62	52	77	525
東口ボ偏差値	49.7	57.8	55.5	50.5	36.2	59.0	52.9	66.3	57.1

(注)5教科8科目文系型(国、数2科目、英(筆記およびリスニング)、地理歴史2科目、理科1科目)での受験者12万582人(受験者総数26万4604人)で集計した偏差値。科目名の下のカッコ内は各教科の配点。英語の(筆)は筆記、(リ)はリスニング。

論が、数学の世界で発見されたときである。理論の準備なしに、ある日突然AIが完成する、というのはSFの世界でのみ起こり得る奇跡だ。

一方で東口ボは、教科書やウィキペディアを丸暗記し、検索をして切り張りすることで、もっともらしい小論文を書いたり、穴埋めをしたりすることは人間以上にうまくできることを示した。東口ボは東大世界史模試の600字の論述式の問題で、人間の受験生に後れを取ることにはなかったのだ(国立情報学研究所ニュースリリース, 2016b)。つまり、論述かマークシートか、が問題なのではない。意味がわからなければ解けないような問題が出題されるかどうか、が問題なのだ。

2013年、オックスフォード大学の研究グループは2030年ごろまでにアメリカの雇用者の半数の職が機械に代替されるという予測を発表した(Frey and Osborne, 2013)。論文で指摘された、銀行や保険の審査や会計処理、スポーツ記事の執筆といった「消える職業」は、抽象概念の操作や他者との協力はそれほど必要ではなく、過去のデータから類似事例を見つけ出し、型どおりに当てはめるという作業が占める割合が圧倒的に多いということなのだろう。

3. 意味がわかる、というスキルの重要性

「絵に描いた餅」にしてはならない

デジタル化の時代の労働市場に求められる人材あるいはスキルとはどのようなものだろう。前述のオックスフォード大学の研究で、機械に代替されるとされた職業の上位には、銀行の融資担当者、保険の審査担当者、パラリーガル、会計士、スポーツ記事の執筆など、高学歴や資格を必要とする職業も並んだ。デジタル化によって大きな影響を受けると予想されている職種が、(人間の目から見ると)多岐にわたるため、ともするとAIは万能なのではないかという誤解を招きやすい。しかし、上述のような職種は、第1章で述べたような機械に代替されや

すい条件をたまたま備えていただけのことである。

AI技術は、論理または確率・統計などの数学の言語に依拠しているため、①入出力が定まらない非定型的労働、②正解・不正解の判定基準を定型化しにくい労働、③極めて少数の事例から学び正しい判断をする労働、④これまでにはないビジネスを生み出す労働、といった数学的言語で処理できない労働をAIに求めても無力である。従って、このような労働力は今後とも需要は増えこそすれ、減ることとはないと考えられる。

しかし、落とし穴もある。前述の「③極めて少数の事例から学び、正しい判断をする労働」や、「④これまでにはないビジネスを生み出す労働」は、高度にクリエイティブな労働であり、これらの労働が希少であることは20世紀のころから社会全体で共有されていた認識である。仮に、定型的教育プログラムによってそのような能力開発が可能であったならば、すでに開発されていたに違いない。現在、声高に叫ばれているアクティブラーニングは、第2次大戦直後の進駐軍時代の日本において「生活単元学習」という名のもとで実践された教育に極めて似ている。しかし、「学んだ知識を、社会の現象に当てはめて問題解決したり、検討・議論したりする」ことを重んじる生活単元学習に対しては、「日本人の学力を著しく低下させている」との批判が巻き起こり、1960年代には系統的な学習指導要領に修正されたとの歴史がある。教育現場の実感としては、生活単元学習の理想は「絵に描いた餅」であり、すでに能力が高い子どもはさらに能力を高めるが、それ以外の子どもは何らスキルを身に着けることができないことが問題視されたのである。例えば「100メートルを10秒未満で走る」ことを求められたとしても、そのような人材は確率的にしか育成することはできない。にもかかわらず、教育システムの目標をそこに設定すると、結局のところ膨大な教育コストに反してほとんど得られるものが何もない。生活単元学習の教訓から学ぶべきであり、同じことを繰り返してはならない。

一方、「①入出力が定まらない非定型的労働」や、「②正

解・不正解の判定基準を定型化しにくい労働」は、状況判断や意味理解、言葉や身体を通じて適切なコミュニケーションができる人間であれば、機械以上の能力を発揮することは決して難しくない。具体的に言い換えるなら、本来初等中等教育が目指していた「よく見、よく読み、よく聞き、よく書き、よく話す能力」、そういう当たり前の能力さえ全員が身に着けることができるなら、人工知能の時代を恐れることはない。なぜなら AIには意味がわからず、真の意味で言葉を運用することができないからだ。

意味を理解しながら読める能力の重要性

インターネットが張り巡らされ、多くの情報がデジタル化される高度知識社会においては、人間同士のやりとりの多くが対面でのコミュニケーションからメール等のドキュメントでのやりとりに軸足が移る。また、デジタル化の進展に合わせて求められるスキルが急速に変わるため、労働者は新しい知識を吸収し続ける必要がある。しかし、伝統的な徒弟的方法で伝達するのはスピードが追い付かないため、多くの企業では、eラーニングなどで自学自習することを労働者に求める。また、自ら起業する場合には、契約書の読み方や法令順守の在り方などを、インターネットから入手したドキュメントから学ぶ必要がある。つまり、学校という場や、担当部署の中で、教師役の人から「教えてもらう」のではなく、ドキュメントから自ら学ぶということが不可避に求められるのである。その際、どうしても避けることができないのが、「意味を理解しながら読める」という能力である。デジタルイゼーションの中で、機械と協働しつつ①や②の能力を発揮して、人間が機械と差別化を図るには、「よく見、よく読み、よく聞き、よく書き、よく話す能力」が必要となる。

これらは、一般的に中等教育、さらには高等教育を受けた者ならば、誰もが身に着けているスキルだと思いがちである。特に、日本のように経済協力開発機構(OECD)の学力調査(PISAやTIMSS)で優秀な成績を収めているような国では、当然にそうだと考えがちである²。

しかし、その期待は実態とは大きくかけ離れている。われわれは日本数学会と協力し、2011年に5000人を超える大学生に対し、高校1年までに習う数学の中でも、特に基本的な項目をどれだけ理解しているかの調査を行った(日本数学会教育委員会, 2013)。その結果、いくつかの課題が見いだされた。例えば、「平均」がもつ意味を正しく理解

している大学生は調査対象者の4分の3にとどまることがわかった。「全部の数値を足して、個数(人数)で割ると平均が求められる」ことを知っており、それを実行できる大学生の割合は100%に近いにもかかわらず、平均から何が結論として導かれるのか理解していないのである。また、特に耳目を集めたのが、「なぜ偶数と奇数をたすと奇数になるかを説明しなさい」という問題への正答率が20%に満たなかった(準正答を含めても34%)ことである。答えは「偶数を $2n$ 、奇数を $2m+1$ と表すと、その和は $2n+2m+1=2(n+m)+1$ になるため奇数になる」である。すべての大学生が、中学校・高校と2度学んだはずの題材であり、特に高度な数学的な知識や計算の正確さは求められない。では、なぜ6割以上の大学生が準正答にさえたどりつけなかったのか。それは、彼らが中学1年生のときに学んだはずの変数の正しい使い方を身に着けていなかったためである。

20世紀の労働市場は、なるべく多くの知識を有し、正確に処理ができる労働者を求め、学校での教育システムはその要求に対応せざるを得なかった。だが、12年間の初等中等教育でそうした知識やスキルを効率よく詰め込もうとするあまり、「意味はわからないがとりあえず学んだ、やれと言われればとりあえずできる」という人材が多数育成されてしまった可能性がある。つまり、AIと最も差別化できるはずの「よく見、よく読み、よく聞き、よく書き、よく話す能力」が教育できておらず、現代のAIに簡単に代替されるような表層的なスキルしか身に着いていないのではないかとの懸念が生じたのである。意味を理解せず表層的に問題を解いただけの東ロボが2015、2016年と2年続けて偏差値57を上回り、高校3年生の上位25%に入ったということも、この仮説が正しいことを示唆するものであった。

そこで、われわれは中高校生が実際にどれだけ教科書に書かれている基本的な文を正しく読むことができるかを直接測定する方法を考案した。それが「リーディングスキルテスト」である。

2 経団連が2016年にまとめた「今後の教育改革に関する基本的考え方」においても、日本は基礎基本の読解力やスキルは高いということを前提として、英語やプログラミング教育、アクティブラーニングの推進をすべきと結論づけている(日本経済団体連合会, 2016)。

第Ⅱ部 デジタル化時代に必要となる人間の能力

1. リーディングスキルテストとは

AIにとって得意・不得意な読解

リーディングスキルテスト(以下RST)とは、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、受験者が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のテストである(国立情報学研究所ニュースリリース, 2016a)。RSTでは、国語の長文読解とは異なり、教科書や新聞から採った50~200字程度の短文を正しく読み解けるかを問う。受験者は与えられた時間の中で、正確に、かつ、できるだけ多く解くよう指示される。現状では、問題は択一または複数選択式で、問題ごとの制限時間は設けていない。出題される問題は、文の表層的な情報を読み取れる能力を測るものと、文の意味を理解し、正しく推論を実行できる能力を測るものの2つのタイプに分かれる。前者は(1)係り受け認識、(2)照応認識、(3)同義文判定に関する問題であり、また、後者は(4)推論、(5)イメージ同定、(6)具体例認識に関するものだ。

本稿ではすべての問題タイプを解説できないので、詳しくは論文等の情報を参照していただきたい(国立情報学研究所ニュースリリース, 2016aおよびArai et. al., 2017)。大まかにいうと、(1)から(3)の問題については、比較的AIでも解ける問題で、すでに人工知能を使った解析を行う研究対象となっているのに対して、(4)から(6)の問題は、不連続かつ劇的なイノベーションがない限りAIでは解決が難しいと考えられるものである。例えば、(4)の「推論」の問題では、「ヨーロッパは日本より相対的に緯度が高いので、夏の昼の時間が長い」という知識が与えられたとき、「ヨーロッパの夏の夜の時間は、日本に比べてどうか」ということを判断できるかどうか等が問われる。人間は、1日は昼と夜で構成されるという常識に基づき論理的に推論することで、「ヨーロッパは日本より相対的に緯度が高いので、夏の夜の時間が短い」ことがわかるが、常識に欠けるAIにはこのような推論は難しい³。人は、推論することで数少ない知識から豊かな世界観を精緻に構築し

ているのである。

次に、(1)の係り受けと、(5)のイメージ同定に関する出題問題を紹介し、正答率の結果をお示ししよう。まず、図1のような問題を考えてみよう。この問題は、文の係り受け関係を正しく認識しているかを問う問題である。

図1 「(1)係り受け」の問題例

以下の文を読みなさい。
仏教は東南アジア、東アジアに、キリスト教はヨーロッパ、南北アメリカ、オセアニアに、イスラム教は北アフリカ、西アジア、中央アジア、東南アジアにおもに広がっている。

この文脈において、以下の文中の空欄にあてはまる最も適切なものを選択肢のうちから一つ選びなさい。
オセアニアに広がっているのは()である。
A. ヒンドゥー教 B. キリスト教
C. イスラム教 D. 仏教

(出典)東京書籍(株) 中学校社会教科書「新しい社会 地理」P.36。

この種の問題は、現時点で開発済みの構文解析プログラムでも正解できる。正解は「B」である。

次に、図2は、(5)のイメージ同定に分類される、文を表象する正しい図を選ぶ問題例である。

正解は「A」である。こうした問題については、よほどフレームを限定しないと機械に解かせることは困難だろう。

人間の理解も不十分

表2は、上記の2つの問題についての生徒の正答率を示したものである。図1の係り受けの問題について高校3年生では94%という高い正答率を示しているものの、図2のイメージ同定の問題は、同じ高校3年生の正答率は45%にとどまっている。問題を解いた高校3年生は、進学率がほぼ100%の高校の生徒であったが、中学2年生で学んだ数学知識があれば解ける問題であるにもかかわらず、問題文を読んで理解することができなかったのである。これは、文系だから、理系だから、という問題で済ませるものではない。

しばしば、本調査のように、成績や入試に直結しないタイプの調査に対しては、「生徒が真面目に取り組まなかったのではないか」との疑念が持たれることがある。もし、多くの高校生が適当に答えを選んでいるなら、図1の問題

3 このような推論を汎用的にさせることを目指して第五世代コンピューターは失敗した。ただし、限定的な範囲ではいわゆるエキスパートシステムとして実現された。

図2 イメージ同定の問題例

下記の文の内容を表す図として適当なものを、A~Dのうちからすべて選びなさい。

原点Oと点(1,1)を通る円がX軸と接している。

表2 公立中学高等学校における図1と図2の問題の正答率(単位: %)

学年	図1	図2
中1	58.2	10.7
中2	48.8	22.2
中3	64.6	25.4
高1	71.8	29.0
高2	84.2	30.0
高3	94.1	45.5

でAの「ヒンドゥー教」を選ぶ高校生が相当数いるはずであるが、実際には1%でしかなかった。つまり、少なくとも進学校の高校生は真面目に提示文を読み、まじめに答えを選んでいいる。しかも、本調査で出題された問題は、「成績がよい人ほどよくでき、成績が悪い人ほどよく間違える」ことが担保されている良問ばかりである。こうした条件下にもかかわらず、図2の問題の正答率が5割に満たなかったのである。これらの問題を解けない生徒に、三角関数を教えてもまったく意味はない。だが、不思議なことに彼らの多くが、なんらかの方法で三角関数を表面的には解けるふりをして(あるいは解けなくても)、大学入試を突破していると考えられる。

2. リーディングスキルテストの理論的な考え方

問題の難易度を推定

ここで議論を先を進める前に、RSTの理論的な考え方について説明しよう。

RSTは各受験者に各問題タイプについて100問以上の問題群から受験者の解答状況に応じて問題が提示されるような適応型テストを目指している。したがって、隣に座っている受験者と自分とは解いている問題が違う。またこれまでに行った調査では、調査対象とした小学6年生から社会人までの1万5290人に対して、ランダムに出題している。RSTはなぜこのような方法で実施しても問題ないのだろうか、またそもそもなぜこのような方法で実施しているのだろうか。

一般に、テストの結果は各問題の合計得点で表される。しかしながら、例えば合計得点が100点満点中の90点であったとしても、それはその生徒の能力が高いことを必ずしも意味しない。なぜなら、そのテストは簡単な問題ばかりかもしれないからだ。合計得点は、生徒の能力だけではなく、出題された問題の難易度にも依存する。そのため、RSTでは、より正しく読解力を測定するため、出題項目の難易度に応じて能力を評価する方法が採られている。その際、必要となる問題の難易度は、事前に受験者の解答データをもとに推定する。したがって、難易度が推定済みの問題を解くことで、問題の難易度に依存しない独立の能力値を推定でき、受験者ごとに異なる問題を解いていたとしても適切な評価が可能となるのである(コラム参照)。

なぜ、RSTではわざわざ受験者ごとに異なる問題を出題しているのだろうか。RSTが目指しているのは、各受験者が問題を解くたびに、その解答データを使って当該受験者にとって適切な問題を出題する形式のテストである⁴。その実施のためには、問題の難易度が推定済みの問題を大量に用意する必要がある。そのため、1万5290人の受験者に別々の問題を出題し解答データを収集することによ

コラム

問題の難易度を推定するための方法を最も基本的なモデルを用いて説明しよう。各問題の正答確率は、 $\theta_i - b_j$ の関数で表される。ここで、 θ_i は受験者*i*の能力値、 b_j は項目*j*の難しさを表す。そして、このモデルでは正答確率は、 $\theta_i > b_j$ のとき50%以上、 $\theta_i = b_j$ のとき50%、 $\theta_i < b_j$ のとき50%以下になる。問題に対する正答確率は、受験者*i*の能力と項目*j*の難しさとの差が重要なファクターであるという考え方を取り入れ、生徒の解答データから、個々の問題の難しさを推定している。 b の値が高いほど、その問題は難しいことを示している。この考え方は、項目反応理論と呼ばれる。

(参考文献)

Lord, F. M. and Novick, M. R. (1968), *Statistical Theories of Mental Test Scores*, Addison-Wesley.

4 これを適応型テストと呼ぶ。詳しくは、Linden and Glas(2010)を参照いただきたい。

て、大量の問題の難易度を推定しているのである。受験者すべてに同じ問題を出題しては難易度が推定された項目を数多く得ることができない。

問題内容の修正

現在、RSTプロジェクトでは適応型テストの実装を目指して、問題の作成・テストの実施・難しさの推定・問題内容の修正を進めている。ここではその方法について説明する。

図3は、受験者の係り受けの能力を4つに分類し、そのグループごとに図1の問題のA～Dの選択肢のうちどれを選んだのか、その割合を示したものである。図1の問題ではBの「キリスト教」が正解なので、Bの折れ線が太字で示されている。図3をみると、係り受けの能力が高いほど、この問題に正答する割合が増えている。このため、この問題は係り受けの能力を適切に反映した良問であると考えることができる。

図4は、同じく係り受けの能力別に、図1とは異なる別の問の選択肢の選択状況を示したものである。この問題の正解はAである。最も能力の高いグループは他のグループよりも正答Aの選択割合が高いため、本問は、高い能力と低中程度の能力の受験生を識別する問題であるといえ

図3 係り受け問題の能力別にみた図1問題の選択率の違い

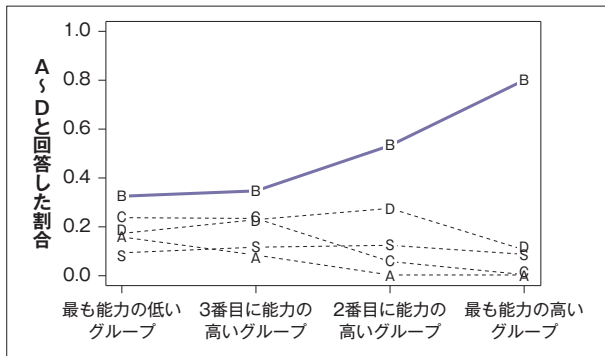
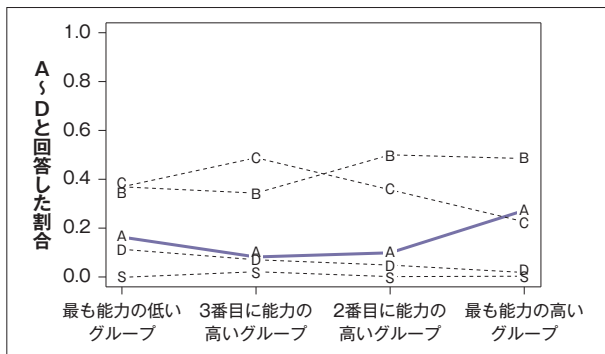


図4 係り受け問題の能力別にみた選択率の違い
一問の修正を要することが明らか



る。しかしながら、最も読解能力の高いグループであったとしても、誤答の選択肢であるBの選択割合が最も高い。ここからいくつかのことが推測できる。1つには、正答ばかりでなく、誤答の選択肢も正答と考えることができってしまう問題である可能性だ。その場合は、当該問題の内容や選択肢の修正が必要となる。あるいは、極めて能力の高い群(「最も能力が高いグループ」の中でさらに能力の高い群)の受験者を識別する問題なのかもしれない。この場合には、正答率が高いと考えられる学校等で実施することが望ましい。このようにして、RSTプロジェクトでは問題内容の修正に取り組んでいる。

3. リーディングスキルテストの結果

AIと同じような読み方をしている学生

RSTではすべての問題が択一または複数選択式なので、ランダムに答えを選んでもある確率で正解しうる。各問題タイプについて、公立中学1年生から3年生および、進学率がほぼ100%の高校1年生から高校3年生までの正答率、および正答率がランダム解答より良いとはいえない生徒(以下、ランダム解答)の割合を統計的に求めたものが表3である。

すべての問題タイプについて、ほぼ、高い学年ほど正答率も高く、ランダム解答の生徒の割合が低いことがわかる。このことは、中高生は学校および生活の中で各タイプの読解能力を伸ばしていることを示すとともに、RSTのテストとしての妥当性を示している⁵。

ここで注目すべきが、AIが射程に入れている(1)から(3)の問題タイプに比べて、意味理解なしに解くことが難しい(4)から(6)の問題タイプの正答率の方が悪く、ランダム解答の生徒の割合が高いという点である。まさに、先に述べたように、AIに代替されやすいような読解力の身に着け方をしている生徒が少なからずいると考えられる。

デジタル化時代に伸ばすべき力

今回、われわれが調査を行うまで、国でも学界でも民間でも、「中高校生は教科書程度の日本語は読めているか」という問いについて1度も調査をしてこなかった。「最近の若い者の書く文章はひどい」とか「新聞をろくに読んでいない」といった批判がちまたにあふれていたにもかかわらず、また、統一の学力状況調査(通称 学テ)を年50億円の経費で毎年実施していたにもかかわらず、である。それで

5 ただし、中3から高1の差は、各生徒の発達によるものというよりは、高校入試によって読解に課題が多い生徒がスクリーニングされた結果と解釈すべきだろう。(より偏差値の低い高校の結果については近日中に公表予定。)

表3 各問題タイプの正答率(カッコ外)およびランダム解答の生徒の割合(カッコ内)(単位:%)

	(1)係り受け	(2)照応認識	(3)同義文判定
中1	56.4 (44.6)	55.4 (46.5)	67.3 (58.1)
中2	58.4 (40.5)	55.9 (42.6)	71.8 (52.5)
中3	64.5 (29.3)	66.8 (24.4)	74.7 (41.8)
高1	82.0 (6.7)	82.1 (5.8)	87.5 (17.0)
高2	86.4 (3.5)	81.3 (10.4)	88.7 (12.5)
高3	87.5 (3.0)	86.9 (2.2)	90.5 (8.5)

	(4)推論	(5)イメージ	(6)具体例
中1	50.6 (68.4)	26.4 (57.7)	23.6 (62.1)
中2	54.0 (58.9)	28.6 (45.4)	25.4 (56.6)
中3	57.9 (49.9)	36.3 (34.5)	33.7 (41.4)
高1	67.9 (32.9)	51.1 (16.8)	49.1 (19.3)
高2	69.7 (29.5)	57.7 (10.4)	50.0 (8.5)
高3	74.9 (19.1)	53.2 (4.4)	51.3 (14.3)

は、教育改革は必ずや失敗するであろう。

今回の学習指導要領の改定により、道徳が教科として位置づけられた。その是非はいったん置くとして、小学校の国語教育が道徳と相当部分重複していることは以前から指摘されていたことである。国語と英語の教育法の違いを見ればよくわかる。英語は「英語」というものを外部の対象物として論理的に分解して教える。しかし国語は、母語であるからという理由から、「自然に身に着くもの」と前提し、その運用を論理的に指導する場面が少なかった。例えば「AならばB」という文が真であるとき「BならばA」が真であるとは限らない、とか、「誰もがAである」という文が真であるとき、「Aでないような人はいない」が真であり、「Aである人もいる」が偽である、ということ教えるような場面は国語ではほとんど見られない。

グラフや表の正しい読み方、箇条書きの正しい書き方、資料の妥当性の検証などの論理的な活動は、筆者の気持ちを押し量ることや、発表するというような活動に比べると圧倒的に少ない。道徳が教科化されたのであれば、国語からは(すべてとは言わないが)半分程度、「気持ちを押し量る」活動を、上記のような論理的活動へとシフトするという案は検討する価値があると思われる。

4. エビデンスに基づく教育改革の必要性

エビデンスの欠けた中で改革を行おうとすると必ず失敗する。旧ソ連や、カンボジアのポル・ポト政権が行おうと

した改革を例に挙げる必要もないだろう。では、日本が今般行おうとしている、あるいは過去に行ってきた教育改革には、必要とされるエビデンスはそろっていたのだろうか。

教育は誰もが受けてきた経験があり、誰もが一家言をもつ。しかし、エビデンスに欠いた提言は、本人がよかれと思っていなくても、しばしば有害である。先に引用した経団連の提言においても強調されていたように、今後は、エビデンスに基づいた教育改革が求められる。

特に、①デジタルライゼーションの可能性と限界について見極めるための調査、②デジタルライゼーションが労働市場に与える影響に関する詳細な分析、③デジタルライゼーションが急速に進む中、日本社会をソフトランディングさせるために何が必要かについての計量的なシミュレーション、④労働市場への影響を念頭においてどのようなスキルをもつ労働者を、どの程度確保する必要があるかについての検討、⑤そのために必要な義務教育および高等教育のデザイン、また、⑥それを達成するために必要な政策(教育の無償化等)の検討、などが急務だといえよう。

中学校卒業までに少なくとも8割以上の生徒が中学校の教科書や新聞程度の文章を、苦にせず正しく読むことができるように教育するために、どこから授業の時間をひねり出す必要がある。プログラミングや英語は、現状の中学生の読解力状況を考えると、無理であり、かつ、無駄と考えられるため、導入については再考が必要だろう。それにしても教育を見直すための議論の時間が足りない。

参考文献

- Arai, N. H., 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." (accepted), Proceedings of the 39th Annual Cognitive Science Society Meeting (CogSci 2017).
- Frey, C. B. and Osborne, M. A. (2013). "The Future of Employment: How Susceptible are Jobs to Computerisation?". http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf
- van der Linden, W. J. and Glas, C. A. W. Eds. (2010). *Elements of Adaptive Testing*. Springer.
- 国立情報学研究所ニュースリリース(2016a)、「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」、7月26日、http://www.nii.ac.jp/userimg/press_20160726.pdf
- 国立情報学研究所ニュースリリース(2016b)、「センター試験模試6科目で偏差値50以上 2年連続で世界史の偏差値が65突破/物理は偏差値59.0に大幅向上 論述式の数学(理系)は全問完全自動解答で偏差値76.2を達成 ~NII人工知能プロジェクト「ロボットは東大に入れるか」~」、11月14日、http://www.nii.ac.jp/userimg/press_20161114.pdf
- 日本経済団体連合会(2016)、「『今後の教育改革に関する基本的考え方』—第3期教育振興基本計画の策定に向けて—」、4月19日、http://www.keidanren.or.jp/policy/2016/030_honbun.pdf
- 日本数学会教育委員会(2013)、「第一回大学生数学基本調査報告書」、3月14日、<http://mathsoc.jp/publication/tushin/1801/chousa-houkoku.pdf>

新井紀子(あらいのりこ)

国立情報学研究所社会情報研究系教授、社会共有知研究センター長。博士(理学)(東京工業大学)。専門は、数理論理学、人工知能、教育工学。

尾崎幸謙(おさき こうけん)

筑波大学大学院ビジネス科学研究科准教授、統計数理研究所客員准教授。博士(文学)(早稲田大学)。専門は統計科学、行動遺伝学、社会調査。



PDFはこちらから

NIRA オピニオンペーパーは、ホームページでもご覧いただけます
<http://www.nira.or.jp/president/opinion/index.html>

NIRA オピニオンペーパー [no.31]

2017年7月5日発行
© 公益財団法人 NIRA 総合研究開発機構 2017
発行人：牛尾治朗

※本誌に関するご感想・ご意見をお寄せください。
E-mail: info@nira.or.jp



公益財団法人 NIRA 総合研究開発機構
〒150-6034 東京都渋谷区恵比寿4-20-3
恵比寿ガーデンプレイスタワー 34階
TEL:03-5448-1710 FAX:03-5448-1744

<http://www.nira.or.jp/index.html>